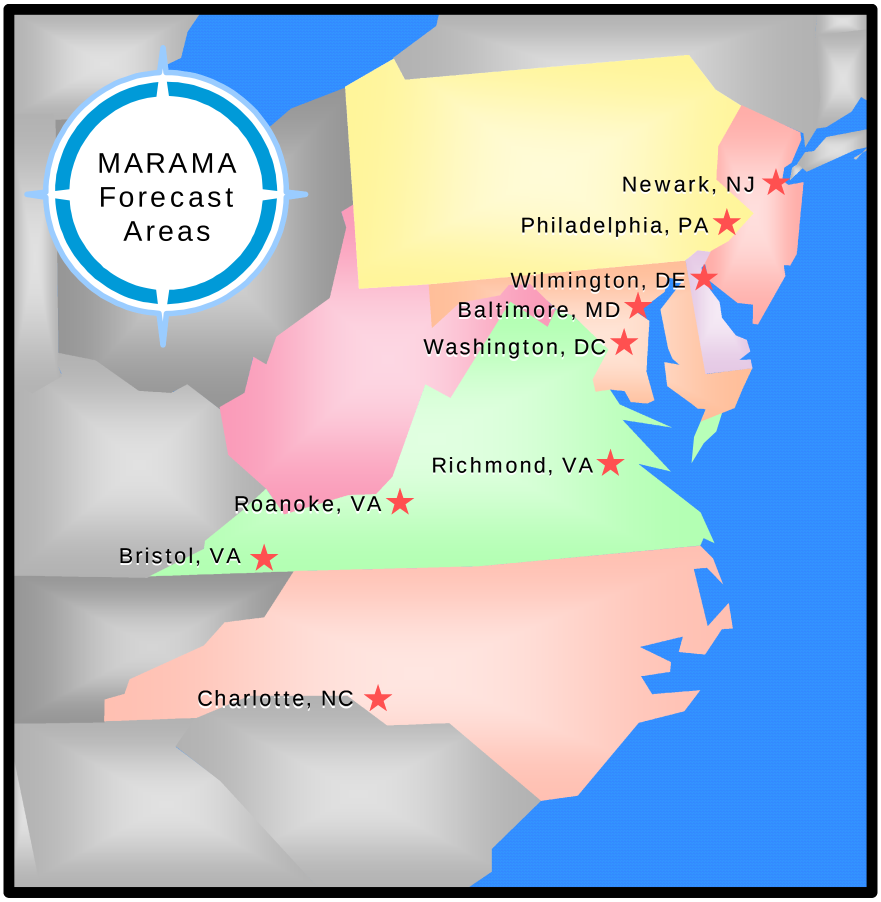
Mid-Atlantic Regional Air
Management Association

The Development of PM$_{2.5}$ Forecasting Tools for Selected Cities in the MARAMA Region

Final Report

September 30, 2004



Prepared by
ICF Consulting/SAI
101 Lucas Valley Road, Suite 260
San Rafael, CA 94093
(415) 507-7108

Prepared for the
Mid-Atlantic Regional Air Management Association

04-046

About MARAMA

The Mid-Atlantic Regional Air Management Association is a voluntary, non-profit association of ten state and local air pollution control agencies. MARAMA's mission is to strengthen the skills and capabilities of member agencies and to help them work together to prevent and reduce air pollution impacts in the Mid-Atlantic Region.

MARAMA provides cost-effective approaches to regional collaboration by pooling resources to develop and analyze data, share ideas, and train staff to implement common requirements.

The following State and Local governments are MARAMA members: Delaware, the District of Columbia, Maryland, New Jersey, North Carolina, Pennsylvania, Virginia, West Virginia, Philadelphia, and Allegheny County, Pennsylvania.

About Systems Applications International, LLC

Systems Applications International (SAI), a division of ICF Consulting, Inc., specializes in air quality data analysis and modeling. SAI was established in 1968 and throughout its history has pioneered the use of advanced analytical and modeling tools to support air quality analysis and modeling assessments of primary and secondary pollutants.

SAI is the principal developer of the Urban Airshed Model modeling systems (UAM and UAM-V) and the REgional Modeling System for Aerosols and Deposition (REMSAD). SAI has also developed statistical data analysis techniques to support the selection and characterization of modeling episode periods and ozone and particulate matter forecasting.

On the Cover: Map showing the forecast areas of the nine $PM_{2.5}$ forecasting tools that were developed.

For copies of this report contact:

MARAMA
Mid-Atlantic Regional Air Management Association
711 West 40th Street
Suite 312
Baltimore, MD 21211

phone 410.467.0170
fax 410.467.1737
http://www.marama.org/

The Development of $PM_{2.5}$ Forecasting Tools for Selected Cities in the MARAMA Region

September 30, 2004

Prepared by
ICF Consulting/SAI
101 Lucas Valley Road, Suite 260
San Rafael, CA 94093
(415) 507-7108

Prepared for the
Mid-Atlantic Regional Air Management Association

04-046

Acknowledgements

This report was written by Sharon Douglas, Johanna Mangahas, Belle Hudischewskyj, and Jay Haney of ICF/SAI. Ana Alvarez, Geoffrey Glass, and Tom Myers of ICF/SAI are acknowledged for their technical contributions to the CART analysis and the development of the forecasting tools. This report was expertly produced by Jim Phillips. Special thanks to Bill Gillespie of MARAMA and all of the state and local forecasters listed below who contributed to the design and evaluation of the study components.

Rob Altenburg, Pennsylvania Department of Environmental Protection

George Bridgers, North Carolina Department of Environment and Natural Resources

Scott Jackson, North Carolina Department of Environment and Natural Resources

Joe Martini, Delaware NREC, Air Quality Management Section

Andy Mikula, New Jersey Department of Environmental Protection

Sean Nolan, Pennsylvania Department of Environmental Protection

Bill Ryan, Penn State University

Dan Salkovitz, Virginia Department of Environmental Quality

Matt Seybold, Maryland Department of the Environment

Ram Tangirala Ph. D, District of Columbia, Department of Health

Mike Woodman, Maryland Department of the Environment

Table of Contents

List of Figures

List of Tables

Executive Summary

The primary objective of the MARAMA $PM_{2.5}$ forecasting assistance project was to develop and evaluate statistical-based tools to support $PM_{2.5}$ forecasting for nine cities in the MARAMA region. The nine cities included Charlotte, North Carolina; Bristol, Roanoke, and Richmond, Virginia; Washington, D.C.; Baltimore, Maryland; Philadelphia, Pennsylvania; Wilmington, Delaware; and Newark/Elizabeth, New Jersey. The study included the analysis of PM2.5 and meteorological data using Classification and Regression Tree (CART) analysis software and the development, testing, and evaluation of interactive forecasting tools for each area. Data and information gathered throughout the course of the project were used, together with the CART analysis results, to describe the relationships between meteorology and PM concentration and, specifically, the conditions associated with high $PM_{2.5}$ events in each forecast area.

CART Analysis

The CART analysis software was applied for each area for a multi-year period (nominally 1999–2002). All days with available data within this period were classified and grouped into bins in accordance with the values of observed and calculated meteorological and air quality parameters that comprised the input dataset. Twenty-four-hour average $PM_{2.5}$ concentration was used as the classification variable for this application and a variety of meteorological and air quality parameters were used as input data.

The air quality data used for this study consisted of measurements of $PM_{2.5}$ from sites located within and potentially upwind of each area of interest. The final dataset used for the CART analysis included Federal Reference Method (FRM) $PM_{2.5}$ data.

The meteorological data used for this study consisted of measurements of various surface and upper-air meteorological parameters for sites located within and near each area of interest.

Each CART classification bin was assigned to one of three classification categories, representing a different range of $PM_{2.5}$ concentration. The three categories were defined according to the EPA established guidelines for PM2.5 forecasting: less than 15.5 (Category 1), 15.5 to less than 40.5 (Category 2), and greater than or equal to 40.5 μgm^{-3} (Category 3). Since only a few data points were in the highest EPA category of greater than or equal to 65 μgm^{-3}, this category was not used in the analysis. The three categories used in this analysis are also referred to as “good”, “moderate”, and “unhealthy for sensitive groups (USG).”

As part of the CART application, more than 20 diagnostic and sensitivity tests were conducted for each area. The first of these included only the meteorological input parameters. The remaining tests examined the use of alternate input parameters, as well as different forms of the classification variable for $PM_{2.5}$.

Key findings from the CART analysis include:

- Different types of $PM_{2.5}$ episodes can be identified for each area based on meteorological and prior day PM indicators.
- Regional $PM_{2.5}$ parameters are more important in classifying the days for smaller/southern urban areas; local $PM_{2.5}$ variables are more important for the larger/more northern areas.
- Stability parameters are important for all areas and more stable conditions are generally associated with higher $PM_{2.5}$ concentrations.

- Temperature is used to segregate the days seasonally and is overall well correlated with the observed $PM_{2.5}$ concentrations.
- Relative humidity is also used to segregate the days but high relative humidity can be associated with both low and high observed $PM_{2.5}$ concentrations
- Wind speed is important in defining classification groupings and lower wind speeds almost always lead to higher $PM_{2.5}$ bins.
- Wind direction is often used by CART to separate and group the days, but does not always vary regularly among the categories.
- For all areas, less precipitation is associated with lower $PM_{2.5}$ but is not frequently used by CART.

The CART results can be characterized in terms of classification accuracy, which is used to quantify the degree to which days within each bin have observed concentrations corresponding to the range assigned to the bin. Misclassification can occur due to a number of reasons including: monitoring network limitations, length (completeness) of the analysis period, use of discrete classification categories, and data errors or missing data.

For this study, two sets of final CART results were produced. The first of these was used to prepare the operational versions of the forecasting tools for each area. For this set of results, the average classification accuracy is 84 percent, ranging between 80 and 91 percent, as presented in Table ES-1.

Table ES-1. CART Classification Accuracy for the Operational Forecasting Tools

	Number of CART Bins	Classification Accuracy (%)
Charlotte	33	81
Bristol	33	90
Roanoke	34	91
Richmond	29	83
Washington	38	80
Baltimore	34	80
Philadelphia	35	82
Wilmington	36	81
Newark	34	86

For forecasting purposes, it is important that higher PM days are correctly classified, and that the number of lower PM days placed into higher PM bins is minimized. For Charlotte, Bristol, Roanoke, and Richmond, there were very few Category 3 days. All Category 3 days were correctly classified. There was some tendency for CART to place Category 1 and 2 days into the Category 3 bins, especially for Charlotte and Bristol.

For the Washington, Baltimore, Philadelphia, Wilmington, and Newark areas, there were more Category 3 days. With the exception of two days for Washington, all of the Category 3 days were correctly classified. However, a significant number of Category 2 days (as well as some Category 1 days) were misclassified as Category 3.

A second set of CART results were produced for research purposes and were used to prepare research versions of the forecasting tools for each area. The research CART results differ from the operational CART results in their use of prior-day $PM_{2.5}$ input parameters. The research tools rely primarily on $PM_{2.5}$ data for one day rather than two days prior to the analysis day. For the research results, the average classification accuracy was 84 percent, ranging from 78 to 91 percent. Although the overall accuracy was similar, these results were generally less promising than the operational results, mostly because more days from Categories 1 and 2 were misclassified into the Category 3 bins. This is somewhat puzzling since more information about prior day $PM_{2.5}$ concentration should improve the classifications rather than degrade them. This issue was not resolved as part of the current project and the research versions of the tools were developed to allow further investigation of this issue and to support future work in this area.

$PM_{2.5}$ Forecasting Tools

The CART results were transformed into forecasting algorithms for each area so that observed and predicted values of the input parameters (for current and future days) could be used to place a future day into a classification bin. Future values for the meteorological parameters were obtained from standard meteorological forecast products, for example, the National Weather Service ETA model, the Global Forecast Systems (GFS) model, or the Nested Grid Model (NGM)). The resulting classification and forecast was determined by the observed and predicted data values and the pathways that comprise the CART classification tree. In this forecast mode, the predicted $PM_{2.5}$ concentration is assigned the value of the classification bin in which the day is placed.

This approach to forecasting has several attributes. Compared to simple regression techniques, the use a CART-based forecasting algorithm accommodates the possibility that different meteorological conditions can lead to the same or similar $PM_{2.5}$ concentration and, most importantly, that there may be multiple pathways to high $PM_{2.5}$. The parameters and parameter values associated with the CART classification tree provide information about the relative importance of these parameters in determining forecast $PM_{2.5}$ concentrations. Thus the CART technique offers physical insight into phenomena being studied. By segregating the data values into classification bins, CART also provides information regarding the frequency of occurrence of the conditions associated with each classification category. In this manner, the likely recurrence rate for a particular type of day and the associated prevalent conditions were obtained.

An important consideration in forecasting is, of course, the availability of real-time data to support forecasting. $PM_{2.5}$ data collected using FRM measurements were used for the CART analysis—as they were expected to provide the most consistent and accurate values. However, forecasters must rely on continuous measurements of $PM_{2.5}$ (which are available on a near real time basis) to provide information about prior day PM levels at local and upwind sites. Continuous $PM_{2.5}$ measurements do not always agree with the FRM measurements. Adjusting the continuous data to an FRM equivalent value may be one way to overcome this limitation.

For each set of CART results, four tools were developed. The four tools were for: 1) Charlotte; 2) Bristol, Roanoke, and Richmond; 3) Baltimore, Washington, Philadelphia, and Wilmington; and 4) Newark. Each tool consisted of an interface for the entry of observed and forecasted data and other parameters, the forecasting algorithms and supporting calculations for one or more areas, and several options for the display, summary, and storage/archival of the input parameters and the forecast results. The operational versions of the tools were used to support the first year of $PM_{2.5}$ forecasting for several of the areas of interest.

Preliminary versions of each tool were evaluated on a real-time basis and using historical data. Meteorologists in six of the nine MARAMA areas tested the draft versions of the operational $PM_{2.5}$ forecasting tools during February and March of 2004. For as many days as possible, each participant entered the measured and forecasted meteorological and air quality data required by the tool to predict the next day's $PM_{2.5}$ level. For the real-time evaluation, prediction accuracy ranged from 55 to 75 percent using strict evaluation criteria, and from 75 to 88 percent when days with observed concentrations very close to the values defining the different categories were considered to be correctly classified within either category. It is important to keep in mind in reviewing these percentages that all of the days forecast exhibited low (good) or moderate $PM_{2.5}$ levels. No high PM days were observed at the continuous monitors in February and March of 2004.

Continuous data were used to evaluate the forecasts during the initial real-time evaluation period. Later, the evaluation statistics were recalculated for four of the areas using FRM data. Forecast accuracy was better for Richmond and Wilmington, but worse for Charlotte and Baltimore when the FRM data were used in place of the continuous data for evaluation. The greatest differences in performance were for Wilmington (where the FRM concentrations tended to be lower than the continuous values) and Baltimore (where the FRM concentrations tended to be higher than the continuous values). Thus, uncertainty in the observed PM concentrations may affect the integrity of the real-time evaluation results.

Overall, the real-time testing of the draft version of the forecasting tools was inconclusive primarily because the period February-March 2004 did not contain any days with high $PM_{2.5}$ concentrations.

Use of historical data for June through August 2003 enabled evaluation of the forecasting tools for all nine areas. Unlike the real-time 2004 evaluation, the summer 2003 period provided ample USG days to test the tools' ability to accurately predict high PM.

The historical evaluation suggested that given perfect forecasts of the meteorological input parameters, $PM_{2.5}$ concentration ranges can be correctly predicted for 50 to 70 percent of the days and correctly predicted using the less strict evaluation criteria for 65 to 85 percent of the days (with the exception of Bristol, which has a 55 percent accuracy even with the less strict criteria).

In the historical evaluation, the two sites with the worst performance were Bristol and Roanoke. These sites had fewer data than the other sites. This outcome suggests that, because of the limited database, the CART results are incomplete with respect to representing all of the types of conditions that might occur at these sites. The implication is that use of a limited dataset may limit the predictive capability of the tools. The limited size of the historical database used to develop the tools limits for forecast performance for all areas.

The false alarm rate was relatively high for all areas, where it could be calculated, and this reflects the tendency for overestimation found in the CART results. With this tendency, the probability of detection is good for most sites, and the bias is positive in all cases for which it could be calculated. This outcome suggests that the meteorological inputs, and consequently the CART results may not sufficiently represent the conditions associated with the day-to-day transition from high to lower PM concentrations. The overpopulation of the higher PM bins with lower PM days (both in the CART results and in the historical forecast results) may also be due to a lack of a sufficient number of high PM days in the dataset. High $PM_{2.5}$ days are needed in the dataset to provided a good representation of the conditions that are associated with these days.

For a first attempt at developing a CART-based forecasting tool for these nine areas, the results are promising. The evaluation statistics are lower than, but not that much lower than those that would be considered good for 8-hour ozone forecasting (and ozone is a simpler pollutant to forecast and has been much more extensively measured and studied).

Factors Influencing $PM_{2.5}$ Concentrations

In describing the factors influencing $PM_{2.5}$ concentrations for each area, we considered 1) the magnitude and spatial and temporal characteristics of the $PM_{2.5}$ concentrations, 2) the meteorological features influencing $PM_{2.5}$ concentrations, and 3) the characteristics of high $PM_{2.5}$ events. The analysis was designed to complement, in a qualitative sense, the forecast information provided by the CART-based $PM_{2.5}$ forecasting tools.

With regard to the spatial and temporal characteristics of the $PM_{2.5}$ concentrations observed in the areas of interest:

- There is a greater incidence of high $PM_{2.5}$ days in the northern part of the MARAMA study area and within the larger metropolitan areas.
- During the period studied, the largest number of observed USG days occurred either during the second or third quarters of the year, encompassing the late spring and summer periods, although some USG days occurred during the fall and winter months as well in some areas.
- Correlations of $PM_{2.5}$ concentrations among the different areas suggest that there is a regional component to $PM_{2.5}$ in the areas of interest from Washington (possibly Richmond) northward, but that on any given day (with a few exceptions) there are also local meteorological and/or emissions influences that affect the areas separately.

- The characteristics of high $PM_{2.5}$ events vary among the areas of interest according to geographical characteristics and local and regional emissions characteristics.

Considering the meteorological features influencing $PM_{2.5}$ concentrations and the characteristics of high $PM_{2.5}$ events:

- A review of the meteorological conditions associated with high $PM_{2.5}$ in the areas of interest reveals that many of these days are influenced by a slow-moving or stationary high pressure system over the area of interest that results in suppressed vertical mixing of emissions/pollutants and low wind speeds or stagnation.
- For most of the areas, there are different types of high $PM_{2.5}$ events and these are distinguished by different stability characteristics and wind directions; the overall characteristics also vary with season.
- CART appears to be able to distinguish and group the USG days quite effectively.

Recommendations

Based on the results and findings of the study, as well as the issues and problems that we encountered in conducting the work, we provide recommendations for future enhancement of the forecasting tools and an improved understanding of $PM_{2.5}$ issues in this section.

All aspects of this study (including the development, refinement, and evaluation of the forecasting tools) emphasize the need for daily FRM and continuous $PM_{2.5}$ data on both a local and regional basis.

As more $PM_{2.5}$ data become available, use of a larger dataset encompassing a longer time period would likely better capture the range of different meteorological/$PM_{2.5}$ conditions that are likely to occur in the future as well as to better characterize the conditions associated with high PM days (which were few in number during the analysis period for several of the areas).

Continued evaluation of the forecasting tools, including an assessment of the different meteorological forecast products, will also provide important information related to improving forecast skill.

The data and results of this study could be used to enhance $PM_{2.5}$ State Implementation Plan (SIP) analyses for the areas of interest. Specifically this study can be used to support the development of a “conceptual description” of $PM_{2.5}$ formation and transport for each area (a required element of a SIP). The study results could also be used in “weight-of-evidence” analyses in which data and modeling results are use to support or corroborate the outcome of an attainment demonstration.

1. Introduction

1.1. Background and Objectives

The recent emphasis on fine particulate matter as an air pollutant of concern is based primarily on epidemiological studies that have indicated a cause and effect relationship between exposure to fine particles and health effects, including respiratory and cardiovascular disease and premature mortality. Particulates are also a primary constituent of regional haze, which limits visibility and thus diminishes the natural beauty of our environment.

Fine particulates in the atmosphere consist of primary particles that are emitted directly from sources and secondary particles that form in the atmosphere through chemical and physical processes. Pollutants that contribute to the formation of secondary aerosols include sulfur dioxide (SO_2), oxides of nitrogen (NO_x), and ammonia (NH_3). Natural sources of fine particulates and precursor pollutants include wind blown dust, sea salt, and forest fires. Anthropogenic contributors include numerous agricultural, mobile, and industrial sources. Meteorology plays an important role in particulate formation and transport and the determining the ambient particulate concentration levels.

The U.S. Environmental Protection Agency (EPA) established new National Ambient Air Quality Standards (NAAQS) for fine particulate matter in 1997. Under these standards, fine particles are defined as those with a diameter of less than 2.5 microns; particles of this size are also referred to as $PM_{2.5}$. The annual $PM_{2.5}$ standard requires the three-year average annual mean concentration to be less than 15 micrograms per cubic meter (μgm^{-3}). The daily $PM_{2.5}$ standard requires the three-year average of the 98th percentile daily average concentration to be less than 65 μgm^{-3}. According to recent data and recommendations by the States and EPA, the Mid-Atlantic Regional Air Management Association (MARAMA) region contains several nonattainment areas for $PM_{2.5}$, based on the annual standard.

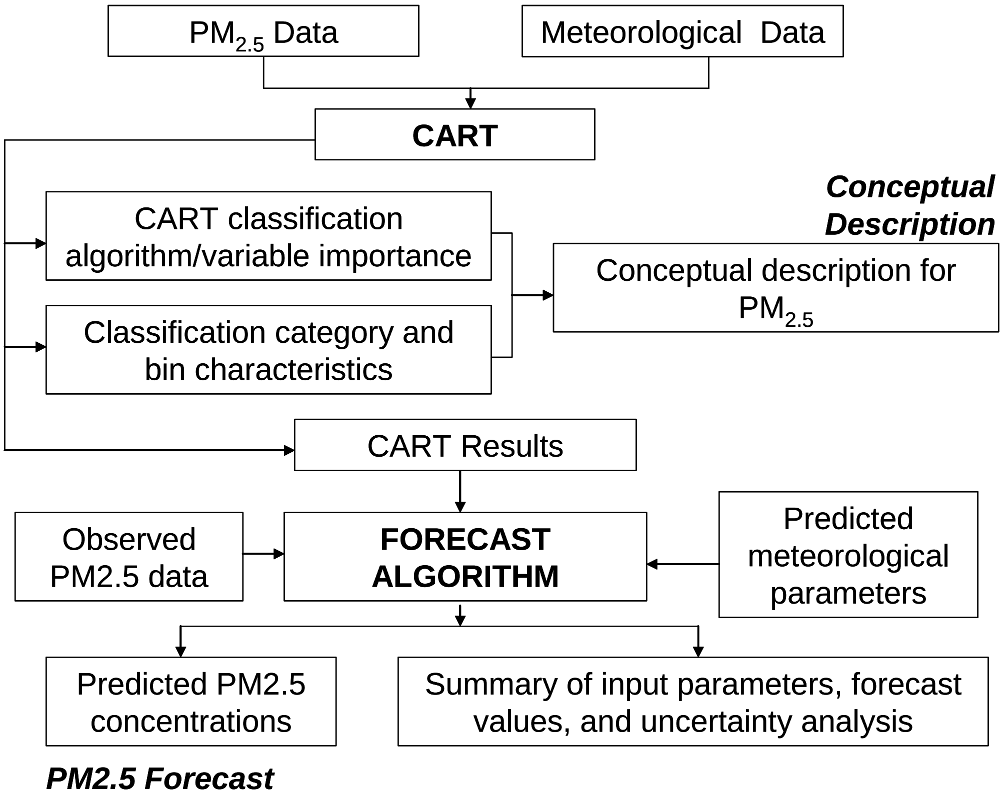
Compliance with these standards requires state and local agencies to monitor $PM_{2.5}$ concentrations within populated areas and, as needed, to develop and implement air quality plans for attainment and maintenance of the standards. To help protect public health, state and local agencies began daily forecasting of $PM_{2.5}$ concentrations in October 2003. Information regarding expected $PM_{2.5}$ concentrations allows the public to make informed decisions about their daily activities and to avoid unnecessary exposure to unhealthful concentrations. This information can also be used by businesses and industries to guide activities related to mitigation of emissions that may contribute to unhealthful particulate levels.

The primary objective of the MARAMA $PM_{2.5}$ forecasting assistance project was to develop and evaluate statistical-based tools to support $PM_{2.5}$ forecasting for nine cities in the MARAMA region. The nine cities include Charlotte, North Carolina; Bristol, Roanoke, and Richmond, Virginia; Washington, D.C.; Baltimore, Maryland; Philadelphia, Pennsylvania; Wilmington, Delaware; and Newark/Elizabeth, New Jersey. A secondary objective was to use available data and the results from the statistical analysis to understand the factors influencing $PM_{2.5}$ formation and transport in each forecast area and the MARAMA region.

1.2. Technical Overview of the Project

In this study, we used available air quality and meteorological data, together with the Classification and Regression Tree (CART) analysis technique, to develop forecasting algorithms as well as a description of the factors influencing $PM_{2.5}$ concentrations within each of the nine areas of interest. The data were obtained from EPA and the National Weather Service (NWS) and were processed and quality assured for use in the CART analysis. The CART technique was then used to examine and extract information from the data, and the resulting information was used to describe each area and to develop the forecasting algorithms. A schematic diagram of the CART-based forecasting and analysis methodology is provided in Figure 1-1.

Figure 1-1. Conceptual Design of the MARAMA PM2.5 Forecasting and Analysis Methodology



CART is a statistical analysis tool that can be used to separate days with different values of a classification variable into different bins. The CART technique accomplishes this task through the growth of a binary decision tree, comprised of a progression of binary spits on the values of a set of input variables. The resulting tree has multiple branches, of varying complexity, each of which represents a path to a specific bin. Each bin is associated with a range of values of the classification variable.

For this analysis, CART was applied for a multi-year period (nominally 1999–2002) and all days within this period were classified and grouped into bins in accordance with the values of observed and calculated meteorological and air quality parameters that comprise the input dataset. We used 24-hour average $PM_{2.5}$ concentration as the classification variable for this application and a variety of meteorological and air quality parameters as input data.

The resulting CART trees were transformed into forecasting algorithms for each area so that observed and predicted values of the input parameters (for current and future days) can be used to place a future day into a classification bin. Future values for the meteorological parameters are obtained from standard meteorological forecast products. Using this approach, the path taken through the CART tree and the resulting classification is determined by the observed and predicted data values and the binary splits that comprise the classification tree. In this forecast mode, the predicted $PM_{2.5}$ concentration is assigned the value of the classification bin in which the day is placed. By providing a basis for estimating $PM_{2.5}$ concentrations using observed (or predicted) values of related variables, CART analysis can be used to forecast $PM_{2.5}$ concentrations.

The CART-based forecasting algorithm relies on the relationships that are identified between the input variables and $PM_{2.5}$ concentration (as derived using observed data). We also used this information in this study to improve our understanding of the factors and processes contributing to high $PM_{2.5}$ values in the areas of interest and throughout the region.

This approach enabled the preparation of useable forecasting tools to support the first year of $PM_{2.5}$ forecasting for several of the areas of interest. However, the ability of the tools to represent the type and range of conditions and the different types of $PM_{2.5}$ events that characterize each area is limited by the data used to develop the tool. Data for 1999–2002 were used, and, for most areas, data were available for only a subset of this period. It is anticipated that the incorporation of new data and information would enhance the performance of the tools as well as our understanding of $PM_{2.5}$ issues.

1.3. Report Contents

A summary of the data used in this project is provided in Section 2 of the report. The CART application is described in detail in Section 3. The factors influencing $PM_{2.5}$ concentrations within each area are discussed in Section 4. The forecasting tools are documented in Section 5, and an evaluation of the tools is presented. Finally, some recommendations for further study are provided in Section 6.

This page intentionally left blank.

2. Project Database

This project relied on historical air quality and meteorological data to support the development and evaluation of the CART-based forecasting tools. The acquisition, processing, and archival of the historical data is described in this section of the report.

2.1. Air Quality Data

The air quality data used for this study consist of measurements of $PM_{2.5}$ for sites located within and potentially upwind of each area of interest. The final dataset used for the CART analysis includes $PM_{2.5}$ data obtained using the Federal Reference Method (FRM) measurement systems. Data collected using one or more continuous measurement systems were obtained and processed as part of an exploratory CART analysis. Data for the precursor species sulfur dioxide (SO_2) and oxides of nitrogen (NO_x) were also obtained and processed for two of the areas of interest.

2.1.1. Data Sources and Initial Processing Steps

All air quality data were obtained from the AIRS (Atmospheric Information Retrieval System; http://www.epa.gov/air/data/index.html) database. This database is updated regularly by EPA and the latest version of the database at the time of data retrieval was used.

In preparing the $PM_{2.5}$ data, we first identified all monitoring sites within and potentially upwind of each area of interest and determined whether the data for each individual site are FRM, continuous, or speciated. We also determined the data collection interval. For upwind sites, we required the availability of both daily FRM and continuous data—the former for use in the CART analysis as an indicator of the prior day's upwind $PM_{2.5}$ concentration and the latter for use in forecasting. We then extracted and reformatted the FRM data for each available site. For most sites, the FRM data are available on a daily basis. For two areas, Bristol and Roanoke, Virginia, the FRM data are available every three days.

During the course of the $PM_{2.5}$ forecasting project, several exploratory CART analyses were performed that used additional air quality data. SO_2 and NO_x data were obtained from AIRS and processed for sites in the Baltimore and Charlotte areas. Continuous $PM_{2.5}$ data were also obtain and processed for all of the areas of interest and associated upwind areas.

To ensure the reliability of the underlying data from the AIRS database as well as the extraction and reformatting steps, we conducted the following quality assurance checks for all data:

- State and county codes for each site were verified.
- Units for all data elements were confirmed.
- Randomly selected values in the re-formatted files were cross-checked against the original data files for accuracy.
- $PM_{2.5}$ (or other species) values for each site were extracted and sorted according to magnitude, to check the range of values for reasonableness (e.g., that all concentration values are positive) and the completeness of the dataset (i.e., that missing values are accounted for and properly indicated).

2.1.2. Summary of Data Sites and Parameters

Table 2–1 lists the air quality data sites for each area of interest, as used for the MARAMA $PM_{2.5}$ forecasting tool development project. Both local and potential upwind sites are listed; only local sites were used in determining the area-wide maximum $PM_{2.5}$ concentration for input to CART.

Table 2-1. Summary of Air Quality Monitoring Sites and Data Used in the MARAMA $PM_{2.5}$ Forecasting Tool Development Project

Area/Site Name	AIRS ID	Pollutant	Measurement Type	Sampling Frequency	Date Commenced *(for period of study)*	Use in CART Analysis
Charlotte						
Kannapolis	370250004	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Gastonia	370710016	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Charlotte #10 Fire Station	371190010	$PM_{2.5}$	FRM	Daily	1/99	Local
Charlotte Plaza	371190034	$PM_{2.5}$	FRM	Daily	1/99–7/99	Local
Charlotte #16 Fire Station	371190040	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Charlotte Garinger	371190041	$PM_{2.5}$/SO_2/ NO_x	FRM/TEOM/ Analyzers	Daily/Hourly	7/99	Local/ Recirculation/ Exploratory
Emerywood Dr.	371190042	$PM_{2.5}$	FRM	1 in 3 days	9/00	Local
HWY 321—Back Field	450910006	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Winston-Salem	370670022	$PM_{2.5}$/SO_2	FRM	Daily/Hourly	1/99	Upwind/ Exploratory
Greenville	450450009	$PM_{2.5}$	FRM	Daily		Upwind
Spartanburg	450450010	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Greenville	450450008	SO_2	Analyzer	Hourly		Exploratory
Bristol						
Sullivan Co, TN	471631007	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Highlands View Elementary School	515200006	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Knoxville—Davanna St.	470931013	$PM_{2.5}$	TEOM	Daily	1/99	Upwind
Knoxville—Vermont Ave.	470931017	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Knoxville—Mildred Dr.	470931020	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Roanoke						
Raleigh Court Library	517700014	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Market Street Fire Station	517750010	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local

Area/Site Name	AIRS ID	Pollutant	Measurement Type	Sampling Frequency	Date Commenced *(for period of study)*	Use in CART Analysis
Winston-Salem	370670022	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Richmond						
Shirley Plantation	510360002	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Bensley Armory	510410003	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Mathematics & Science Center	510870014	$PM_{2.5}$	FRM/TEOM	Daily	1/99	Local/ Recirculation
DEQ Regional Office	510870015	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
DEQ Air Monitoring Office	517600020	$PM_{2.5}$	FRM	Daily	1/99	Local
McMillan/DC	110010043	$PM_{2.5}$	FRM/TEOM	Daily	1/99	Upwind
Winston-Salem	370670022	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Washington, D.C.						
River Terrace School	110010041	$PM_{2.5}$	FRM	Daily	2/99	Local
Ohio Drive	110010042	$PM_{2.5}$	FRM	1 in 3 days	2/99	Local
McMillan Reservoir	110010043	$PM_{2.5}$	FRM	Daily	1/99	Local/ Recirculation
Rockville	240313001	$PM_{2.5}$	FRM	1 in 3 days	7/99	Local
Goddard Space Center	240330002	$PM_{2.5}$	FRM	1 in 3 days	7/02	Local
Suitland	240338001	$PM_{2.5}$	FRM	1 in 3 days	8/99	Local
Aurora Hills Vis. Ctr.	510130020	$PM_{2.5}$	FRM	1 in 3 days		Local
Lee District Park	510290030	$PM_{2.5}$	FRM	Daily	1/99	Local
Steven Corners	510591004	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Lewinsville	510595001	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Broad Run High School	511071005	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Old Town (Baltimore)	245100040	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Gettysburg	420010001	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Math & Science Center (Richmond)	510870014	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Baltimore						
Davidsonville	240030014	$PM_{2.5}$	FRM	1 in 3 days	8/99	Local
Ft. Meade	240030019	$PM_{2.5}$	FRM	1 in 3 days	2/99	Local
Glen Burnie	240031003	$PM_{2.5}$	FRM	1 in 3 days	11/99	Local
Riviera Beach	240032002	$PM_{2.5}/SO_2$	FRM/ Analyzer	1 in 3 days	2/99	Local/ Exploratory

Area/Site Name	AIRS ID	Pollutant	Measurement Type	Sampling Frequency	Date Commenced *(for period of study)*	Use in CART Analysis
Padonia	240051007	$PM_{2.5}$	FRM	1 in 3 days	1/00	Local
Essex	240053001	$PM_{2.5}$	FRM	Daily	8/99	Local
Edgewood	240251001	$PM_{2.5}$	FRM	1 in 3 days	8/99	Local
NEPS	245100006	$PM_{2.5}$	FRM	1 in 3 days	8/99	Local
NWPS	245100007	$PM_{2.5}$	FRM	1 in 3 days	8/99	Local
SE Police Station	245100008	$PM_{2.5}$	FRM	1 in 3 days	6/01	Local
FMC	245100035	$PM_{2.5}$	FRM	Daily	8/99	Local
Old Town	245100040	$PM_{2.5}$/ NO_x	FRM/TEOM	Daily	1/99	Local/ Recirculation/ Exploratory
Westport	245100049	$PM_{2.5}$	FRM	1 in 3 days		Local
Fire Stn. #50	245100052	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
McMillan Reservoir (Washington)	110010043	$PM_{2.5}$	FRM	Daily	1/99	Recirculation
Gettysburg	420010001	$PM_{2.5}$	FRM	Daily	1/99	Recirculation
Math & Sci. Center (Richmond)	517600020	$PM_{2.5}$	FRM	Daily	1/99	Upwind
River Terrace School	110010041	SO_2	FRM	Hourly	2/99	Exploratory
Sci. Museum	517600024	SO_2	FRM	Hourly	1/99	Exploratory
Philadelphia						
AMS Lab	421010004	$PM_{2.5}$	FRM	Daily	2/99	Local
Belmont Water Treatment	421010020	$PM_{2.5}$	FRM	1 in 3 days	3/99	Local
Northeast Airport	421010024	$PM_{2.5}$	FRM	1 in 3 days	2/99	Local
Community Health Services	421010047	$PM_{2.5}$	FRM	1 in 3 days	2/99	Local
Elmwood	421010136	$PM_{2.5}$	FRM	Daily	2/99	Local
Roxy Water Pump	421010014	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Camden Lab	340070003	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Pennsauken	340071007	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Gibbstown	340155001	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Bristol	420170012	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Chester	420450002	$PM_{2.5}$	FRM	1 in 3 days		Exploratory
Norristown	420910013	$PM_{2.5}$	FRM	1 in 3 days		Exploratory

Area/Site Name	AIRS ID	Pollutant	Measurement Type	Sampling Frequency	Date Commenced *(for period of study)*	Use in CART Analysis
McMillan Reservoir (Washington)	110010043	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Old Town (Baltimore)	245100040	$PM_{2.5}$ /NO_x	FRM	Daily	1/99	Upwind
Gettysburg	420010001	$PM_{2.5}$	FRM	Daily	1/99	Upwind
New Castle—MLK	100032004	$PM_{2.5}$	FRM	Daily	2/99	Recirculation
Camden	340070003	$PM_{2.5}$	FRM/TEOM	Daily	1/99	Recirculation
Wilmington						
Bellefonte	100031003	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Lums Pond	100031007	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Newark UD	100031011	$PM_{2.5}$	FRM	1 in 3 days	3/99–11/99	Local
Newark	100031012	$PM_{2.5}$	FRM	1 in 3 days	12/99	Local
New Castle—MLK	100032004	$PM_{2.5}$	FRM	Daily	2/99	Local/ Recirculation
Fairhill	240150003	$PM_{2.5}$	FRM	1 in 3 days	11/99	Local
McMillan Reservoir (Washington)	110010043	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Old Town (Baltimore)	245100040	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Gettysburg	420010001	$PM_{2.5}$	FRM	Daily	1/99	Upwind
Newark						
Fort Lee	340030003	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Newark	340130011	$PM_{2.5}$	FRM/TEOM	1 in 3 days	1/99	Local
Willis Center	340130015	$PM_{2.5}$	FRM	1 in 3 days	4/99	Local
Lexington	340130016	$PM_{2.5}$	FRM	1 in 3 days	7/01	Local
Ryders Lane	340230006	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Elizabeth Lab	340390004	$PM_{2.5}$	FRM	Daily	1/99	Local/ Recirculation
Elizabeth—Mitchell	340390006	$PM_{2.5}$	FRM	1 in 3 days	1/99	Local
Rahway	340392003	$PM_{2.5}$	FRM	1 in 3 days	12/99	Local
MLK (New Castle)	100032004	$PM_{2.5}$	FRM/TEOM	Daily	2/99	Upwind
Camden	340070003	$PM_{2.5}$	FRM/TEOM	Daily	1/99	Upwind
Bethelehem-Freemansburg	420950025	$PM_{2.5}$	FRM	Daily	1/99	Upwind

Only the entries labeled local and upwind were used in the final CART analyses. The data for the local sites were used to calculate the daily maximum $PM_{2.5}$ concentration for the areas of interest. The data for the upwind sites were used to provide information about possible transport or recirculation of PM. For each area of interest with more than one local $PM_{2.5}$ monitoring site, the maximum over all local sites was determined and used to represent the daily $PM_{2.5}$ concentration for that area. Similarly, for upwind areas with more than one $PM_{2.5}$ monitoring site, the maximum over all sites was used. In the exploratory analyses, data for the individual sites were used independently.

The local $PM_{2.5}$ concentration for each area provided the classification parameter for the CART analysis. Specifically, the classification variable for each area was assigned a value of 1, 2 or 3 based on the value of the local daily maximum concentration. Each classification category represents a different range of PM2.5 concentration. The three categories were defined based on the EPA established guidelines for PM2.5 forecasting as follows: less than 15.5 (Category 1), 15.5 to less than 40.5 (Category 2), and greater or equal to 40.5 μgm^{-3} (Category 3). Since only a few data points were in the highest EPA category of greater than or equal to 65 μgm^{-3}, this category was not used in the analysis. The three categories used in this analysis are also referred to by the colors: green, yellow, and orange and by the descriptors “good”, “moderate”, and “unhealthy for sensitive groups (USG).”

The specific air quality parameters used in the final CART analysis for each area are listed and described in Table 2-2. In this table and throughout the discussion of the CART analysis, the “analysis” day is the day that is classified by CART and the two-days-prior day is the day two days prior to the analysis day. Note that later in the report, the “analysis” day is the “forecast” day. In both cases, it is the day for which the classification analysis or the forecast is being made.

Table 2-2. Summary of $PM_{2.5}$ Parameters Used in the Final CART Analysis to Support the MARAMA $PM_{2.5}$ Forecasting Tool Development

Forecast Area	Parameter Name	Description	Units
Charlotte			
	bpm_c	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	y2dpm_gs	The 24-hour average PM2.5 concentration for two days prior at Greenville-Spartanburg.	μgm^{-3}
	y2dpm_me	The 24-hour average PM2.5 concentration for two days prior at Mecklenberg.	μgm^{-3}
	y2dpm_ws	The 24-hour average PM2.5 concentration for two days prior at Winston-Salem.	μgm^{-3}
Bristol			
	bpm_br	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	y2dpm_kn	The 24-hour average PM2.5 concentration for two days prior at Knoxville.	μgm^{-3}
Roanoke			

Forecast Area	Parameter Name	Description	Units
	bpm_ro	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	y2dpm_ws	The 24-hour average PM2.5 concentration for two days prior at Winston-Salem.	μgm^{-3}
Richmond			
	bpm_r	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	y2dpm_mc	The 24-hour average PM2.5 concentration for two days prior at Washington D.C. (McMillian).	μgm^{-3}
	y2dpm_rh	The 24-hour average PM2.5 concentration for two days prior at Richmond.	μgm^{-3}
	y2dpm_ws	The 24-hour average PM2.5 concentration for two days prior at Winston-Salem.	μgm^{-3}
Washington D.C.			
	bpm_dc	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	mxmcotgy	The maximum 24-hour average PM2.5 concentration for two days prior at McMillian (Washington D.C), Old Town (Baltimore), and Gettysburg.	μgm^{-3}
	y2dpm_rh	The 24-hour average PM2.5 concentration for two days prior at Richmond.	μgm^{-3}
Baltimore			
	bpm_b	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	mxmcotgy	The maximum 24-hour average PM2.5 concentration for two days prior at McMillian (Washington D.C), Old Town (Baltimore), and Gettysburg.	μgm^{-3}
	y2dpm_rh	The 24-hour average PM2.5 concentration for two days prior at Richmond.	μgm^{-3}
Philadelphia			
	bpm_p	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	mxcanw	The maximum 24-hour average PM2.5 concentration for two days prior at Camden and New Castle.	μgm^{-3}
	mxmcotgy	The maximum 24-hour average PM2.5 concentration for two days prior at McMillian (Washington D.C), Old Town (Baltimore), and Gettysburg.	μgm^{-3}
Wilmington			
	bpm_w	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	mxmcotgy	The maximum 24-hour average PM2.5 concentration for two days prior at McMillian (Washington D.C), Old Town (Baltimore), and Gettysburg.	μgm^{-3}

Forecast Area	Parameter Name	Description	Units
	y2dpm_nw	The 24-hour average PM2.5 concentration for two days prior at New Castle.	μgm^{-3}
Newark			
	bpm_n	*The classification parameter.* It has a value of 1, 2, or 3 such that each value corresponds to a 24-hour average PM2.5 concentration for the analysis day that is <15.5, 15.5 to < 40.5, or ≥ 40.5 μgm^{-3}.	none
	mxcanw	The maximum 24-hour average PM2.5 concentration for two days prior at Camden and New Castle.	μgm^{-3}
	y2dpm_ez	The 24-hour average PM2.5 concentration for two days prior at Elizabeth.	μgm^{-3}

In the data files that accompany this report, the site-specific portions of the parameter names are defined as follows:

gs = Greenville-Spartanburg, SC
me = Charlotte, NC (Mecklenburg Co.)
ws = Winston-Salem, NC
kn = Knoxville, TN
rh = Richmond, VA
mc = Washington, D.C. (McMillan Reservoir)
ot = Baltimore, MD (Old Town)
nw = New Castle, DE
ca = Camden, NJ
ez = Elizabeth, NJ
gy = Gettysburg, PA

2.1.3. Problems and Limitations

A key limitation of the study is related to the availability of historical $PM_{2.5}$ data for use in the CART analysis. As indicated in Table 2-1, $PM_{2.5}$ monitoring began during 1999 or 2000 for most sites/areas and data completeness ranged from approximately 65 to 100 percent for the dependent variable, based on the full period of 1999–2002. For the Bristol and Roanoke sites in Virginia, data are available only every three days. Use of data for a three- to four-year period of record with few high $PM_{2.5}$ values may limit the ability of CART to identify the key high $PM_{2.5}$ regimes or distinguish the complete set of conditions that lead to the various $PM_{2.5}$ levels—simply because the high PM days and/or the full range of meteorological conditions are not represented by a sufficient number of days in the historical database.

An important consideration in the use of the historical data to develop a real-time forecasting tool is, of course, the availability of real-time data to support the forecasting. $PM_{2.5}$ data collected using the FRM measurement systems were used for the CART analysis—as they are expected to provide the most consistent and accurate concentration values. It follows that these data are best suited for establishing meaningful relationships between meteorological parameters and PM concentration. However, because they are collected using filters, data are typically not available until several weeks after the sampling date. Instead, forecasters must rely on continuous measurements of $PM_{2.5}$ (which are available on a real time basis) to provide information about prior day PM levels at local and upwind sites and to support the forecasting. There are several different types of instruments used to collect continuous data, and these do not always agree with the FRM measurements. The level of disagreement varies from site to site, and typically from season to season (with temperature and humidity), as discussed in some detail by Gillespie et al. (2004). The issue for the CART-based forecasting project is that the real-time data from continuous measurement systems may be different enough from the FRM data under some circumstances to cause an erroneous forecast. For most areas, prior day PM2.5 concentrations were important to the CART analysis and thus to the forecasts - increasing the possibility that differences in the data types could contribute to forecast errors. Adjusting the continuous data to an FRM equivalent value is an option for the forecasters to use to overcome this limitation.

2.2. Meteorological Data

The meteorological data used for this study consist of measurements of various surface and upper-air meteorological parameters for sites located within and nearby each area of interest. To represent the local- and regional-scale meteorological conditions for each area, we selected one local surface meteorological monitoring site and one or more nearby upper-air monitoring site(s). Upper-air data collected using profiler measurement systems were also obtained and processed for several areas as part of an exploratory analysis.

2.2.1. Data Sources and Initial Processing Steps

The historical surface and upper-air data meteorological data were obtained from the National Climatic Data Center (NCDC), either via the Internet or from published CD databases. Profiler data were obtained from the National Oceanic and Atmospheric Administration (NOAA).

To ensure the reliability of the meteorological data as well as the extraction and reformatting steps, we conducted the following quality assurance checks for all data:

- All source codes used to collect and reprocess data from the original format to that used by CART were specifically reviewed before application to confirm the suitability of the data processing software for the data type/format.
- The units for all data elements and for all sites were confirmed.
- The range of time over which the data are available and the time stamp for each data element were reviewed.
- For data elements that are used directly by CART, several (at least ten) random dates and times were selected and the values of the meteorological data elements were spot-checked against the original data files.
- For data elements that are computed from the original values, several (at least 10) random dates and times were selected and the values of the derived quantities were checked against independent calculations using the original data.
- The values of the meteorological parameters for each site were sorted according to magnitude, to check the range of values for reasonableness (e.g., that all values are within

expected ranges for each parameter) and the completeness of the dataset (i.e., that missing values are accounted for and properly indicated).

2.2.2. Summary Tables

Table 2-3 lists the surface meteorological data sites for each area of interest, as used for the MARAMA $PM_{2.5}$ forecasting tool development project.

Table 2-3. Summary of Surface Meteorological Monitoring Sites and Data Used in the MARAMA $PM_{2.5}$ Forecasting Tool Development Project

Area/Site Name	WBAN Number	Parameters*	Sampling Frequency	Availability During 1999–2002
Charlotte				
CLT—Charlotte Douglas Intl. Airport	13881	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Bristol				
TRI—Bristol Tri Cities Airport	13877	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Roanoke				
ROA—Roanoke Regional Airport	13741	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Richmond				
RIC—Richmond International Airport	13740	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Washington, D.C.				
DCA—Washington Regan National Airport	13743	T, RH, WS, WD, Precip	Hourly	1/99–12/02
IAD—Washington D.C. Dulles Intl. Airport	93738	WS, WD	Hourly	1/99–12/02
Baltimore				
BWI - Baltimore Washington Intl. Airport	93721	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Philadelphia				
PHL - Philadelphia Intl. Airport	13739	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Wilmington				
Wilmington New Castle County Airport	13781	T, RH, WS, WD, Precip	Hourly	1/99–12/02
Newark				
EWR—Newark Intl. Airport	14734	T, RH, WS, WD, Precip	Hourly	1/99–12/02

**T= temperature, RH = relative humidity, WS = wind speed, WD = wind direction…*

The NWS surface meteorological datasets were largely complete. Missing data were appropriately flagged in the CART datasets.

The upper-air meteorological data sites for each area of interest, as used for the MARAMA $PM_{2.5}$ forecasting tool development project are listed and summarized in Table 2-4. The upper-air monitoring sites were matched to the areas of interest based on proximity and in an attempt to best represent the regional airflow patterns within the surrounding area. Location relative to geographic features, including the coastline, was also considered.

Table 2-4. Summary of Upper-Air Meteorological Monitoring Sites and Data Used in the MARAMA $PM_{2.5}$ Forecasting Tool Development Project

Area/Site Name	WBAN Number	Parameters*	Sampling Frequency	Availability During 1999–2002
Charlotte				
GSO—Greensboro	13723	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Bristol				
RNK—Roanoke/Blackburg	53829	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Roanoke				
RNK—Roanoke/Blackburg	53829	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Richmond				
IAD—Sterling/Washington D.C./Dulles	93734	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Washington, D.C.				
IAD—Sterling/Washington D.C./Dulles	93734	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Baltimore				
IAD—Sterling/Washington D.C./Dulles	93734	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Philadelphia				
IAD—Sterling/Washington D.C./Dulles	93734	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Wilmington				
IAD—Sterling/Washington D.C./Dulles	93734	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02
Newark				
OKX—Broohkaven	94793	T, RH-Cloud, WS, WD, ϕ	Twice per day (0 and 12Z)	1/99–12/02

**T= temperature, RH-Cloud = cloud index based on relative humidity, WS = wind speed, WD = wind direction,* ϕ *= geopotential height*

The specific surface meteorological parameters used in the final CART analysis for each area are listed and described in Table 2-5. In this table and throughout the discussion of the CART analysis, the “analysis” day is the day that is classified by CART and the prior day is the day prior to the analysis day. Note that later in the report, the “analysis” day is the “forecast” day. In both cases, it is the day for which the classification analysis or the forecast is being made.

Table 2-5. SUMMARY of Surface Meteorological Parameters Used in the Final CART Analysis to Support the MARAMA PM2.5 Forecasting Tool Development

Variable names are generic and vary slightly for each monitoring site.

Parameter Name	Description	Units
tmax_xx	Daily maximum surface temperature for the analysis day.	°C
tmin_xx	Daily minimum surface temperature for the analysis day.	°C
rh24_xx	Average relative humidity for the analysis day based on temperatures and dew point temperatures at 6Z, 9Z, 12Z, 15Z, 18Z, 21Z, 0Z, 3Z	%
pflg4_xx	Number of 6-hourly periods with rainfall greater that 0.1 inches for the analysis day.	unitless (value of 0–4)
wb24_xx	Average (vector) wind direction bin for the analysis day based on values at 6Z, 9Z, 12Z, 15Z, 18Z, 21Z, 0Z, 3Z . Binned such that 1= N, 2 = E, 3 = S, 4 = W, 5 = Calm). Not used for Philadelphia, Wilmington, or Newark.	unitless (value of 1–5)
wb24_xx2	Average (vector) wind direction bin for the analysis day based on values at 6Z, 9Z, 12Z, 15Z, 18Z, 21Z, 0Z, 3Z. Binned such that 1= NE, 2 = SE, 3 = SW, 4 = NW, 5 = Calm). Used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
ws24_xx	Average (vector) wind speed for the analysis day based on values at 6Z, 9Z, 12Z, 15Z, 18Z, 21Z, 0Z, 3Z.	ms^{-1}

In Table 2-5 and data files that accompany this report, the site-specific portions (*xx*) of the parameter names are defined as follows:

c = Charlotte, NC
br = Bristol, VA
ro = Roanoke, VA
r = Richmond, VA
dc = Washington, D.C. (Reagan/National Airport)
d = Washington, D.C. (Dulles Airport)
b = Baltimore, MD (Old Town)
p = Philadelphia, PA
w = Wilmington, DE
ne = Newark, NJ

The upper-air meteorological parameters are listed and described in Table 2-6.

Table 2-6. Summary of Upper-Air Meteorological Parameters Used in the Final CART Analysis to Support the MARAMA PM2.5 Forecasting Tool Development

Variable names are generic and vary slightly for each monitoring site.

Parameter Name	Description	Units
t85amxx	850 mb temperature corresponding to the morning sounding (12Z) on the analysis day.	°C
t85pmx	850 mb temperature corresponding to the evening sounding (0Z) on the analysis day.	°C
delt950x	Difference in temperature between the 950 mb temperature corresponding to the morning sounding (12Z) on the analysis day and that at the surface of the same sounding. Not used for Bristol or Roanoke.	°C
delt900x	Difference in temperature between the 900 mb temperature corresponding to the morning sounding (12Z) on the analysis day and that at the surface of the same sounding.	°C
delt850x	Difference in temperature between the 850 mb temperature corresponding to the morning sounding (12Z) on the analysis day and that at the surface of the same sounding.	°C
htthty7x	Height difference computed as the difference of the average 700 mb geopotential height on the current day (corresponding to the morning and evening soundings) and the average 700 mb geopotential height of the day prior to the analysis day (corresponding to the morning and evening soundings of that day).	m
cloudx	Cloud index defined as the maximum of the cloud indexes determined from the relative humidity of the morning and evening soundings. CLOUDAMx = equal to1, 2, or 3 as follows, using the relative humidity (RH) corresponding to the morning (AM) sounding at 850 mb and 700 mb = 1 if RH 850 AM < 80 and RH 700 AM <65 = 2 if RH 850 AM >=80 and RH 700 AM <65 = 2 if RH 850 AM < 80 and RH 700 PM >=65 =3 if RH 850 AM >=80 and RH 700 AM >=65 CLOUDPMx = equal to1, 2, or 3 as follows, using the relative humidity (RH) corresponding to the morning (AM) sounding at 850 mb and 700 mb = 1 if RH 850 PM < 80 and RH 700 PM <65 = 2 if RH 850 PM >=80 and RH 700 PM <65 = 2 if RH 850 PM < 80 and RH 700 PM >=65 =3 if RH 850 PM >=80 and RH 700 PM >=65 Cloudx is then the maximum of cloudamx and cloudpmx	none
ywb85pmx	850 mb wind direction corresponding to the evening sounding (0Z) of the day prior to the analysis day. Binned such that 1= N, 2 = E, 3 = S, 4 = W, 5 = Calm).	unitless (value of 1–5)
ywb85pmx2	850 mb wind direction corresponding to the evening sounding (0Z) of the day prior to the analysis day. Binned such that 1= NE, 2 = SE, 3 = SW, 4 =NW, 5 = Calm). Used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)

Parameter Name	Description	Units
ywb70pmx	700 mb wind direction bin corresponding to the evening sounding (0Z) of the day prior to the analysis day (binned such that 1= N, 2 = E, 3 = S, 4 = W, 5 = Calm). Not used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
ywb70pmx2	700 mb wind direction bin corresponding to the evening sounding (0Z) of the day prior to the analysis day (binned such that 1= NE, 2 = SE, 3 = SW, 4 = NW, 5 = Calm). Used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
wb85amx	850 mb wind direction bin corresponding to the morning sounding (12Z) of the analysis day. Binned such that 1= N, 2 = E, 3 = S, 4 = W, 5 = Calm. Not used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
wb85amx2	850 mb wind direction corresponding to the morning sounding (12Z) of the analysis day. Binned such that 1= NE, 2 = SE, 3 = SW, 4 = NW, 5 = Calm. Used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
wb85pmx	850 mb wind direction bin corresponding to the evening sounding (0Z) of the analysis day. Binned such that 1= N, 2 = E, 3 = S, 4 = W, 5 = Calm. Not used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
wb85pmx2	850 mb wind direction bin corresponding to the evening sounding (0Z) of the analysis day. Binned such that 1= NE, 2 = SE, 3 = SW, 4 = NW, 5 = Calm. Used for Philadelphia, Wilmington, and Newark.	unitless (value of 1–5)
vawb85x	850 mb vector average wind direction determined from morning (12 Z) and evening (0 Z) soundings at 850 mb (binned such that 1= NE, 2 = SE, 3 = S, W 4 = NW, 5 = Calm). Used for Charlotte (Roanoke).	unitless (value of 1–5)
yws85pmx	850 mb wind speed corresponding to the evening sounding (0Z) of the day prior to the analysis day	ms-1
yws70pmx	700 mb wind speed corresponding to the evening sounding (0Z) of the day prior to the analysis day	ms-1
ws85amx	850 mb wind speed corresponding to the morning sounding (12Z) of the analysis day	ms-1
ws85pmx	850 mb wind speed corresponding to the evening sounding (0Z) of the analysis day	ms-1
vaws85x	850 mb vector average wind speed determined from morning (12 Z) and evening (0 Z) soundings on the analysis day. Used for Charlotte (Roanoke).	ms-1

In Table 2-6 and data files that accompany this report, the site-specific (*xx*) portions of the parameter names are defined as follows:

G = Greensboro, NC

R = Roanoke, VA

D = Washington, D.C. (Dulles Airport)

B = Brookhaven, NY

In addition to the surface and upper-air meteorological data and the air quality data, an additional variable, seas3 was used. This variable was set equal to “1” if the analysis day was in the month of January, February, March, November, or December. The variable was set equal to “2” if the analysis day was in the month of April, May, September, or October. And lastly was set equal to “3” if the analysis day was in the month of June, July, or August.

2.2.3. *Problems and Limitations*

For this analysis we used primarily routine NWS data and data quality and completeness was generally very good. We encountered one issue with the surface data. For Washington D.C., the surface wind data for the Dulles Airport monitor were substituted during the course of the analysis for the surface wind data for Reagan/National Airport. Although the Reagan/National Airport is located closer to the urban area, the location of the wind monitor relative to the Potomac River is expected to cause the winds from this monitor to be unrepresentative of the area. Thus the surface winds for the Dulles Airport monitor, located in an open area to the west of the city, were used instead.

One issue regarding the use of the upper-air data is that with the exception of Roanoke and Washington, D.C., there are no upper-air monitoring sites within the areas of interest. Thus we were required to use data for the nearest upper-air monitoring sites to describe the upper-air conditions. The assignments, as given in Table 2-4, were based on proximity. Location relative to geographic features, including the coastline, was also considered. In general, good matches were achieved with either nearby or similarly located sites. Nevertheless, the lack of local upper-air data is a limitation for the analysis.

One possible solution to the lack of local upper-air data is the use of profiler data, where available. As part of this study, we investigated the use of profiler data for Baltimore using the Ft. Meade profiler data. Because of a lack of moisture measurements and temperature data coupled in time/space, we used only the wind data available from Ft. Meade. Moisture, temperature, and geopotential height data for the CART simulations were based on measurements from the Dulles soundings for the CART runs. We found that the results using the Ft. Meade wind data were similar to those resulting from the use of Dulles sounding data. Since nothing appeared to be gained from the use of these data, the use of the more standard, readily available data from Dulles was chosen.

Key issues with the use of the profiler data were that moisture data were not available and temperature data were either not available, or were not coupled in time. Also, for the most part, data at the sites of interest were not available for the entire analysis period.

2.3. Electronic Datasets

The CART input datasets for each area are provided as an electronic attachment to this report (Attachment A). The air quality data were processed using Microsoft ACCESS and EXCEL on personal computers (PCs). The meteorological data were initially processed using UNIX Fortran programs on main-frame computers and the data were then passed to PCs where they were converted to EXCEL format. The air quality and meteorological data from the various sources were then merged into EXCEL spreadsheets for each area of interest. These data files were then converted into systat (*.sys) format using DBMS/Copy for Windows. It was at this point that additional “computed” parameters were added (i.e. for cloud, season, maximum PM), final missing data were set/flagged consistently, the databases were “stripped” of days not meeting criteria for a given area, and final QA/QC was performed. Days with missing dependent variables were not specifically stripped out and as a result the final CART-ready databases do contain days with missing dependent variables. CART itself handles these days appropriately. CART was run using these final *.sys formatted files. Data files provided to the MARAMA participants are EXCEL files that have been created from these final *.sys files used in CART. All blank (missing) cells have been replaced with “-999.”

This page intentionally left blank.

3. CART Analysis Methods and Results

In this section we discuss the application of CART for each of the areas of interest. We begin with a brief overview of the CART program.

3.1. Overview of CART

The CART analysis software (Brieman et al., 1984; Steinberg et al., 1997) is a statistical analysis tool that partitions a dataset into discrete subgroups based on the value of a user-defined classification variable (e.g., 24-hour average $PM_{2.5}$ concentration). The remaining variables in the database are selected as to whether or not they provide a segregation of the data for different values of the classification variable. The analysis procedure assumes that there is a causal relationship between the independent variables and the dependent variable. Consequently, it is necessary to construct a database of independent variables such that this relationship can be identified.

The CART technique is designed to segregate objects or, in the case of air quality analysis, days with different values of a classification variable into different bins or terminal nodes. The CART technique accomplishes this task through the development of a binary decision tree, comprised of a progression of binary splits on the values of the independent variables. At each split, or node, the data are divided according to their value for one of the independent variables, in a way that improves their segregation by the dependent variable. The end of a branch—called a terminal node, or bin—corresponds to a subset of the data with predominantly one value for the classification variable, characterized by independent variable ranges defined along the path to that bin. Thus the tree identifies parameter conditions frequently associated with values of the dependent variable. The user specifies the desired complexity of the tree, that is, the degree of branching and resulting number of terminal bins.

The parameter and parameter values associated with the CART classification tree provide information on the relative importance of the various air quality and meteorological parameters to the air quality conditions as represented by the dependent variable. Thus the CART technique not only segregates the days, but does so in a manner that provides physical insight into the classified days. This physical insight allows the analyst to examine whether the data partitioning is meaningful.

By segregating the data values into the classification bins, CART also provides information regarding the frequency of occurrence of the conditions associated with each classification category. In this manner, the likely recurrence rate for a particular type of day and the associated prevailing conditions are obtained.

3.2. CART Application Procedures

The primary goal of this project was to use the results of the CART application to develop a forecasting algorithm for each area of interest. CART was applied for a multi-year period (nominally 1999–2002) and all days with available data within this period were classified and grouped into bins in accordance with the values of observed and calculated meteorological and air quality parameters that comprise the input dataset.

As discussed in detail in Section 2 of this report, we used 24-hour average $PM_{2.5}$ concentration as the classification variable for this application. The classification variable for each area was assigned a value of 1, 2 or 3 based on the value of the local daily maximum concentration. The categories were defined based on the EPA established guidelines for PM2.5 forecasting as follows: less than 15.5 (Category 1), 15.5 to less than 40.5 (Category 2), and greater or equal to 40.5 μgm^{-3} (Category 3). Since only a few data points were in the highest EPA category of greater than or equal to 65 μgm^{-3}, this category was not used in the analysis. In applying CART we also included a variety of meteorological and air quality parameters as input data, as discussed in detail in Section 2. CART was applied separately for each area of interest.

CART requires the specification of "costs" associated with the misclassification of days into bins corresponding to a different category than indicated by the observed data. For this application we assigned the misclassification costs so that misclassification by two categories was twice as costly as misclassification by one category (the costs are applied on a relative basis). Misclassification can occur due to a number of reasons including: monitoring network limitations (the highest PM concentration in an area may not be observed), use of discrete classification categories (days with PM values near the category boundaries may be misplaced into a lower or higher category, but in this case the concentration difference is only slight), the complexity of the inter-variable relationships, the completeness of the dataset with respect to defining these relationships, and data errors or missing data. The misclassification costs are used in optimizing the trees, considering both classification accuracy and the number of terminal bins.

For this study, we selected trees comprised of approximately 30 to 35 terminal bins, with the best accuracy within this size range. We examined the results with respect to classification accuracy and physical reasonableness. As discussed in Section 4 of this report, we also used the results to examine the factors influencing $PM_{2.5}$ concentrations within each area of interest. Specific review tasks included:

- The input variables and CART input parameters were checked and verified.
- The matrix representing the statistical goodness of the classification (as created by CART), was examined for serious misclassification.
- The relative importance of the input parameters was reviewed.
- The overall structure of the classification tree and number of classification bins were checked to ensure that the pathways to the different classification bins are distinct and that the bins provide a reasonable segregation of the days based on the daily $PM_{2.5}$ values.
- The values used to determine the branching of the CART output classification trees were checked to ensure that the values are reasonable and consistent with the input data.
- All splits in the decision tree were examined to ensure that the parameters and values used to develop the classification tree are physically meaningful (i.e., consistent with basic conceptual models of $PM_{2.5}$ formation and transport).
- One or more bins representing each classification category were selected and the decision pathways leading to those bins were explicitly checked to ensure that they are physically reasonable.

As a final step in the review of the CART results, we also prepared tabular summaries of the mean values of the input variables for each category and each key bin.

3.3. CART Results

As a guide to the summary of the CART results that follow, we provide a brief summary of the components of the CART application:

The CART results that are presented in the greatest detail in this section are those for which the most accurate classification was achieved for all sites, using a consistent set of data and assumptions. These CART results were used to create the final “operational” versions of the $PM_{2.5}$ forecasting tools for each area.

Prior to the preparation of the final “operational” versions of the tool for this project, draft versions of each tool were prepared and distributed to MARAMA and the state forecasters. The CART results used in the preparation of the draft tools were evaluated and refined in preparing the final results for the operational tools, and are also presented in this section.

As part of the CART application, more than 20 diagnostic and sensitivity tests were conducted for each area. The first of these included only the meteorological input parameters. The remaining tests examined the use of alternate input parameters, as well as different forms of the dependent variable for $PM_{2.5}$. The key findings from the diagnostic and sensitivity tests are discussed in this section.

As part of this exploratory analysis, a “research” version of the forecasting tool was prepared for each area that includes an additional prior day’s PM concentration parameter. The “research-version” CART results are also briefly presented.

3.3.1. Summary and Key Findings from the Diagnostic and Sensitivity Testing

Meteorological and $PM_{2.5}$ Input Parameters

As first test of CART, we used only meteorological parameters. The purpose of this test was twofold, to: 1) quality assure the input datasets and ensure their readiness for CART, and 2) obtain information about the relative importance of the various meteorological parameters in the construction of the CART trees. Please note that several sensitivity tests were performed as part of the meteorological parameters only applications, to refine the meteorological inputs. A key refinement was the use of the number of six-hour periods of precipitation versus total precipitation; this was done to represent both the magnitude and the temporal (and potentially geographical) extent of the precipitation. The change in 700 mb geopotential height (from the prior day to the analysis day) replaced the twice-daily 700 mb geopotential height variables as a potentially better indicator of regional-scale pressure patterns. Both of these changes to the meteorological input parameters resulted in a slight improvement to the CART results for most areas.

We then added the prior day (two-days-ago) $PM_{2.5}$ concentrations to the input files. These results provided insight into the relative importance of the prior-day PM inputs and whether this information improved the ability of the CART tool to correctly classify the historical days and develop meaningful relationships. The results for each area follow. Please note that several sensitivity tests were conducted for each area to determine the best approach for including the prior day $PM_{2.5}$ values—these addressed which local and upwind sites/areas to include and how/whether to combine the values over multiple sites.

The results of the “met only” and combined meteorological and air quality parameters CART applications for each area are as follows:

- For Charlotte, the key meteorological parameters are the 900 mb to surface temperature difference, relative humidity, and 850 mb temperature. Surface wind speed is also a factor. Including the two-days-ago PM concentrations for Charlotte, Greenville-Spartanburg, and Winston-Salem changes the relative importance of the parameters slightly, but the meteorological parameters listed above remain most important. The overall classification accuracy of 78 percent is not improved, but the misclassification of Category 1 and 2 days into Category 3 bins is improved dramatically (reduced from 15 to 6).
- For Bristol, the key meteorological parameters are the 850 mb temperature, 900 mb to surface temperature difference, and relative humidity. Surface temperature is also somewhat important. Including the two-days-ago PM concentrations for Knoxville changes the relative importance of the parameters such that surface wind direction, surface temperature, and the two-days-ago Knoxville parameters are most important. Overall classification accuracy is improved from 83 to 88 percent, and the misclassification of Category 1 and 2 days into Category 3 bins is also improved (reduced from 11 to 4 days).
- For Roanoke, the key meteorological parameters are wind speed aloft, surface temperature, and, 850 mb temperature. Including the two-days-ago PM concentrations for Winston-Salem changes the relative importance of the parameters slightly. Overall classification accuracy is improved (from 88 to 92 percent), but the two Category 3 days are misclassified as Category 2. Including the two-days-ago information for Richmond results in almost no change to the classification tree and slight worse results; this parameter was not retained in subsequent CART applications for Roanoke.
- For Richmond, the key meteorological parameters are surface temperature and geopotential height; 900 to surface temperature difference and surface wind speed are somewhat important. Including the two-days-ago PM concentrations for Richmond, Winston-Salem, and Washington, D.C. changes the relative importance of the parameters slightly, and improves the overall classification accuracy slightly (from 83 to 84 percent). The misclassification of Category 1 and 2 days into Category 3 bins is improved (reduced from 6 to 0 days).
- For Washington, D.C., the key meteorological parameters are the 900 mb to surface temperature difference, 850 mb temperature, and surface wind speed. Surface wind direction and relative humidity area also factors. Including the two-days-ago PM concentrations for Washington, Baltimore, Gettysburg areas (the maximum over the three areas) and Richmond changes the order of importance of the key parameters. Overall classification accuracy is improved only slightly (from 76 to 77 percent), but the misclassification of Category 1 and 2 days into Category 3 bins is improved dramatically (reduced from 38 to 17 days). As we move into the Northeast Corridor, please note that there are quite a few more high $PM_{2.5}$ days.
- For Baltimore, the key meteorological parameters are the 900 mb to surface temperature difference, surface wind speed, and surface temperature. Surface wind direction and relative humidity are also important. Including the two-days-ago PM concentrations for Washington, Baltimore, Gettysburg areas (the maximum of the three areas) and Richmond changes the order of importance of the key parameters, and the PM values for the three areas moves to fourth in importance. Overall classification accuracy is improved from 74 to 80 percent and the misclassification of Category 1 and 2 days into Category 3 bins is reduced from 33 to 25 days.

- For Philadelphia, the key meteorological parameters are the surface temperature and 900 mb to surface temperature difference. Surface wind speed and relative humidity are also somewhat important. Including the two-days-ago PM concentrations for Camden and New Castle (the maximum over the two areas) and Washington, Baltimore, and Gettysburg (the maximum of the three areas) does not change the relative importance of the parameters, but the Camden-New Castle PM value takes on some importance. Overall classification accuracy is improved only slightly (from 80 to 82 percent), and the misclassification of Category 1 and 2 days into Category 3 bins is reduced from 24 to 16 days.
- For Wilmington, the key meteorological parameters are 850 mb temperature, geopotential height, and surface wind speed. Including the two-days-ago PM concentrations for Camden and New Castle (the maximum over the two areas) and Washington, Baltimore, and Gettysburg (the maximum of the three areas) does not change the relative importance of the parameters, but the Camden-New Castle PM value takes on some importance. Overall classification accuracy is unchanged from 78 percent, and the misclassification of Category 1 and 2 days into Category 3 bins is reduced from 41 to 36 days.
- For Newark, the key meteorological parameters are relative humidity, 900 mb to surface temperature difference, 850 mb temperature, and surface wind speed. Including the two-days-ago PM concentrations for Elizabeth and Camden-New Castle (the maximum over the two sites) changes the relative importance of the parameters slightly. Both PM parameters take on some importance. Overall classification accuracy is improved from 80 to 84 percent, but the misclassification of Category 1 and 2 days into Category 3 bins is worse (increased from 10 to 14 days).

The resulting CART trees using the combined meteorological and $PM_{2.5}$ parameters were designated the “Regional 1” series of trees. Key findings from the CART results at this stage of the project included:

- Different types of $PM_{2.5}$ episodes can be identified based on meteorological and prior day PM indicators.
- Regional $PM_{2.5}$ variables are more important for smaller/southern urban areas; local $PM_{2.5}$ variables are more important for the larger/more northern areas.
- Stability parameters are important for all areas.
- Temperature tends to be used as a splitter early in the tree (segregating the days seasonally).
- Relative humidity is used to segregate the days but does not have a straightforward categorical tendency.
- Wind speed is important and lower wind speeds almost always lead to higher $PM_{2.5}$ bins.
- Wind direction is often used as a split parameter, but does not always vary regularly among the categories.
- For all areas, less precipitation is associated with lower $PM_{2.5}$ but is not frequently used by CART.

Refinement of Meteorological Input Parameters

Additional sensitivity tests involved some refinement and modification of the meteorological parameters. Specifically, 700 mb wind data for the analysis day were omitted from the CART application. Those for the day prior to the analysis day were retained. Note that the 700 mb pressure level is typically at a height of approximately 3000 m. The reasoning here was that while the higher-level winds may influence the transport of pollutants into an area on the day prior, the local weather and transport conditions for the day in question are better described by the 850 mb winds. The variable use of the winds for both levels also suggested some redundancy in the information. Overall, the CART results were improved slightly when the 700 mb winds for the analysis day were omitted (mostly with regard to the reasonableness of the splits defining the pathways to the bins).

In addition, a new parameter was added to the CART analysis to indicate the time of year or season. This was primarily an attempt to represent the known variations in the amount of biogenic emissions that are present in the atmosphere and that may contribute to secondary aerosol formation. To account for seasonal variations in vegetative cover, three periods were defined. The winter period includes November, December, January, February, and March. The transitional period includes April, May, September, and October. The summer period includes June, July, and August. Including this parameter did not significantly change the CART results. Instead, surface temperature was more frequently used by CART to separate the days seasonally. Nevertheless, this parameter was retained for possible future refinement.

With these additional refinements, the resulting CART trees were designated the “Regional 2” series of trees. These were used in preparing the preliminary version of the “operational” forecasting tools.

Following an evaluation of the preliminary tools, using both real-time and historical data, additional sensitivity tests were designed and conducted to include some additional meteorological information that seemed relevant to some missed forecasts, and to incorporate some new ideas related to the use of prior day PM data.

The relative importance of stability and specifically the 900 mb to surface temperature difference parameter for most areas led us to consider whether additional stability parameters would be helpful in capturing inversions or other stability related features with different depths. In addition, for one area, Philadelphia, a missed forecast for a winter day with high observed $PM_{2.5}$ concentration seemed to be due to the presence of a very shallow surface inversion (B. Ryan, personal communication). To test the use of additional stability parameters, we defined two new parameters: the 850 mb to surface temperature difference and the 950 mb to surface temperature difference. We also defined a parameter that was the maximum of the three stability parameters, thinking that this would capture the inversion strength, regardless of the depth of the inversion. We then tested the use of these parameters in CART, first by substituting the maximum value parameter for the 900 mb difference, and then by adding all three of the difference parameters. Use of the maximum value parameter degraded the CART results for almost all areas. In hindsight, this is likely because the parameter represented different things on different days and thus it was difficult for CART to establish relationships among all days included in the dataset. Use of all three parameters, neither significantly improved nor degraded the results. CART tended to make use of all three of the parameters in various parts of the trees, and the relationships seemed reasonable. The three (separate) stability parameters were retained in subsequent of the CART applications. The resulting CART trees were designated the “Regional 3” series of trees. These were used in preparing the final version of the “operational” forecasting tools.

A second missed high $PM_{2.5}$ forecast (also for the Philadelphia area) appeared to be related to the regional-scale recirculation of pollutants (from Philadelphia to over the Atlantic Ocean, and then back again) over a three-day period (B. Ryan, personal communication). To account for this type of event we experimented with two different recirculation indexes. In both cases, the recirculation index was defined based on the 850 mb wind data already included in the CART analysis. This parameter was assigned a value of 0 or 1, with 1 indicating a potential for recirculation aloft. In the first of these, recirculation was defined according to: (1) the difference in wind direction at the 850 mb level between the previous day's evening sounding and the analysis day's morning sounding and (2) the average 850 mb wind speed (average of the evening and morning soundings). A day was classified as a recirculation day if the difference in 850 mb wind direction from the previous afternoon to the current morning was within 15 degrees of 180 degrees (i.e., almost directly opposite) or if the average wind speed at 850 mb was less than or equal to 3 ms^{-1}. In the second of these, two-day recirculation was also considered—using the same definition as above—and if either one-day or two-day recirculation was indicated, the index was set equal to 1. These parameters were included separately in CART but yielded no change in the CART results. In both cases, the index was not considered important by CART and the parameter was not retained for subsequent CART applications.

Prior-Day $PM_{2.5}$ Input Parameters

A final series of diagnostic and sensitivity simulations were conducted to examine the use of $PM_{2.5}$ data for one day prior to the analysis day (rather than two days prior). Of course, this is problematic from a forecasting perspective, since forecasts need to be made around midday and hourly data would only be available through approximately noon. There are several approaches that have been developed to estimate the air quality index using only 12 hours of hourly $PM_{2.5}$ data. Three of these are discussed and evaluated by McMillan (2004). For this study, we assumed that one or more of these approaches would be used and we included the prior day's value in CART.

For this series of tests, we prepared the prior day $PM_{2.5}$ input data three different ways, based on: 1) FRM data, 2) noon-to-noon 24-hour average of the continuous data, and 3) 12-hour average of the continuous data. These additional PM inputs were prepared for the same sites that were used to specify the two-days-ago values. In preparing the data, we found that the use of continuous data resulted in major data gaps, in many cases because the continuous monitors came on line during the mid to latter part of the analysis period. Use of these data in CART gave poor results. Instead we focused on the use of the FRM data, with the assumption that forecasters would use some methodology to estimate the prior day values.

Several alternative prior-day $PM_{2.5}$ parameters were tested. First, the prior day value was simply added to the dataset. It was used both in conjunction with the two-days-ago value, and as a replacement to the two-days-ago value. In both cases, the use of the prior-day value increased the tendency for Category 1 and 2 days to be placed into Category 3 bins. One possible explanation for this is that the meteorological parameters used in CART are not able to fully describe the conditions that would lead to a decrease in $PM_{2.5}$ (such as a cold front passage; or afternoon thundershowers). Conditions resulting in a decrease in PM are often more sudden or dramatic than those associated with an increase in PM. Thus use of a prior day value that is relatively high, frequently results (in CART) in a high value on the analysis day.

To try to mitigate the importance of the prior-day $PM_{2.5}$ concentration (as well as the need for forecasters to correctly estimate the exact value of the prior-day concentration) we also used a binned version of the concentration as an input parameter. This new parameter was assigned a value of 1 through 4, corresponding to the following ranges in $PM_{2.5}$ concentration: less than 15.5, 15.5 to less than 28, 28 to less than 40.5, and greater than 40.5 μgm^{-3}.

Further, we calculated an adjusted prior-day $PM_{2.5}$ concentration that accounted for tendencies in the concentration. Specifically, if the difference between the prior-day and two-days-ago is positive (increasing PM concentration) no adjustment is made. If the difference between the prior-day and two-days-ago is negative (decreasing PM concentration) the prior-day value is lowered by the same percentage amount. This adjusted prior-day value was also used directly and as a binned input parameter (using the same bin structure as given above).

The results of the tests using the prior-day $PM_{2.5}$ concentrations are summarized as follows:

- Use of the prior-day $PM_{2.5}$ concentration increases the overall accuracy of the CART analysis for several areas of interest but in general the results are characterized by a greater tendency to place Category 1 and 2 days into Category 3 bins
- Binning the prior-day concentration mitigates the tendency for overestimation and lessens the importance of the parameter in the construction of the CART tree.
- Adjusting the value for decreasing PM from two-days-ago to the prior day also mitigates this tendency (by allowing for an observed decreasing tendency in $PM_{2.5}$ to be accounted for).
- Binning the adjusted prior-day concentrations gives the best results overall, for the greatest number of areas (among our areas of interest).

Other considerations also favor the binned form of the parameter. The use of a binned value alleviates the need for a forecaster to correctly estimate the value (only the range needs to be correct). Although we use three bins for the classification variable, we used four bins for the prior-day value in order to distinguish between low and high Category 2 days and account for tendencies within this rather broad category.

The resulting CART trees were designated the “Research” series of trees. These were used in preparing the “research” version of the forecasting tools.

3.3.2. Preliminary and Final “Operational” CART Results

The CART results for the preliminary and final “operational” versions of the forecasting tools are presented and compared in this section. These are designated the “Regional 2” and “Regional 3” CART trees, respectively. As noted earlier, we selected CART trees with around 30–35 bins, with the best classification accuracy possible for that range of complexity. The classification accuracy is the percentage of correctly classified days, that is, days whose concentration levels match the concentration levels of the bins in which they fall. The classification accuracy for the Regional 2 trees is 83 percent on average, ranging between 78–91 percent. For the Regional 3 trees, the average classification accuracy is 84 percent, ranging between 80 and 91 percent.

Please note that only the classification results are presented in this section of the report. A more detailed analysis of the final, operational CART results for each area is provided in Section 4.

Overall classification accuracy for the Regional 2 trees is summarized in Table 3-1. The distribution of correctly and incorrectly classified days for each classification category is provided in Table 3-2.

Table 3-1. CART Classification Matrices for Regional 2 Trees

	Number of CART Bins	Classification Accuracy (%)
Charlotte	35	82
Bristol	33	88
Roanoke	34	91
Richmond	31	83
Washington	39	78
Baltimore	33	80
Philadelphia	35	82
Wilmington	37	81
Newark	34	85

Table 3-2. CART Classification Matrices for Regional 2 Trees.

	Regional 2 CART		
Actual Class	Category 1	Category 2	Category 3
Charlotte			
1	512	93	2
2	111	434	3
3	0	0	7
Bristol			
1	187	24	0
2	17	149	4
3	0	0	8
Roanoke			
1	228	20	0
2	17	168	1
3	0	0	2
Richmond			
1	694	106	0
2	124	391	0
3	0	0	7
Washington			
1	588	139	5
2	146	474	18

	Regional 2 CART		
Actual Class	Category 1	Category 2	Category 3
3	0	2	33
Baltimore			
1	384	79	2
2	89	330	18
3	0	0	31
Philadelphia			
1	641	103	3
2	116	412	17
3	0	0	28
Wilmington			
1	562	115	2
2	106	434	18
3	0	0	26
Newark			
1	350	47	3
2	51	251	9
3	0	0	18

For Charlotte, Bristol, Roanoke, and Richmond, there are very few Category 3 days. Overall classification accuracy is good to very good, and all Category 3 days are correctly classified. There is some tendency for CART to place Category 1 and 2 days into the Category 3 bins, in particular for Charlotte and Bristol.

For the Washington, Baltimore, Philadelphia, and Wilmington areas, there are more Category 3 days and overall classification accuracy is less good, but still around 80 percent for all four areas. With the exception of two days for Washington, all of the Category 3 days are correctly classified. However, a significant number of Category 2 days (as well as some Category 1 days) are misclassified as Category 3. In general, these tend to have concentrations that are near the high end of the Category 2 range, but not in all cases. Note that the number of bins is also quite large for Washington and Wilmington, as needed to get near our target of 80 percent accuracy.

Overall classification accuracy for the Regional 3 trees is summarized in Table 3-3. The distribution of correctly and incorrectly classified days for each classification category is provided in Table 3-4.

Table 3-3. CART Classification Matrices for Regional 3 Trees

	Number of CART Bins	Classification Accuracy (%)
Charlotte	33	81
Bristol	33	90
Roanoke	34	91
Richmond	29	83
Washington	38	80
Baltimore	34	80
Philadelphia	35	82
Wilmington	36	81
Newark	34	86

Table 3-4. CART Classification Matrices for Regional 3 Trees

	Regional 3 CART		
Actual Class	Category 1	Category 2	Category 3
Charlotte			
1	486	118	3
2	91	453	4
3	0	0	7
Bristol			
1	189	21	1
2	16	151	3
3	0	0	8
Roanoke			
1	223	25	0
2	14	171	1
3	0	0	2
Richmond			
1	649	86	0
2	117	349	1
3	0	0	7

	Regional 3 CART		
Actual Class	Category 1	Category 2	Category 3
Washington			
1	596	128	8
2	141	472	25
3	0	2	33
Baltimore			
1	377	85	3
2	83	339	15
3	0	0	31
Philadelphia			
1	641	103	3
2	116	412	17
3	0	0	28
Wilmington			
1	565	110	4
2	108	437	13
3	0	0	26
Newark			
1	360	36	4
2	48	251	12
3	0	0	18

Compared to the Regional 2 trees, overall classification accuracy is about the same or slightly better and the number of bins is the same or slightly lower. The tendency for overestimation is the same for Charlotte, Bristol, Roanoke, Richmond, and Philadelphia; worse for Newark and Washington; and slightly better for Baltimore and Wilmington. In preparing the Regional 3 trees, we noted and corrected a discrepancy in our approach to omitting days from the dataset based on missing data, and for all areas consistently omitted days for which the two-days-ago PM values were missing for any of the local or upwind sites used in the CART analysis. Thus the number of days is different between the Regional 2 and Regional 3 trees for some of the areas; this is especially apparent for Richmond.

Complete listings of the CART results for the Regional 3 trees are provided as an electronic attachment to this report (Attachment B).

3.3.3. "Research" CART Results

The CART results for the "research" version of the forecasting tools are presented in this section. These are designated the "Research" CART trees. As discussed in Section 3.3.1, the Research CART trees differ from the operational CART trees in their use of prior-day $PM_{2.5}$ input parameters. The Research trees rely primarily on $PM_{2.5}$ data for one day rather than two days prior to the analysis day. The data values are adjusted using the two-day prior data to account for tendencies in the concentration and binned according to specified concentration ranges.

For the Research trees, the average classification accuracy is 84 percent, ranging between 78 and 91 percent. Although the overall accuracy is similar, these results were generally less promising than either of the "Regional" tree sets, mostly because even more days from Categories 1 and 2 were misplaced into the Category 3 bins. This is somewhat puzzling—since it makes sense that more information about prior day $PM_{2.5}$ concentrations would improve the classification rather than degrade it. This issue was not resolved as part of the current project and the research versions of the tools were developed to allow further investigation of this issue and to support future work in this area.

Overall classification accuracy for the Research trees is summarized in Table 3-5. The distribution of correctly and incorrectly classified days for each classification category is provided in Table 3-6.

Table 3-5. CART Classification Matrices for Research Trees

	Number of CART Bins	Classification Accuracy (%)
Charlotte	34	86
Bristol	29	89
Roanoke	33	91
Richmond	34	86
Washington	37	78
Baltimore	34	80
Philadelphia	33	80
Wilmington	34	82
Newark	35	87

Table 3-6. CART Classification Matrices for Research Trees

	Research CART		
Actual Class	Category 1	Category 2	Category 3
Charlotte			
1	518	75	0
2	77	446	5
3	0	0	7
Bristol			
1	182	18	4
2	14	144	7
3	0	0	8
Roanoke			
1	211	27	0
2	11	171	1
3	0	0	2
Richmond			
1	628	73	0
2	82	352	0
3	0	0	6
Washington			
1	574	148	4
2	122	478	29
3	0	0	35
Baltimore			
1	375	87	3
2	72	343	20
3	0	0	29
Philadelphia			
1	639	107	1
2	116	389	38
3	0	0	28
Wilmington			
1	537	93	5
2	92	415	20

	Research CART		
Actual Class	Category 1	Category 2	Category 3
3	0	0	24
Newark			
1	328	40	2
2	31	250	14
3	0	0	16

For all areas, all Category 3 days are correctly classified. However, a significant number of Category 2 days (as well as some Category 1 days) are misclassified as Category 3. Days for which either the prior day or two-days-ago PM values were missing for any of the local or upwind sites were omitted from the dataset. Because the criteria are applied to both prior days rather than only two-days-ago the number of is sometimes different from the Regional 3 trees.

Complete listings of the CART results for the Research trees are provided as an electronic attachment to this report (Attachment C).

This page intentionally left blank.

4. Factors Influencing $PM_{2.5}$ Concentrations

In this section, we summarize the observed data used for the CART application and use these data along with other supporting information to describe the meteorological and transport conditions associated with different $PM_{2.5}$ levels in each of the areas of interest and throughout the MARAMA region.

The description of the factors influencing $PM_{2.5}$ concentrations for each area includes 1) an analysis of the magnitude and spatial and temporal characteristics of the $PM_{2.5}$ concentrations, 2) a summary of the meteorological features influencing $PM_{2.5}$ concentrations (based on weather maps, wind distribution diagrams, local knowledge, and categorical summaries of the CART input data), and 3) a detailed analysis of the characteristics of high $PM_{2.5}$ events.

This approach to describing the $PM_{2.5}$ is designed to complement, in a qualitative sense, the forecast information provided by the CART-based $PM_{2.5}$ forecasting tool.

4.1. An Overview of the Formation, Transport, and Deposition of Fine Particulate Matter ($PM_{2.5}$) in the Atmosphere

Before we present the details of the analysis characterizing the relationships between meteorological conditions and fine particulate matter ($PM_{2.5}$) within the MARAMA region, (and the statistical analysis tool developed from this analysis to assist in forecasting $PM_{2.5}$), this section provides a brief overview of the regulatory requirements for addressing $PM_{2.5}$, and a summary of the formation, transport, and deposition processes that affect ambient concentration levels.

4.1.1. Background

As measured in the atmosphere, “fine” particles are defined as particles with diameters less than 2.5 µm, while “coarse” particles are those with observed size ranges less than 10 µm (referred to as PM_{10}). In July 1997, the National Ambient Air Quality Standard (NAAQS) for particulate matter was revised by EPA. At this time, the original annual standard for PM_{10} was retained, while a new 24-hour average PM_{10} standard was added. In addition, new annual and 24-hour $PM_{2.5}$ standards were set. As required by the Clean Air Act, EPA performed a review of the original 1997 standard in 2002 and issued a formal review in August 2003. As recommended by EPA, this review maintains the original form of the $PM_{2.5}$ standard and states that a new proposal regarding the standard will be issued in March 2005 and finalized by December 2005. As a result of these regulations, states are mandated to monitor $PM_{2.5}$ concentrations, and those that weren’t already monitoring at this time began doing so in 1999 or early 2000. To assist states in monitoring for $PM_{2.5}$, a national workshop, sponsored by EPA, was held in 1998 to address and discuss the status of PM measurement research (EPA, 1998). On the basis of data collected from 2001-2003 EPA announced, in June 2004, a list of proposed $PM_{2.5}$ nonattainment areas. Following a response by the states, final designations are expected to be provided by EPA in December 2004.

Because of the link between $PM_{2.5}$ and respiratory illness, mortality, visibility impairment, and the deposition effects on water bodies and ecosystems, much effort has been expended in recent years at both the local and national levels to assess the state of fine particle concentrations throughout the U.S., and to advance the knowledge and science of $PM_{2.5}$

formation. These efforts have been undertaken to investigate the physical and chemical processes leading to $PM_{2.5}$ formation, to establish statistical relationships between meteorology and $PM_{2.5}$ formation, and to further develop and refine existing air quality models, which will be used as planning tools to develop and evaluate control strategies for meeting the applicable standards.

In July 1999, EPA finalized a new regional haze regulation, which is aimed at protecting and improving visibility in 156 Class I areas (Wilderness Areas and National Parks). Five Regional Planning Organizations (RPOs) have been established in various parts of the country to address the requirements of the regional haze regulations. Activities being undertaken by these groups include enhanced data collection (including chemical speciation of particulate matter), data analysis, emission inventory development, and air quality modeling, which is required to show future-year improvements in visibility as a result of expected changes in precursor emissions. Using available information and the known state of the science, regional assessments have been conducted to guide certain of the RPOs' in activities aimed at addressing the regional haze rule (AER, 2001; DRI, 2002). In addition, recent reports are available that summarize the knowledge and policy implications for addressing visibility (Malm, 1999; Watson, 2002). These publications summarize the current state of knowledge and discuss the challenges to be faced in lowering $PM_{2.5}$ concentrations and improving future visibility throughout the US.

4.1.2. Formation of $PM_{2.5}$ in the Atmosphere

Fine particles (also referred to as aerosols) in the atmosphere are emitted from a variety of man-made and biogenic sources (referred to as “primary” particulates) and are formed in the atmosphere from the interaction of organic and inorganic precursor gases (referred to as “secondary” particulates). They are responsible for adverse health effects and cause the most degradation in visibility. Primary fine particulates include water droplets, dust, smoke, and soot. Emission sources include open burning, power plants, automobiles, and residential wood combustion. Secondary particulates include sulfates and nitrates, which are formed in chemical reactions of sulfur dioxide (SO_2), nitrous oxides (NO_x), reactive organic gases, ammonia, etc., which are emitted by fuel combustion sources (power plants, automobiles, heaters, boilers, etc.) and other natural sources. The chemical composition, size, and ambient levels of $PM_{2.5}$ vary widely throughout the US. Nitrates and elemental carbon make up most of the fine particle mass in the West (with sulfate a smaller constituent), while sulfate constitutes the dominant fraction in the East (followed by nitrate and carbon). Heavier particles have resident lifetimes in the atmosphere of hours (due to gravitational settling), while smaller particles have resident lifetimes of days to weeks. Smaller particles are easily inhaled into the human respiratory system and may cause physiological damage. Mercury or cadmium particles deposited out of the atmosphere are toxic to living organisms and nitrates and sulfates are corrosive to building materials and vegetation. Deposited nitrates and ammonium contribute to the eutrophication of water bodies.

The major factors that affect the concentration and distribution of $PM_{2.5}$ aerosols include:

- Spatial and temporal distribution of toxic and particulate emissions including sulfur dioxide (SO_2), oxides of nitrogen (NO_x), volatile organic compounds (VOC), and ammonia (NH_3) (both anthropogenic and nonanthropogenic),
- Size composition of the emitted PM,

- Spatial and temporal variations in the wind fields,
- Dynamics of the boundary layer, including stability and the level of mixing,
- Chemical reactions involving PM, SO_2, NO_x and other important precursor species,
- Diurnal variations of solar insulation and temperature,
- Loss of primary and secondary aerosols and toxics by dry and wet deposition, and
- Ambient air quality immediately upwind and above the region of study.

A number of reactions take place in the gas phase that lead to the formation of gases that are precursors to aerosols. Secondary aerosols are formed from gases in the atmosphere by three processes: condensation, nucleation, and coagulation. Condensation involves gases condensing on smaller nuclei, nucleation involves the interaction of gases and particles to form larger particles, and coagulation involves particle growth by collision. Relative humidity plays a key role in particle growth, especially for sulfates and nitrates.

Gaseous NO_x reacts in the atmosphere with reactive hydrocarbons and organic particulates in a very complex set of reactions resulting in secondary organic particles, nitric acid, and ammonium nitrate. Nitric acid can be a precursor to PM, but HNO_3 itself is fairly volatile and highly prone to deposition on surfaces other than PM. When ammonia is present, ammonia and nitric acid can react to form ammonium nitrate. This reaction may take place in gas phase at low humidity (forming solid particles), but it is more likely to take place in aqueous phase, in tiny water droplets (aerosols) suspended in the atmosphere. This would seem to be a straightforward process for forming PM, but the presence of sulfate (formed from SO_2, as discussed below) can cause volatile HNO3 to reform from the ammonium nitrate (Seinfeld, 1986). Therefore, the amount of PM derived from NO_x is a function not only of the rate of formation of nitric acid, but also of how much ammonia and how much sulfate is present in the atmosphere.

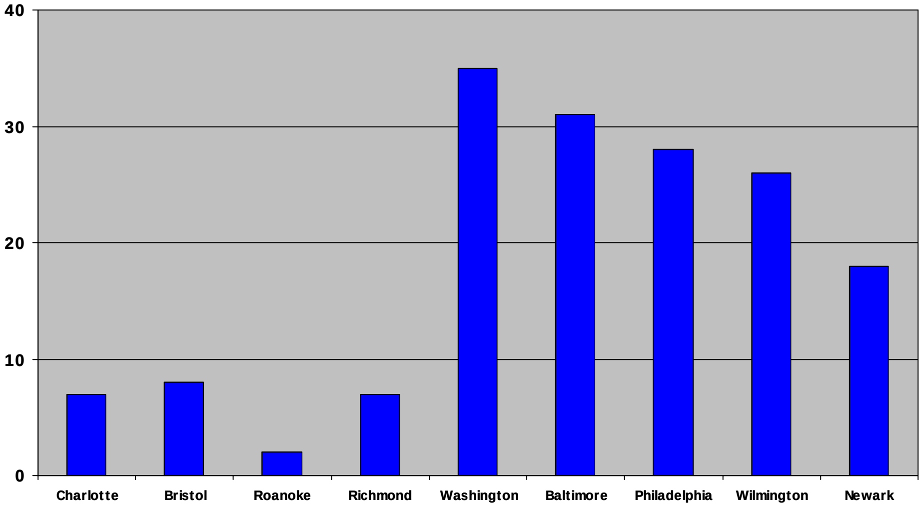
The processes leading to PM formation from SO_2 are comparatively straightforward. In the reaction with OH, SO_2 is oxidized to SO_3. When hydrated, this becomes H_2SO_4 (sulfuric acid). If ammonia gas is present, the sulfuric acid will react with it to form ammonium sulfate. Since both sulfuric acid and ammonium sulfate are strongly hygroscopic compounds, they will almost always exist in the atmosphere with a coating of water in aerosol droplets. Sulfate, therefore, exists almost exclusively in the atmosphere as PM. In the commonly used terminology, all SO_4 is referred to collectively as sulfate, whether it exists as sulfuric acid or ammonium sulfate (or other sulfate compounds).

Once fine particles are formed in the atmosphere, their small size and mass allow them to be suspended for long periods of time (days to weeks) and transported by synoptic- and meso-scale weather systems long distances from where they were originally formed. Fine dust from the Saharan Desert has been measured in the U.S., while smoke from wildfires in Central and South America and Northern Canada has also impacted areas of the U.S. It is also suspected that aerosols formed from industrial emissions in Asia travel across the Pacific to North America adding to the observed “background” aerosol concentration. Over time, depending on a number of physical factors (e.g., weather conditions, land use, etc.) fine particles deposit out of the atmosphere by both dry and wet deposition processes. Dry deposition involves settling or impaction with water bodies or other surfaces, while wet deposition includes uptake by water droplets within clouds, and subsequent rainout and washout of particles below precipitating clouds.

4.2. Regional Overview of $PM_{2.5}$

The number of days with $PM_{2.5}$ concentrations within the Unhealthy for Sensitive Groups (USG) range for the 1999–2002 CART analysis period is shown in Figure 4-1. Note that the exact number of days may be different from observed, based on our application of missing data criteria for CART, and that this chart is intended only to be used for qualitative assessment. For the four southernmost cities of Charlotte, Bristol, Roanoke, and Richmond there are a small number of USG days, ranging from 2 for Roanoke to 8 for Bristol. Keep in mind the data were only collected every third day during the analysis period for both Bristol and Roanoke, so the number of USG days for these two areas is likely somewhat higher. There is a big jump in the number of USG days as we consider the more northern sites (along the Northeast Corridor). This number drops off again further northward into New Jersey. Some missing data and different data collection start dates for the sites/areas prevent a detailed, quantitative comparison of the number of USG days, but qualitatively there seems to be a greater incidence of high $PM_{2.5}$ days in the northern part of the MARAMA study area and within the larger metropolitan areas.

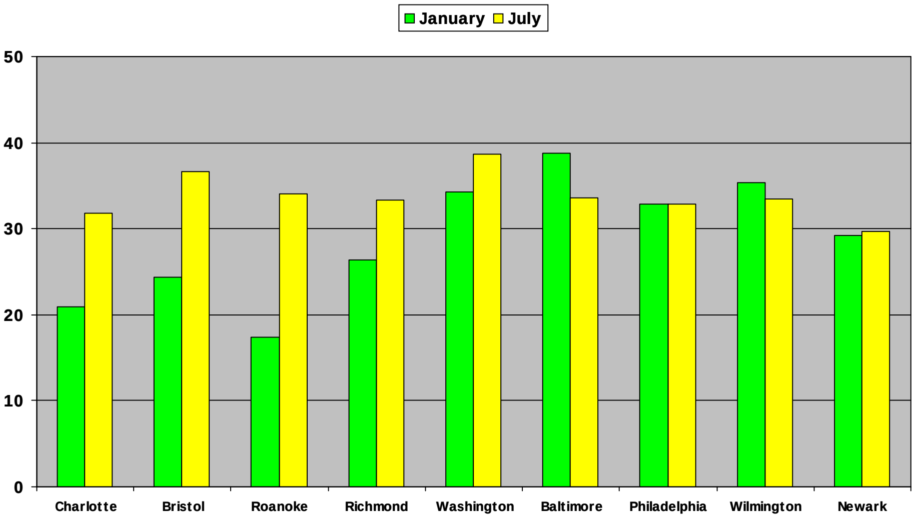
Figure 4-1. Number of USG Days During the 1999–2002 CART Analysis Period in Each of the Areas of Interest



In terms of the seasonal distribution of $PM_{2.5}$ concentrations, summer is the worst season for all nine areas and spring is almost always the best. Good concentration days are not the majority or are barely the majority for winter in Wilmington, Washington, Philadelphia, Newark, and Baltimore. Most sites see lowest PM in early spring and fall.

For the highest days (considering the 90th percentile values), the $PM_{2.5}$ concentration levels are relatively consistent throughout the region during the spring, summer and fall months, but quite different during the winter months. Figure 4-2 shows the 90th percentile values for January and July for each area of interest. For the more northern sites, the concentrations are higher during the winter months and the 90th percentile values are consistent during these two peak periods. For the more southern sites, there are larger differences between the January and July values, with the summertime values on the order of 5 to 15 μgm^{-3} higher.

Figure 4-2. 90th Percentile Daily Maximum PM2.5 Concentration (μgm-3) for the 1999–2002 CART Analysis Period for Each of the Areas of Interest: January and July



Because determination of compliance with the $PM_{2.5}$ standard relies on quarters rather than meteorologically based seasons, it is also instructive to summarize the data in terms of the distribution of high $PM_{2.5}$ or USG days by quarter. Table 4-1 presents a general summary of the observed USG days for each of the areas of interest. Included in the table are the total number of available days of valid data contained in the datasets, and the quarterly distribution of these days. Note that the number of days given in this table is generally greater than in the final CART datasets due to missing data issues and the need to remove days with missing data for the application of CART. Because some of the PM monitors in the MARAMA area weren't deployed until 2000, and because data are available only every third day for a two of the monitors, data availability during this period varies widely from region to region, with Roanoke and Bristol having the least amount of data and the Charlotte and Washington D.C. area monitors having the most.

Table 4-1. Summary of Data Availability and Number of Observed Category 3 (Unhealthy for Sensitive Groups) Days in the MARAMA Region for the Period 1999–2002

Area of Interest	Number of days with valid PM data	Number of Observed Category 3 (USG) Days	Quarter 1 (Jan-Mar)	Quarter 2 (Apr—Jun)	Quarter 3 (Jul—Sep)	Quarter 4 (Oct—Dec)
Charlotte	1453	7	1	0	5	1
Bristol	491	7	0	2	2	3
Roanoke	469	2	0	0	2	0
Richmond	1325	7	0	0	7	0
Washington	1430	35	6	6	17	6
Baltimore	1084	31	6	10	10	5
Wilmington	1401	26	6	9	11	0
Philadelphia	1313	28	7	7	12	2
Newark	971	18	1	7	7	3

As noted in the table, the largest number of observed USG days during this period in the MARAMA region occurs either during the second and third quarters of the year, encompassing the late spring and summer periods, although some USG days occurred during the fall and winter months as well in some of the areas.

Table 4-2 summarizes the correlation between all sites, considering maximum daily $PM_{2.5}$ concentrations for all areas. R-squared values greater than 0.5 are shaded to highlight the areas of agreement. Observed concentrations for Charlotte do not appear to be well correlated with those for any of the other areas. There is some correlation between Bristol and Roanoke as well as between Roanoke and Richmond, indicating some consistency in the same-day concentrations across Virginia. Again the limited datasets for Bristol and Roanoke may limit the extent to which the R-squared values represent the similarities among these areas.

There is a slightly greater degree of correlation for Richmond and Washington, D.C. and even greater correlation for the four urban areas of Washington, Baltimore, Philadelphia, and Wilmington. The highest R-squared value is for Philadelphia and Wilmington, which are nearby to one another. There is some correlation between PM levels for Newark and those for Philadelphia, Wilmington, and to a lesser degree, Baltimore. These results suggest that there is a regional component to $PM_{2.5}$ in the areas of interest from Washington (possibly Richmond) northward, but that on any given day (with a few exceptions) there are also local meteorological and/or emissions influences that affect the areas separately. Note that these values represent same-day correlations, and do not provide the basis for discerning transport.

Table 4-2. Correlations Among the Areas of Interest:
R-Squared Values Calculated Using All Daily Maximum $PM_{2.5}$ Concentrations

	Charlotte	Bristol	Roanoke	Richmond	Washington	Baltimore	Philadelphia	Wilmington	Newark
Charlotte	1.00	0.49	0.48	0.48	0.25	0.18	0.13	0.14	0.09
Bristol		1.00	0.53	0.28	0.23	0.16	0.14	0.14	0.13
Roanoke			1.00	0.53	0.45	0.36	0.28	0.32	0.23
Richmond				1.00	0.62	0.53	0.39	0.47	0.24
Washington					1.00	0.75	0.58	0.64	0.42
Baltimore						1.00	0.68	0.74	0.51
Philadelphia							1.00	0.86	0.66
Wilmington								1.00	0.58
Newark									1.00

The magnitude and distribution of $PM_{2.5}$ concentrations throughout the MARAMA region is determined in part by the prevailing meteorological conditions. Overall the location and movement of the regional-scale high- and low-pressure systems relative to an area determines the prevailing wind and dispersion conditions and thus the source-receptor relationships that characterize a $PM_{2.5}$ event, whereas the persistence and strength of the system influence/determine episode severity. A review of the meteorological conditions for days with high $PM_{2.5}$ in the areas of interest reveals that many of these days are influenced by a slow-moving or stationary high pressure system over the area of interest that results in suppressed vertical mixing of emissions/pollutants and low wind speeds or stagnation. The characteristics of high $PM_{2.5}$ events, however, vary among the areas of interest according to geographical characteristics, local and regional emissions characteristics, and the location of each area relative to other areas in combination with pollutant-transport-conducive meteorological conditions. They also vary with season. Consequently, high $PM_{2.5}$ events occur under a variety of regional- scale and local meteorological conditions and prevailing wind directions.

In the remainder of this section, we explore the $PM_{2.5}$ concentrations and meteorological conditions influencing those concentrations for each area of interest.

4.3. Factors Influencing $PM_{2.5}$ Concentrations for Charlotte, NC

The area-wide daily maximum $PM_{2.5}$ concentration, categorized into three levels of severity, serves as the "characteristic variable" for the CART analysis and the forecasted entity for the tool. The area-wide maximum $PM_{2.5}$ for the Charlotte area was defined for this study as the maximum value over all of the sites listed as the local Charlotte sites in Table 2-1.

4.3.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

The eight FRM monitors used to determine maximum PM2.5 concentrations for the Charlotte MSA come from Cabarrus, Gaston, and Mecklenburg Counties in North Carolina, and York County in South Carolina. The monitor at the Gaston site is collocated with a second monitor, which was used to fill in data missing from the first. The dataset for Charlotte is nearly complete, (all days before August, 1999, were dropped due to missing data for a previous-day $PM_{2.5}$ variable, which was more narrowly defined). Figure 4-3 shows how days of different PM severity are distributed over the seasons. In this case the winter season is defined as December through February, spring is March through May, Summer is June through August, and Fall is September through November. "Good" days have maximum $PM_{2.5}$ concentrations less than 15.5 μgm^{-3}, "moderate" days have concentrations greater than or equal to 15.5 and less than 40.5 μgm^{-3}, and "USG" days have concentrations of 40.5 μgm^{-3} or above. USG days appear predominantly in the summer; these high-PM days are less than 1 percent of the total. Most summer days are moderate, whereas concentrations are good for most of the days in the other seasons. Figure 4-4 shows the highest 90th percentile concentrations in the summer months. There are also some relatively high values in the fall (especially October) and winter months. The lowest values tend to occur during the spring.

Figure 4-3. Distribution of 1999–2002 Days by Season and Severity: Charlotte

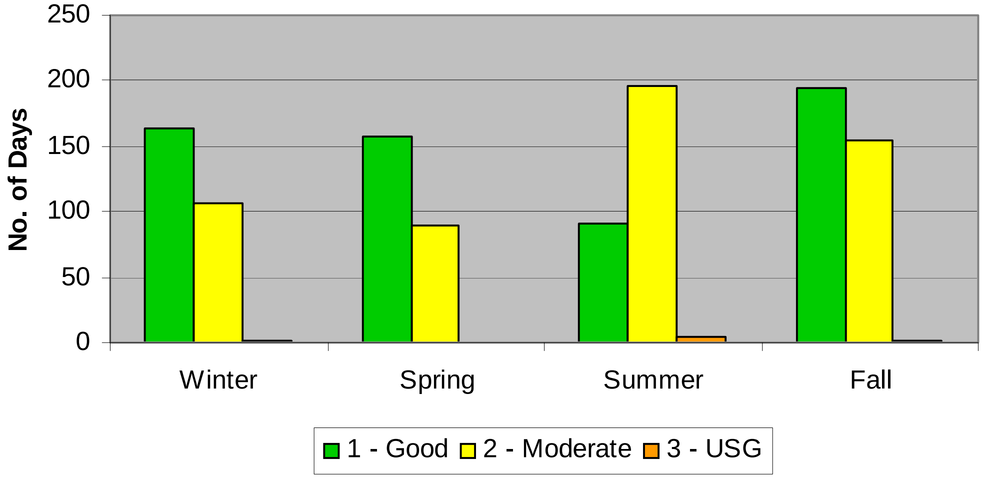
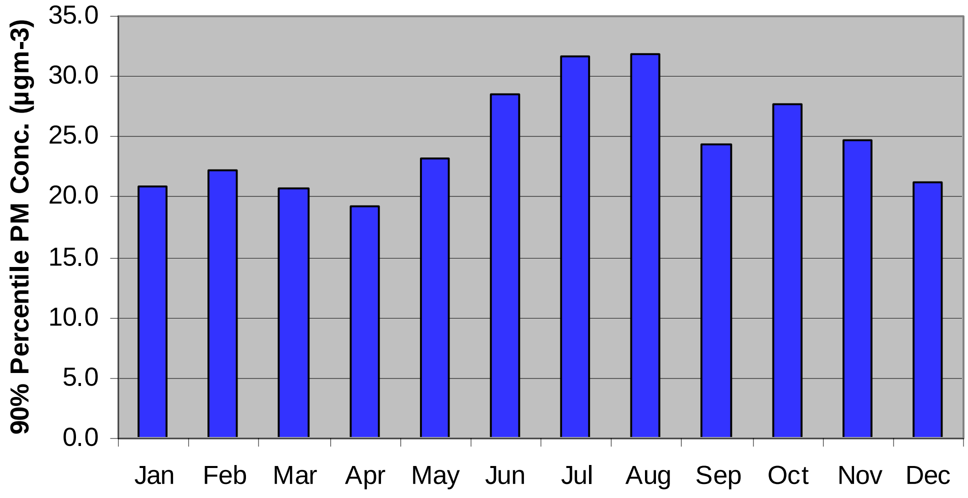


Figure 4-4. 90th Percentile Concentrations by Month (1999–2002): Charlotte



4.3.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Charlotte area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all low, moderate, and high $PM_{2.5}$ days for the Charlotte area are presented in Appendix A (Figure A-1). For consistency with the forecasting, low $PM_{2.5}$ days have maximum concentrations less than 15.5 µgm^{-3}, moderate days have concentrations greater than or equal to 15.5 and less than 40.5 µgm^{-3}, and high days have concentrations of 40.5 µgm^{-3} or above. The wind information in these plots is for the Greensboro, NC upper-air monitoring site. In these diagrams, wind direction is defined as the direction from which the wind is blowing. The length of the bar within that wind-direction sector indicates the frequency of occurrence of a particular wind direction. The shading indicates the distribution of wind speeds.

Upper-air winds for the 850 mb level (approximately 1500 m above ground) are available twice per day, at approximately 0700 and 1900 EST. Distinguishing features in the wind plots (also called wind rose diagrams) for the high $PM_{2.5}$ days, when contrasted to those with other observed concentration ranges, may help to define the wind and/or transport patterns leading to high $PM_{2.5}$.

The wind roses for Charlotte are based on the Greensboro sounding data. Upper-level winds during the low PM days for Charlotte tend to be southwesterly through northwesterly for both the morning and evening soundings. Wind directions are similar for moderate PM days, with somewhat lower wind speeds, especially at the time of the evening sounding. Wind speeds are even lower for the high PM days and there is a greater tendency for easterly wind components at the time of the morning sounding.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-3 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5-40.5, and ≥40.5 μgm^{-3}.

Table 4-3. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Charlotte

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Charlotte (μgm^{-3})	10.8	22.1	44.0
Two-days-ago 24-hour $PM_{2.5}$ for Charlotte (μgm^{-3})	13.7	16.6	23.8
Two-days-ago 24-hour $PM_{2.5}$ for Greenville-Spartanburg (μgm^{-3})	14.6	16.9	21.6
Two-days-ago 24-hour $PM_{2.5}$ for Winston-Salem (μgm^{-3})	14.2	17.0	21.3
Surface Meteorological Parameters			
Maximum surface temperature (°C)	19.4	24.2	29.8
Minimum surface temperature (°C)	8.9	11.8	16.1
Surface relative humidity (%)	66.9	67.4	65.7
Surface wind speed (ms^{-1})	2.4	1.7	1.2
Surface wind direction (degrees)	344	151	180

	Category 1	Category 2	Category 3
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.1	0.0
Upper-Air Meteorological Parameters (Greensboro)			
850 mb temperature (AM) (°C)	7.2	11.2	16.7
850 mb temperature (PM) (°C)	7.7	11.9	18.3
Temperature gradient (850 mb to surface; AM) (°C)	-2.6	-1.0	-1.9
Temperature gradient (900 mb to surface; AM) (°C)	-0.5	1.7	2.1
Temperature gradient (950 mb to surface; AM) (°C)	0.6	3.1	4.7
24-hour difference in 700 mb geopotential height (m)	-4.2	3.8	6.6
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	14.4	10.3	6.7
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	10.1	6.8	6.9
850 mb wind speed (AM) (ms^{-1})	10.9	7.3	3.4
850 mb wind speed (PM) (ms^{-1})	9.6	7.4	5.6
Yesterday’s 700 mb wind direction (PM) (degrees)	269	290	333
Yesterday’s 850 mb wind direction (PM) (degrees)	261	281	315
850 mb wind direction (AM) (degrees)	286	291	135
850 mb wind direction (PM) (degrees)	278	265	225
Estimated cloud cover (range of 1 to 3)	1.9	1.7	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	2

Table 4-3 provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Charlotte. A column-by-column comparison of the values reveals some clear tendencies in several of the air quality and meteorological parameters and

High $PM_{2.5}$ in the Charlotte area is associated with relatively high $PM_{2.5}$ two-days prior—Charlotte as well as Greenville-Spartanburg and Winston-Salem. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences), lower surface wind speeds, and less precipitation. Surface wind directions tend toward southerly for the higher ranges of $PM_{2.5}$. There is no clear tendency for relative humidity.

The upper-air meteorological parameters (based here on the Greensboro sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also a tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. This is especially true for the 900 and 950 mb temperature differences. The difference in geopotential height (defined such that a positive number indicates increasing height (pressure) over the Charlotte area) is also positively correlated with higher PM concentrations.

Lower wind speeds aloft (especially for the analysis day) and a tendency for more southerly wind directions aloft are also aligned with higher $PM_{2.5}$ concentrations.

Finally, the cloud cover and season parameters do not vary much across the three categories.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include two-days-ago $PM_{2.5}$ for Charlotte, surface temperature, 850 mb temperature, the 950 to surface temperature difference, and 850 mb wind speed at the time of the morning sounding. All of these are also well correlated with the $PM_{2.5}$ concentration for the analysis day.

4.3.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Charlotte area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-4 considers the input parameter values for the key USG bins. For Charlotte there is only one key bin and it contains four of the seven USG days.

Table 4-4. Summary of Mean Air Quality and Meteorological Parameters for the Key USG CART Classification Bin: Charlotte.

	Bin 30
Number of days	4
$PM_{2.5}$ Parameters	
24-hour $PM_{2.5}$ for Charlotte (μgm^{-3})	43.8
Two-days-ago 24-hour $PM_{2.5}$ for Charlotte (μgm^{-3})	30.9
Two-days-ago 24-hour $PM_{2.5}$ for Greenville-Spartanburg (μgm^{-3})	25.3
Two-days-ago 24-hour $PM_{2.5}$ for Winston-Salem (μgm^{-3})	26.7
Surface Meteorological Parameters	
Maximum surface temperature (°C)	35.0
Minimum surface temperature (°C)	21.3
Surface relative humidity (%)	60.2
Surface wind speed (ms^{-1})	1.3
Surface wind direction (degrees)	*
Number of six hour periods with precipitation (range is 1 to 4)	0.0

	Bin 30
Upper-Air Meteorological Parameters (Greensboro)	
850 mb temperature (AM) (°C)	18.3
850 mb temperature (PM) (°C)	20.4
Temperature gradient (850 mb to surface; AM) (°C)	-3.6
Temperature gradient (900 mb to surface; AM) (°C)	2.1
Temperature gradient (950 mb to surface; AM) (°C)	4.3
24-hour difference in 700 mb geopotential height (m)	4.3
Yesterday's 700 mb wind speed (PM) (ms^{-1})	5.3
Yesterday's 850 mb wind speed (PM) (ms^{-1})	5.6
850 mb wind speed (AM) (ms^{-1})	3.2
850 mb wind speed (PM) (ms^{-1})	5.3
Yesterday's 700 mb wind direction (PM) (degrees)	270
Yesterday's 850 mb wind direction (PM) (degrees)	225
850 mb wind direction (AM) (degrees)	180
850 mb wind direction (PM) (degrees)	198
Estimated cloud cover (range of 1 to 3)	1.8
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3

Wind directions are evenly divided between N and S.

Since there are so few USG days, the mean characteristics of days within Bin 30 match fairly closely those for Category 3, as presented in Table 4-3 above. Even higher $PM_{2.5}$ concentrations two-days-prior, slightly lower winds speeds aloft, and a more dominant southerly wind component distinguish the Bin 30 days from the other USG/Category 3 days contained in the dataset.

While table 4-4 provides an overall summary of the mean characteristics for the key high $PM_{2.5}$ bin, it is also useful to examine the conditions associated with each day or episode.

For the Charlotte area, seven USG days occurred during the 1999–2002 period. The specific dates, including the observed $PM_{2.5}$ concentration (μgm^{-3}), are presented in Table 4-5. Included in the table is information about whether these dates are also USG days for other areas within the MARAMA region. The CART classification bin is also provided, so that the reader can link the weather summaries to the bins and characteristics discussed above.

Table 4-5. USG Days for Charlotte: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	Orange Day for Other Areas
August 7, 1999	Saturday	30	41.1	
August 13, 1999	Friday	30	44.0	
January 1, 2000	Saturday	9	45.2	Washington
November 2, 2000	Thursday	28	46.9	Bristol
August 15, 2001	Wednesday	11	40.5	
July 17, 2002	Wednesday	30	45.2	
July 18, 2002	Thursday	30	44.8	Baltimore, Washington, Richmond, Newark, Philadelphia

The observed concentrations for these days generally fall in the lower end of the USG classification range. The majority of the days occur in the summer months, with two of the seven days occurring during winter.

For the August 1999 days, the Charlotte area was influenced by typical summertime surface and upper-level high-pressure systems, which affected a good portion of the Southeast. Maximum temperatures on both days approached 100°F, with mostly sunny skies and no rainfall.

On January 1, 2000, the Charlotte area was situated between a surface high-pressure system centered off the coast of Delaware and a weak low-pressure system over Mississippi. The morning lows in the area were in the upper 30's while the maximum temperatures reached the mid-60's, with light winds throughout the day and no precipitation. Since this was the first day of the new millennium, the $PM_{2.5}$ concentrations may have been influenced by early morning fireworks in the Charlotte area.

For November 2, 2000, the area was influenced by a strong upper-level ridge that affected the entire eastern seaboard. Skies were generally clear and winds were very light throughout the day, with minimum temperatures in the area in the upper 30's and maximums in the low 70's, with no precipitation.

On August 15, 2001, the weather in the Charlotte area was influenced by a weak summertime upper-level ridge and a moderately strong surface high pressure system centered over Pennsylvania. Winds were light throughout the day with minimum temperatures in the upper 60's and maximums around 90. Shallow fog conditions with 3 miles visibility were reported in the early morning hours.

For the July 17–18, 2002 period, a relatively strong upper level high was located over the southeast, with a strong surface high-pressure area over Georgia. Winds in the upper levels above Charlotte were generally very light, with a northwesterly direction. Lows during these days were near 70 with highs reaching 93 on both days. Hazy conditions were reported in the early morning on both days, with mostly sunny skies and no precipitation occurring in the area on either day. As indicated by the fact that USG days were also measured in the Baltimore, Washington, Richmond, Newark, and Philadelphia areas on July 18, the synoptic conditions causing high PM concentrations were widespread throughout the MARAMA region and persisted for several days.

This review of the meteorological conditions indicates the high PM concentrations occur under a variety of synoptic situations, but nearly all of these include high pressure over or to the north of the Charlotte area and light winds. The day-specific conditions discussed above are consistent with the categorical and CART-based average conditions for all and the subset of USG days, indicating that the CART bin captures the key characteristics of the majority of USG days and that the information contained in the categorical summaries can be used independently to guide the preparation of $PM_{2.5}$ forecasts.

4.4. Factors Influencing $PM_{2.5}$ Concentrations for Bristol, VA

The area-wide maximum $PM_{2.5}$ concentration for the Bristol area was defined for this study as the maximum value over all of the sites listed as the local Bristol sites in Table 2-1.

4.4.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

The area-wide daily maximum PM2.5 concentrations for the Bristol MSA are the daily maximums over two FRM monitors: one in Sullivan County, Tennessee, and one in the city of Bristol. A second monitor in Sullivan Co. was used as backup in the event of missing data for the first. These monitors record fine mass every three days. Two percent of the days are USG, and Figure 4-5 shows that these days only occurred in the summer and fall. Concentrations are worst in summer, which has more moderate days than good. Figure 4-6 shows the 90th percentile concentrations for each month; again, summer months have the highest value, but the November concentration follows close behind.

Figure 4-5. Distribution of 1999–2002 Days by Season and Severity: Bristol

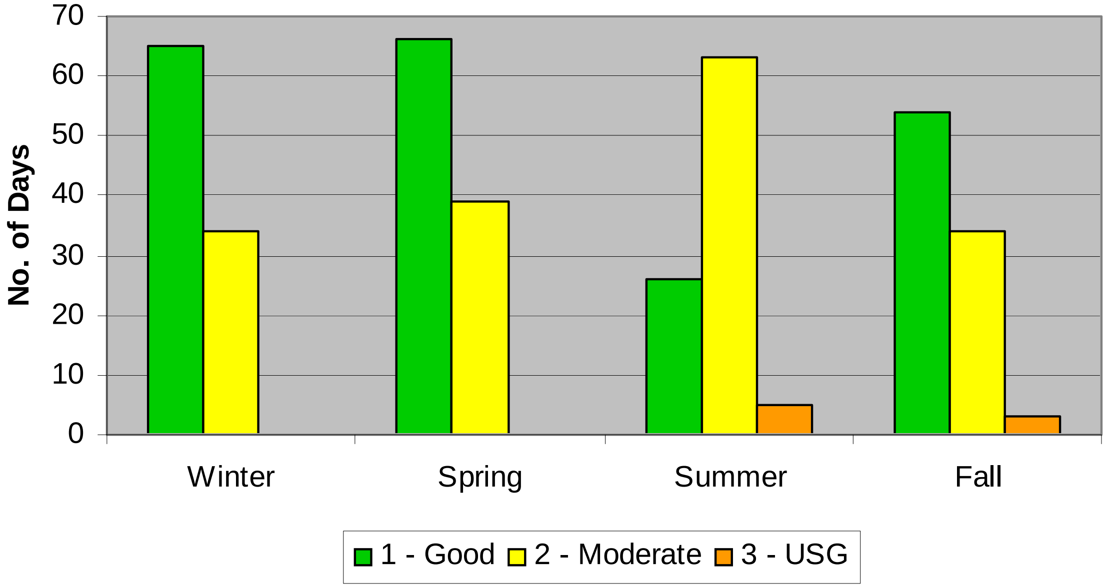
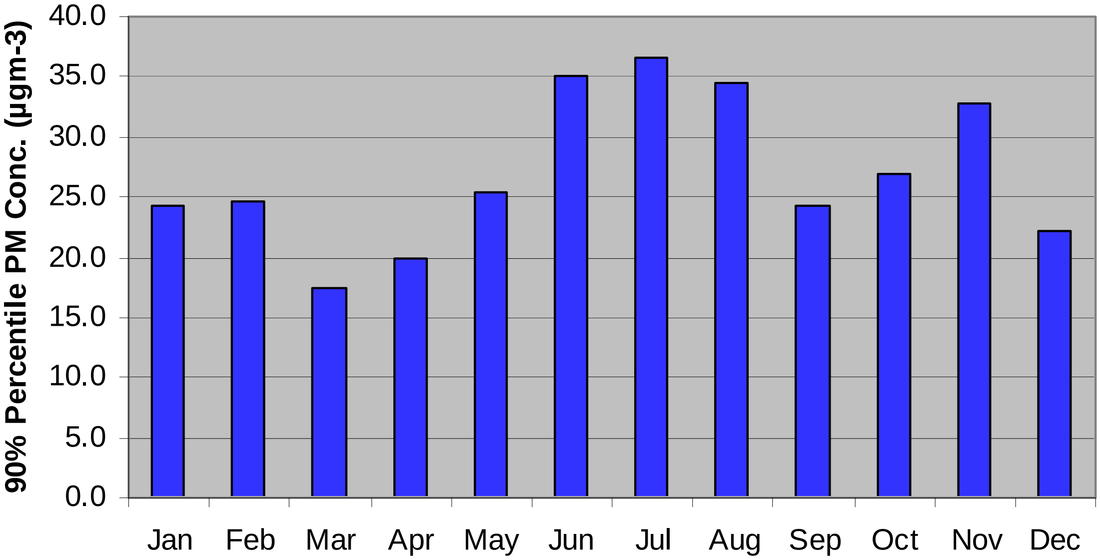


Figure 4-6. 90th Percentile Concentrations by Month (1999–2002): Bristol



4.4.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Bristol area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Bristol area are presented in Appendix A. The wind information in these plots is for the Roanoke upper-air monitoring site. The plots use the same format and contain the same information described above for the Charlotte area.

The wind roses for Bristol (Figures A-3 and A-4) are based on the Roanoke sounding data. The upper-level winds for the low PM days for Bristol tend to be westerly to northwesterly, but there are also southwesterly winds on some portion of the days. When moderate PM is observed, wind speeds are lower than for the low PM days. Compared to the low PM days, the winds are similarly directed in the morning, and there is a greater percentage of days with southwesterly winds during the evening. The highest PM days are dominated by northwesterly to northerly winds at the time of both the morning and evening sounding.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-6 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-6. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Bristol

	Category 1	Category 2	Category 3
PM$_{2.5}$ Parameters			
24-hour PM$_{2.5}$ for Bristol (μgm^{-3})	10.4	22.5	45.0
Two-days-ago 24-hour PM$_{2.5}$ for Knoxville (μgm^{-3})	16.4	21.2	29.9
Surface Meteorological Parameters			
Maximum surface temperature (°C)	16.7	22.7	26.5
Minimum surface temperature (°C)	5.4	9.7	13.1
Surface relative humidity (%)	68.7	71.4	70.9
Surface wind speed (ms^{-1})	2.0	1.1	1.0
Surface wind direction (degrees)	278	252	270
Number of six hour periods with precipitation (range is 1 to 4)	0.4	0.2	0.1
Upper-Air Meteorological Parameters (Roanoke)			
850 mb temperature (AM) (°C)	4.1	9.9	14.0
850 mb temperature (PM) (°C)	5.1	11.2	14.7
Temperature gradient (850 mb to surface; AM) (°C)	-1.2	1.3	2.6
Temperature gradient (900 mb to surface; AM) (°C)	1.1	3.1	5.4
Temperature gradient (950 mb to surface; AM) (°C)	-7.4	3.0	-5.8
24-hour difference in 700 mb geopotential height (m)	15.6	11.8	6.8
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	10.9	7.5	5.6
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	12.0	9.7	8.9
850 mb wind speed (AM) (ms^{-1})	10.1	7.5	6.4
850 mb wind speed (PM) (ms^{-1})	274	289	270
Yesterday’s 700 mb wind direction (PM) (degrees)	259	273	315
Yesterday’s 850 mb wind direction (PM) (degrees)	278	278	333
850 mb wind direction (AM) (degrees)	279	263	338
850 mb wind direction (PM) (degrees)	2.0	1.9	2.0
Estimated cloud cover (range of 1 to 3)	2	2	3
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	2

A column-by-column comparison of the values in table 4-6 reveals some clear tendencies in several of the air quality and meteorological parameters.

High $PM_{2.5}$ in the Bristol area is associated with relatively high $PM_{2.5}$ in the Knoxville area two-days prior. Thus, a regional day-to-day build up is indicated.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures, lower surface wind speeds, and less precipitation. There is no clear tendency for relative humidity or surface winds directions (which tend to be westerly, on average, for all three categories).

The upper-air meteorological parameters (based here on the Roanoke sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also a tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. There is no clear tendency for the change in geopotential height.

Lower wind speeds aloft and a tendency for more northerly wind directions aloft are also aligned with higher $PM_{2.5}$ concentrations.

Finally, the cloud cover parameter does not vary much across the three categories, and the seasonal indicator suggests that the higher PM days tend to be during the summer months.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include two-days-ago $PM_{2.5}$ for Knoxville, surface temperature, surface wind speed, 850 mb temperature, and 850 mb wind speed at the time of the previous evening sounding. As noted earlier, these tend also to show the greatest differences among the classification categories.

4.4.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Bristol area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-7 considers the input parameter values for the key USG bins. For Bristol there is only one key bin and it contains four of the seven USG days.

Table 4-7. Summary of Mean Air Quality and Meteorological Parameters for the Key USG CART Classification Bin: Bristol

	Bin 29
Number of days	6
PM$_{2.5}$ Parameters	
24-hour PM$_{2.5}$ for Bristol (µgm^{-3})	43.2
Two-days-ago 24-hour PM$_{2.5}$ for Knoxville (µgm^{-3})	34.1
Surface Meteorological Parameters	
Maximum surface temperature (°C)	26.5
Minimum surface temperature (°C)	12.5
Surface relative humidity (%)	69.0
Surface wind speed (ms^{-1})	0.9
Surface wind direction (degrees)	270
Number of six hour periods with precipitation (range is 1 to 4)	0.2
Upper-Air Meteorological Parameters (Roanoke)	
850 mb temperature (AM) (°C)	14.4
850 mb temperature (PM) (°C)	14.8
Temperature gradient (850 mb to surface; AM) (°C)	4.4
Temperature gradient (900 mb to surface; AM) (°C)	7.2
Temperature gradient (950 mb to surface; AM) (°C)	na
24-hour difference in 700 mb geopotential height (m)	-11.9
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	6.1
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	5.2
850 mb wind speed (AM) (ms^{-1})	9.8
850 mb wind speed (PM) (ms^{-1})	7.2
Yesterday’s 700 mb wind direction (PM) (degrees)	315
Yesterday’s 850 mb wind direction (PM) (degrees)	315
850 mb wind direction (AM) (degrees)	333
850 mb wind direction (PM) (degrees)	326
Estimated cloud cover (range of 1 to 3)	1.8
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2

Since there are so few USG days, the mean characteristics of days within Bin 29 match very closely those for Category 3, as presented in Table 4-6 above. Even higher PM$_{2.5}$ concentrations two-days-prior distinguish the Bin 29 days from the other USG/Category 3 days contained in the dataset. These days also tend to occur during the transitional seasons, rather than in summer.

It is also useful to examine the conditions associated with each day or episode.

For the Bristol area, seven orange days occurred during the 1999–2002 period. The specific dates, including the observed $PM_{2.5}$ concentration (µg/m^3), are listed in Table 4-8.

Table 4-8. USG Days for Bristol: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (µgm^{-3})	USG Day for Other Areas?
June 2, 2000	Friday	29	40.8	Baltimore
June 11, 2000	Sunday	29	42.2	Baltimore, Wilmington, Philadelphia, Newark
July 26, 2000	Wednesday	28	56.2	
October 24, 2000	Tuesday	29	43.0	
October 27, 2000	Friday	29	43.8	Washington, Newark
November 2, 2000	Thursday	29	43.6	Charlotte
July 18, 2001	Wednesday	29	45.7	

The observed concentrations for these days generally fall in the lower end of the USG classification range, with the exception of July 26. The USG days are distributed over the summer and fall months, with three of the eight days occurring during the fall.

For June 2, 2000, the area was influenced by a weak upper-level ridge and a surface high centered over Tennessee. Minimum temperatures were in the upper 60’s, with highs in the upper 80’s to low 90’s. Upper-level winds were very light and westerly and no precipitation was reported in the area.

For June 11, 2000, Bristol’s weather was dominated by a weak ridge aloft and a strong surface Bermuda high-pressure system centered offshore of North Carolina. Winds were light and variable on the surface throughout the day and light and southerly aloft. Shallow fog was reported in the area in the early morning hours, with lows in the upper 60’s and high near 90, and no precipitation. PM concentrations were also measured in the USG range at sites in Baltimore, Wilmington, Philadelphia, and Newark, reflecting region-wide stagnation conditions across the area.

On July 26, 2000, the southeast was under the influence of a very weak upper-level ridge system, with very light winds. A surface low-pressure system was located over the Baltimore-Washington area, but hazy skies and light winds persisted in the Bristol area. Maximum temperatures were in the mid-80’s, with minimums in the upper 60’s. No precipitation was reported in the general area on this day.

For the October 24 and 27, 2000 days, the Bristol area was under the influence of a relatively strong upper-level ridge and strong surface high-pressure system centered over the southeast. Winds aloft on these days were very light and northwesterly. Lows were in the mid-50s and highs were near 80, with shallow fog reported both mornings and no precipitation reported either day.

For November 2, 2000, the area was influenced by a strong upper-level ridge that influenced the entire eastern seaboard. Skies were generally clear and winds were very light throughout the day, with minimum temperatures in the area in the upper 30’s and maximums in the low 70’s, with no precipitation. This day was also a USG day for the Charlotte area.

For July 18, 2001, Bristol's weather was influenced by a strong upper-level ridge centered over the mid-plains, and a moderately strong surface high-pressure system over Georgia. Lows were near 70 and highs approached 90 throughout the area. Winds aloft were very light and northwesterly. Hazy conditions and limited visibility were reported during the morning hours and precipitation was reported in the Roanoke area, northeast of Bristol.

This review of the meteorological conditions indicates the high PM concentration occur under a variety of synoptic situations, but nearly all of these manifest themselves as stagnation conditions near the surface. This is consistent with the very light wind speeds indicated by the categorical and CART-based average conditions for all and the subset of USG days. CART finds this parameter to be important and thus appears to capture the key characteristics of the majority of USG days. This consistency also suggests that the categorical summaries for Bristol can be used independently to guide the preparation of $PM_{2.5}$ forecasts.

4.5. Factors Influencing $PM_{2.5}$ Concentrations for Roanoke, VA

The area-wide maximum $PM_{2.5}$ for the Roanoke area was defined for this study as the maximum value over all of the sites listed as the local Roanoke sites in Table 2-1.

4.5.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

The 436 days for the Roanoke daily maximum PM concentrations come from the maximum of two FRM monitors, one in the city of Roanoke and the other in the city of Salem, Virginia. Half a percent of these days are USG, and these all occur in the summer, as Figure 4-7 shows. Most summer days are moderate and most days in the other seasons are good; the profile of monthly 90th percentile concentrations shown in Figure 4-8 peaks relatively gently in July, with a minor peak in February.

Figure 4-7. Distribution of 1999–2002 Days by Season and Severity: Roanoke

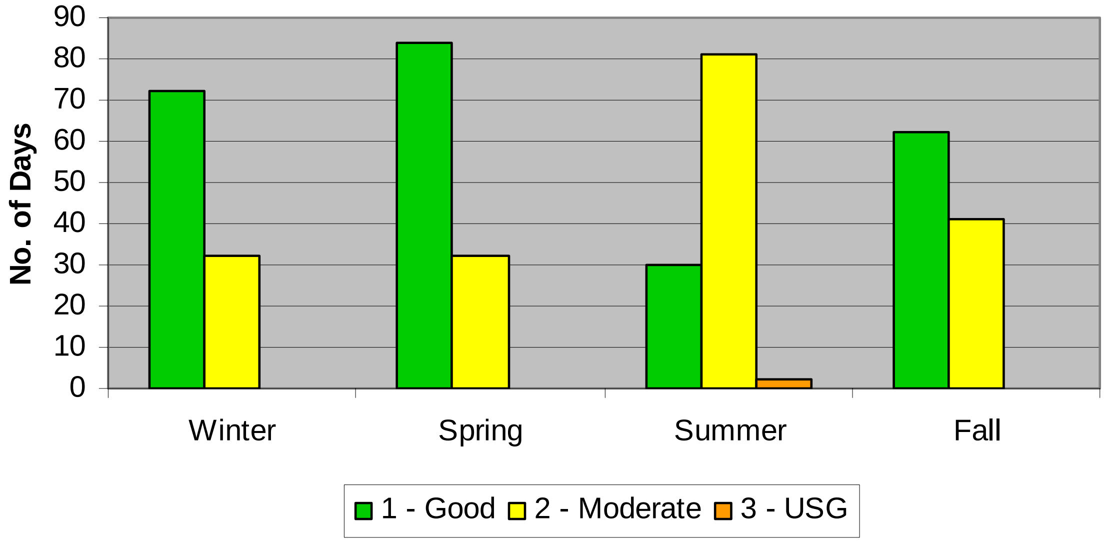
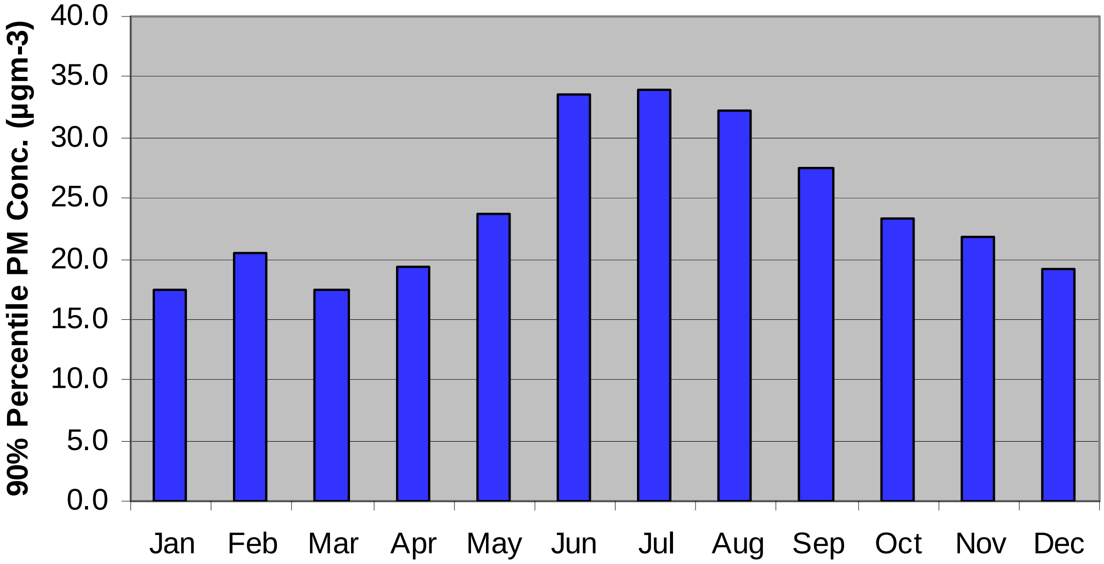


Figure 4-8. 90th Percentile Concentrations by Month (1999–2002): Roanoke



4.5.2. *Meteorological Factors Influencing $PM_{2.5}$ Concentrations*

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Roanoke area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Roanoke area are presented in Appendix A. The wind information in these plots is for the Roanoke upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Roanoke (Figures A-5 and A-6) are based on the Roanoke sounding data. The upper-level winds are predominately westerly to northwesterly for the low PM days. There is a notable increase in the incidence of southwesterly winds for the moderate PM days. At the time of the evening sounding, southwesterly winds dominate the wind rose. Since there are only two high PM days for Roanoke, wind roses were not prepared for this concentration level.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-9 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-9. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Roanoke

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Roanoke (μgm^{-3})	10.1	22.7	46.8
Two-days-ago 24-hour $PM_{2.5}$ for Winston-Salem (μgm^{-3})	13.7	18.2	20.9
Surface Meteorological Parameters			
Maximum surface temperature (°C)	16.4	23.9	33.6
Minimum surface temperature (°C)	6.2	12.4	21.4
Surface relative humidity (%)	59.3	64.9	62.2
Surface wind speed (ms^{-1})	3.0	1.8	1.4
Surface wind direction (degrees)	277	230	315
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.0
Upper-Air Meteorological Parameters (Roanoke)			
850 mb temperature (AM) (°C)	4.0	11.4	18.6
850 mb temperature (PM) (°C)	5.0	12.8	20.0
Temperature gradient (850 mb to surface; AM) (°C)	-0.9	1.7	0.4
Temperature gradient (900 mb to surface; AM) (°C)	1.1	3.7	4.0
Temperature gradient (950 mb to surface; AM) (°C)	na	na	na
24-hour difference in 700 mb geopotential height (m)	-3.7	3.7	-11.5
Yesterday's 700 mb wind speed (PM) (ms^{-1})	15.8	10.1	4.4
Yesterday's 850 mb wind speed (PM) (ms^{-1})	10.9	6.6	4.1
850 mb wind speed (AM) (ms^{-1})	12.2	9.0	10.3
850 mb wind speed (PM) (ms^{-1})	10.0	7.3	3.1
Yesterday's 700 mb wind direction (PM) (degrees)	278	287	0
Yesterday's 850 mb wind direction (PM) (degrees)	270	269	0
850 mb wind direction (AM) (degrees)	284	275	0
850 mb wind direction (PM) (degrees)	277	263	270
Estimated cloud cover (range of 1 to 3)	2.0	1.8	1.0
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	3

Table 4-9 provides an overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Roanoke. A column-by-column comparison of the values reveals some clear tendencies in several of the air quality and meteorological parameters.

High $PM_{2.5}$ in the Roanoke area is associated with relatively high $PM_{2.5}$ two-days prior—in Roanoke and to a lesser extent Winston-Salem. Thus, the regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days, with emphasis on a local build up or recirculation.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures, lower surface wind speeds, and less precipitation. Surface wind directions tend toward northwesterly for the higher ranges of $PM_{2.5}$. There is no clear tendency for relative humidity.

The upper-air meteorological parameters (based here on the Roanoke sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also a tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. There is no clear tendency for the difference in geopotential height.

Lower wind speeds aloft (with the exception of the 850 mb winds for the morning of the analysis day) are aligned with higher $PM_{2.5}$ concentrations. Wind directions veer from westerly to northerly with the higher PM values.

Finally, cloud cover is less for the high PM days, and the season index indicates that the highest concentrations tend to occur during the summer months.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. For Roanoke, the most important parameters are surface temperature and 850 mb temperature. Wind speeds aloft are next most important. All of these are also very well correlated with the $PM_{2.5}$ concentration for the analysis day.

4.5.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Roanoke area. Since there are only two high $PM_{2.5}$ days in the dataset for Roanoke, we did not prepare a separate table of the characteristics of the USG bins for this area.

Only two USG days occurred during the 1999–2002 period in the Roanoke area, although as noted above, the available data are limited for this site. The specific dates, including the observed $PM_{2.5}$ concentration (µg/m^3), are as follows:

Table 4-10. USG Days for Roanoke: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (µgm^{-3})	USG Day for Other Areas?
August 8, 2001	Wednesday	32	52.8	Baltimore, Washington, Richmond
July 1, 2002	Monday	33	40.7	Baltimore, Washington (7/2)

Both of these USG days occur in the summer with observed concentrations for August 8, 2001 in the middle of the USG range, while the concentration for July 1, 2002 is just within the USG category.

On August 8, 2001, the Roanoke area was influenced by a broad upper-level ridge centered over the central U.S., and a weak surface high pressure system over Georgia. Surface winds throughout the day were calm with upper-level winds very light and southerly. The low for Roanoke was 74°F while the high for this day was 93, with no precipitation reported. Similar conditions persisted throughout the region leading to high PM concentrations in the Baltimore, Washington, and Richmond areas on this day.

For July 1, 2002, the Roanoke weather was dominated by a strong upper-level ridge centered over the Midwest, and a strong, broad surface high pressure system centered directly over the Roanoke area. Upper-level winds on this day were very light and variable, while surface winds were light and variable. The low for Roanoke was 68, while the high for the day was 88. This day was also a USG day for the Baltimore and Washington areas and was the start of the multi-day PM episode across the MARAMA region which lasted through July 4th, as the upper-level ridge built further over the area, strengthening the persistent surface high. USG days occurred in the Baltimore, Washington, Philadelphia, and Newark areas on July 2 and 3, and in Washington on July 4.

This review of the meteorological conditions indicates the high PM concentration occurs in conjunction with surface high pressure and light winds, allowing for the multi-day build up of particulates in the area. This is consistent with the very light wind speeds indicated by the categorical averages. CART appears to capture the effects of the high pressure using the 850 mb temperature as a key parameter in distinguishing the high PM days. There are really not enough high PM days for Roanoke to say much more about the characteristics of the high PM days.

4.6. Factors Influencing $PM_{2.5}$ Concentrations for Richmond, VA

The area-wide maximum $PM_{2.5}$ for the Richmond area was defined for this study as the maximum value over all of the sites listed as the local Richmond sites in Table 2-1.

4.6.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

Maximum PM2.5 concentrations over five FRM monitors in Charles City, Richmond City, Chesterfield County, and Henrico County determined the area-wide maximum for Richmond. Of the days with available data from the 1999–2002 period, about half a percent had USG concentrations, and all of these occurred in summer, as shown in Figure 4-9. The majority of summer days were moderate, whereas good days dominated the other seasons. In Figure 4-10, which shows the 90th percentile concentrations, one sees highest concentrations in the summer months, and the next highest in January. Richmond is characterized by a more distinct annual profile than many of the other areas included in the analysis.

Figure 4-9. Distribution of 1999–2002 Days by Season and Severity: Richmond

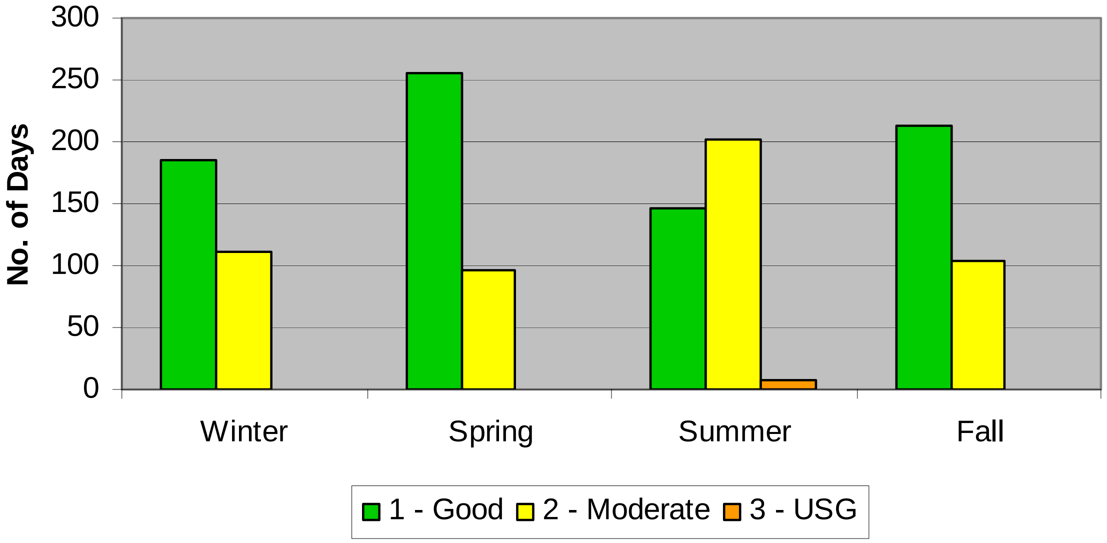
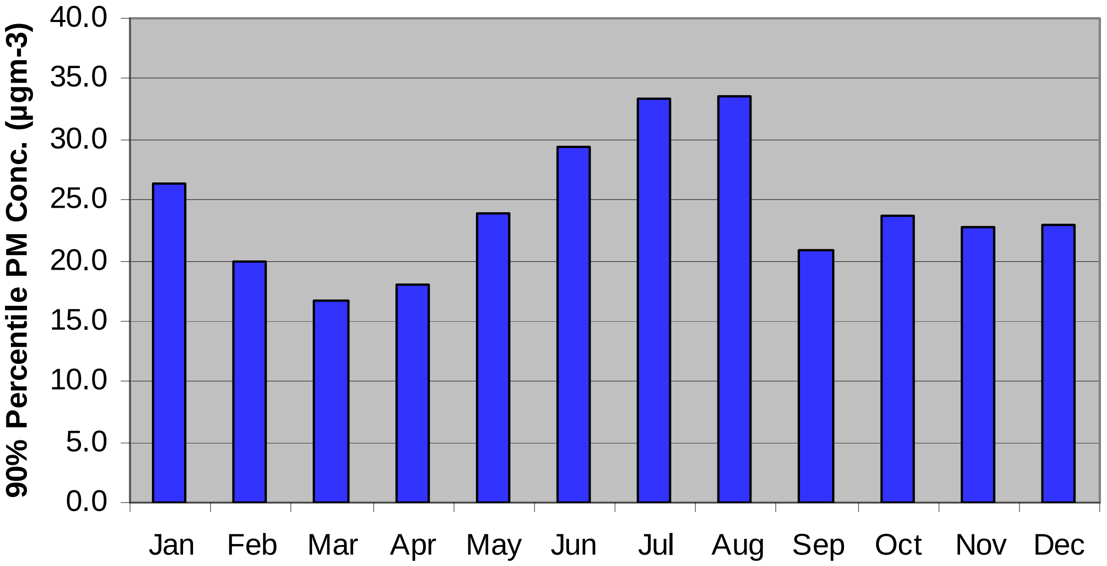


Figure 4-10. 90th Percentile Concentrations by Month (1999–2002): Richmond



4.6.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Richmond area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Richmond area are presented in Appendix A. The wind information in these plots is for the Dulles Airport (Sterling, VA) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Richmond (Figures A-7 and A-8) are based on the Dulles Airport sounding data. The upper-level winds are predominately southwesterly to northerly for the low PM days, for both the morning and evening sounding. For many of the days, the directions fall within the westerly to northwesterly portion of this range. There is a notable increase in the incidence of southwesterly winds for the moderate PM days; wind speeds are also lower for the moderate days. At the time of the evening sounding, southwesterly winds dominate the wind rose. For the highest PM days, winds have either a northerly or southerly component. Given the small number of days, a wind pattern does not emerge. Wind speeds are much lower than for the other PM concentration levels.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-11 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-11. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Richmond

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	10.1	22.1	46.1
Two-days-ago 24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	13.5	16.2	25.3
Two-days-ago 24-hour $PM_{2.5}$ for Washington, D.C. (μgm^{-3})	15.0	18.1	31.2
Two-days-ago 24-hour $PM_{2.5}$ for Winston-Salem (μgm^{-3})	14.5	17.6	27.5
Surface Meteorological Parameters			
Maximum surface temperature (°C)	18.7	23.8	35.8
Minimum surface temperature (°C)	8.2	12.0	23.2
Surface relative humidity (%)	67.2	70.4	63.6
Surface wind speed (ms^{-1})	3.1	2.3	1.8
Surface wind direction (degrees)	318	188	171
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.1

	Category 1	Category 2	Category 3
Upper-Air Meteorological Parameters (Dulles Airport)			
850 mb temperature (AM) (°C)	4.5	9.9	18.8
850 mb temperature (PM) (°C)	5.3	10.6	19.6
Temperature gradient (850 mb to surface; AM) (°C)	-2.9	-1.0	-3.0
Temperature gradient (900 mb to surface; AM) (°C)	-0.5	2.1	2.1
Temperature gradient (950 mb to surface; AM) (°C)	0.2	3.1	3.6
24-hour difference in 700 mb geopotential height (m)	-3.9	2.6	-18.4
Yesterday's 700 mb wind speed (PM) (ms^{-1})	11.3	8.3	4.5
Yesterday's 850 mb wind speed (PM) (ms^{-1})	10.6	8.7	5.1
850 mb wind speed (AM) (ms^{-1})	14.6	11.7	5.2
850 mb wind speed (PM) (ms^{-1})	10.9	8.2	6.7
Yesterday's 700 mb wind direction (PM) (degrees)	293	280	315
Yesterday's 850 mb wind direction (PM) (degrees)	283	269	270
850 mb wind direction (AM) (degrees)	281	286	315
850 mb wind direction (PM) (degrees)	274	281	270
Estimated cloud cover (range of 1 to 3)	1.9	1.8	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	3

Table 4-11 shows how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Richmond. A column-by-column comparison of the values reveals some clear tendencies in several of the air quality and meteorological parameters.

High $PM_{2.5}$ in the Richmond area is associated with relatively high $PM_{2.5}$ two-days prior—in Richmond, Washington, D.C., and Winston-Salem. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high PM days.

The surface meteorological parameters indicate that higher $PM_{2.5}$ concentrations occur with higher temperatures (primarily reflecting seasonal differences), lower surface wind speeds, lower relative humidity, and less precipitation. Surface wind directions are northwesterly, on average for the low PM days, and tend toward southerly for the higher ranges of $PM_{2.5}$.

The upper-air meteorological parameters (based here on the Dulles Airport sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures and somewhat greater stability (although the relationships between PM and stability is less well defined than for some of the other areas, possibly due to distance and location of the upper-air monitoring site). There is no clear trend in the difference in geopotential height parameter.

There is a very clear tendency for lower wind speeds aloft (for both the day prior to the analysis day and the analysis day) but little difference in wind directions aloft among the categories.

The cloud cover parameters do not vary much, and the seasonal index show that most of the USG days occur during summer.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include two-days-ago $PM_{2.5}$ for Winston-Salem, surface temperature, and 850 mb temperature. Surface wind speed is also somewhat important. The upper-level wind speeds appear to vary directly with PM, but are of lesser importance in the construction of the CART tree.

4.6.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Richmond area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-12 considers the input parameter values for the key USG bins. For Richmond there is only one key bin and it contains all of the seven USG days.

Table 4-12. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Richmond

	Bin 27
Number of days	7
$PM_{2.5}$ Parameters	
24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	46.1
Two-days-ago 24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	25.3
Two-days-ago 24-hour $PM_{2.5}$ for Washington, D.C. (μgm^{-3})	31.2
Two-days-ago 24-hour $PM_{2.5}$ for Winston-Salem (μgm^{-3})	27.5
Surface Meteorological Parameters	
Maximum surface temperature (°C)	35.8
Minimum surface temperature (°C)	23.2
Surface relative humidity (%)	63.6
Surface wind speed (ms^{-1})	1.8
Surface wind direction (degrees)	171

	Bin 27
Number of six hour periods with precipitation (range is 1 to 4)	0.1
Upper-Air Meteorological Parameters (Dulles Airport)	
850 mb temperature (AM) (°C)	18.8
850 mb temperature (PM) (°C)	19.6
Temperature gradient (850 mb to surface; AM) (°C)	-3.0
Temperature gradient (900 mb to surface; AM) (°C)	2.1
Temperature gradient (950 mb to surface; AM) (°C)	3.6
24-hour difference in 700 mb geopotential height (m)	-18.4
Yesterday's 700 mb wind speed (PM) (ms^{-1})	4.5
Yesterday's 850 mb wind speed (PM) (ms^{-1})	5.1
850 mb wind speed (AM) (ms^{-1})	5.2
850 mb wind speed (PM) (ms^{-1})	6.7
Yesterday's 700 mb wind direction (PM) (degrees)	315
Yesterday's 850 mb wind direction (PM) (degrees)	270
850 mb wind direction (AM) (degrees)	315
850 mb wind direction (PM) (degrees)	270
Estimated cloud cover (range of 1 to 3)	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3

Since the seven USG days are all contained in Bin 27, the characteristics of this bin are identical to those for the Category 3 days, as discussed above

Next we examine the conditions associated with each high PM day.

Data retrieval for the Richmond area was high for the period 1999–2002, and only seven USG days occurred during this period. The specific dates, including the observed $PM_{2.5}$ concentration (μgm^{-3}), are listed in Table 4-13.

Table 4-13. USG days for Richmond: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
July 6, 1999	Tuesday	27	48.5	
August 8, 2001	Wednesday	27	51.5	Roanoke, Baltimore, Washington
August 9, 2001	Thursday	27	41.5	Baltimore, Washington, Philadelphia, Wilmington
July 2, 2002	Tuesday	27	41.6	Baltimore, Washington, Philadelphia, Newark
July 3, 2002	Wednesday	27	50.5	Baltimore, Washington, Philadelphia, Newark
July 18, 2002	Thursday	27	46.2	Baltimore, Washington, Charlotte, Newark
August 13, 2002	Tuesday	27	42.2	Baltimore, Washington, Wilmington, Philadelphia

All of the USG days for the Richmond area occurred during the summer months, with observed concentrations in the middle of the range for the USG category. Except for the July 6, 1999 period, the meteorological conditions in the MARAMA region were widespread and persistent enough to cause high PM concentrations throughout the entire domain on the USG days measured in the Richmond area.

For July 6, 1999, the weather in the Richmond area was influenced by a broad, relatively flat upper-level ridge, and by a surface high-pressure system centered over Mississippi. Winds aloft were weak and southwesterly while surface winds were light and variable. The low temperature for the day at Richmond was 75, while the high was 98. Hazy skies and limited visibility were reported in the early morning hours and no precipitation occurred in the area on this day.

For the August 8–9, 2001 period, the Richmond area and the State of Virginia were influenced by a broad upper-level ridge centered over the central U.S., and a weak surface high-pressure system over Georgia. Surface winds throughout the day were calm with upper-level winds very light and southerly. The low temperatures on these two days in Richmond were in the low 70’s, while the highs were in the upper 90’s, with no precipitation reported in the area on either of these days. August 8, 2001 was also a USG day for the Roanoke area, so conditions conducive to the buildup of PM were pervasive across the state.

As noted above, the July 1–4, 2002 period exhibited high PM conducive conditions throughout the MARAMA region, with USG days measured from the Richmond area and at all sites north during this multi-day episode. Conditions are discussed in an earlier section.

The July 18–19, 2002 period exhibited severe, PM conducive conditions during which USG days were measured throughout the MARAMA region. The region was under the influence of a broad summertime upper-level ridging pattern that was transitioning to weak zonal flow. A surface high-pressure system centered over Georgia resulted in stagnant winds, high temperatures, and mostly clear, hazy skies throughout the region. High temperatures were in the mid-90s, while lows were measured in the low 70’s. Hazy skies and limited visibility were reported in the Richmond area during this period.

This review of the meteorological conditions indicates the high PM concentration occur when the region is under the influence of a high-pressure system; conditions near the surface are characterized by high temperatures and low wind speeds. These conditions are consistent with categorical and CART-based average conditions for the USG days. CART primarily uses surface temperature, 850 mb temperature, and surface wind speed to represent these conditions. CART also picks up on the regional-scale build up or PM as a precursor of USG days.

4.7. Factors Influencing $PM_{2.5}$ Concentrations for Washington, D.C.

The area-wide maximum $PM_{2.5}$ for the Washington area was defined for this study as the maximum value over all of the sites listed as the local Washington sites in Table 2-1.

4.7.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

Eleven FRM monitors, plus two additional monitors used to fill in missing data points for their respective collocated monitors, determined the area maximum for Washington DC. Of the days examined, 2.5% are USG, and these are spread over all seasons, with half occurring in summer, winter and fall each taking about a quarter, and one lone high PM day appearing in the spring. Figure 4-11 visualizes this distribution, and also shows closely matched quantities of good and moderate days in the winter, a prevalence of good days in the spring, mostly moderate days in the summer, and mostly good days in the fall. The profile of 90th percentile concentrations shown in Figure 4-12 is triple-peaked as for some of the other areas, with the highest values in June and July, followed by January, August, and October, and the lowest values in March and September.

Figure 4-11. Distribution of 1999–2002 Days by Season and Severity: Washington

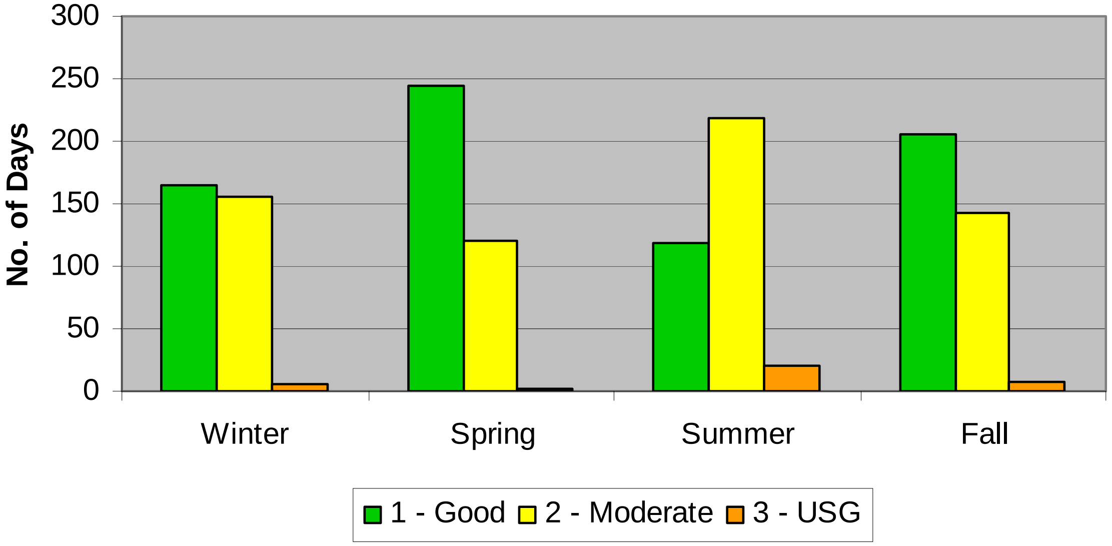
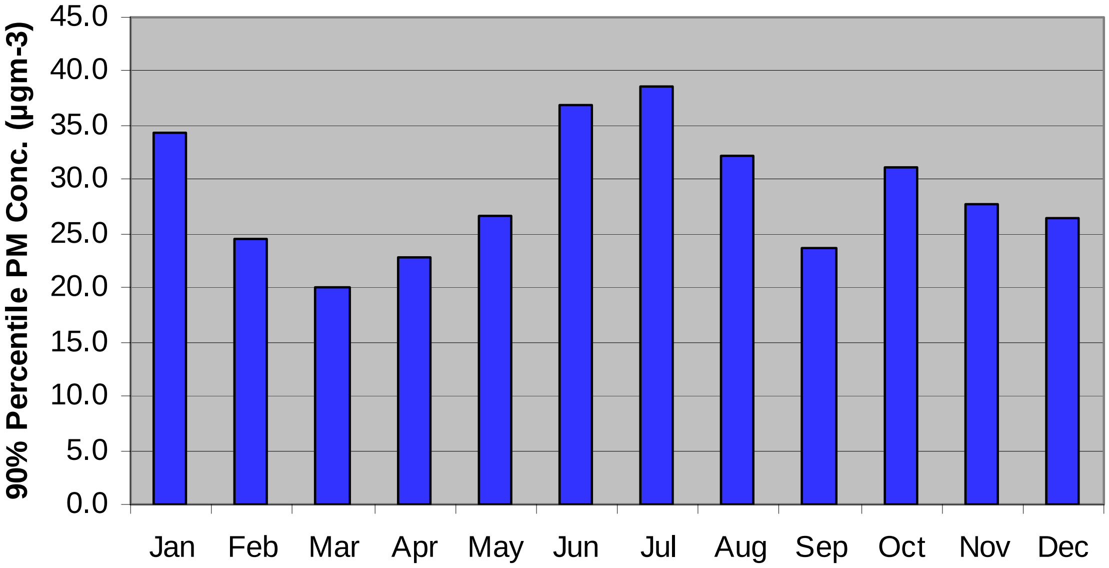


Figure 4-12. 90th Percentile Concentrations by Month (1999–2002): Washington



4.7.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Washington, D.C. area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Washington area are presented in Appendix A. The wind information in these plots is for the Dulles Airport (Sterling, VA) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Washington (Figures A-9 and A-10) are based on the Dulles Airport sounding data. The upper-level winds are predominately southwesterly to northwesterly for the low PM days, at the time of the morning sounding and west-southwesterly to northerly at the time of the evening sounding. For both sounding times, wind back slightly for the moderate PM days, with a shift to dominant southwesterly winds in the morning and westerly winds in the evening. For the highest PM days, wind speeds are much lower than for the other PM concentration levels and the wind directions are southwesterly, westerly, and northwesterly on the various days.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-14 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour PM$_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 µgm^{-3}.

Table 4-14. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Washington, D.C.

	Category 1	Category 2	Category 3
PM$_{2.5}$ Parameters			
24-hour PM$_{2.5}$ for Washington (µgm^{-3})	10.5	23.0	48.1
Two-days-ago maximum 24-hour PM$_{2.5}$ for Washington, Baltimore, and Gettysburg (µgm^{-3})	16.1	19.3	26.5
Two-days-ago 24-hour PM$_{2.5}$ for Richmond (µgm^{-3})	13.0	15.7	20.9
Surface Meteorological Parameters			
Maximum surface temperature (°C)	17.1	21.8	26.2
Minimum surface temperature (°C)	8.5	12.4	16.8
Surface relative humidity (%)	61.4	68.8	67.6
Surface wind speed (ms^{-1})	3.7	2.6	2.1
Surface wind direction (degrees)	308	235	249
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.1
Upper-Air Meteorological Parameters (Dulles Airport)			
850 mb temperature (AM) (°C)	3.7	9.1	13.9
850 mb temperature (PM) (°C)	4.3	9.9	15.4
Temperature gradient (850 mb to surface; AM) (°C)	-3.3	-0.7	-1.8
Temperature gradient (900 mb to surface; AM) (°C)	-1.1	2.2	2.4
	-0.3	3.0	3.9
24-hour difference in 700 mb geopotential height (m)	-3.0	1.0	0.1
Yesterday's 700 mb wind speed (PM) (ms^{-1})	15.2	11.9	7.8
Yesterday's 850 mb wind speed (PM) (ms^{-1})	11.5	8.4	6.1
850 mb wind speed (AM) (ms^{-1})	11.7	8.9	5.9
850 mb wind speed (PM) (ms^{-1})	10.6	9.2	7.4
Yesterday's 700 mb wind direction (PM) (degrees)	285	282	297
Yesterday's 850 mb wind direction (PM) (degrees)	281	276	273
850 mb wind direction (AM) (degrees)	301	277	283
850 mb wind direction (PM) (degrees)	290	266	288
Estimated cloud cover (range of 1 to 3)	1.8	1.9	1.6
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	2

Table 4-14 summarizes how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Charlotte. A column-by-column comparison of the values reveals some possible relationships between $PM_{2.5}$ and several of the air quality and meteorological parameters.

High $PM_{2.5}$ in the Washington area is associated with relatively high $PM_{2.5}$ two-days prior—both in the Washington area and in Richmond. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days. Note, however, that neither of the prior-day PM parameters are of high importance to CART.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences), lower surface wind speeds, and less precipitation. Surface wind directions tend toward southwesterly for the higher ranges of $PM_{2.5}$, compared to northwesterly for the lowest range. Relative humidity is, on average, slightly higher with higher PM.

The upper-air meteorological parameters (based here on the Dulles Airport sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also a tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. This is especially true for the 900 and 950 mb temperature differences. The difference in geopotential height does not vary regularly across the categories.

Considering the upper-air wind parameters, lower wind speeds aloft characterize the higher PM days. There is no well defined tendency with regard to wind direction aloft, and, on average, westerly winds prevail.

Finally, the cloud cover is less for higher PM, but there and the season parameter does not vary across the three categories.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include surface temperature, relative humidity, 850 mb temperature, relative humidity, and 950 to surface temperature difference. All of these are also well correlated (directionally) with the $PM_{2.5}$ concentration for the analysis day.

4.7.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Washington, D.C. area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-15 considers the input parameter values for the key USG bins. For Washington, four bins contain 66 percent of the USG days.

Table 4-15. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Washington, D.C.

	Bin 34	Bin 11	Bin 19	Bin 37
Number of days	12	4	4	3
$PM_{2.5}$ Parameters				
24-hour $PM_{2.5}$ for Washington (μgm^{-3})	48.1	48.3	50.7	47.4
Two-days-agomaximum 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (μgm^{-3})	28.0	14.9	29.7	38.1
Two-days-ago 24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	21.9	13.6	21.6	20.4
Surface Meteorological Parameters				
Maximum surface temperature (°C)	33.3	5.0	27.5	33.7
Minimum surface temperature (°C)	23.2	-1.8	18.5	23.7
Surface relative humidity (%)	61.1	77.2	75.6	66.8
Surface wind speed (ms^{-1})	2.7	1.2	2.1	3.3
Surface wind direction (degrees)	259	45	225	225
Number of six hour periods with precipitation (range is 1 to 4)	0.0	0.3	0.0	0.3
Upper-Air Meteorological Parameters (Dulles Airport)				
850 mb temperature (AM) (°C)	17.2	-1.0	14.7	17.3
850 mb temperature (PM) (°C)	19.1	2.9	14.7	18.8
Temperature gradient (850 mb to surface; AM) (°C)	-2.5	3.4	-3.7	-5.4
Temperature gradient (900 mb to surface; AM) (°C)	2.1	3.7	2.2	-1.4
Temperature gradient (950 mb to surface; AM) (°C)	4.0	3.2	3.7	0.3
24-hour difference in 700 mb geopotential height (m)	-0.5	-5.8	-3.2	-26.8
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	7.9	10.9	3.5	8.1
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	6.0	6.4	5.5	5.5
850 mb wind speed (AM) (ms^{-1})	5.1	7.0	5.3	12.1
850 mb wind speed (PM) (ms^{-1})	6.7	10.8	8.0	6.5
Yesterday’s 700 mb wind direction (PM) (degrees)	315	270	315	243
Yesterday’s 850 mb wind direction (PM) (degrees)	301	288	225	207
850 mb wind direction (AM) (degrees)	306	243	243	315
850 mb wind direction (PM) (degrees)	286	270	270	333
Estimated cloud cover (range of 1 to 3)	1.5	1.5	2.0	2.0
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3	1	2	3

While many of the characteristics are similar for the exceedance bins, there are some differences. These provide insight into the factors influencing the high PM days within each bin.

Days within Bins 34 and 19 have similar values for previous day PM concentration, whereas Bin 11 days have much lower values, on average, and Bin 37 days have much higher values, on average, in the local Washington area than days within the other bins. Thus these bins are characterized by regional-scale build up of PM (Bins 34 and 19), rapid build up of PM (Bin 11), and persistent high values in the local area (Bin 37). From the temperatures, as well as from the seasonal index, the bins represent different times of the year—with Bin 11 for winter days, Bin 19 for transitional season days, and Bins 34 and 37 for summer days.

In addition to the lowest temperatures and prior-day PM values, the days within Bin 11 are characterized by the lowest surface wind speeds and the deepest stable layers. Surface wind directions from the northeast are also unique to this bin.

Bin 19 is comprised primarily of transitional season days and these days have the second lowest wind speeds, on average, but otherwise conditions that tend to be intermediate to the other bins.

The two bins comprised mostly of summer days have slightly higher surface wind speeds, on average, and lower relative humidity, than the other two bins. They differ from one another in the stability characteristics such that days within Bin 37 are much less stable. Days within this bin also show decreasing heights and high wind speeds aloft during the morning hours, compared to days within Bin 34. Thus, there appear to be two different summertime regimes with different synoptic characteristics.

Next, we examine the conditions associated with each day or episode.

Data retrieval and availability for the Washington area were high for the period 1999–2002, and thirty five orange days occurred during this period. Of all the areas of interest in the MARAMA region, the Washington area experienced the largest number of USG days during this period. The specific dates, including the observed $PM_{2.5}$ concentration (μgm^{-3}), are listed in Table 4-16.

Table 4-16. USG days for Washington, D.C.: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
July 22, 1999	Thursday	37	56.3	
September 27, 1999	Monday	19	67.0	
January 1, 2000	Saturday	26	46.0	Charlotte
June 10, 2000	Saturday	34	42.1	Baltimore, Philadelphia, Newark
June 23, 2000	Friday	26	47.7	
July 1, 2000	Saturday	23	47.0	
July 9, 2000	Sunday	34	41.2	
October 16, 2000	Monday	12	40.5	
October 26, 2000	Thursday	17	50.3	Baltimore
October 27, 2000	Friday	19	44.3	
November 8, 2000	Wednesday	12	47.0	Baltimore
November 9, 2000	Thursday	22	42.1	
January 8, 2001	Monday	11	48.2	
January 13, 2001	Saturday	11	49.4	Baltimore

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
January 18, 2001	Thursday	11	40.6	
January 23, 2001	Tuesday	11	54.9	Baltimore, Wilmington, Philadelphia
January 24, 2001	Wednesday	10	49.4	Baltimore, Wilmington, Philadelphia
May 4, 2001	Friday	19	41.0	Baltimore
June 13, 2001	Wednesday	26	46.1	Baltimore
June 29, 2001	Friday	19	50.5	Baltimore, Wilmington, Philadelphia
August 6, 2001	Monday	37	43.8	Baltimore, Wilmington, Philadelphia, Newark
August 7, 2001	Tuesday	34	44.8	
August 8, 2001	Wednesday	34	48.7	Roanoke, Baltimore, Richmond, Wilmington
August 9, 2001	Thursday	34	50.5	Baltimore, Richmond, Roanoke, Wilmington, Philadelphia
November 16, 2001	Friday	12	45.2	
June 25, 2002	Tuesday	34	56.1	Baltimore, Wilmington
July 2, 2002	Tuesday	34	55.5	Baltimore, Richmond, Wilmington, Philadelphia, Newark
July 3, 2002	Wednesday	34	49.7	Baltimore, Richmond, Wilmington, Philadelphia
July 4, 2002	Thursday	30	59.1	
July 7, 2002	Sunday	34	43.6	
July 8, 2002	Monday	34	49.5	
July 9, 2002	Tuesday	37	42.0	Philadelphia
July 18, 2002	Thursday	34	52.2	Baltimore, Richmond, Charlotte, Wilmington, Newark
July 19, 2002	Friday	34	43.1	Baltimore
August 13, 2002	Tuesday	31	47.8	Baltimore, Richmond, Wilmington, Philadelphia

For the Washington area, the greatest number of USG days occurs in the summer months, but overall the days are distributed among all quarters of the year. The observed concentration levels on the USG days fall into the low to mid-range for the category, with the exception of the 67 μgm^{-3} measured on September 27, 1999, which actually falls into the red category. Given the number of orange days for the Washington area, rather than discuss each orange day individually, the discussion will include groups of days by season, or specific multi-day episodes.

For the USG days measured during the summer months, the Washington area experiences similar meteorological conditions that lead to high PM concentrations: light winds, high temperatures, limited mixing, high humidity, and high solar radiation. Important features that influence regional PM formation are the location and strength of the upper level ridges that affect the regional wind, temperature, stability fields, as well as the cloud and precipitation fields, which are important influences on solar radiation and its role in the photochemistry of

PM formation. Another important aspect is the location and strength of the surface high-pressure system and the resulting influence on surface winds, temperatures, cloud cover, humidity, precipitation, and local dispersion characteristics. On many of the observed summer USG days for the Washington area, the upper-level ridge is located directly over the area or is very weak, reflecting typical summer conditions in the upper atmosphere. With an upper-level ridge in this position, the temperatures aloft increase and the wind speeds decrease, leading to a buildup of PM over multiple days. On many of these days, skies are relatively clear (hazy) and precipitation is also suppressed in the area, which allows for further buildup of PM. The August 6–9, 2001 and the July 1–4, 2002 periods exhibited multiple USG days throughout the region and are good examples of widespread, persistent summertime conditions that lead to high PM in the area.

As noted above, observed USG days for the Washington area occur in every quarter of the year. The summer months experience the highest PM concentrations in the MARAMA region (a large portion of this being sulfate) because of the enhancement in sulfate formation due to photochemistry and the availability of moisture compared to drier wintertime conditions.

The wintertime conditions for observed USG days in the Washington area, such as those that occurred during January 2000 and 2001 indicate a number of features are important in influencing the buildup of PM concentrations. The locations of the upper-level ridges (and troughs) that migrate across the area in the winter months influence the strength of the surface features. During January 2001, for example, the Washington area (and the entire East Coast) was under the influence of a cold air mass from Canada. This air mass was associated with a subsidence aloft, and a strong surface high-pressure system, which resulted in inversions throughout the area that limited dispersion and allowed PM concentrations to build up over the area. These conditions persisted until the upper level features moved across the area, bringing unsettled weather, precipitation, and other conditions not conducive to a build up of PM. The January 23–24, 2001 period is a good example of widespread, persistent wintertime conditions leading to high observed PM at multiple sites throughout the region.

During the spring and fall months of the year, regional weather patterns that limit wind speeds and dispersion occur in the Washington area and, on occasion, are enough to result in high PM concentrations that fall into the USG category.

This review of the meteorological conditions indicates the high PM concentrations occur under a variety of synoptic situations, and that these vary by season. Interestingly, the CART-based classification strongly replicates this and most days within the key high PM bins correspond to the same seasonal periods.

The results for this area are a good example of how very different conditions can lead to high PM concentrations. In this case, the categorical summaries should not be used to guide the forecasting, and instead the bin by bin characteristics must be considered.

4.8. Factors Influencing $PM_{2.5}$ Concentrations for Baltimore, MD

The area-wide maximum $PM_{2.5}$ for the Baltimore area was defined for this study as the maximum value over all of the sites listed as the local Baltimore sites in Table 2-1.

4.8.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

The Baltimore-area daily maximum PM2.5 variable was defined as the maximum over fourteen FRM sites in Anne Arundel and Harford Counties as well as the city of Baltimore. Data from two additional FRM monitors were also used whenever data were missing from their collocated monitors in the primary set of fourteen. Figure 4-13 shows how days of different PM severity are distributed over the seasons. Although USG days appear in all seasons, they most often occur in the summer, when most days are moderate or worse. Overall, three percent of the days are USG and about half of these occur in summertime, and another quarter in winter. Figure 4-14 shows the fine mass concentrations at the 90th percentile for each month. The summer months are high, as one would expect, but the highest 90th percentile value actually occurs in January.

Figure 4-13. Distribution of 1999–2002 Days by Season and Severity: Baltimore

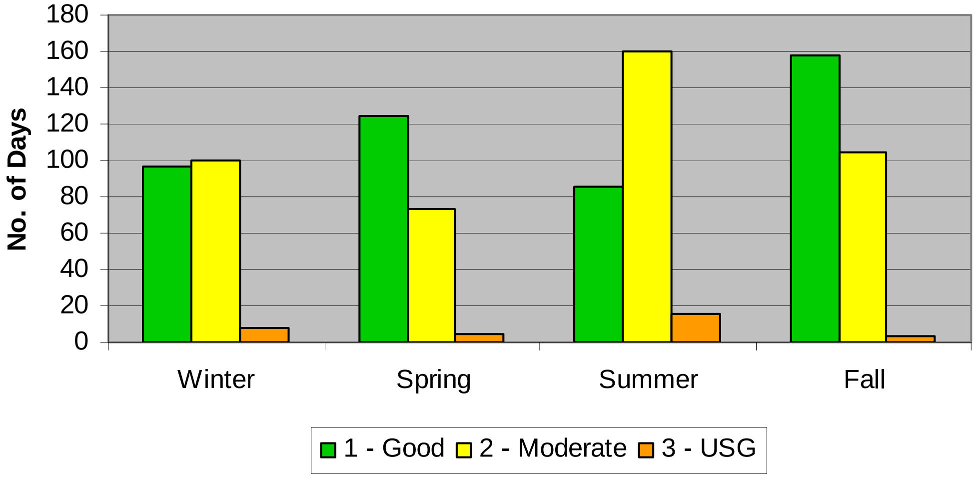
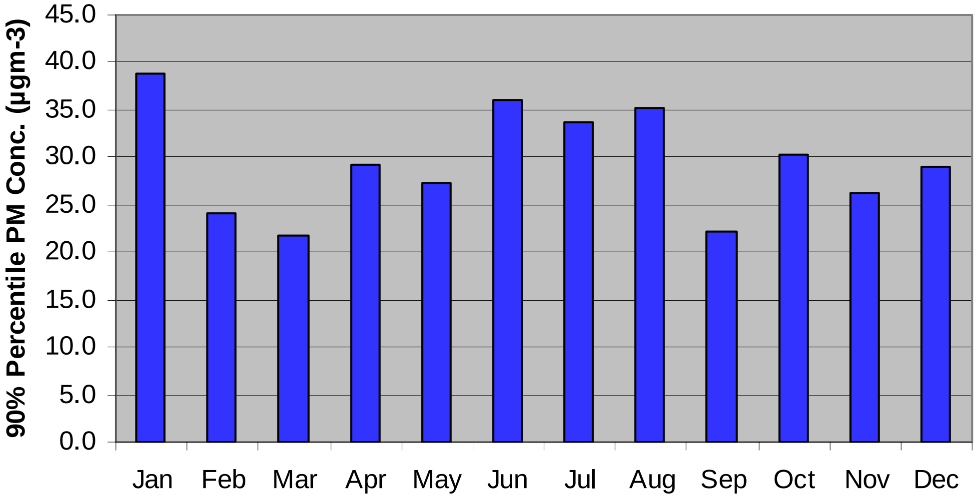


Figure 4-14. 90th Percentile Concentrations by Month (1999–2002): Baltimore



4.8.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Baltimore area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Baltimore area are presented in Appendix A. The wind information in these plots is for the Dulles Airport (Sterling, VA) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Baltimore (Figures A-11 and A-12) are based on the Dulles Airport sounding data. The upper-level winds are predominately westerly to northwesterly for the low PM days, at the time of both the morning and evening soundings, but there are some days with southwesterly winds during the evening hours in this category. For both sounding times, wind directions, on average, back to a more southwesterly direction for the moderate PM days, with lower wind speeds than for the lower PM days. For the highest PM days, wind speeds are much lower than for the other PM concentration levels and the wind directions are west-southwesterly to northwesterly at the time of the morning sounding and southerly to northwesterly at the time of the evening sounding.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-17 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-17. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Baltimore

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Baltimore (µgm^{-3})	10.5	23.2	49.2
Two-days-ago maximum 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (µgm^{-3})	17.1	18.9	26.5
Two-days-ago 24-hour $PM_{2.5}$ for Richmond (µgm^{-3})	13.6	15.5	19.4
Surface Meteorological Parameters			
Maximum surface temperature (°C)	16.8	21.7	24.1
Minimum surface temperature (°C)	7.0	10.3	12.9
Surface relative humidity (%)	64.1	70.4	69.9
Surface wind speed (ms^{-1})	3.0	1.9	1.7
Surface wind direction (degrees)	278	218	202
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.1
Upper-Air Meteorological Parameters (Dulles Airport)			
850 mb temperature (AM) (°C)	4.1	9.3	13.4
850 mb temperature (PM) (°C)	4.7	10.2	14.3
Temperature gradient (850 mb to surface; AM) (°C)	-3.5	-0.7	-0.3
Temperature gradient (900 mb to surface; AM) (°C)	-1.2	2.2	3.1
Temperature gradient (950 mb to surface; AM) (°C)	-0.3	3.1	3.8
24-hour difference in 700 mb geopotential height (m)	-5.1	1.9	-1.8
Yesterday's 700 mb wind speed (PM) (ms^{-1})	15.4	11.9	8.5
Yesterday's 850 mb wind speed (PM) (ms^{-1})	11.4	8.3	6.5
850 mb wind speed (AM) (ms^{-1})	11.1	8.6	6.6
850 mb wind speed (PM) (ms^{-1})	10.5	9.0	7.8
Yesterday's 700 mb wind direction (PM) (degrees)	282	282	295
Yesterday's 850 mb wind direction (PM) (degrees)	279	276	267
850 mb wind direction (AM) (degrees)	296	275	289
850 mb wind direction (PM) (degrees)	292	266	262
Estimated cloud cover (range of 1 to 3)	1.9	1.9	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	2

Table 4-17 provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Baltimore. A column-by-column comparison of the values reveals some clear tendencies in several of the air quality and meteorological parameters.

High $PM_{2.5}$ in the Baltimore area is clearly associated with relatively high $PM_{2.5}$ two-days prior—in the Baltimore-Washington area and to a lesser extent the Richmond area. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences), lower surface wind speeds, and less precipitation. Surface wind directions tend toward southerly (from westerly) for the higher ranges of $PM_{2.5}$. Relative humidity is slightly higher, on average, for the higher PM categories.

The upper-air meteorological parameters (based here on the Dulles Airport sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also a tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. This is especially true for the 900 and 950 mb temperature differences.

Lower wind speeds aloft also distinguish the higher $PM_{2.5}$ concentration days. There is no pronounced difference in average wind direction among the categories.

Finally, the cloud cover and season parameters do not vary much across the three categories.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include two-days-ago $PM_{2.5}$ for the Baltimore-Washington area, surface temperature, surface wind speed, 850 mb temperature, the 900 to surface temperature difference, and 850 mb wind speed at the time of the morning sounding. All of these vary regularly with $PM_{2.5}$ concentration for the analysis day.

4.8.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Baltimore area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-18 considers the input parameter values for the key USG bins. For Baltimore there are four key bins and these contain 77 percent of the USG days.

Table 4-18. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Baltimore

	Bin 29	Bin 16	Bin 34	Bin 18
Number of days	13	4	4	3
$PM_{2.5}$ Parameters				
24-hour $PM_{2.5}$ for Baltimore (μgm^{-3})	49.0	49.6	53.3	51.0
Two-days-ago maximum 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (μgm^{-3})	32.7	22.7	19.6	18.9
Two-days-ago 24-hour $PM_{2.5}$ for Richmond (μgm^{-3})	21.8	17.5	15.3	16.4

	Bin 29	Bin 16	Bin 34	Bin 18
Surface Meteorological Parameters				
Maximum surface temperature (°C)	34.1	5.0	32.9	17.2
Minimum surface temperature (°C)	21.2	-7.6	21.9	9.6
Surface relative humidity (%)	63.1	76.3	73.6	83.9
Surface wind speed (ms^{-1})	1.8	0.4	1.7	2.1
Surface wind direction (degrees)	230	45	225	90
Number of six hour periods with precipitation (range is 1 to 4)	0.0	0.0	0.3	0.0
Upper-Air Meteorological Parameters (Dulles Airport)				
850 mb temperature (AM) (°C)	18.5	1.9	17.8	11.9
850 mb temperature (PM) (°C)	18.2	3.3	18.1	12.9
Temperature gradient (850 mb to surface; AM) (°C)	-1.6	11.3	-2.7	-0.7
Temperature gradient (900 mb to surface; AM) (°C)	3.6	11.8	0.3	1.5
Temperature gradient (950 mb to surface; AM) (°C)	5.4	8.0	2.0	2.9
24-hour difference in 700 mb geopotential height (m)	3.1	-12.3	-24.5	16.2
Yesterday's 700 mb wind speed (PM) (ms^{-1})	5.3	7.3	13.9	12.3
Yesterday's 850 mb wind speed (PM) (ms^{-1})	4.9	6.2	6.0	10.6
850 mb wind speed (AM) (ms^{-1})	4.5	9.1	6.0	7.4
850 mb wind speed (PM) (ms^{-1})	5.1	7.8	11.7	9.6
Yesterday's 700 mb wind direction (PM) (degrees)	321	270	297	270
Yesterday's 850 mb wind direction (PM) (degrees)	270	288	270	243
850 mb wind direction (AM) (degrees)	306	270	270	297
850 mb wind direction (PM) (degrees)	279	252	270	243
Estimated cloud cover (range of 1 to 3)	1.7	1.3	1.8	2.0
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3	1	3	2

While many of the characteristics are similar for the high PM bins, there are some differences. These provide insight into the factors influencing the concentration that characterize each bin.

Days within Bin 29 are characterized by the highest two-days-ago day PM concentrations, whereas days within the other three key bins have lower and more consistent values, both for the Baltimore-Washington and Richmond areas. From the temperatures, as well as from the seasonal index, the bins represent different times of the year—with Bin 16 for winter days, Bin 18 for transitional season days, and Bins 29 and 34 for summer days.

In addition to the lowest temperatures, the days within Bin 16 are characterized by very stable temperature differences that are much larger than for the other key bins. The stable layer is also deeper for this bin and extends through the 850 mb level. Days within this bin have the lowest wind speeds overall, with an average that is nearly zero. Surface wind directions from the northeast are also unique to this bin.

Bin 18 is comprised primarily of transitional season days. Wind speeds tend to be higher, on average, than for the other bins, both near the surface and aloft. The change in geopotential height is most positive for days within this bin. Surface wind directions are, on average, from the east, which is unique to this bin. Cloud cover is the greatest over all key bins.

The two bins comprised mostly of summer days have higher temperatures and intermediate surface wind speeds when compared to the other key bins. Days within these bins also exhibit southwesterly surface wind directions. Relative humidity is higher for Bin 34. The bins also differ from one another in the stability characteristics such that days within Bin 34 are less stable. Days within this bin also show decreasing heights and high wind speeds aloft during the morning hours, compared to days within Bin 29. Thus, there appear to be two different summertime regimes with different synoptic characteristics.

Next, we examine the conditions associated with each day or episode.

Data retrieval and availability for the Baltimore area were relatively high (although not as high as for Washington) for the period 1999–2002, and 31 USG days occurred during this period, resulting in the second largest number of orange days during this period in the MARAMA region. The specific dates, including the observed $PM_{2.5}$ concentration (μg/m^3), are as follows:

Table 4-19. USG Days for Baltimore: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
June 2, 2000	Friday	34	43.3	Bristol
June 11, 2000	Sunday	29	42.0	Washington, Bristol
October 26, 2000	Thursday	18	53.4	Washington
November 8, 2000	Wednesday	13	41.3	Washington
December 16, 2000	Saturday	15	50.4	
January 12, 2001	Friday	3	41.4	
January 13, 2001	Saturday	16	53.6	Washington
January 14, 2001	Sunday	16	45.5	
January 23, 2001	Tuesday	16	56.7	Washington, Wilmington, Philadelphia
January 24, 2001	Wednesday	5	43.2	Washington, Wilmington, Philadelphia
January 25, 2001	Thursday	12	63.7	
April 7, 2001	Saturday	18	52.5	
April 10, 2001	Tuesday	18	47.1	
May 3, 2001	Thursday	29	40.9	
May 4, 2001	Friday	29	43.8	Washington
June 12, 2001	Tuesday	28	40.6	Washington
June 28, 2001	Thursday	29	42.9	Wilmington, Philadelphia
June 29, 2001	Friday	29	62.1	Washington, Wilmington, Philadelphia

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
August 5, 2001	Sunday	29	51.3	
August 6, 2001	Monday	29	45.1	Washington, Wilmington, Newark
August 8, 2001	Wednesday	29	46.7	Washington, Richmond, Roanoke, Wilmington
August 9, 2001	Thursday	29	53.4	Washington, Richmond, Wilmington, Philadelphia
June 25, 2002	Tuesday	34	59.6	Washington
July 2, 2002	Tuesday	29	54.1	Washington, Richmond, Wilmington, Philadelphia, Newark
July 3, 2002	Wednesday	29	50.7	Washington, Richmond, Wilmington, Philadelphia
July 18, 2002	Thursday	29	50.5	Washington, Richmond, Charlotte, Wilmington, Newark
July 19, 2002	Friday	34	46.3	Washington, Richmond
August 13, 2002	Tuesday	29	52.7	Washington, Richmond
August 24, 2002	Saturday	34	64.2	
October 4, 2002	Friday	26	41.8	
December 10, 2002	Tuesday	16	42.7	

The high PM days in the Baltimore area during this period are distributed more evenly across the seasons than for Washington. Although high PM occurs more often in the summer months, high PM days occurred during all quarters of the year. Due to their proximity, the Baltimore and Washington areas encounter very similar weather conditions leading to high PM concentrations throughout the year. The January 23–24, 2001 wintertime conditions leading to high PM in Washington also caused high PM in the Baltimore area, extending to Wilmington and Philadelphia as well.

A rather severe summertime episode occurred during the period August 5–9, 2001. This episode was dominated by a large upper-level ridge extending over the entire U.S. with a strong surface high-pressure system centered over the mid-Atlantic states. This pattern persisted for several days. High temperatures were in the upper 80's at the beginning of the period to near 100 at the end of the period. Winds at the upper levels were light and westerly, while surface winds were light and variable. Skies were reported hazy during this period with partly cloudy conditions and little precipitation. The combination of persistent stagnant conditions led to a regional buildup of PM throughout the MARAMA region with USG days reported at seven of the nine areas of interest during one or more days of this episode.

The results for this area are another example of how different conditions can lead to high PM concentrations. CART effective separates into different types of high PM events that share seasonal characteristics and then separates them further into bins based on other differences in the parameters. The is not just one pathway to high $PM_{2.5}$. In this case, the categorical summaries should not be used to guide the forecasting, and instead the bin by bin characteristics must be considered.

4.9. Factors Influencing $PM_{2.5}$ Concentrations for Philadelphia, PA

The area-wide maximum $PM_{2.5}$ for the Philadelphia area was defined for this study as the maximum value over all of the sites listed as the local Philadelphia sites in Table 2-1.

4.9.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

Five FRM monitors in the greater Philadelphia area determine the area-wide maximum $PM_{2.5}$ concentrations. Two percent of the days with available data were USG, and as Figure 4-15 shows, most of these days, as usual, appeared in the summer, although six days, or one-third the summer total, appeared in winter. Unlike the other areas, good and moderate summer days are closely matched in quantity, and good days are in the minority in winter. Figure 4-16 shows the 90th percentile concentrations, which are highest in June but about equal in January and July, which share the second-highest rank.

Figure 4-15. Distribution of 1999–2002 Days by Season and Severity: Philadelphia

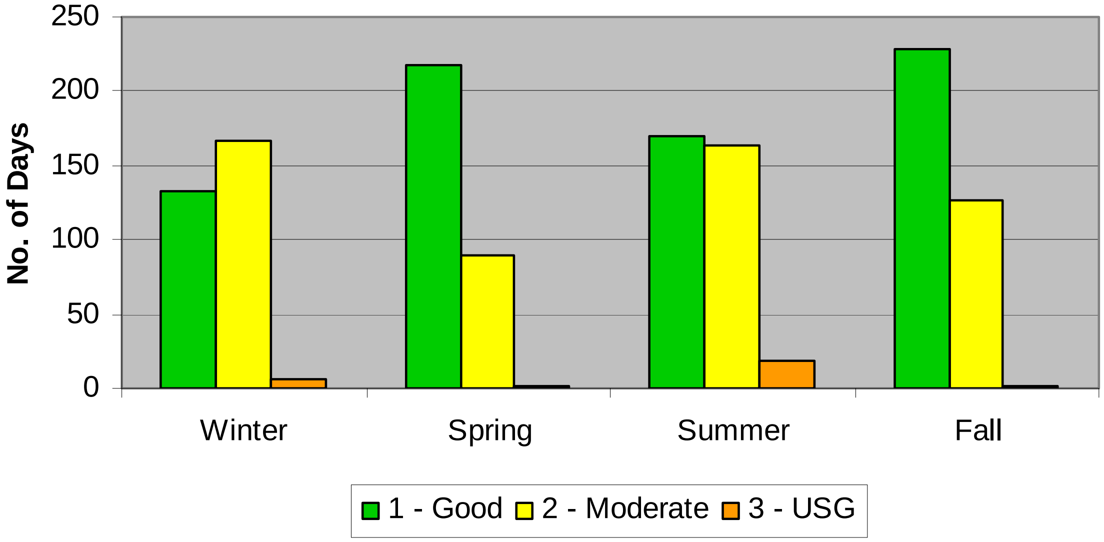
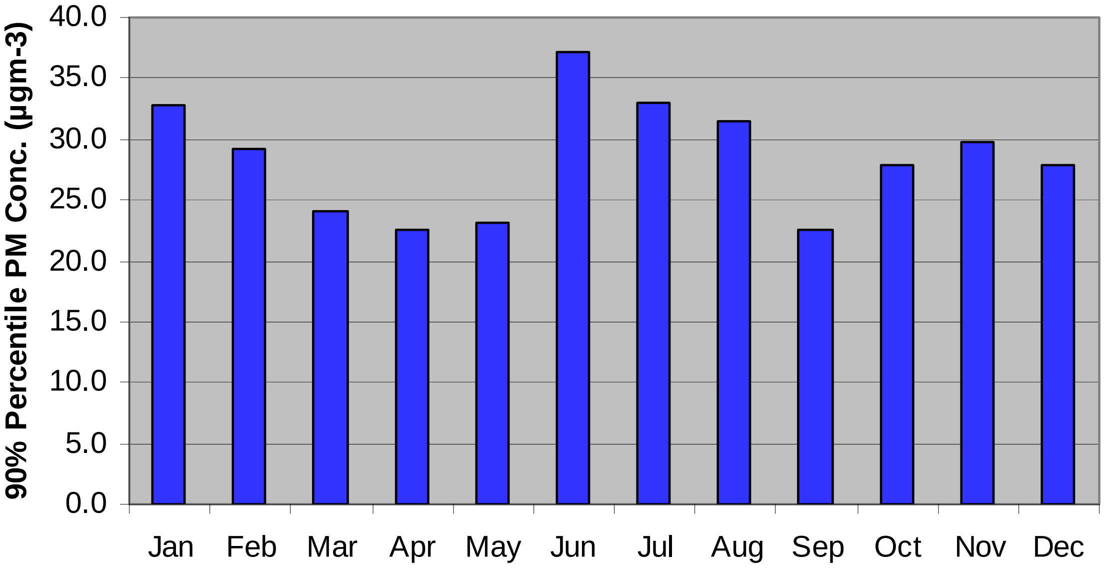


Figure 4-16. 90th Percentile Concentrations by Month (1999–2002): Philadelphia



4.9.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Philadelphia area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Philadelphia area are presented in Appendix A. The wind information in these plots is for the Dulles Airport (Sterling, VA) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Philadelphia (Figures A-13 and A-14) are based on the Dulles Airport sounding data. The upper-level winds are predominately westerly to northwesterly for the low PM days, at the time of both the morning and evening soundings. Northwesterly winds characterize the greatest number of days. For both sounding times, wind directions, on average, back to a more southwesterly direction for the moderate PM days, with lower wind speeds than for the lower PM days. For the highest PM days, wind speeds are much lower than for the other PM concentration levels and the wind directions range from southwesterly to northwesterly; wind predominantly westerly wind directions at the time of the morning sounding and predominantly southwesterly wind directions at the time of the evening sounding.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-20 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-20. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Philadelphia

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Philadelphia (μgm^{-3})	9.8	22.9	46.8
Two-days-ago maximum 24-hour $PM_{2.5}$ for Camden and New Castle (μgm^{-3})	15.5	17.5	26.7
Two-days-ago 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (μgm^{-3})	16.9	18.6	30.2
Surface Meteorological Parameters			
Maximum surface temperature (°C)	17.2	19.4	26.7
Minimum surface temperature (°C)	8.9	9.9	15.5
Surface relative humidity (%)	62.8	69.7	68.8
Surface wind speed (ms^{-1})	3.9	3.0	2.8
Surface wind direction (degrees)	274	188	191
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.1
Upper-Air Meteorological Parameters (Dulles Airport)			
850 mb temperature (AM) (°C)	4.9	7.9	14.1
850 mb temperature (PM) (°C)	5.6	8.7	14.7
Temperature gradient (850 mb to surface; AM) (°C)	-3.4	-0.4	0.3
Temperature gradient (900 mb to surface; AM) (°C)	-1.2	2.6	4.1
Temperature gradient (950 mb to surface; AM) (°C)	-0.3	3.4	5.5
24-hour difference in 700 mb geopotential height (m)	-1.3	-1.5	-18.7
Yesterday's 700 mb wind speed (PM) (ms^{-1})	14.6	12.4	7.7
Yesterday's 850 mb wind speed (PM) (ms^{-1})	11.0	8.8	6.1
850 mb wind speed (AM) (ms^{-1})	10.9	9.7	6.9
850 mb wind speed (PM) (ms^{-1})	10.2	9.8	7.6
Yesterday's 700 mb wind direction (PM) (degrees)	244	242	263
Yesterday's 850 mb wind direction (PM) (degrees)	245	224	227
850 mb wind direction (AM) (degrees)	256	234	232
850 mb wind direction (PM) (degrees)	248	222	215
Estimated cloud cover (range of 1 to 3)	1.9	1.8	1.6
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	2

Table 4-20 provides an overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Philadelphia.

High $PM_{2.5}$ in the Philadelphia area is associated with relatively high $PM_{2.5}$ two-days prior—in both the Philadelphia (Camden-Wilmington) and Baltimore-Washington areas. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences) and less precipitation. Surface wind directions tend toward southerly for the higher ranges of $PM_{2.5}$, compared to westerly for the lowest range of concentration. There is no clear tendency for relative humidity and surface wind speed.

The upper-air meteorological parameters (based here on the Dulles Airport sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures and more stable (positive) lapse rates. The difference in geopotential height is much more negative for the higher PM days.

Considering the upper-air wind data, the higher PM days are characterized by lower wind speeds aloft. Winds aloft are, on average, southwesterly, for all three categories.

High PM is associated with slightly less cloud cover; overall, the season parameters do not distinguish the categories at this most general level.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include two-days-ago for the Camden-New Castle area, surface temperature, 850 mb temperature, and 900 to surface temperature difference. All of these are also well correlated with the $PM_{2.5}$ concentration for the analysis day.

4.9.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different 8-hour ozone concentration levels for the Philadelphia area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-21 considers the input parameter values for the key USG bins. For Philadelphia, there are two key bins containing 17 and 7, respectively, of the 28 USG days.

Table 4-21. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Philadelphia.

	Bin 34	Bin 26
Number of days	17	7
$PM_{2.5}$ Parameters		
24-hour $PM_{2.5}$ for Philadelphia (μgm^{-3})	46.2	47.3
Two-days-ago maximum 24-hour $PM_{2.5}$ for Camden and New Castle (μgm^{-3})	30.9	20.0

	Bin 34	Bin 26
Two-days-ago 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (µgm^{-3})	33.9	26.7
Surface Meteorological Parameters		
Maximum surface temperature (°C)	35.0	12.5
Minimum surface temperature (°C)	23.9	0.8
Surface relative humidity (%)	65.1	77.0
Surface wind speed (ms^{-1})	3.5	1.3
Surface wind direction (degrees)	184	214
Number of six hour periods with precipitation (range is 1 to 4)	0.2	0.0
Upper-Air Meteorological Parameters (Dulles Airport)		
850 mb temperature (AM) (°C)	18.4	7.1
850 mb temperature (PM) (°C)	19.1	7.1
Temperature gradient (850 mb to surface; AM) (°C)	-3.3	8.3
Temperature gradient (900 mb to surface; AM) (°C)	1.0	10.7
Temperature gradient (950 mb to surface; AM) (°C)	3.2	10.2
24-hour difference in 700 mb geopotential height (m)	-14.1	-24.8
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	6.7	8.4
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	5.6	6.6
850 mb wind speed (AM) (ms^{-1})	5.7	9.9
850 mb wind speed (PM) (ms^{-1})	6.6	10.7
Yesterday’s 700 mb wind direction (PM) (degrees)	260	270
Yesterday’s 850 mb wind direction (PM) (degrees)	238	217
850 mb wind direction (AM) (degrees)	232	214
850 mb wind direction (PM) (degrees)	221	189
Estimated cloud cover (range of 1 to 3)	1.8	1.1
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3	1

The two key high PM bins represent winter and summer types of PM events.

Days within Bin 26 (containing a majority of winter time days) are associated with lower two-days-ago PM concentrations, yet higher concentrations, on average, on the analysis days, compared to days within Bin 34 (the summertime bin). Temperatures and surface wind speeds are much lower for days within Bin 26. The days within this bin are also distinguished by very stable lapse rates and a deep stable layer. Wind speeds aloft are greater for Bin 26 than for Bin 34, as expected during wintertime synoptic conditions.

Data retrieval and availability for the Philadelphia area were high for the period 1999–2002, and 28 USG days occurred during this period. The specific dates, including the observed $PM_{2.5}$ concentration (μgm^{-3}), are listed in Table 4-22.

Table 4-22. USG Days for Philadelphia: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
July 19, 1999	Monday	34	50.5	Wilmington
July 24, 1999	Saturday	34	46.9	Wilmington
July 31, 1999	Saturday	34	42.3	Wilmington
October 30, 1999	Saturday	26	41.8	
February 4, 2000	Friday	2	49.2	
February 10, 2000	Thursday	26	48.9	Wilmington
February 11, 2000	Friday	26	48.0	
March 9, 2000	Thursday	26	41.7	
June 10, 2000	Saturday	34	41.5	Washington, Newark
June 11, 2000	Sunday	34	44.6	Wilmington, Baltimore, Bristol, Newark
January 14, 2001	Sunday	26	45.5	
January 23, 2001	Tuesday	26	52.9	Baltimore, Washington, Wilmington
January 24, 2001	Wednesday	15	41.9	Baltimore, Washington, Wilmington
May 4, 2001	Friday	30	46.2	Baltimore, Washington
June 28, 2001	Thursday	34	42.9	Wilmington
June 29, 2001	Friday	34	49.2	Baltimore, Washington, Wilmington
June 30, 2001	Saturday	34	51.8	Wilmington, Newark
August 6, 2001	Monday	34	46.5	Baltimore, Washington, Wilmington, Newark
August 9, 2001	Thursday	34	50.4	Baltimore, Washington, Richmond, Wilmington
August 10, 2001	Friday	34	41.2	
November 18, 2001	Sunday	26	52.1	
June 9, 2002	Sunday	13	57.2	
July 2, 2002	Tuesday	34	42.8	Baltimore, Washington, Wilmington, Richmond
July 3, 2002	Wednesday	34	45.4	Baltimore, Washington, Richmond, Wilmington
July 9, 2002	Tuesday	34	44.0	Washington
July 18, 2002	Thursday	34	46.3	Baltimore, Washington, Richmond, Charlotte, Wilmington, Newark
July 19, 2002	Friday	34	58.5	Baltimore, Washington, Wilmington, Newark
August 13, 2002	Tuesday	34	40.9	Baltimore, Washington, Richmond, Wilmington

High PM concentrations were measured during all quarters at the Philadelphia sites with the maximum number of high days occurring during the summer months and the minimum number of high PM days occurring during the first quarter of the year. The Philadelphia area also measured high PM concentrations during the wintertime episode of January 23–24, 2001, and the summertime episodes discussed above: August 6–9, 2001, July 1–4, 2002, and July 18–19, 2002. Another widespread but short-term event occurred on August 13, 2002. During this period, a moderately strong upper-level ridge centered over the eastern states resulted in very light southwesterly winds aloft, and a moderately strong surface high-pressure system centered over Virginia. Minimum temperatures in the Philadelphia area were in the low 70's, while maximum temperatures were in the upper 90's. Hazy skies and fog were reported in the early morning hours at multiple sites throughout the region. These conditions led to high concentrations at sites extending from Richmond to Philadelphia. A USG day was observed in the Newark area on August 14. Meteorological conditions changed in the region on August 15, in advance of an approaching cold front, resulting in lower measured PM concentrations throughout the region.

This review of the meteorological conditions indicates that high PM concentrations occur under a variety of synoptic situations, but in general (and as indicated by the CART results) the majority of summertime events are associated with regional-scale build up and transport of PM, while the wintertime events seem to be driven by local meteorological conditions and can be isolated, depending upon the geographical extent of the PM conducive meteorological conditions. CART quite clearly distinguishes the winter- and summertime events and places a majority of these into two key bins. Other high PM days are placed in other high PM bins. CART thus appears to be able to distinguish and group the USG days quite effectively. Because of these differences, the categorical summaries should not be used to guide the forecasting, and instead the bin by bin characteristics must be considered.

4.10. Factors Influencing $PM_{2.5}$ Concentrations for Wilmington, DE

The area-wide maximum $PM_{2.5}$ for the Wilmington area was defined for this study as the maximum value over all of the sites listed as the local Wilmington sites in Table 2-1.

4.10.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

New Castle County in Delaware and Cecil County in Maryland provide data for Wilmington from six FRM monitors, plus three additional collocated monitors, each used as a back-up for the other monitor at its site. Two percent of these days are USG, with 19 occurring in the summer, six in the winter and one in the spring, as shown in Figure 4-17. Both summer and winter have fewer good than moderate days; Figure 4-18 shows peak monthly 90th percentile values in June and January, and the lowest concentrations in March and September.

Figure 4-17. Distribution of 1999–2002 Days by Season and Severity: Wilmington

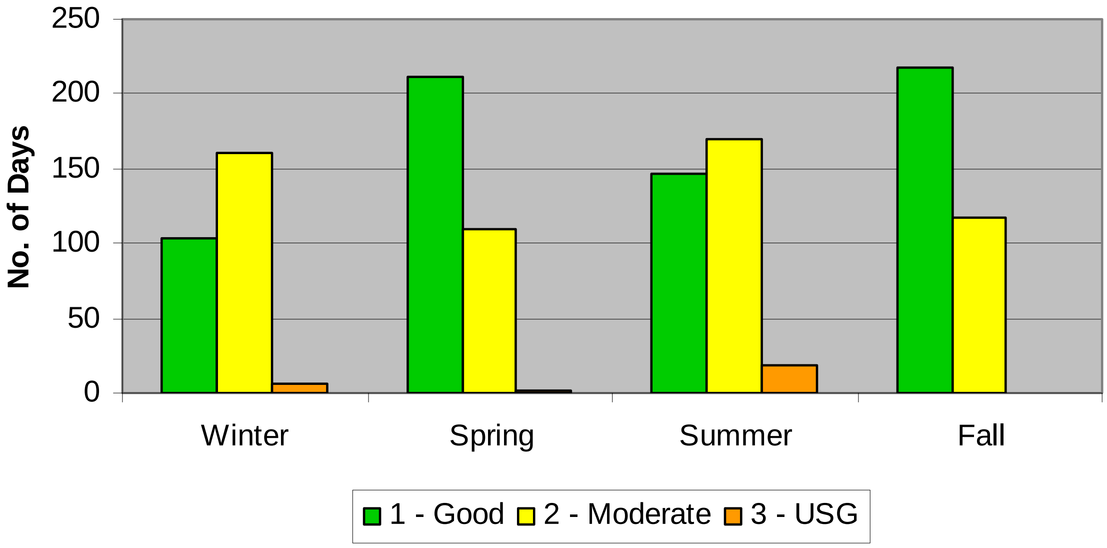
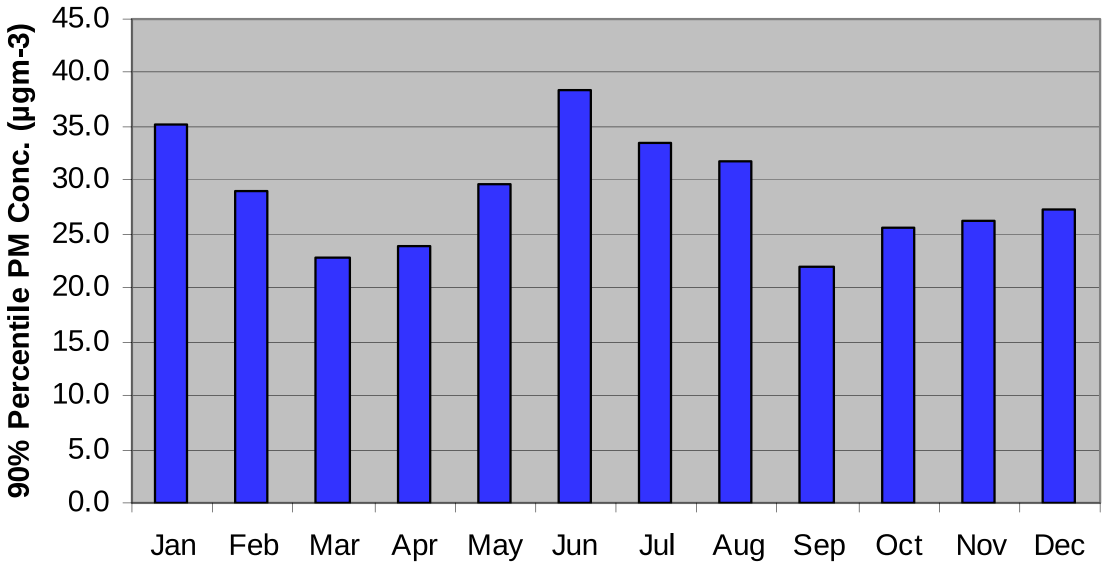


Figure 4-18. 90th Percentile Concentrations by Month (1999–2002): Wilmington



4.10.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Wilmington area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Wilmington area are presented in Appendix A. The wind information in these plots is for the Dulles Airport (Sterling, VA) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Wilmington (Figures A-15 and A-16) are based on the Dulles Airport sounding data. The upper-level winds are predominately westerly to northwesterly for the low PM days at the time of the morning sounding, and southwesterly to northwesterly winds at the time of the evening sounding. For both sounding times, wind directions, on average, back to a more southwesterly direction for the moderate PM days, with lower wind speeds than for the lower PM days. For the highest PM days, wind speeds are much lower than for the other PM concentration levels and the wind directions generally range from southwesterly to northwesterly; at the time of the evening sounding many different directions are represented.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-23 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 µgm^{-3}.

Table 4-23. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Wilmington

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Wilmington (µgm^{-3})	10.3	22.9	47.4
Two-days-ago 24-hour $PM_{2.5}$ for New Castle (µgm^{-3})	15.2	17.1	21.5
Two-days-ago maximum 24-hour $PM_{2.5}$ for Washington, Baltimore, and Gettysburg (µgm^{-3})	17.0	18.8	27.1
Surface Meteorological Parameters			
Maximum surface temperature (°C)	17.0	19.2	26.5
Minimum surface temperature (°C)	8.0	9.3	16.0
Surface relative humidity (%)	65.7	69.8	70.1
Surface wind speed (ms^{-1})	3.7	2.6	2.1
Surface wind direction (degrees)	276	186	184
Number of six hour periods with precipitation (range is 1 to 4)	0.3	0.2	0.0
Upper-Air Meteorological Parameters (Dulles Airport)			
850 mb temperature (AM) (°C)	5.3	7.8	15.3
850 mb temperature (PM) (°C)	6.0	8.8	15.5

	Category 1	Category 2	Category 3
Temperature gradient (850 mb to surface; AM) (°C)	-3.3	-0.8	-1.8
Temperature gradient (900 mb to surface; AM) (°C)	-1.1	2.3	1.9
Temperature gradient (950 mb to surface; AM) (°C)	-0.3	3.2	3.9
24-hour difference in 700 mb geopotential height (m)	-1.4	-0.6	-17.3
Yesterday's 700 mb wind speed (PM) (ms^{-1})	14.2	12.7	6.8
Yesterday's 850 mb wind speed (PM) (ms^{-1})	10.5	9.0	6.1
850 mb wind speed (AM) (ms^{-1})	10.6	9.6	5.8
850 mb wind speed (PM) (ms^{-1})	10.1	9.6	6.9
Yesterday's 700 mb wind direction (PM) (degrees)	242	245	277
Yesterday's 850 mb wind direction (PM) (degrees)	242	229	257
850 mb wind direction (AM) (degrees)	253	237	247
850 mb wind direction (PM) (degrees)	246	227	243
Estimated cloud cover (range of 1 to 3)	1.9	1.8	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	3

Table 4-23 provides an overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for Wilmington. The results for Wilmington are very similar to those for Philadelphia.

High $PM_{2.5}$ in the Wilmington area is associated with relatively high $PM_{2.5}$ two-days prior—in both Wilmington and the Baltimore-Washington area. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences) and less precipitation. Surface wind directions tend toward southerly for the higher ranges of $PM_{2.5}$, compared to westerly for the lowest range of concentration. There is no clear tendency for relative humidity and surface wind speed.

The upper-air meteorological parameters (based on the Dulles Airport sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures and generally more stable (positive) lapse rates. The difference in geopotential height is much more negative for the higher PM days.

Considering the upper-air wind data, the higher PM days are characterized by lower wind speeds aloft. Winds aloft are, on average, southwesterly, for all three categories.

High PM is associated with slightly less cloud cover and tends to occur during the summer.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include surface temperature, surface wind speed, 850 mb temperature, and 950 to surface temperature difference. All of these are also well correlated with the $PM_{2.5}$ concentration for the analysis day.

4.10.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different $PM_{2.5}$ concentration levels for the Wilmington area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-24 considers the input parameter values for the key USG bins. For Wilmington there are two key bins that contain 15 and 5, respectively, of the 26 USG days.

Table 4-24. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Wilmington

	Bin 29	Bin 2
Number of days	15	5
PM2.5 Parameters		
24-hour PM2.5 for Wilmington (µgm-3)	48.3	48.3
Two-days-ago 24-hour PM2.5 for New Castle (µgm-3)	23.6	15.3
Two-days-ago maximum 24-hour PM2.5 for Washington, Baltimore, and Gettysburg (µgm-3)	32.7	14.2
Surface Meteorological Parameters		
Maximum surface temperature (°C)	33.9	4.5
Minimum surface temperature (°C)	23.0	-5.2
Surface relative humidity (%)	64.2	81.2
Surface wind speed (ms-1)	2.7	0.6
Surface wind direction (degrees)	190	270
Number of six hour periods with precipitation (range is 1 to 4)	0.0	0.0
Upper-Air Meteorological Parameters (Dulles Airport)		
850 mb temperature (AM) (°C)	18.1	-0.9
850 mb temperature (PM) (°C)	18.9	0.1
Temperature gradient (850 mb to surface; AM) (°C)	-3.4	6.2
Temperature gradient (900 mb to surface; AM) (°C)	1.0	7.4
Temperature gradient (950 mb to surface; AM) (°C)	3.1	8.4
24-hour difference in 700 mb geopotential height (m)	-20.8	-36.2
Yesterday's 700 mb wind speed (PM) (ms-1)	5.7	8.8
Yesterday's 850 mb wind speed (PM) (ms-1)	5.3	7.3
850 mb wind speed (AM) (ms-1)	5.5	8.2
850 mb wind speed (PM) (ms-1)	6.8	9.1
Yesterday's 700 mb wind direction (PM) (degrees)	275	270
Yesterday's 850 mb wind direction (PM) (degrees)	252	270
850 mb wind direction (AM) (degrees)	248	270
850 mb wind direction (PM) (degrees)	246	252
Estimated cloud cover (range of 1 to 3)	1.7	1.5
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3	1

The two key high PM bins represent winter and summer types of PM events.

Days within Bin 2 (containing a majority of winter time days) are associated with lower two-days-ago PM concentrations, yet similar concentrations, on average, on the analysis days, compared to days within Bin 29 (the summer time bin). Temperatures and surface wind speeds are much lower for days within Bin 2. Surface wind directions are also different for the two bins and are westerly for Bin 2 (winter) and southerly for Bin 29 (summer). The days within Bin 2 are also distinguished by more stable lapse rates and a deeper stable layer, than days within Bin 29—typical of wintertime conditions. Wind speeds aloft are greater for Bin 2 than for Bin 29, as expected during wintertime synoptic conditions.

Next we explore, the conditions associated with the USG events.

Data retrieval and availability for the Wilmington area were high for the period 1999–2002, and 26 USG days occurred during this period. The specific dates, including the observed $PM_{2.5}$ concentration (μgm^{-3}), are presented in Table 4-25.

Table 4-25. USG Days for Wilmington: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
June 7, 1999	Monday	29	40.9	
June 8, 1999	Tuesday	29	45.5	
July 19, 1999	Monday	29	46.0	
July 24, 1999	Saturday	29	44.7	
July 31, 1999	Saturday	36	43.9	
January 1, 2000	Saturday	2	42.6	Charlotte, Washington
February 4, 2000	Friday	2	45.2	
February 10, 2000	Thursday	2	53.4	
June 11, 2000	Sunday	29	43.5	Baltimore, Bristol, Philadelphia, Newark,
January 13, 2001	Saturday	2	40.8	
January 23, 2001	Tuesday	2	59.6	Baltimore, Washington
January 24, 2001	Wednesday	4	58.5	Baltimore, Washington
May 19, 2001	Saturday	10	40.9	
June 14, 2001	Thursday	15	41.4	Newark
June 28, 2001	Thursday	29	41.9	Philadelphia
June 29, 2001	Friday	29	51.8	Baltimore, Washington, Philadelphia
June 30, 2001	Saturday	29	44.6	Newark, Philadelphia
August 6, 2001	Monday	29	44.9	Baltimore, Washington, Philadelphia, Newark
August 8, 2001	Wednesday	29	50.0	Baltimore, Washington, Richmond, Roanoke
August 9, 2001	Thursday	29	53.1	Baltimore, Washington, Richmond, Roanoke, Philadelphia

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG Day for Other Areas?
June 25, 2002	Tuesday	34	42.1	Baltimore, Washington, Philadelphia, Newark
July 2, 2002	Tuesday	29	57.8	Baltimore, Washington, Richmond, Philadelphia, Newark
July 3, 2002	Wednesday	29	46.1	Baltimore, Washington, Richmond, Philadelphia
July 18, 2002	Thursday	29	56.0	Baltimore, Washington, Richmond, Charlotte, Newark
July 19, 2002	Friday	29	57.6	Baltimore, Washington, Newark, Philadelphia, Wilmington
August 13, 2002	Tuesday	34	40.6	Baltimore, Washington, Richmond, Philadelphia

For the Wilmington area, the majority of high PM days were measured during the summer months, and no USG days occurred in the fourth quarter of the year (October-December). Due to its proximity to the Washington, Baltimore, and Philadelphia areas, the Wilmington area experiences similar meteorological conditions that lead to high PM concentrations. As noted above, the Wilmington area experienced multiple USG days during the January 23–24, 2001 wintertime episode, and during the widespread summertime episode periods of August 5–9, 2001 and July 1–4, 2002.

Another widespread episode that occurred in the MARAMA region was the July 17–19, 2002 period. Similar to the other summertime episodes, a strong upper-level ridge was centered over the Midwest during this period, with a surface high-pressure system centered over Georgia. Upper-level winds were light and southwesterly, while surface winds were light and variable. Maximum temperatures were in the upper 90's, while minimum temperatures were in the low 70's. Hazy skies and limited visibility were reported during the morning hours throughout the region. The meteorological conditions of this episode are very similar to those of the July 1–4, 2002 period. High PM was measured at six of the nine areas of interest, from Charlotte to the south extending to Newark to the north.

This review of the meteorological conditions indicates that high PM concentrations occur under a variety of synoptic situations. As for Philadelphia, CART distinguishes the winter- and summertime events and places a majority of these into two key bins. Other high PM days are placed in other high PM bins. CART thus appears to be able to distinguish and group the USG days quite effectively. Because of these differences, the categorical summaries should not be used to guide the forecasting, and instead the bin by bin characteristics must be considered.

4.11. Factors Influencing $PM_{2.5}$ Concentrations for Newark, NJ

The area-wide maximum $PM_{2.5}$ for the Newark/Elizabeth area was defined for this study as the maximum value over all of the sites listed as the local Newark sites in Table 2-1.

4.11.1. Summary of Observed $PM_{2.5}$ Data (1999–2002)

The data for Newark come from ten FRM monitors in three New Jersey counties in the Newark MSA: Essex, Middlesex, and Union. Two additional monitors are collocated with two others and only used if data from the primary monitors are missing. Only 2.5 percent of these days are USG, all but three occur in the summer months. Figure 4-19 shows the distribution of days by season and severity. Winter has almost as many moderate days as good, though only one very high USG day; fall and spring have mostly good days. Figure 4-20 shows the 90th percentile concentrations by month, with the highest occurring in June and August, but second highest in October, followed by July, followed closely by January.

Figure 4-19. Distribution of 1999–2002 Days by Season and Severity: Newark

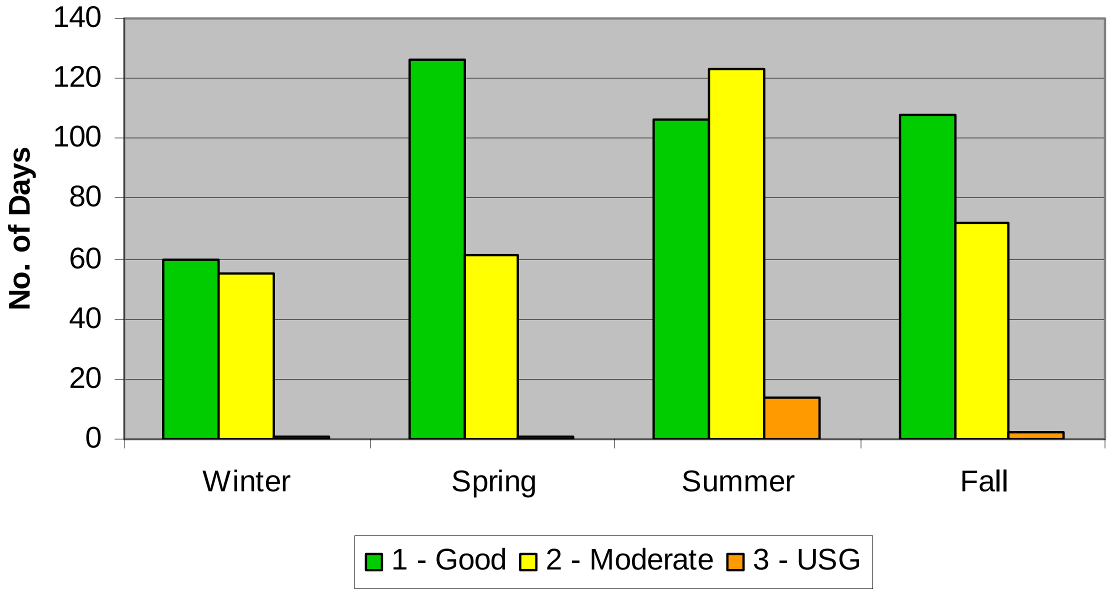
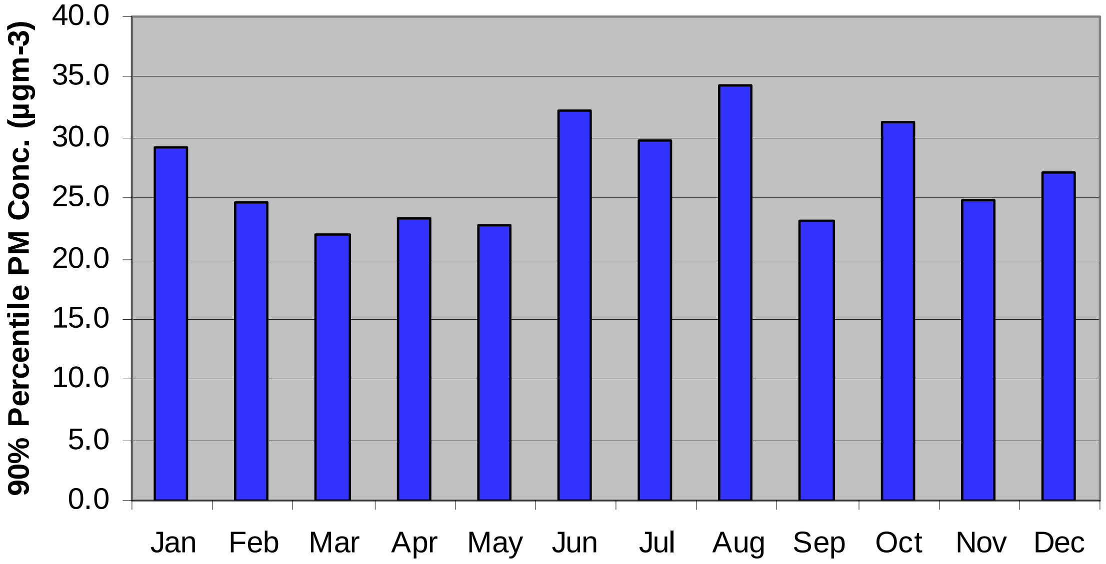


Figure 4-20. 90th Percentile Concentrations by Month (1999–2002): Newark



4.11.2. Meteorological Factors Influencing $PM_{2.5}$ Concentrations

The meteorological conditions associated with the different ranges of $PM_{2.5}$ concentration and specifically the highest PM days for the Newark area are discussed in this subsection.

Wind Patterns Associated with High $PM_{2.5}$

Plots comparing the frequency of wind directions and speeds for all days and for low, moderate, and high $PM_{2.5}$ days for the Newark area are presented in Appendix A. The wind information in these plots is for the Brookhaven (Long Island, NY) upper-air monitoring site. The plots use the same format and contain the same information as for the other areas (described earlier in this section).

The wind roses for Newark (Figures A-17 and A-18) are based on the Brookhaven sounding data. The upper-level winds are predominately west-southwesterly to northerly for the low PM days at the time of both the morning and evening soundings, northwesterly winds characterize the greatest number of days for the evening hour. For both sounding times, wind directions, on average, back to a more southwesterly direction for the moderate PM days, with lower wind speeds than for the lower PM days. The range in wind direction is southwesterly to northwesterly, and the greatest number of days with westerly winds. For the highest PM days, there is a further shift toward southwesterly and the predominant range in wind direction is southwesterly to westerly. Wind speeds are lower than for the other PM concentration levels.

Categorical Summaries

A comparison of the meteorological characteristics for different ranges of $PM_{2.5}$ concentration in Table 4-26 provides a basis for further distinguishing days within the different categories based on the values of meteorological parameters. In preparing this table, we used the comprehensive meteorological and PM dataset compiled for the CART application. Key meteorological parameters, as used by CART to construct the classification tree, are shaded in this table so that we can focus on the differences in these key parameters as well the differences found throughout the dataset. Categories 1 to 3 represent the standard three ranges of 24-hour $PM_{2.5}$ concentration: <15.5, 15.5–40.5, and ≥40.5 μgm^{-3}.

Table 4-26. Summary of Mean Air Quality and Meteorological Parameters for Each CART Classification Category: Newark

	Category 1	Category 2	Category 3
$PM_{2.5}$ Parameters			
24-hour $PM_{2.5}$ for Newark/Elizabeth (μgm^{-3})	9.6	23.4	45.4
Two-days-ago 24-hour $PM_{2.5}$ for Elizabeth (μgm^{-3})	14.9	15.8	22.8
Two-days-ago 24-hour $PM_{2.5}$ for Bethlehem (μgm^{-3})	13.2	15.2	22.0
Two-days-ago maximum 24-hour $PM_{2.5}$ for Camden and New Castle (μgm^{-3})	15.6	17.1	24.9
Surface Meteorological Parameters			
Maximum surface temperature (°C)	18.2	22.2	29.9
Minimum surface temperature (°C)	10.0	12.8	19.4
Surface relative humidity (%)	59.6	67.2	67.8

	Category 1	Category 2	Category 3
Surface wind speed (ms^{-1})	3.9	3.1	3.2
Surface wind direction (degrees)	268	176	169
Number of six hour periods with precipitation (range is 1 to 4)	0.2	0.2	0.2
Upper-Air Meteorological Parameters (Brookhaven)			
850 mb temperature (AM) (°C)	5.0	9.0	14.7
850 mb temperature (PM) (°C)	5.3	10.1	16.5
Temperature gradient (850 mb to surface; AM) (°C)	-4.4	-2.5	-3.9
Temperature gradient (900 mb to surface; AM) (°C)	-2.1	0.5	0.7
Temperature gradient (950 mb to surface; AM) (°C)	-1.2	1.7	2.5
24-hour difference in 700 mb geopotential height (m)	-0.3	4.1	1.1
Yesterday’s 700 mb wind speed (PM) (ms^{-1})	14.3	12.7	10.8
Yesterday’s 850 mb wind speed (PM) (ms^{-1})	11.1	9.3	7.9
850 mb wind speed (AM) (ms^{-1})	10.5	9.6	8.4
850 mb wind speed (PM) (ms^{-1})	10.2	10.3	10.0
Yesterday’s 700 mb wind direction (PM) (degrees)	238	248	264
Yesterday’s 850 mb wind direction (PM) (degrees)	252	243	250
850 mb wind direction (AM) (degrees)	257	228	236
850 mb wind direction (PM) (degrees)	260	237	204
Estimated cloud cover (range of 1 to 3)	1.8	1.9	1.6
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	2	2	3

Table 4-26 provides an overview of how average conditions for each classification category for Newark.

High $PM_{2.5}$ in the Newark area is associated with relatively high $PM_{2.5}$ two-days prior—in the Newark-Elizabeth area, as well as in the Camden-New Castle area and in Bethlehem, PA. Thus, a regional day-to-day build up of $PM_{2.5}$ is indicated for high $PM_{2.5}$ days.

The surface meteorological parameters indicate a correlation between higher $PM_{2.5}$ concentrations and higher temperatures (primarily reflecting seasonal differences), lower surface wind speeds, and higher relative humidity. Surface wind directions tend toward southerly, compared to westerly for the lowest PM range. There is no clear tendency with respect to wind speed or precipitation.

The upper-air meteorological parameters (based here on the Brookhaven, NY sounding) indicate that higher $PM_{2.5}$ concentrations occur with higher 850 mb temperatures. There is also some tendency for more stable (positive) lapse rates to be associated with higher $PM_{2.5}$ days. This is especially true for the 900 and 950 mb temperature differences.

Considering the upper-air winds, wind speeds are slightly lower aloft (especially for the analysis day); wind directions are similar for all three categories and, on average, southwesterly.

Finally, the cloud cover is less for the high PM days, the majority of which tend to occur, based on the season index, during the summer months.

The input parameters that are most used by CART in the construction of the classification tree (either to define the splits/branching structure) or as surrogates to the primary variables in this regard are highlighted in the table. These include surface temperature, relative humidity, 850 mb temperature, and 900 to surface temperature difference. All of these are also correlated with the $PM_{2.5}$ concentration for the analysis day. Newark is one of the few area for which relative humidity is a key CART parameter and varies regularly among the categories (increasing with increasing PM concentration).

4.11.3. Characteristics of High $PM_{2.5}$ Events

The categorical summary table provides a general overview of how average conditions vary across (and potentially lead to) different 8-hour ozone concentration levels for the Newark area. Within the high $PM_{2.5}$ categories, there are other key differences among the parameters that result in different types of high $PM_{2.5}$ events. We have used the CART results to examine these differences.

Only certain of the CART bins are frequently associated with $PM_{2.5}$ concentrations that are in the USG range or Category 3. Of these, we identified the bins with the most number of days as key bins. Table 4-27 considers the input parameter values for the key USG bins. For Newark there are two key bins and these contain 89 percent of the USG days.

Table 4-27. Summary of Mean Air Quality and Meteorological Parameters for Key USG CART Classification Bins: Newark

	Bin 34	Bin 13
Number of days	13	3
$PM_{2.5}$ Parameters		
24-hour $PM_{2.5}$ for Newark/Elizabeth (μgm^{-3})	43.0	46.9
Two-days-ago 24-hour $PM_{2.5}$ for Elizabeth (μgm^{-3})	21.8	22.8
Two-days-ago 24-hour $PM_{2.5}$ for Bethlehem (μgm^{-3})	22.8	18.1
Two-days-ago maximum 24-hour $PM_{2.5}$ for Camden and New Castle (μgm^{-3})	27.4	17.5
Surface Meteorological Parameters		
Maximum surface temperature (°C)	34.4	19.4
Minimum surface temperature (°C)	22.8	11.5
Surface relative humidity (%)	61.9	82.6
Surface wind speed (ms^{-1})	4.1	0.4
Surface wind direction (degrees)	175	0
Number of six hour periods with precipitation (range is 1 to 4)	0.2	0.0

	Bin 34	Bin 13
Upper-Air Meteorological Parameters (Brookhaven)		
850 mb temperature (AM) (°C)	17.0	9.4
850 mb temperature (PM) (°C)	18.8	10.1
Temperature gradient (850 mb to surface; AM) (°C)	-5.2	-1.8
Temperature gradient (900 mb to surface; AM) (°C)	0.2	1.3
Temperature gradient (950 mb to surface; AM) (°C)	2.0	2.7
24-hour difference in 700 mb geopotential height (m)	3.1	0.0
Yesterday's 700 mb wind speed (PM) (ms^{-1})	10.4	9.9
Yesterday's 850 mb wind speed (PM) (ms^{-1})	8.7	6.7
850 mb wind speed (AM) (ms^{-1})	9.4	4.1
850 mb wind speed (PM) (ms^{-1})	9.9	7.9
Yesterday's 700 mb wind direction (PM) (degrees)	270	180
Yesterday's 850 mb wind direction (PM) (degrees)	264	207
850 mb wind direction (AM) (degrees)	252	207
850 mb wind direction (PM) (degrees)	212	180
Estimated cloud cover (range of 1 to 3)	1.6	1.7
Seasonal indicator (1 = winter, 2 = transitional, 3 = summer)	3	2

The two key high PM bins represent transitional period and summer of PM events.

Days within Bin 3 (containing a majority of transitional period days) are associated with lower two-days-ago PM concentrations at the upwind sites but higher values at the local site. The PM concentrations are also higher, on average, on the analysis days, compared to days within Bin 34 (the summertime bin). Temperatures and surface wind speeds are much lower for days within Bin 3. Surface wind directions are also different for the two bins and are northerly for Bin 2 (winter) and southerly for Bin 29 (summer). The days within Bin 2 are also distinguished by slightly more stable lapse rates than days within Bin 34. Wind speeds aloft are lower for Bin 3 than for Bin 34. Wind directions aloft are southerly, on average, for days in Bin 3, and westerly to southwesterly for days within Bin 34. These differences are similar to the winter/summer differences for the key bins for the Philadelphia and Wilmington area, but less dramatic.

Next, we examine the conditions associated with each day or episode.

Data retrieval and availability for the Newark area were moderate for the period 1999-2002, and 18 USG days occurred during this period. The specific dates, including the observed $PM_{2.5}$ concentration (µgm^{-3}), are listed in Table 4-28.

Table 4-28. USG Days for Newark: 1999–2002

Date	Day of Week	CART bin	$PM_{2.5}$ (μgm^{-3})	USG day for other areas?
June 2, 2000	Friday	34	41.6	
June 10, 2000	Saturday	34	45.0	Philadelphia
June 11, 2000	Sunday	34	41.6	Baltimore, Bristol, Wilmington, Philadelphia
October 26, 2000	Thursday	13	54.6	Baltimore, Washington
October 27, 2000	Friday	30	77.7	Washington, Bristol
December 11, 2000	Monday	13	44.9	
June 14, 2001	Thursday	34	43.4	Wilmington
June 30, 2001	Saturday	34	46.4	Wilmington, Philadelphia
August 6, 2001	Monday	34	41.0	Baltimore, Washington, Wilmington
August 10, 2001	Friday	34	42.4	
March 15, 2002	Friday	30	40.6	
June 11, 2002	Tuesday	34	42.5	
June 26, 2002	Wednesday	34	42.1	
July 2, 2002	Tuesday	34	41.8	Baltimore, Washington, Richmond, Wilmington, Philadelphia
July 18, 2002	Thursday	34	43.4	Baltimore, Washington, Richmond, Charlotte, Wilmington, Philadelphia
July 19, 2002	Friday	13	41.7	Baltimore, Washington, Wilmington, Philadelphia
August 13, 2002	Tuesday	34	43.9	Baltimore, Washington, Richmond, Wilmington
August 14, 2002	Wednesday	34	44.0	

Although data retrieval for the Newark area was less than that for Philadelphia, Wilmington, Baltimore, and Washington, a number of high $PM_{2.5}$ events were measured during the 1999–2002 period.

The Newark area experienced high PM concentrations during the summertime episodes discussed above: August 6–10, 2001, July 1–4, 2002, July 17–19, 2002, and August 13–14, 2002.

Very high PM was measured during one fall episode in the MARAMA region during the period October 24–27, 2000. During this period, the area was influenced by a moderately strong upper-level ridge centered over the Midwest. A strong, persistent surface high-pressure system was centered directly over the mid-Atlantic states and gradually weakened and moved northeastward by the end of the period. Low temperatures were in the mid 50's, with highs in the low 70's. Partly cloudy skies and fog were reported in the early morning hours throughout the region. Surface winds were very light, reflecting stagnation conditions, allowing for a buildup of PM concentrations throughout the region. In addition to the Newark area, USG level concentrations were measured at Baltimore and Washington on October 26 and at Baltimore, Washington, and Bristol on October 27.

This review of the meteorological conditions indicates the high PM concentration occur under a variety of synoptic situations, that vary according to season. There are two key USG bins for Newark and these represent summertime and transitional-period conditions. Other high $PM_{2.5}$ days are placed in other high PM bins. CART thus appears to be able to distinguish and group the USG days quite effectively. Because of these differences, the categorical summaries should not be used to guide the forecasting, and instead the bin by bin characteristics must be considered.

This page intentionally left blank.

5. $PM_{2.5}$ Forecasting Tools

CART-based forecasting tools were developed for each of the areas of interest. The forecasting algorithms were based on the CART binary decision trees and supporting information. Each tool consists of an interface for the entry of observed and forecasted data and other parameters, the forecasting algorithms and supporting calculations for one or more areas, and several options for the display, summary, and storage/archival of the input parameters and the forecast results.

Tools were developed for each of three sets of CART results. These included a draft version of the “operational” tools (based on the Regional 2 CART analysis), a final version of the “operational” tools (based on the Regional 3 CART results), and a “research” version of the tools (based on the Research CART results).

For each of the three sets of CART results, four tools were developed for: 1) Charlotte; 2) Bristol, Roanoke, and Richmond; 3) Baltimore, Washington, Philadelphia, and Wilmington; and 4) Newark. When multiple areas are included, the user must select the forecast area and forecasts are prepared one area at a time. The combined tools facilitate the preparation of forecasts for multiple areas (using only one program) and also allow the upper-air data that is input for one area to be used in preparing the forecast for another without reentry.

The forecast tools are described in this section. An evaluation of the tools using real-time and historical data is also presented.

5.1. Description of the CART-Based Forecasting Tools

The following description of the CART-based $PM_{2.5}$ forecasting tools includes an overview of the concepts, input requirements, features, and output summaries.

5.1.1. Conceptual Overview

By providing detailed information about the classification of historical days into bins with different $PM_{2.5}$ concentration ranges based on the values of related meteorological and air quality parameters, the CART trees provide a basis for similarly classifying future days based on the observed and predicted values of these same parameters. Specifically, the observed data and forecast parameters corresponding to a future day are compared with the decision points that define the CART tree and assigned to one of the classification bins. The path taken through the CART tree and the resulting classification is determined by the values of the observed data and forecast parameters and the binary splits that comprise the classification tree. The forecasted $PM_{2.5}$ concentration is assigned the value of the CART bin into which the day is classified.

This approach to forecasting has several attributes. Compared to simple regression techniques, the use a CART-based forecasting algorithm accommodates the possibility that different meteorological conditions can lead to the same or similar $PM_{2.5}$ concentrations and, most importantly, that there may be multiple pathways to high $PM_{2.5}$. The parameter and parameter values associated with the CART classification tree provide information on the relative importance of the various air quality and meteorological parameters to the air quality conditions as represented by the dependent variable. Thus the CART technique offers additional physical insight into phenomena being studied. By segregating the data values into the classification bins, CART also provides information regarding the frequency of occurrence of the conditions associated with each classification category. In this manner, the likely recurrence rate for a particular type of day and the associated prevalent conditions are obtained.

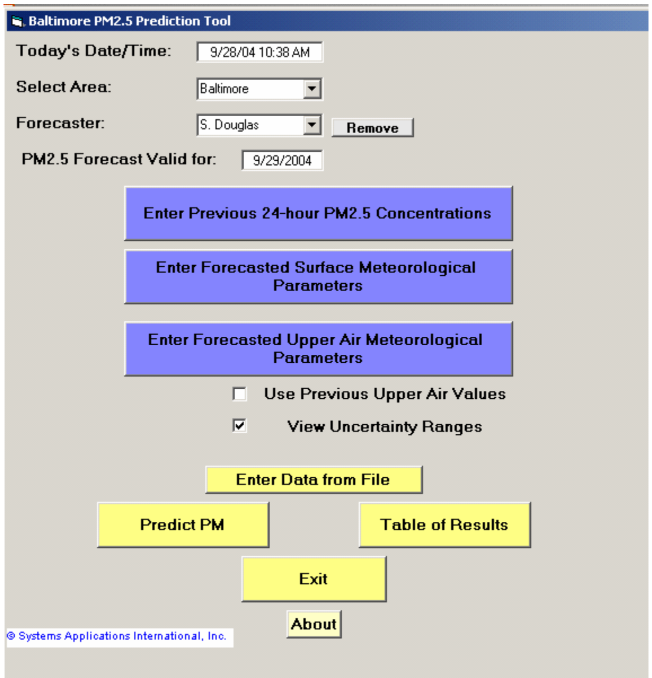
Two key assumptions come into play in the use of the CART result in this way. First, we assume that the relationships identified by CART and defined by the classification tree are physically meaningful. Our review and quality assurance of the CART outputs helps to ensure this, but it is important to keep in mind that CART is a statistical tool and not all of the identified statistical relationships can be confirmed to be physically meaningful (in part due to the complex nature of $PM_{2.5}$ formation and transport, and in part due to the complexity of the CART results). Next we assume that the CART application is complete with respect to representation of both the full range of different $PM_{2.5}$ regimes as well as the full set of input parameters needed to characterize the different regimes. Use of a limited dataset (in this case, a three- to four-year dataset) affects our ability to represent the range of regimes. The robustness of the input parameters is limited by the number and type of measurements, the spatial and temporal resolution of the “data”, and the quality of the “data” in both the historical and forecast modes.

5.1.2. Input Requirements

In discussing the input requirements, we begin with some basic information that is either supplied by the tool or must be supplied by the forecaster. Basic forecast elements such as the date and time at which the forecast is made “Today’s Date and Time” and the date for which the forecast is valid “PM2.5 Forecast Valid For” are supplied automatically by the tool. The forecast valid date is automatically set to tomorrow’s date but can be changed by the forecaster. The user may enter his or her name “Forecaster” and for the multiple-area tools, must select an area “Select Area.” The initial input screen for an example application for Baltimore is displayed in Figure 5-1.

All other input parameters are described in some detail in Section 2 of this report and in more operational terms in the next few subsections.

Figure 5-1. Initial Input Screen for the PM2.5 Forecasting Tool: Example for Baltimore



Today's Date/Time: 9/28/04 10:38 AM

Select Area: Baltimore

Forecaster: S. Douglas

PM2.5 Forecast Valid for: 9/29/2004

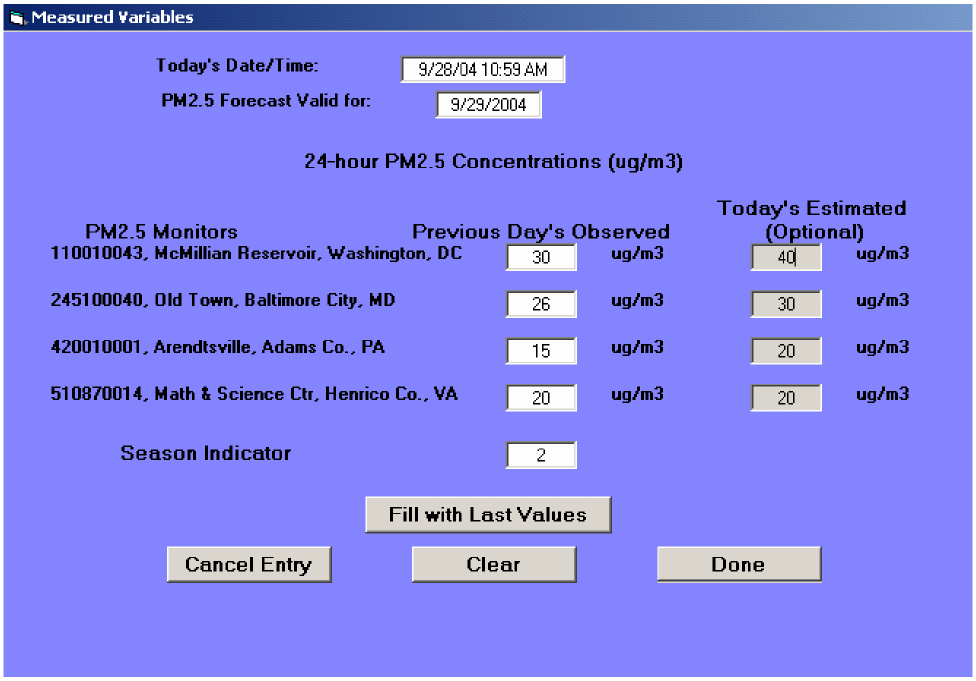
☐ Use Previous Upper Air Values

☑ View Uncertainty Ranges

PM and Other Input Parameters

The first input screen is for entry of the “Previous 24-hour PM2.5 Concentrations”. An example of this input screen is given in Figure 5-2. These inputs must be entered by hand. The user must provide the observed values for $PM_{2.5}$ for each site listed, for two days prior to the forecast day. We have also included a second column for estimated $PM_{2.5}$ values for the day prior to the forecast day. This information is required for the research version of the tool and is optional for the operational version of the tool. We suggest that consideration and entry of the one-day prior values may help with the review, interpretation and subsequent use of the CART-based forecast results.

Figure 5-2. Example Input Screen for PM2.5 Data



Today's Date/Time: 9/28/04 10:59 AM
PM2.5 Forecast Valid for: 9/29/2004

24-hour PM2.5 Concentrations (ug/m3)

PM2.5 Monitors	Previous Day's Observed		Today's Estimated (Optional)	
110010043, McMillian Reservoir, Washington, DC	30	ug/m3	40	ug/m3
245100040, Old Town, Baltimore City, MD	26	ug/m3	30	ug/m3
420010001, Arendtsville, Adams Co., PA	15	ug/m3	20	ug/m3
510870014, Math & Science Ctr, Henrico Co., VA	20	ug/m3	20	ug/m3

Season Indicator 2

It is important to keep in mind that the CART-based forecasting tools were developed using PM2.5 data from FRM measurement systems—as they are expected to provide the most consistent and accurate concentration values. However, because they are collected using filters, FRM data are typically not available until several weeks after the sampling date. Thus, forecasters must rely on continuous measurements of $PM_{2.5}$ (which are available on a real time basis) to provide information about prior day PM levels at local and upwind sites and to support the forecasting. There are several different types of instruments used to collect continuous data, and these do not always agree with the FRM measurements. The level of disagreement varies from site to site, and typically from season to season (with temperature and humidity), as discussed in some detail by Gillespie et al. (2004). The use of the real-time data from continuous measurement systems may be different enough from the FRM data under some circumstances to cause an erroneous forecast. For most areas, prior day PM2.5 concentrations were important to the CART analysis and thus to the forecasts - increasing the possibility that differences in the data types could contribute to forecast errors.

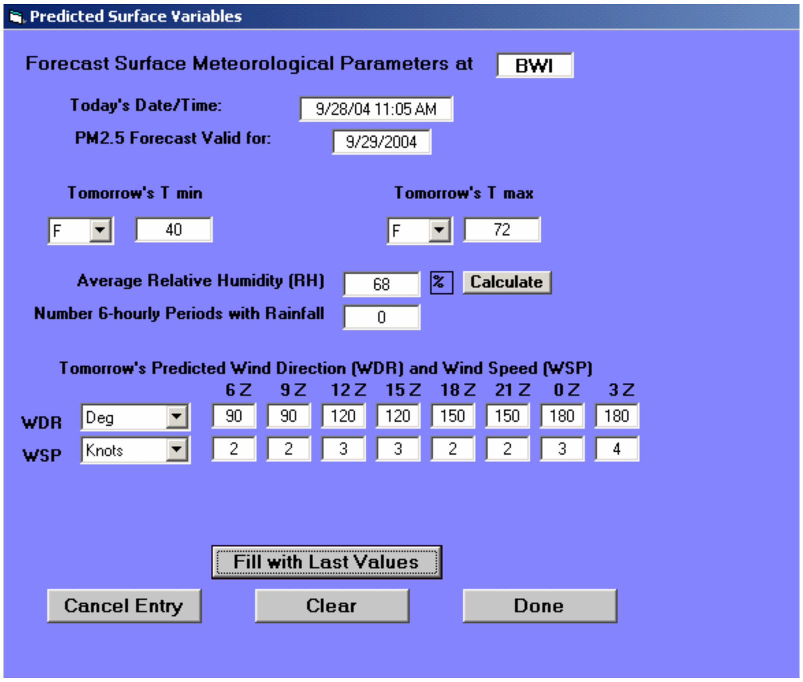
In specifying the prior-day $PM_{2.5}$, the forecaster should consider the whether the TEOM (or other real-time) data should be adjusted to account for differences between these data and the FRM data (as used in the underlying CART analysis).

The user must also specify the seasonal period of the forecast day. To account for seasonal variations in vegetative cover, there are three periods to choose from. The winter period includes November, December, January, February, and March. The transitional period includes April, May, September, and October. The summer period includes June, July, and August. This is an input rather than automatically generated to allow the user to choose different periods than appropriate for the date, for example, during transitional times or to accommodate unusual meteorological conditions such as drought.

Surface Meteorological Parameters

The second input screen is for entry of “Forecasted Surface Meteorological Parameters”. An example of this input screen is given in Figure 5-3. These inputs may either be entered by hand or using the automated data entry feature, as discussed in the next section on features of the tool. The surface meteorological inputs are listed and described in Table 2-5. Care should be taken to specify the correct units for each parameter, as appropriate. Relative humidity (daily average) can either be entered directly or calculated based on 3-hourly values of temperature and dew-point temperature. Note that the typical forecast products provide the surface values at three-hourly intervals. The expected meteorological monitoring site will appear at the top of the screen. For Washington, D.C., surface winds from Dulles Airport (IAD) are recommended.

Figure 5-3. Example Input Screen for Surface Meteorological Data



Predicted Surface Variables

Forecast Surface Meteorological Parameters at BWI

Today's Date/Time: 9/28/04 11:05 AM

PM2.5 Forecast Valid for: 9/29/2004

Tomorrow's T min: F 40

Tomorrow's T max: F 72

Average Relative Humidity (RH) 68 %

Number 6-hourly Periods with Rainfall 0

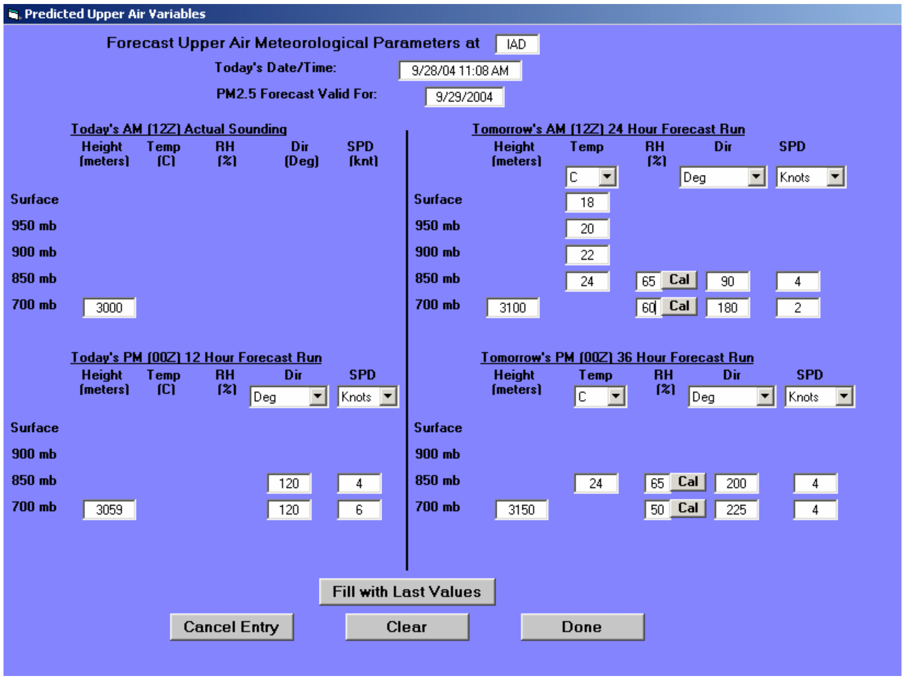
Tomorrow's Predicted Wind Direction (WDR) and Wind Speed (WSP)

		6 Z	9 Z	12 Z	15 Z	18 Z	21 Z	0 Z	3 Z
WDR	Deg	90	90	120	120	150	150	180	180
WSP	Knots	2	2	3	3	2	2	3	4

Upper-Air Meteorological Parameters

The third input screen is for entry of “Forecasted Upper-Air Meteorological Parameters”. An example of this input screen is given in Figure 5-4. These inputs may either be entered by hand or using the automated data entry feature, as discussed in the next section on features of the tool. The upper-air meteorological inputs are listed and described in Table 2-6. Care should be taken to specify the correct units for each parameter, as appropriate. Relative humidity can either be entered directly or calculated based on predicted values of temperature and dew-point temperature. The expected meteorological monitoring site will appear at the top of the screen. The entries are organized in chronological order and then by level (with increasing vertical height) for each required variable.

Figure 5-4. Example Input Screen for Upper-Air Meteorological Data



Forecast Upper Air Meteorological Parameters at IAD

Today's Date/Time: 9/28/04 11:08 AM

PM2.5 Forecast Valid For: 9/29/2004

Today's AM (12Z) Actual Sounding

	Height (meters)	Temp (C)	RH (%)	Dir (Deg)	SPD (knt)
Surface					
950 mb					
900 mb					
850 mb					
700 mb	3000				

Tomorrow's AM (12Z) 24 Hour Forecast Run

	Height (meters)	Temp	RH (%)	Dir	SPD
		C		Deg	Knots
Surface		18			
950 mb		20			
900 mb		22			
850 mb		24	65	90	4
700 mb	3100		60	180	2

Today's PM (00Z) 12 Hour Forecast Run

	Height (meters)	Temp (C)	RH (%)	Dir	SPD
				Deg	Knots
Surface					
900 mb					
850 mb				120	4
700 mb	3059			120	6

Tomorrow's PM (00Z) 36 Hour Forecast Run

	Height (meters)	Temp	RH (%)	Dir	SPD
		C		Deg	Knots
Surface					
900 mb					
850 mb		24	65	200	4
700 mb	3150		50	225	4

5.1.3. *Features*

Automated Data Entry

The surface and upper-air meteorological inputs can be entered by hand or can be read in from external data files. For the MARAMA project, surface and upper-air meteorological inputs are prepared on a daily basis by meteorologists from the Pennsylvania Department of Environmental Protection (DEP) and posted to a MARAMA forecaster’s web site (S. Nolan, personal communication). There are currently three options for the obtaining the surface input parameters from the web site. The parameters are derived from the output for three different models including the NWS ETA model, the Global Forecast Systems (GFS) model, and the Nested Grid Model (NGM). The upper-air parameters are currently available for the ETA model only. The parameters, levels, and units are designed to match those required by the forecast tools.

Other Input Related Features

For tools that contain multiple areas, the upper-air data for a given upper-air monitoring site need not be entered twice. Instead, the user can check the box on the first form that is labeled “Previous Upper Air Values” to use the last entered data for the assigned upper-air site.

At the bottom of each data input screen is a box labeled “Fill with Last Values.” This option allows the user to quickly make changes to one or more of the previously entered input parameters. This feature allows the forecaster to explore how small changes in one or more of the input parameters affect the forecast result.

Once the data for each category have been entered the box on the first screen will change color. When the inputs for all three categories have been provided by the user, the tool is ready to prepare a forecast.

Forecast Probabilities

The CART-based probabilities associated with the forecast bins are reported as part of the forecast. These characterize the probability for a day within the bin to belong to the classification category to which that bin is assigned or to belong to another classification category. This takes into account the number of days within the bin, weighted by the observed data distribution and the misclassification costs.

Forecast Range

The forecasting accuracy will depend upon the accuracy of the input data and, in particular, the meteorological forecasts. Errors or uncertainties in the meteorological forecasts will translate into errors or uncertainties in the $PM_{2.5}$ forecasts. To address the issue of uncertainty in the meteorological input data and its effect on the $PM_{2.5}$ forecast, we have included an uncertainty feature. This feature can be selected by checking the “View Uncertainty Ranges” box on the first form.

The uncertainty feature allows the user to run the forecast and obtain results for two alternate forecast scenarios. For the “High” forecast, the parameters are adjusted to be generally more conducive to higher $PM_{2.5}$ concentrations as follows:

- Wind speeds reduced by 0.5 ms^{-1}
- Temperatures increased by 1.5°C
- Temperature differences (stability parameters) increased by 0.5°C.

For the “Low” forecast, the parameters are adjusted to be generally more conducive to higher $PM_{2.5}$ concentrations as follows:

- Wind speeds increased by 0.5 ms^{-1}
- Temperatures lowered by 1.5°C
- Temperature differences (stability parameters) decreased by 0.5°C.

The objective of this feature is to allow the user to assess the potential uncertainty of the forecast due to uncertainties in the meteorological forecasts and rounding of the meteorological forecast data. A result matching the main prediction result indicates that small uncertainties in the meteorological forecast will not affect the predicted $PM_{2.5}$ level, but a change in either low or high $PM_{2.5}$ colors and bins with respect to the main prediction indicates that the prediction is subject to uncertainty. This feature is intended to provide perspective regarding the sensitivity of the forecast to small errors or uncertainties in the meteorological forecasts.

5.1.4. Outputs

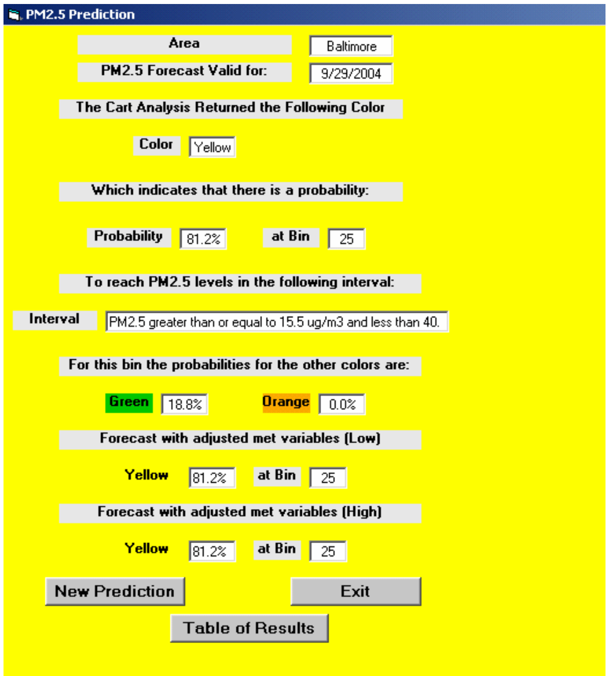
Once the data are entered, select “Predict PM” to obtain a forecast for 24-hour $PM_{2.5}$ for tomorrow (using the color-based air quality index (AQI)). The inputs and results will be presented on the screen and also summarized in a table. These tables can be used to check the inputs and to record the inputs and outputs.

Forecast Result

A primary output of the tool is the CART bin number (the bin into which the forecast day was placed) and the corresponding $PM_{2.5}$ concentration range for that bin. The forecast colors and ranges are as follows: Green (less than 15.5 μgm^{-3}), Yellow (15.5 to less than 40.5 μgm^{-3}), and Orange (greater than or equal to 40.5 μgm^{-3}). These correspond to “Good”, “Moderate”, and “USG”, forecasts. The colors and ranges are indicated in the output. An example forecast result in given in Figure 5-5.

The forecast also includes the probabilities associated with the bin and, if requested by the user, the bin number and corresponding $PM_{2.5}$ range for the high and low forecasts.

Figure 5-5. Example Forecast Result Screen



Area: Baltimore

PM2.5 Forecast Valid for: 9/29/2004

The Cart Analysis Returned the Following Color

Color: Yellow

Which indicates that there is a probability:

Probability: 81.2% at Bin: 25

To reach PM2.5 levels in the following interval:

Interval: PM2.5 greater than or equal to 15.5 ug/m3 and less than 40.

For this bin the probabilities for the other colors are:

Green: 18.8% Orange: 0.0%

Forecast with adjusted met variables (Low)

Yellow: 81.2% at Bin: 25

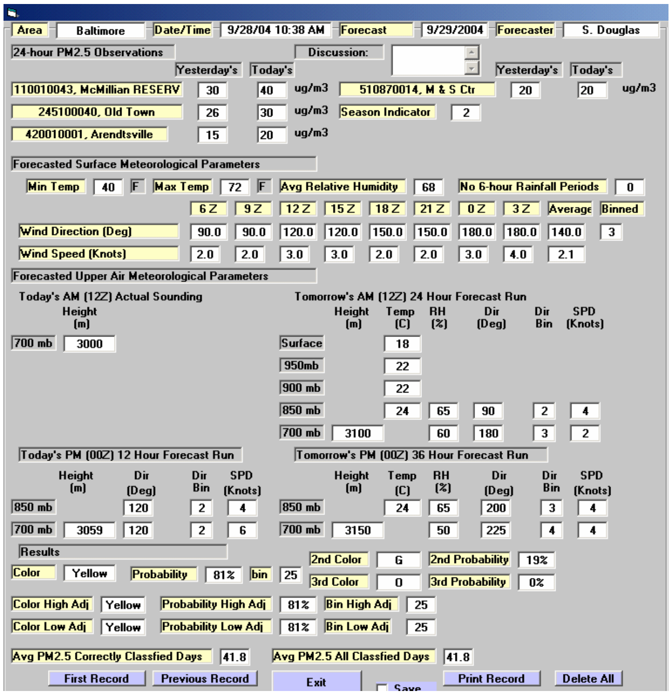
Forecast with adjusted met variables (High)

Yellow: 81.2% at Bin: 25

Summary of Results Table

The results table summarizes the various input parameters as well as the forecast result and supporting information. An example summary of results table is given in Figure 5-6. In addition, the average values for (1) all correctly classified days within the bin and (2) all days within the bin are given in the summary table. These values are based on the historical days that were place in that bin, and may provide some additional perspective to the forecast range. Space is provided for the user to enter forecast notes into the summary of results table.

Figure 5-6. Example Summary of Results Table



Area: Baltimore
Date/Time: 9/28/04 10:38 AM
Forecast: 9/29/2004
Forecaster: S. Douglas

24-hour PM2.5 Observations

	Yesterday's	Today's	
110010043, McMillian RESERV	30	40	ug/m3
245100040, Old Town	26	30	ug/m3
420010001, Arendtsville	15	20	ug/m3
510870014, M & S Ctr	20	20	ug/m3

Discussion:

Season Indicator: 2

Forecasted Surface Meteorological Parameters

Min Temp: 40 F
Max Temp: 72 F
Avg Relative Humidity: 68
No 6-hour Rainfall Periods: 0

	6 Z	9 Z	12 Z	15 Z	18 Z	21 Z	0 Z	3 Z	Average	Binned
Wind Direction (Deg)	90.0	90.0	120.0	120.0	150.0	150.0	180.0	180.0	140.0	3
Wind Speed (Knots)	2.0	2.0	3.0	3.0	2.0	2.0	3.0	4.0	2.1	

Forecasted Upper Air Meteorological Parameters

Today's AM (12Z) Actual Sounding

	Height (m)
700 mb	3000

Tomorrow's AM (12Z) 24 Hour Forecast Run

	Height (m)	Temp (C)	RH (%)	Dir (Deg)	Dir Bin	SPD (Knots)
Surface		18				
950mb		22				
900 mb		22				
850 mb		24	65	90	2	4
700 mb	3100		60	180	3	2

Today's PM (00Z) 12 Hour Forecast Run

	Height (m)	Dir (Deg)	Dir Bin	SPD (Knots)
850 mb		120	2	4
700 mb	3059	120	2	6

Tomorrow's PM (00Z) 36 Hour Forecast Run

	Height (m)	Temp (C)	RH (%)	Dir (Deg)	Dir Bin	SPD (Knots)
850 mb		24	65	200	3	4
700 mb	3150		50	225	4	4

Results

Color: Yellow
Probability: 81%
bin: 25
2nd Color: G
2nd Probability: 19%
3rd Color: 0
3rd Probability: 0%
Color High Adj: Yellow
Probability High Adj: 81%
Bin High Adj: 25
Color Low Adj: Yellow
Probability Low Adj: 81%
Bin Low Adj: 25
Avg PM2.5 Correctly Classfied Days: 41.8
Avg PM2.5 All Classfied Days: 41.8

Archiving the Outputs

The tabular summaries are automatically saved within the database tool for each forecast. From the tool, the user may view the summary tables for previous forecasts, by paging through the archive of summary tables.

Selected tabular summaries can be exported to an Excel file, by checking the “Save” box for each table of information that is to be exported and then clicking on “Save As.” Only those tables that are checked will be exported.

All summary tables can be deleted using the “Delete All” button. This will clear all outputs from the tool.

5.2. Evaluation

In this section, we describe the methods and results of the evaluation of the draft version of the operational tools that was performed to air their refinement and the subsequent development of the final versions of the tools. The evaluation discussed here concerns the tool's predictive accuracy, which depends on the CART tree itself, and not the user interface or other elements not based on CART but also evaluated and improved throughout the project.

5.2.1. Real-time Evaluation

Meteorologists in six of the nine MARAMA areas tested the draft versions of the operational $PM_{2.5}$ forecasting tools during February and March of 2004. For as many days as possible, each participant input the measured and forecasted meteorological and air quality data required by the tool to predict the next day's $PM_{2.5}$ level. "Good" days have maximum $PM_{2.5}$ concentrations less than 15.5 μgm^{-3}, "moderate" days have concentrations greater than or equal to 15.5 and less than 40.5 μgm^{-3}, and "USG" days have concentrations of 40.5 μgm^{-3} or above. The CART predictions were recorded and sent to a single person, who consolidated the results for each site and compared these to $PM_{2.5}$ observations from continuous monitors within each area. In this comparison, one continuous monitor was selected to represent each area: Garinger for Charlotte, Math & Science Center for Richmond, McMillan Reservoir for Washington, D.C., Old Town for Baltimore, Camden for Philadelphia, and MLK for Wilmington. Later, the CART predictions were also compared to quality-assured FRM data, which is compiled some time later than the continuous data. The data used for the evaluation are area-wide maximums over several FRM monitors within the area, as similar as possible to the area-wide maximums used to characterize each site during the pre-tool CART analysis. Thus this second comparison is closest to evaluating what the CART trees were originally designed to predict. At the time of the study, four MARAMA areas had sufficient first-quarter 2004 FRM data to undergo this second evaluation: Baltimore, Charlotte, Richmond, and Wilmington.

Several metrics were used to compare the $PM_{2.5}$ forecasting tool predictions to the observed continuous or FRM data. A simple matrix tallied how many days observed in each $PM_{2.5}$ category were forecast into each level. Accuracy, false alarm rate, probability of detection, critical success index (threat score), and bias statistics were derived from this information. The false alarm rate equals the percent of predicted USG days that did not turn out to be USG. The probability of detection equals the percent of observed USG days that were predicted to be USG. The critical success index is the number of successfully predicted USG days divided by the sum of false USG predictions and unpredicted USG days, and the bias is the ratio of number of predicted USG days over the number of observed USG days. In practice, these last three metrics were rarely of use since USG observations only occurred in two instances, both for Baltimore using FRM data. Therefore the false alarm rate and accuracy were the most informative measures, and the latter for the most part measured the tools' ability to tease out Good and Moderate days. A tool for the calculation of these metrics was provided by M. Seybold from the Maryland Department of the Environment (MDE).

The metrics described above were applied to the six areas in two different ways. The first, "strict" evaluation is a straightforward comparison of predicted and observed PM levels using the metrics described above. The second, "fuzzy-border" evaluation represents a best-case scenario by counting predictions as correct if the observed PM concentration fell within a designated border zone between the observed and predicted PM levels. For example, a Moderate prediction would be counted as correct even if the observed value is 14 $\mu g/m^3$, a little below the cut-off of 15.5 $\mu g/m^3$. The border zones are defined as follows: Good and Moderate predictions are both correct for concentrations greater than or equal to 13.5 $\mu g/m^3$ and less than or equal to 17.5 $\mu g/m^3$; Moderate and USG predictions are both correct for concentrations greater than or equal to 36.5 $\mu g/m^3$ and less than or equal to 44.5 $\mu g/m^3$.

Results of the “strict” and “fuzzy-border” evaluations are described below for the evaluation with continuous $PM_{2.5}$ observations and for the evaluation with FRM observations.

Real-time Evaluation Using Continuous PM Observations

Table 5-1 below provides statistics for the six sites evaluated for their ability to predict $PM_{2.5}$ levels indicated by a local continuous monitor. In addition to the accuracy and false alarm statistics, the table lists the number of days evaluated and the percentage of these days with “Good” PM levels. No USG days were observed, so the remainder of the days are all Moderate. Because no USG days were observed, the bias, critical success index, and probability of detection metrics were not included in the chart.

Table-5-1. Evaluation Metrics for PM Tools Using Continuous PM Observations

MARAMA Area—Monitor	No. Days	% Good Days	Accuracy (Strict)	False Alarm Rate (Strict)	Accuracy (Fuzzy)	False Alarm Rate (Fuzzy)
Charlotte, NC—Garinger	35	80%	66%	na*	71%	na*
Richmond, VA—Math & Sci. Ctr	38	79%	74%	na*	87%	na*
Washington, DC—McMillan	32	84%	75%	100%	88%	100%
Baltimore, MD—Oldtown	34	59%	68%	na*	77%	na*
Philadelphia, PA—Camden	29	59%	55%	na*	75%	na*
Wilmington, DE—MLK	37	35%	73%	na*	81%	na*

**No USG days predicted or observed*

Prediction accuracy ranges from 55 to 75 percent under the strict evaluation, and from 75 to 88 percent under the fuzzy evaluation. It is important to keep in mind in reviewing these percentages that all of the days exhibited low (good) or moderate $PM_{2.5}$ levels. One way to evaluate the predictive ability of the tools is to compare the accuracy to the accuracy if one had simply predicted all Good days (or all Moderate, in the case of Wilmington). Compared to the results using only one consistent forecast, the forecasting tool for Wilmington does a good job of predicting PM levels, the Baltimore tool does fairly well, the Philadelphia, Richmond, and Washington tools do barely well, and Charlotte does not do well at all. But this is a naive measure since prediction of very high PM days most concerns the forecaster, rather than the distinction between Good and Moderate. No high PM days were observed at the continuous monitors in February and March of 2004, fortunately for air quality but unfortunately for tool evaluation.

Real-time Evaluation Using FRM PM Observations

Table 5-2 below provides statistics for the four sites evaluated for their ability to predict $PM_{2.5}$ levels indicated by the maximum $PM_{2.5}$ concentration over several FRM monitors selected from the area. As in the previous section, the table gives the percentage of Good days according to the FRM data. No USG days were predicted for these four areas during the period, so false alarm rates are not shown.

Observed USG days appeared only for Baltimore; these two days were classified as Moderate by the forecasting tool so the table shows 0 percent as the detection probability; the critical success index and bias for Baltimore are also zero under strict evaluation, and nonexistent under fuzzy evaluation as the two USG days were below 44 µg/m^3 and therefore are almost Moderate. Fifty-nine percent of Baltimore's days had Moderate PM levels, according to the FRM data.

Table 5-2. Evaluation Metrics for PM Tools Using FRM PM Observations

MARAMA Area	No. Days	% Good Days	Accuracy (Strict)	Detection Prb. (Strict)	Accuracy (Fuzzy)	Detection Prb. (Fuzzy)
Charlotte, NC	35	60%	57%	na*	69%	na*
Richmond, VA	38	78%	78%	na*	89%	na*
Baltimore, MD	34	35%	50%	0%	65%	na*
Wilmington, DE	37	52%	84%	na*	95%	na*

**No USG days predicted or observed*

The PM$_{2.5}$ forecasting tools for Richmond and Wilmington appear to do a genuinely good job, although during this period their ability to predict high PM days remained untested. Agreement with the FRM data is better than with the continuous data in both cases. Forecasting ability is fair for Baltimore and Charlotte, regardless whether strict or fuzzy-border evaluations are considered. Agreement with the FRM data is worse than with the continuous data in both cases. The greatest changes in performance when the FRM data area used appear for Wilmington and Baltimore. The PM$_{2.5}$ levels for Wilmington tended to be lower according to the FRM monitors than according to the continuous monitor, whereas the opposite is true for Baltimore; this suggests uncertainty in actual PM concentrations, something to consider while evaluating PM$_{2.5}$ forecasting tools in real time.

5.2.2. Historical Period Evaluation

Historical data enabled evaluation of the forecasting tools for all nine areas. The same "strict" and "fuzzy-border" procedures described above were applied to the period of June through August, 2003, by running the data for these months through the classification tree using CART software rather than the forecasting tool. The summer 2003 data were prepared for CART in almost the same way the 1999–2002 data were prepared in creation of the original CART trees. The only difference was that some alternate FRM sites were used for the 2003 dependent value data, in instances where the original FRM monitor was shut down and replaced with another. So the observed data in this evaluation are more like the FRM data than the continuous data of the real-time comparisons described above; the 2003 and 2004 FRM-based PM datasets mirror as closely as possible the original 1999–2002 PM$_{2.5}$ data classified by CART.

The advantage of this method is that one can swiftly evaluate the tree using many datapoints (around ninety days for most of the areas). On the other hand, the evaluation is not exactly the same as if it were conducted using the PM forecasting tool, because of CART's use of "surrogate splits." The PM forecasting tools are based on the "primary splits" at the nodes of the decision trees created by CART. However, the CART tree also stores information on surrogate splits, which are rules for classification that are applied if the meteorological or air quality variable used at the primary split is missing. In a real-time forecasting context, there are no missing variables because the forecaster can fill in datapoints with predictions or estimates. For the historical period evaluation described here, missing datapoints were not filled in and so CART resorted to surrogate data when necessary. This should be kept in mind when assessing the results of this subsection. Although the historical period evaluation may not use the exact same predictions the tools would have yielded in a real-time application, the predictions are probably similar. The results presented here are also of interest because, unlike the real-time 2004 evaluation, the summer 2003 period provides ample USG days to test the tools' ability to accurately predict high PM; furthermore, this assessment covers all nine MARAMA areas.

Table 5-3 provides several metrics for both the strict and fuzzy-border evaluations. Since the summer 2003 days were better distributed over Good, Moderate, and USG, the percentage of Good days is not given in the table as it does not provide the most useful comparison in this case. Because there are USG days for most areas, the accuracy, false alarm rate, probability of detection, critical success index, and bias are all informative measures of the tools' predictive utility. The table also provides the number of days evaluated for each area, as well as the number of strictly USG days.

Table 5-3. Evaluation Metrics for CART Historical Period Evaluation

The metrics are: Accuracy (Acc), FAR (False Alarm Rate), DetP (Probability of Detection), CSI (Critical Success Index), and Bias.

MARAMA Area	Days / USG Days	Strict Evaluation					Fuzzy-border Evaluation				
		Acc	FAR	DetP	CSI	Bias	Acc	FAR	DetP	CSI	Bias
Charlotte, NC	90 / 2	59%	na*	0%	0.00	0.00	72%	na*	na*	na*	na*
Bristol, VA	32 / 1	47%	100%	0%	0.00	1.00	53%	100%	0%	0.00	1.00
Roanoke, VA	31 / 1	52%	100%	0%	0.00	1.00	65%	100%	na*	na*	na*
Richmond, VA	89 / 2	71%	na*	0%	0.00	0.00	85%	na*	na*	na*	na*
Washington, DC	92 / 5	62%	88%	40%	0.11	3.40	71%	82%	50%	0.18	2.83
Baltimore, MD	92 / 8	57%	75%	25%	0.17	1.00	75%	50%	80%	0.80	1.60
Philadelphia, PA	85 / 4	66%	57%	75%	0.60	1.75	74%	57%	100%	0.75	2.33
Wilmington, DE	82 / 1	65%	88%	100%	0.14	8.00	85%	75%	100%	0.33	4.00
Newark, NJ	70 / 3	56%	80%	67%	0.22	3.33	63%	80%	100%	0.25	5.00

**Measure cannot be applied since no USG days were predicted and/or observed*

Predictive accuracy ranges from 47 to 71 percent if strict PM classifications are used, and 53 to 85 percent in the best-case scenario where borderline observations do not count against the tool. All areas except Charlotte have a bias (ratio of predicted to observed USG days) greater than one, and thus tend to overprediction, a fact also evident in the high false alarm rates. On the other hand, the probability of detection is fair to good for Newark, Philadelphia, and Wilmington, but problematic for the other areas. If only very high USG days are considered (and borderline USG days are counted as Moderate, according to the border zone definitions given earlier in this section), the probability of detection is good or inapplicable for most sites, but still a problem for Bristol and Washington.

5.2.3. Conclusions

For a first attempt at developing a CART-based forecasting tool for these nine areas—the results are promising. The evaluation statistics are lower than but not that much lower than those that would be considered good for 8-hour ozone forecasting (and ozone is a simpler and much more extensively measured/studied pollutant).

The real-time testing of the draft version of the forecasting tools was inconclusive primarily because the period February-March 2004 did not contain any days with high $PM_{2.5}$ concentrations.

The historical evaluation suggests that given perfect forecasts of the meteorological input parameters, the $PM_{2.5}$ concentration ranges can be correctly predicted for 50 to 70 percent of the days and nearly corrected predicted (using the “fuzzy-border” adjustment) for 65 to 85 percent of the days (with the exception of Bristol, which has a 55 percent accuracy even with the adjustment).

In this historical evaluation, two of the sites with the worst performance are Bristol and Roanoke and these both had fewer data (with an every three day collection interval) than the other sites. Yet the CART trees for two sites had some of the best overall classification accuracy. This outcome suggests that the CART results, while good for characterizing the days in the dataset, are incomplete with respect to representing all of the types of conditions that might occur at these sites. The implication is that use of a limited dataset may limit the predictive ability of the tools, if conditions that are not represented in the dataset occur. This could extend to all areas and the use of the nominal three- to four-year analysis period.

The false alarm rate was relatively high for all areas, where it could be calculated, and this reflects the tendency for overestimation found in the CART trees. With this tendency, the probability of detection is good for most sites, and the bias is positive in all cases for which it could be calculated. This outcome suggests that the meteorological inputs and consequently the CART results may not sufficiently represent the conditions associate with the day-to-day transition from high to lower PM concentrations. The overpopulation of the higher PM bins with lower PM days (both in the CART results and in the historical forecast results) may also be due to a lack of a sufficient number of high PM days in the dataset, as needed to allow a good sampling and representation of the conditions that are associated with these days.

5.3. Operational and Research Versions of the Tools

The evaluation results, per se, did not lead to major revision of the tools. However, practice in using the tools, further consideration of the input parameters, and a few case studies by the state forecasters resulted in a few additions to the inputs. This further development of the tools is discussed in Section 2 of this report (CART diagnostic and sensitivity analysis) and resulted in a revised operational version of the tools as well as a research version of the tools that includes an estimated $PM_{2.5}$ concentration for the day prior to the forecast day.

This page intentionally left blank.

6. Summary and Recommendations

In this study, we developed a series of CART-based $PM_{2.5}$ forecasting tools for nine areas of interest in the MARAMA region including: Charlotte, Bristol, Roanoke, Richmond, Washington, Baltimore, Philadelphia, Wilmington, and Newark. The study included the application of CART and the development, testing, and evaluation of interactive forecasting tools for each area. Data and information gathered throughout the course of the project were used, together with the CART analysis results, to describe the relationships between meteorology and PM concentration and, specifically, the conditions associated with high $PM_{2.5}$ events in each of the areas. Based on the results and findings of the study, as well as the issues and problems that we encountered in conducting the work, we provide recommendations for future enhancement of the forecasting tools and an improved understanding of PM2.5 issues in this section.

The following recommendations pertain to the application of CART for $PM_{2.5}$:

- Update the input datasets to include additional years/seasons in order to better capture the range of different meteorological/$PM_{2.5}$ conditions that are likely to occur in the future as well as to better characterize the conditions associated with the high PM days (which were few in number during the analysis period for several of the areas).
- Using the expanded dataset with more high PM days, conclusively explore the use of alternative prior-day $PM_{2.5}$ concentration parameters for local and upwind sites, using both two-days-ago measured concentrations and prior-day estimated concentrations. It is intuitive that more information about the prior-day PM concentrations should improve the forecasting ability of CART, but our current work found the use of this information problematic (and resulted in the overestimation of $PM_{2.5}$ concentrations).

Additional recommendations pertain to the CART-based forecasting tools:

- Consistently (across the areas of interest) evaluate the forecasting tools for a longer period of time than was accommodated by this study. With a longer evaluation period, we may be able to identify specific patterns or types of PM events that are consistently missed by the CART-based forecasting tools. Combining and inter-comparing the evaluation results for the various areas of interest will aid the identification of missing parameters or information that is needed to capture the types of events that are consistently missed.
- Use the forecast evaluation results to reassess the uncertainty ranges used in the forecasting tools. These account for uncertainties in the input data (especially the meteorological forecasts) and their potential effects on the forecast.
- Evaluate and compare the use of the different meteorological forecast products (for example, ETA, GFS, and NGM).
- Add the capability for multi-day forecasts.
- Conduct detailed case-study analyses for as many of the high PM days as possible and compare the meteorologist perspective on important processes and parameters for the event with those used by CART to classify each day (i.e. generate the forecast)

Additional recommendations concern the improved understanding of the factors influencing $PM_{2.5}$ concentrations within each area of interest:

- Intermittently update the data summaries to include additional years/seasons of data.

- Examine, using available STN data, variations in species distributions among the CART bins and/or other groupings of the high PM days. This would need to be done using a larger dataset than that used for the current study – due to the more limited availability of STN data for the areas of interest.

Our final recommendations address the possible use of the data and results of this study to enhance $PM_{2.5}$ State Implementation Plan (SIP) analysis. For starters, the data analysis results for this study provide the basis for developing a conceptual description of $PM_{2.5}$ formation and transport for each area, which is a required component of a SIP. In addition, a key element of a $PM_{2.5}$ attainment demonstration is the “weight-of-evidence” analysis, in which data and modeling results are use to support or corroborate the outcome of the demonstration. The data analysis and CART results could be used to support the following types of weight-of-evidence analyses:

- Characterization of actual or proposed modeling episode periods in terms of their ability to represent typical meteorological conditions for each of the areas of interest. This would be determined based on the analysis of factors influencing PM2.5 in each area and the CART-based frequency of occurrence of the different types of meteorological conditions. This information could be used to guide the selection of an appropriate simulation period for the application of regional-scale particulate models, the selection of subset modeling episode periods for detailed analysis of certain areas, and the application of the modeled attainment test for $PM_{2.5}$.
- Analysis of data-based and meteorologically adjusted trends (adjusted using CART-based meteorological frequency information). Meteorologically adjusted trends, coupled with information about changes in emissions during the analysis period, could be used to assess the reasonableness of modeling results (i.e., the response of the model to similar emissions changes) and to project future changes in PM concentrations for the region.
- Calculation of meteorologically adjusted PM2.5 design values for use in the application of the $PM_{2.5}$ attainment test for each area of interest. Information from the CART analysis could be used to define a typical year (based on the frequency of occurrence of certain types of meteorological conditions) and the $PM_{2.5}$ design values corresponding to a typical multi-year period (based on actual observations).
- Use available Speciated Trends Network (STN) data in conjunction with the CART results to determine the species compositions of the most frequently occurring types of high PM2.5 events. This could help guide the identification of effective control options for the areas of interest or, in the context of weight-of-evidence, the interpretation and use of any modeling results.

References

AER, (2001) “Scoping Study for Regional Haze in the Upper Midwest,” Prepared for the Lake Michigan Air Directors Consortium, Des Plaines, IL, (Document No. CP089-01-1), February 2001.

Brieman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees.* Wadsworth, Belmont, California.

Deuel, H. P., and S. G. Douglas. 1998. “Episode Selection for the Integrated Analysis of Ozone, Visibility, and Acid Deposition for the Southern Appalachian Mountains.” Systems Applications International, Inc., San Rafael, California (SYSAPP-98/07r1).

Douglas, S. G., J. B. Mangahas, A. B. Hudischewskyj, G. L. Glass, and W. S. Hartley. 2003. “Examination of Inter-Variable Relationships and Implications for Modeling Episode Selection Using the 1999/2000 SEARCH Database,” Systems Applications International, Inc., San Rafael, California (03-060).

EPA. (1998) “Atmospheric Observations: Helping Build the Scientific Basis for Decisions Related to Airborne Particulate Matter,” Report of the PM Measurements Research Workshop, Chapel Hill, North Carolina, 22–23 July 1998.

EPA. 2000. “Guidance for Demonstrating Attainment of Air Quality Goals for $PM_{2.5}$ and Regional Haze.” Draft Version 1.1.

Gillespie, W. G. 2004. “Correlating Federal Reference Method and Continuous $PM_{2.5}$ Monitors in the MARAMA Region.“ Final Draft Report, MARAMA, Baltimore, Maryland (040131).

Malm, W. C., (1999) *Introduction to Visibility* Cooperative Institute for Research in the Atmosphere (Cooperative Agreement CA2350-97-001), NPS Visibility Program, Colorado State University, Fort Collins, CO, May 1999.

McMillan, N. 2004. *Evaluation of* $PM_{2.5}$ *Surrogate Methods,* Battelle, Presented at the EPA 2004 National Air Quality Conference, Baltimore, Maryland, 22–25 February.

Seinfeld, J.H. (1986) *Atmospheric Chemistry and Physics of Air Pollution* John Wiley & Sons.

Steinberg, Dan and Phillip Colla. “CART—Classification and /regression Trees.” San Diego, CA: Salford Systems, 1997.

Watson J. G., (2002) *2002 Critical Review—Visibility: Science and Regulation,* J. of Air and Waste Management Association, Volume 52, 626-713.

This page intentionally left blank.