Guidelines for Developing an Air Quality (Ozone and PM$_{2.5}$) Forecasting Program

Guidelines for Developing an Air Quality (Ozone and PM$_{2.5}$) Forecasting Program

DISCLAIMER

This report was prepared as a result of work sponsored, and paid for, in whole or part, by the U.S. Environmental Protection Agency (EPA). The opinions, findings, conclusions, and recommendations are those of the authors and do not necessarily represent the views of the EPA. The EPA, its officers, employees, contractors, and subcontractors make no warranty, expressed or implied, and assume no legal liability for the information in this report. The EPA has not approved or disapproved this report, nor has the EPA passed upon the accuracy of the information contained herein.

ACKNOWLEDGMENTS

This page is intentionally blank.

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Concluded)

This page is intentionally blank.

LIST OF TABLES

This page is intentionally blank.

LIST OF ACRONYMS

Term	Meaning
AQI	Air Quality Index
A_{ref}	Accuracy of a reference forecast
b_{abs}	Particle absorption
BAM	Beta Attenuation Monitor
CALGRID	California Grid Model
CAMx	Comprehensive Air Quality Model with Extensions
CART	Classification and Regression Tree
CBL	Convective Boundary Layer
CMAQ	Community Multiscale Air Quality Model
CSI	Critical Success Index
DGV	Geometric mean diameter by volume
EC	Elemental Carbon
EMS-95	Emissions Modeling System
EPA	U.S. Environmental Protection Agency
EPS 2.0	Emission Processing System
FAR	False Alarm Rate
FRM	Federal reference method
GOES	Geostationary Operational Environmental Satellites
H_2O	Water vapor
HY-SPLIT CheM	Hybrid Single-Particle Lagrangian Integrated Trajectories with a generalized non-linear Chemistry Module
MAQSIP	Multiscale Air Quality Simulation Platform
MM5	Penn State/NCAR Mesoscale Model Version 5
MODIS	Moderate Resolution Imaging Spectroradiometer
MOPITT	Measurements of Pollution in the Troposphere
MOS	Model output statistics
NAAQS	National Ambient Air Quality Standards
NBL	Nocturnal Boundary Layer
NCEP	National Center for Environmental Prediction
NH_3	Ammonia
NO	Nitric oxide
NO_2	Nitrogen dioxide
NOAA	National Oceanic and Atmospheric Administration
NO_x	Nitrogen oxides
NWS	National Weather Service
O_2	Oxygen
O_3	Ozone
OC	Particulate organic carbon
OC/EC	Organic and elemental carbon
PAMS	Photochemical Assessment Monitoring Stations
PM	Particulate matter
PM_{10}	Particulate matter with an aerodynamic diameter less than 10 micrometers

LIST OF ACRONYMS (Concluded)

Term	**Meaning**
$PM_{2.5}$	Particulate matter less than 2.5 μm in diameter
POD	Probability of Detection
POES	Polar Orbiting Satellites
ppb	Parts per billion
RAMS	Regional Atmospheric Modeling System
RH	Relative humidity
RL	Residual Layer
SAQM	SARMAP Air Quality Model
SMOKE	Sparse Matrix Operator Kernel Emissions
SO_2	Sulfur dioxide
SS	Skill Score
SVOCs	Semi-volatile organic compounds
TEOM	Tapered Element Oscillating Microbalance
UAM-AERO	Urban Airshed Model with Aerosols
UAM-IV	Urban Airshed Model with Carbon Bond IV Chemistry
UAM-V	Variable Grid Urban Airshed Model
VOCs	Volatile organic compounds
WRAC	Wide range aerosol classifier
WSPs	Weather Service Providers

1. INTRODUCTION AND GUIDE TO DOCUMENT

1.1   INTRODUCTION

Air pollution is a contamination of the atmosphere by gaseous, liquid, or solid wastes or by-products that have a serious affect on human health and the biosphere, reduce visibility, and damage materials. The major pollutants affecting the United States and other countries throughout the world are ozone, particulate matter, lead, carbon monoxide, nitrogen dioxide, sulfur dioxide, and toxic compounds.

Air quality forecasts provide the public with air quality information with which they can make daily lifestyle decisions to protect their health. This information allows people to take precautionary measures to avoid or limit their exposure to unhealthy levels of air quality. In addition, many communities use forecasts for initiating air quality "action" or "awareness" days, which seek voluntary participation from the public to reduce pollution and improve local air quality. Current air quality forecasting efforts focus on predicting ozone and $PM_{2.5}$.

Ozone is a reactive oxidant that forms in trace amounts in two parts of the atmosphere: the stratosphere (the layer between 20-30 km above the earth's surface) and the troposphere (ground level to 15 km). Stratospheric ozone, also known as "the ozone layer," is formed naturally and shields life on earth from the harmful effects of the sun's ultraviolet radiation. Near the earth's surface, ground-level ozone can be harmful to human health and vegetation and is created in part by pollution from man-made (anthropogenic) and natural (biogenic) sources. Because ground-level ozone accumulates in or near large metropolitan areas during certain weather conditions, it typically exposes tens of millions of people every week during the summer to unhealthy ozone concentrations (Paul et al., 1987).

Particulate matter (PM) is a complex mixture of solid and liquid particles that vary in size and composition, and remain suspended in the air. Over the past decade, many health effect studies have shown an association between exposure to PM and increases in daily mortality and symptoms of certain illnesses (Dockery and Pope, 1994; Health Effects Institute, 2002; Schwartz, 1994). Sources of PM are numerous; naturally occurring processes and human activities all contribute to total PM concentrations. Some sources are natural, such as dust from the earth's surface (crustal material), sea salt in coastal areas, and biologic material (pollen, spores, and plant and animal debris). Periodic events like forest fires and dust storms can produce large amounts of PM. In cities, PM is mainly a product of combustion from mobile sources such as cars, buses, ships, trucks, and construction equipment, and from stationary sources such as heating furnaces, power plants, and factories. Some PM is emitted directly into the atmosphere as particles (primary particles), while some particles are produced by chemical reactions in the atmosphere (secondary particles).

The size of ambient air particles ranges over a wide scale, from approximately 0.005 to 100 µm in aerodynamic diameter (from the size of just a few atoms to about the thickness of a human hair). Particles fall into three basic size modes: ultrafine particles (smaller than about 0.1 µm in diameter), fine particles (between 0.1 and 2.5 µm), and coarse particles (larger than 2.5 µm). PM_{10} is defined as particulate matter with an aerodynamic diameter less than

10 micrometers. $PM_{2.5}$ is a subset of PM_{10} and includes those particles with an aerodynamic diameter less than 2.5 µm. Cut points (2.5 µm and 10 µm) are not perfectly sharp for these PM indicators; instruments that collect $PM_{2.5}$ and PM_{10} samples collect some particles larger than the cut point while some particles smaller than the cut point are not retained. PM can vary greatly in size, composition, and concentration depending on the sources generating the particles and such factors as geographic location, season, day, time of day, and weather conditions.

In light of the health effects of ground-level ozone, many air quality agencies have been forecasting ozone concentrations to warn the public of unhealthy air and to encourage people to avoid exposure to unhealthy air and voluntarily reduce emissions. Fewer agencies have forecasted PM_{10} or $PM_{2.5}$. From 1978 to 1997, ozone forecasts were based on the 1-hr National Ambient Air Quality Standard (NAAQS) for ozone, which was 125 parts per billion (ppb). In 1997, the U.S. Environmental Protection Agency (EPA) revised the NAAQS to reflect more recent health-effects studies that suggest that respiratory damage can occur at lower ozone concentrations. Under the more-stringent revised standard, regions exceed the NAAQS when the three-year average of the annual fourth highest 8-hr average ozone concentrations is at or above 85 ppb. Likewise in 1999, the EPA implemented a new NAAQS for $PM_{2.5}$. The NAAQS for $PM_{2.5}$ is a 24-hour average concentration of 65 µg/m^3 and an annual standard of 15 µg/m^3.

1.2   DOCUMENT OBJECTIVES

This document provides guidance to help air quality agencies develop, operate, and evaluate ozone and $PM_{2.5}$ forecasting programs. This guidance document provides:

- Background information about ozone and $PM_{2.5}$ and the weather's effect on these pollutants.

- A list of how air quality forecasts are currently used.

- A summary and evaluation of methods currently used to forecast ozone and $PM_{2.5}$.

- Steps to develop and operate an air quality forecasting program.

- Information on the level of effort needed to set up and operate a forecasting program.

The intended audience of this document is project managers, meteorologists, air quality analysts, and data analysts. The information presented in this document is based on literature reviews and on interviews with air quality forecasters throughout the country.

1.3   GUIDE TO THIS DOCUMENT

This document is divided into six sections with the following contents:

Section 2: Processes Affecting Air Quality Concentrations describes the principal chemical processes that produce ozone, $PM_{2.5}$, and their precursor emissions. It also describes how atmospheric phenomena affect ozone and $PM_{2.5}$ concentrations.

Section 3: **Forecasting Applications and Needs** discusses how agencies throughout the United States use air quality forecasts.

Section 4: **Developing Forecasting Methods** explains the different approaches used to forecast air quality. It also describes each method and compares its strengths and limitations, thus allowing forecasters to select the methods that meet their agency's needs and resources.

Section 5: **Steps for Developing an Air Quality Forecasting Program** identifies the steps to develop, operate, and evaluate an ozone or $PM_{2.5}$ forecasting program.

Section 6: **References** provides a list of references cited in this report.

2. PROCESSES AFFECTING AIR QUALITY CONCENTRATIONS

Air quality concentrations are strongly affected by weather. Developing a basic understanding of how ozone and PM forms and where emissions originate will help air quality agencies forecast the effects of weather on ozone, PM, and their precursor emissions.

This section provides a background on ozone (Section 2.1) and PM (Section 2.2) including a summary of chemical reactions and sources of precursor emissions. Section 2.3 explains generally how weather affects pollutant formation, transport, and dispersion. A discussion of how to develop a more detailed understanding of the chemical and meteorological processes that control air pollution is presented in Section 5.2.

2.1   OZONE

Ozone (O_3) is not emitted directly into the air; instead it forms in the atmosphere as a result of a series of complex chemical reactions between oxides of nitrogen (NO_x) and hydrocarbons, which together are precursors of ozone. Ozone precursors have both anthropogenic (man-made) and biogenic (natural) origins. Motor vehicle exhaust, industrial emissions, gasoline vapors, and chemical solvents are some of the major sources of NO_x and hydrocarbons. Many species of vegetation including trees and plants emit hydrocarbons, and fertilized soils release NO_x.

2.1.1   Basic Ozone Chemistry

In the presence of ultraviolet radiation ($h\boldsymbol{n}$), oxygen (O_2) and nitrogen dioxide (NO_2) react in the atmosphere to form ozone and nitric oxide (NO) through the reactions given in Equations 2-1 and 2-2.

$$NO_2 + h\boldsymbol{n} \rightarrow NO + O \qquad\qquad (2\text{-}1)$$

$$O + O_2 \rightarrow O_3 \qquad\qquad (2\text{-}2)$$

Resultant ozone, however, is quickly reacted away to form nitrogen dioxide by the process given in Equation 2-3. This conversion of ozone by NO is referred to as titration. In the absence of other species, a steady state is achieved through the reactions shown by Equations 2-1 through 2-3. Even without anthropogenic emissions, these reactions normally result in a natural background ozone concentration of 25 to 45 ppb (Altshuller and Lefohn, 1996).

$$O_3 + NO \rightarrow NO_2 + O_2 \qquad\qquad (2\text{-}3)$$

Ozone cannot accumulate further unless volatile organic compounds (VOCs), which include hydrocarbons, are present to consume or convert NO back to NO_2 as shown by Equation 2-4.

$$VOC + NO \rightarrow NO_2 + \text{other products} \qquad\qquad (2\text{-}4)$$

This equation is a simplied version of many complex chemical reactions (see National Research Council, 1991, for details). As NO is consumed by this process, it is no longer available to titrate ozone. When additional VOC is added to the atmosphere, a greater proportion of the NO is oxidized to NO_2, resulting in greater ozone formation. Additionally, anthropogenic sources of NO result in greater levels of NO_2 in the atmosphere. This NO_2 is then available for photolysis to NO and O (Equation 2-1) and, ultimately, for conversion to NO_2 (Equation 2-4) and ozone (Equation 2-2).

The formation and increase in ozone concentrations occur over a period of a few hours as shown in **Figure 2-1**. Shortly after sunrise, NO and VOCs react in sunlight to form ozone. Throughout the morning, ozone concentrations increase while NO and VOCs are depleted. Eventually, either the lack of sunlight, NO, or VOCs limit the production of ozone. This diurnal cycle varies greatly depending on site location, emission sources, and weather conditions.
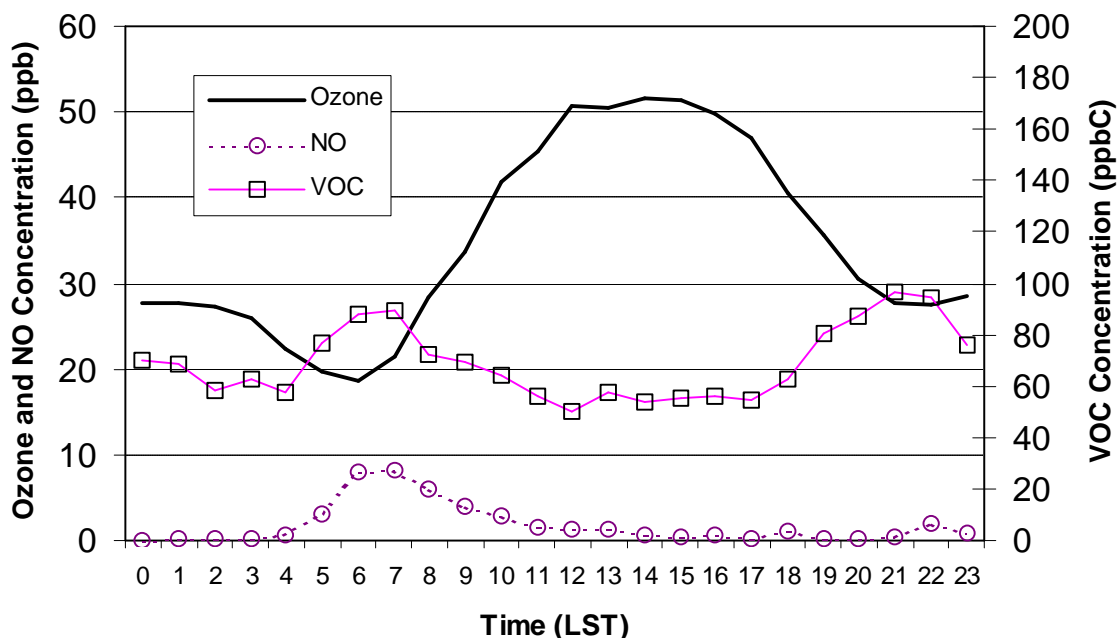
Figure 2-1. Average diurnal profile of ozone, NO, and VOC concentrations for August 1995 at an urban site in Lynn, Massachusetts.

2.1.2   Ozone Precursor Emissions

Precursor emissions of NO and VOC are necessary for ozone to form in the troposphere. Understanding the nature of when and where ozone precursors originate may help forecasters factor day-to-day emissions changes into their forecasts. For example, if a region's emissions are dominated by mobile sources, emissions, and hence ozone that forms, may depend on the day-of-week commute patterns. This section provides a brief overview of the sources and spatial distribution of VOC and NO_x (NO and NO_2) emissions.

Table 2-1 summarizes the total anthropogenic VOC and NO_x emissions in the United States for 1994. The dominant NO_x producers are combustion processes, including industrial and electrical generation processes, and mobile sources such as automobiles. Mobile sources also account for a large portion of VOC emissions. Industries such as the chemical industry or others that use solvents also account for a large portion of VOC emissions.

Table 2-1. Summary of total anthropogenic VOC and NO_x emissions in the United States during 1994 (U.S. Environmental Protection Agency, 1996). Note that 1 short ton equals 2000 pounds.

Source Type	NO_x Emissions		VOC Emissions	
	Emissions (thousand short tons)	Percentage of Total	Emissions (thousand short tons)	Percentage of Total
Fuel Combustion, Electric Utility	7795	33.0	36	0.2
Fuel Combustion, Industrial	3206	13.6	135	0.6
Fuel Combustion, Other	727	3.1	715	3.1
Chemical & Allied Product Manufacturing	291	1.2	1577	6.8
Metals Processing	84	0.4	77	0.3
Petroleum & Related Industries	95	0.4	630	2.7
Other Industrial Processes	328	1.4	411	1.8
Solvent Utilization	3	0.01	6313	27.2
Storage & Transport	3	0.01	1773	7.7
Waste Disposal & Recycling	85	0.4	2273	9.8
On-Road Vehicles	7580	31.9	6295	27.2
Non-Road Sources	3095	13.1	2255	9.7
Miscellaneous	374	1.6	685	3.0
Total Emissions	23,666	—	23,175	—

Anthropogenic VOC and NO_x emissions are highest near urban areas. **Figures 2-2 and 2-3** show the anthropogenic VOC and NO emissions by county across the United States. Notice that emissions levels correlate well with population levels, which are larger in the eastern third of the United States and near metropolitan areas.

Along with anthropogenic emissions, the EPA also estimates annual biogenic emissions. **Figure 2-4** shows that biogenic VOC emissions occur mostly in the forested regions of the United States (Southeast, Northeast, and West Coast regions). Biogenic VOC emissions include the highly reactive compound isoprene. Biogenic VOC emissions from forested and vegetative areas may impact urban ozone formation in some parts of the country. Biogenic NO_x emissions levels are typically much lower than anthropogenic NO_x emissions levels.
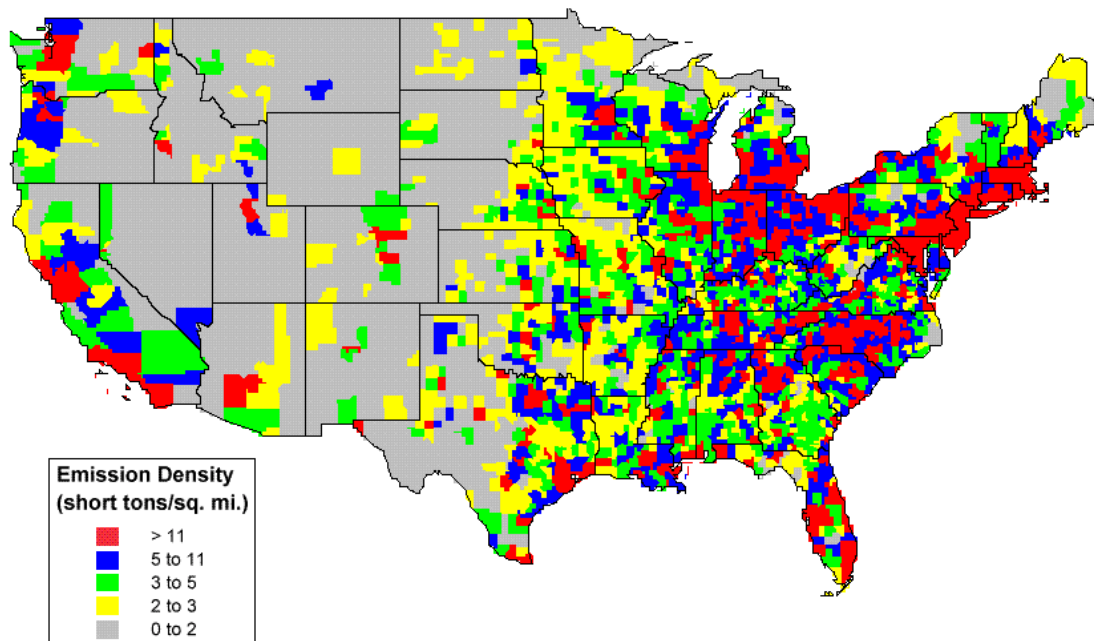
Figure 2-2.   1996 VOC emissions from anthropogenic sources by county
(U.S. Environmental Protection Agency, 1997a).
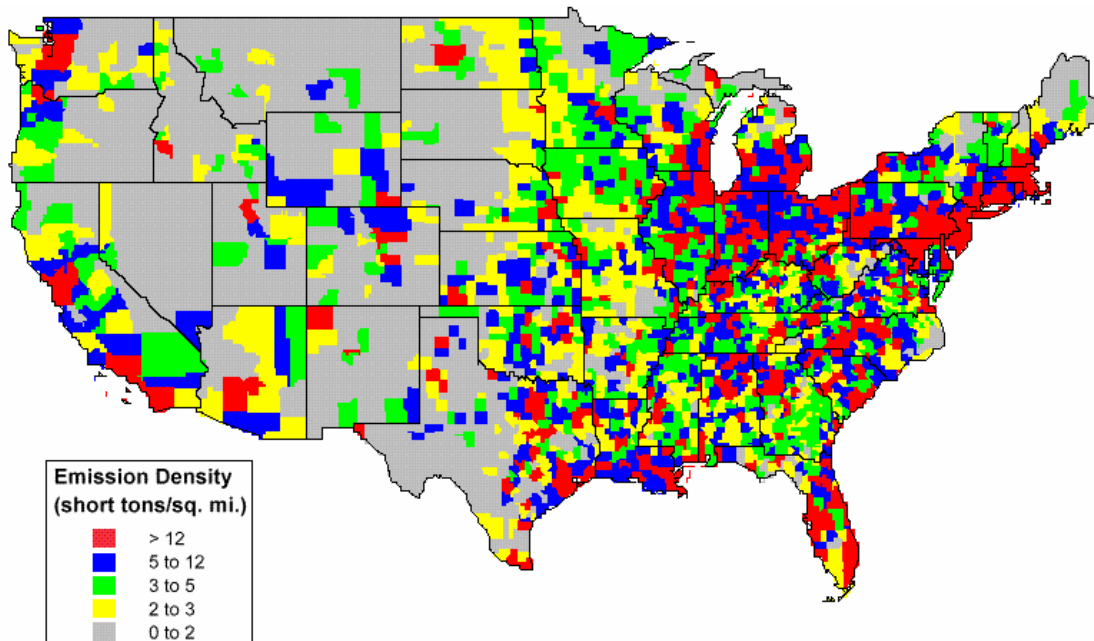
Figure 2-3.   1996 NO emissions from anthropogenic sources by county
(U.S. Environmental Protection Agency, 1997a).
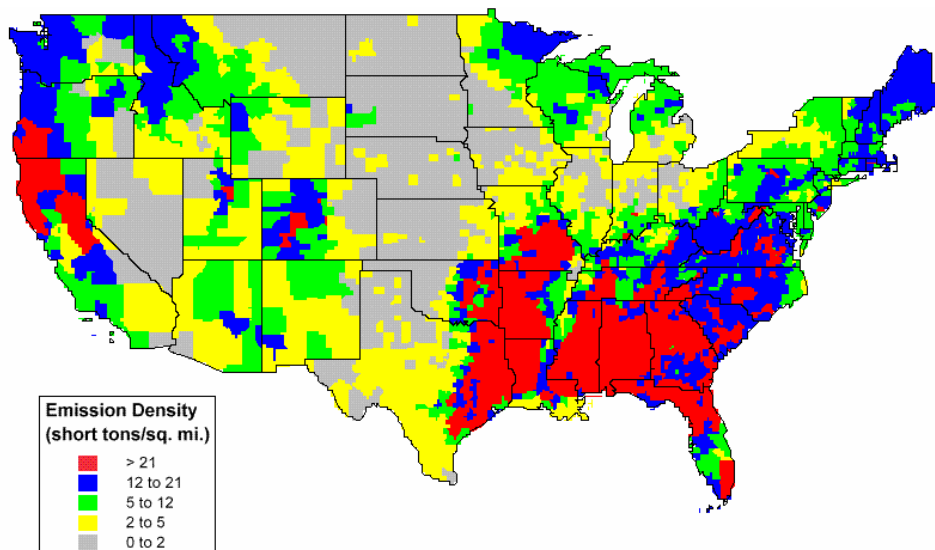
Figure 2-4. 1996 VOC emissions from biogenic sources by county (U.S. Environmental Protection Agency, 1997a).

2.2   PARTICULATE MATTER

Particulate matter is the general term used for a mixture of solid particles and liquid droplets found in air. Numerous studies show association between morbidity/mortality and high PM concentrations. Studies also indicate that short-term exposure to acute PM concentrations can lead to long-term health effects. The negative health effects associated with high PM concentrations and the public's desire for accurate air quality information has produced a need for PM forecasting programs that warn the public one or two days in advance of high PM concentrations. Since the EPA promulgated a new NAAQS for $PM_{2.5}$ (PM less than 2.5 μm in diameter) in 1997 this guidance document has been updated to include information about forecasting $PM_{2.5}$ concentrations.

Much of the material summarized in the following sections was drawn from the $PM_{2.5}$ Data Analysis Workbook (Main and Roberts, 2001), the PM criteria document (U.S. Environmental Protection Agency, 2001), Seinfeld and Pandis (1998), and documents referenced therein.

2.2.1   Basic Particulate Matter Chemistry

Particulate matter, unlike ozone, is not a specific chemical entity but is a mixture of particles of different sizes, shapes, compositions, and chemical, physical, and thermodynamic properties. Atmospheric $PM_{2.5}$ results from primary fine particle emissions (emitted directly from sources) and emissions of gaseous compounds that form secondary aerosols. Secondary particles are formed from gases through chemical reactions in the atmosphere involving atmospheric oxygen (O_2) and water vapor (H_2O); reactive species such as ozone (O_3); radicals such as the hydroxyl and nitrate radicals; and pollutants such as sulfur dioxide (SO_2), nitrogen

oxides (NO_x), ammonia (NH_3), and volatile organic compounds (VOCs) from natural and anthropogenic sources. SO_2 forms sulfates, NO_x forms nitrates, NH_3 forms ammonium compounds, and VOCs form organic carbon compounds. Some particles are liquid; some are solid. Others may contain a solid core surrounded by liquid. Atmospheric particles contain inorganic ions (e.g., nitrate, sulfate, sodium), metallic compounds, elemental carbon (EC), organic compounds, and crustal compounds (e.g., iron, calcium). Some atmospheric particles are hygroscopic and contain particle-bound water. The organic portion of PM is especially complex, containing hundreds of organic compounds.

The particle formation process includes nucleation[1] of particles from gases emitted from sources or formed in the atmosphere by chemical reactions, condensation of gases on existing particles, and coagulation of particles (**Figure 2-5**). Formation, transport, and removal rates are all a function of the particle size, chemical constituents of the particles, and meteorological processes (see **Table 2-2**).
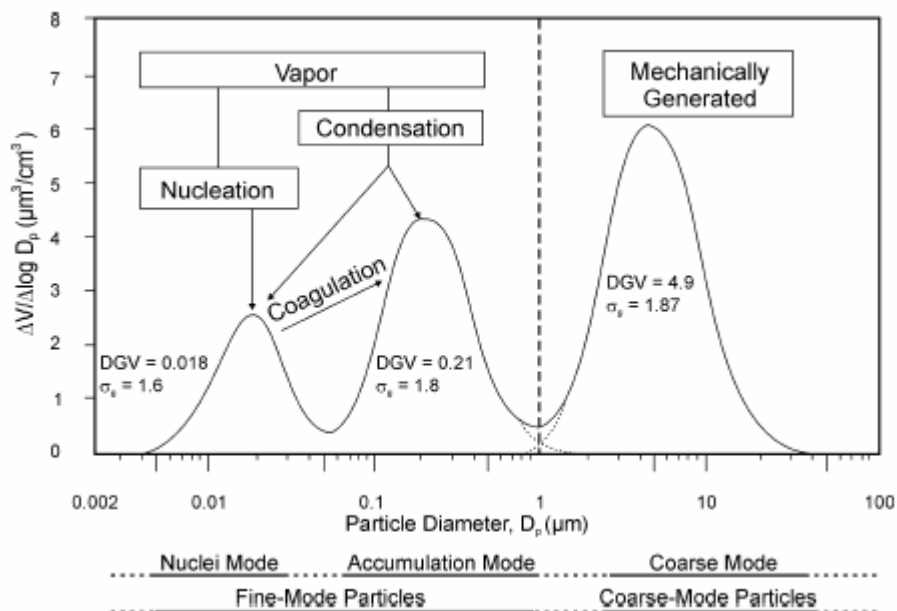
Figure 2-5. Volume size distribution measured in traffic showing fine (including nuclei and accumulation modes) and coarse particle modes (Wilson and Suh, 1997). The geometric mean diameter by volume (DGV), equivalent to volume median diameter, and geometric standard deviation (σ_2) are shown for each mode. Also shown are transformation and growth mechanisms (e.g., nucleation, condensation, and coagulation).

---

[1] The formation of new particles.

Table 2-2. Summary of formation pathways, composition, sources, and atmospheric lifetimes of fine and coarse particulate matter (from Seinfeld and Pandis, 1998).

	Fine	Coarse
Formation pathway	Chemical reaction, nucleation, condensation, coagulation, cloud/fog processes	Mechanical disruption, suspension of dust
Composition	Sulfate, nitrate, ammonium, hydrogen ion, elemental carbon, organics, water, metals (lead, cadmium, vanadium, nickel, copper, zinc, manganese, iron…)	Resuspended road dust; coal and oil flyash; crustal elements (silicon, aluminum, titanium, iron, usually as oxides); calcium carbonate, salt; pollen, mold, spores; plant and animal debris
Sources	Combustion (coal, oil, gasoline, diesel, wood); gas-to-particle conversion of NO_x, SO_x, and VOCs; smelters, mills, etc.	Resuspension of industrial soil and dust; suspension of soil (farming, mining, unpaved roads), biological sources, construction/demolition, ocean spray
Atmospheric lifetime	Days to weeks	Minutes to days
Travel distance	100 to 1000+ km	Generally < 100 km

The size of PM ranges from about tens of nanometers (nm) (which corresponds to molecular aggregates) to tens of microns (70 µm $\cong$ the size of a human hair) (see **Figure 2-6**). The smallest particles are generally more numerous, and the number distribution of particles generally peaks below 0.1 µm. The size range below 0.1 µm is also referred to as the ultrafine range. The largest particles (0.1-10 µm) are small in number but contain most of the PM volume (mass). The peak of the PM surface area distribution is always between the number and the volume peaks.
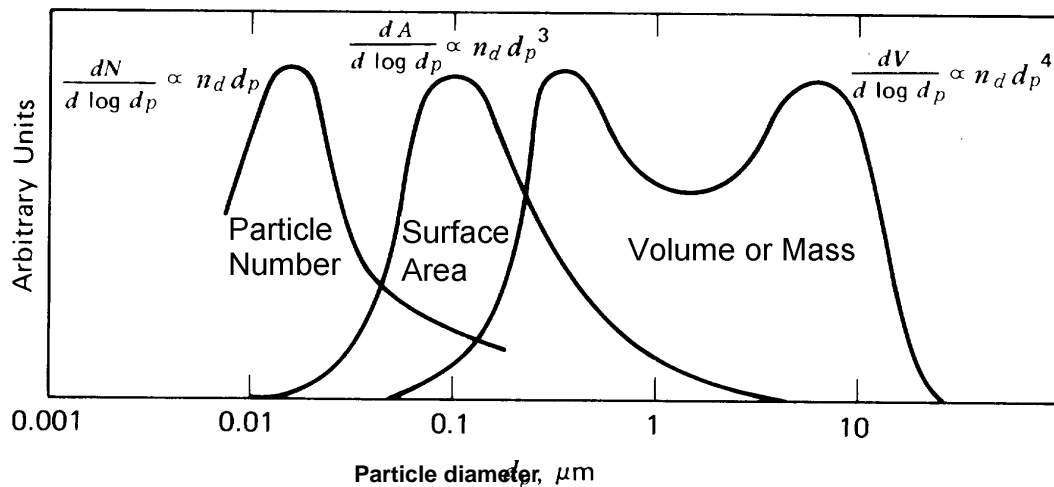
Figure 2-6. Distribution of particle number, surface area, and volume or mass with respect to size (adapted from Husar, 1998).

PM in the 0.1 to 1 micron size range has the longest residence time (days to weeks) because it neither settles nor coagulates quickly. Particles in this size range are the most efficient at penetrating deep into the lung. In addition, the light scattering efficiency per PM mass is highest at about 0.5 μm (see **Figure 2-7**). This is why, for example, 10 μg of fine particles scatter over 10 times more than 10 μg of coarse particles. Thus, $PM_{2.5}$ is important to investigations of both human health and visibility impairment. Figure 2-7 also shows the absorption efficiency; there is little variation of the absorption efficiency as a function of particle size.
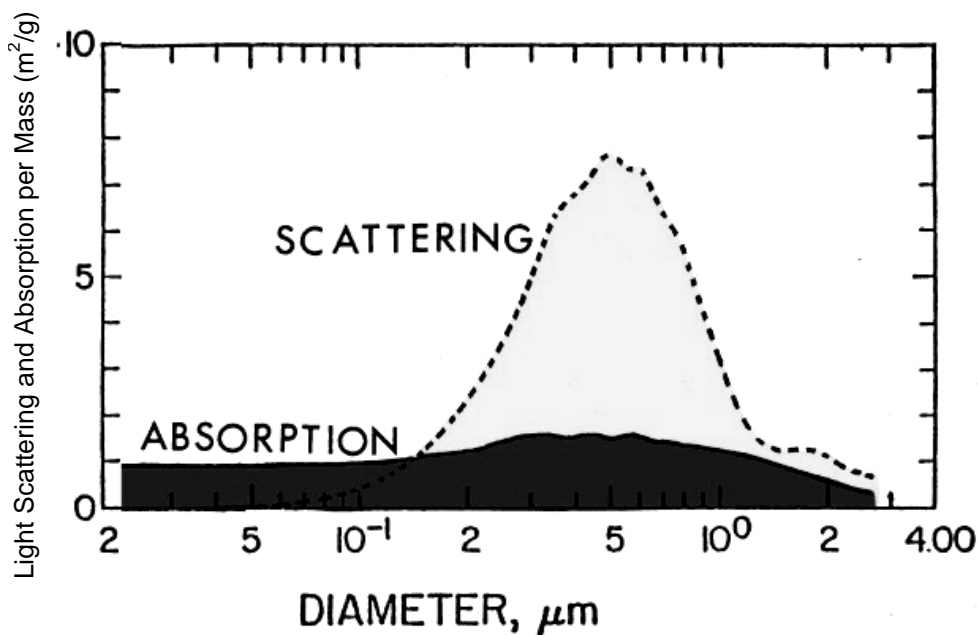
Figure 2-7. Relationship between light scattering, absorption, and particle diameter (Husar, 1998).

Most secondary fine PM is formed from condensable vapors generated by chemical reactions of gas-phase precursors (i.e., vapors generated by chemical reactions condense to form particles). Secondary formation processes can result in either the formation of new particles or the addition of particulate material to pre-existing particles. Most of the sulfate and nitrate and a portion of the organic compounds in atmospheric particles are formed by chemical reactions in the atmosphere. Secondary aerosol formation depends on numerous factors including:

- The concentrations of precursors (which are a function of the proximity of emissions, wind speed, and mixing height).

- The concentrations of other gaseous reactive species such as ozone, hydroxyl radical, peroxy radicals, or hydrogen peroxide (which are a function of solar radiation and temperature).

- Atmospheric conditions, including solar radiation and relative humidity (RH).

- The interactions of precursors and pre-existing particles within cloud or fog droplets or in the liquid film on solid particles.

As a result, it is considerably more difficult to relate ambient concentrations of secondary species to sources of precursor emissions than it is to identify the sources of primary particles. A diagram of these interactions is provided in **Figure 2-8**. Details about the chemical reactions can be found in Seinfeld and Pandis (1998), for example.
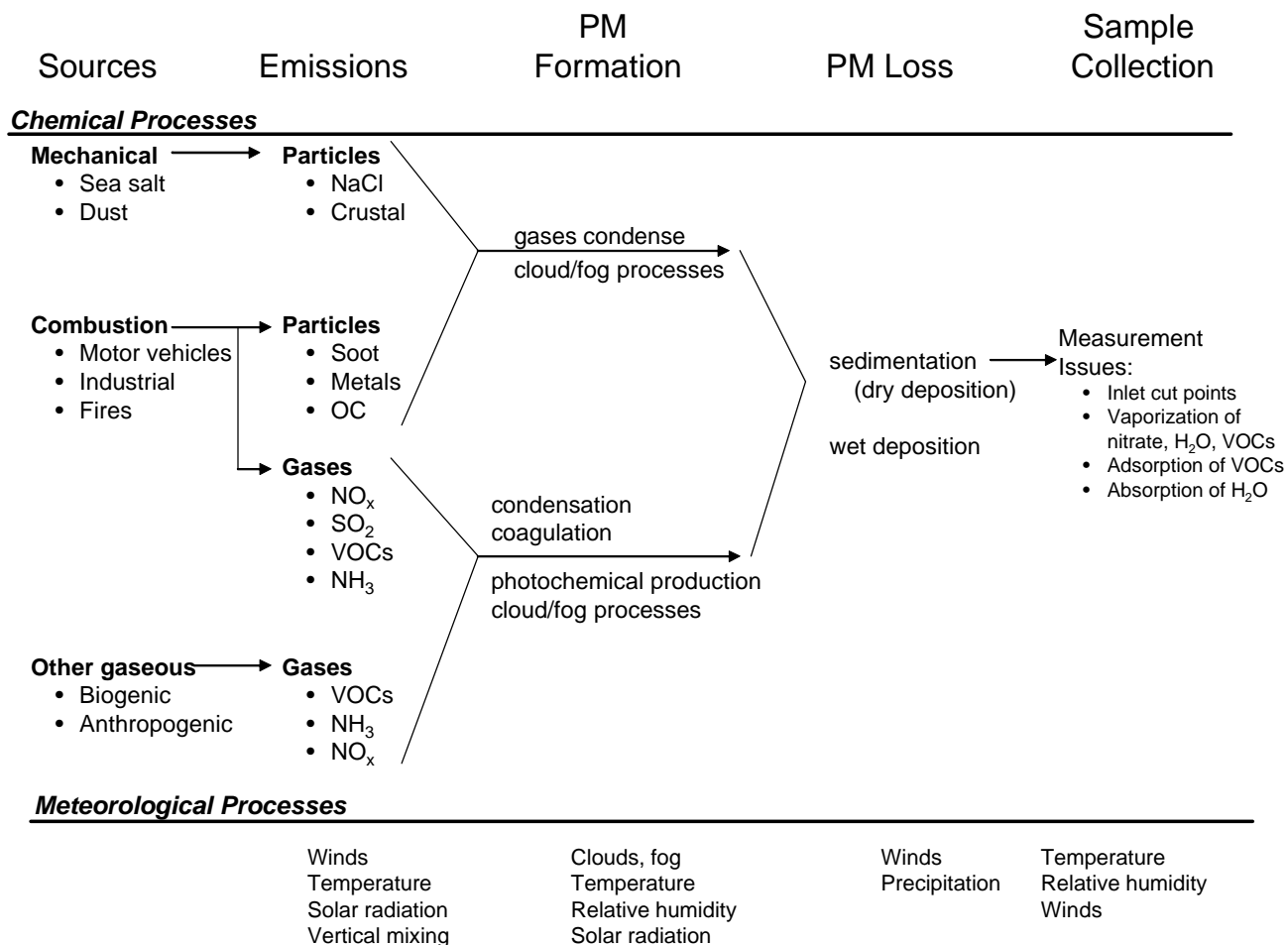
Figure 2-8. Sources of precursor gases and primary particles, PM formation processes, and removal mechanisms. Important meteorological measures are provided for each process.

Sulfates

Sulfates constitute about half of the $PM_{2.5}$ in the eastern United States. Virtually all the ambient sulfate is secondary, formed within the atmosphere from SO_2. About half of the conversion of SO_2 to sulfate occurs in the gas phase through photochemical oxidation in the daytime. NO_x and VOC emissions tend to enhance the photochemical oxidation rate. At least half of the SO_2 oxidation takes place in cloud droplets as air molecules pass through convective clouds. Within clouds, the soluble pollutant gases, such as SO_2, combine with water droplets and

rapidly oxidize to form sulfate. SO_2-to-sulfate transformation rates peak in the summer due to enhanced summertime photochemical oxidation and SO_2 oxidation in clouds.

Conversion of SO_2 to sulfate occurs at about 1% per hour in cloud-free air, but can convert to sulfate at 50% per hour in clouds and fog. Removal rates for SO_2 (mostly by dry deposition) and sulfate (mostly by wet deposition) are about 2 to 3% per hour each. This gives sulfur (as SO_2 and sulfate) an atmospheric residence time of from 1 to 5 days, depending on season, geography, and weather conditions.

Nitrates

Nitrates are a principal component of $PM_{2.5}$ in the western United States and, like sulfates, are nearly all formed within the atmosphere from nitrogen oxide emissions. About one-third of anthropogenic NO_x emissions in the United States are estimated to be removed by wet deposition. NO_2 is converted to nitric acid by reaction with hydroxyl radicals during the day (oxidation). The reaction of hydroxyl radical with NO_2 is 10 times faster than the oxidation of SO_2. The peak daytime conversion rate of NO_2 to nitric acid in the gas phase is about 10 to 50% per hour. During the nighttime, NO_2 is converted into nitric acid by a series of reactions involving ozone and the nitrate radical. Nitric acid reacts with ammonia to form particulate ammonium nitrate. Thus, PM nitrate can be formed at night and during the day. Thermodynamically, nitrate formation is favored in cold, moist conditions. Thus, nitrate formation is enhanced in the winter compared to the summer.

Organic and elemental carbon compounds

Elemental carbon (EC), also called black carbon, is emitted directly into the atmosphere through combustion processes. Particulate organic carbon (OC) is both directly emitted and formed in secondary reactions. OC comprises a significant portion of the $PM_{2.5}$ throughout the United States. Atmospheric reactions involving VOCs yield organic compounds with low vapor pressures at ambient temperature (i.e., the vapor or gas condenses to form a liquid). These reactions can occur in the gas phase, in fog or cloud droplets or in aqueous aerosols. Reaction products from the oxidation of VOCs also may nucleate to form new particles or condense on existing particles to form secondary organic PM. Both biogenic and anthropogenic sources contribute to primary and secondary organic particulate matter. Although the mechanisms and pathways for forming inorganic (i.e., sulfate, and nitrate) secondary particulate matter are fairly well known, those for forming secondary organic PM are not as well understood.

Ozone and the hydroxyl radical are thought to be major contributing reactants. However, other radicals, including nitrate and organic radicals, are thought to contribute to secondary organic PM formation. Experimental studies of the production of secondary organic PM in ambient air have focused on the Los Angeles Basin. Evidence shows that secondary PM formation occurs during periods of photochemical ozone formation in Los Angeles and that as much as 70% of the organic carbon in ambient PM was secondary in origin during a smog episode in 1987 (see Turpin et al., 1991). Other experiments showed that 20 to 30% of the total organic carbon in fine PM in the Los Angeles airshed is secondary in origin on an annually averaged basis (Schauer et al., 1996). Thus, photochemical reactions are important to secondary organic carbon PM formation.

Another formation pathway is the adsorption of semi-volatile organic compounds (SVOCs, e.g., including polycyclic aromatic hydrocarbons)[2] onto existing solid particles. This pathway can be driven by diurnal and seasonal temperature and humidity variations at any time of the year. Higher temperatures generally favor the gaseous phase of the SVOCs.

2.2.2 PM$_{2.5}$ Emissions and Sources

The major constituents of atmospheric PM are sulfate, nitrate, ammonium, and hydrogen ions; particle-bound water; elemental carbon; a variety of organic compounds; and crustal material (see **Table 2-3**). These constituents can be primary or secondary. PM is called "primary" if it is in the same chemical form in which it was emitted into the atmosphere. PM is called "secondary" if it is formed by chemical reactions in the atmosphere. Primary fine particles are emitted from sources either directly as particles or as vapors that rapidly condense to form ultrafine particles (diameters < 0.1 micron) including: soot from diesel engines, a variety of organic compounds condensed from incomplete combustion or cooking, and metal compounds that condense from vapor formed during combustion or smelting.

Table 2-3. Major PM$_{2.5}$ components.

Geological Material – suspended dust consists mainly of oxides of aluminum, silicon, calcium, titanium, iron, and other metal oxides.	**NaCl** – salt is found in PM near sea coasts, open playas, and after de-icing materials are applied. The chloride ion can be replaced by nitrate as a result of reaction during long-range transport.
Sulfate – results from conversion of SO$_2$ gas to sulfate-containing particles.	**Organic Carbon (OC)** – consists of hundreds of separate compounds containing mainly carbon, hydrogen, and oxygen.
Nitrate – results from a reversible gas/particle equilibrium between ammonia, nitric acid, and particulate ammonium nitrate.	**Elemental Carbon (EC)** – composed of carbon without much hydrocarbon or oxygen. EC is black, often called soot.
Ammonium – ammonium bisulfate, sulfate, and nitrate are most common.	**Liquid Water** – soluble nitrates, sulfates, ammonium, sodium, other inorganic ions, and some organic material absorb water vapor from the atmosphere.

There are both anthropogenic and natural sources of PM$_{2.5}$. Anthropogenic emissions that contribute to ambient PM$_{2.5}$ concentrations include the following:

- Mobile sources. Gasoline- and diesel-fueled vehicles; resuspended road dust from vehicle activity on paved and unpaved roads; vehicle tire and brake wear; and off-road mobile sources such as trains, marine vessels, and farm machinery.

---

[2] The phase in which an organic compound exists in the atmosphere is largely dependent on its vapor pressure. Nonvolatile compounds, such as polychlorinated biphenyls, exist almost exclusively on particulate matter (i.e., in the "particle phase"), whereas highly volatile compounds, such as small alkanes, remain in the gas phase. However, due to their intermediate volatility, the SVOCs partition between the gas and particle phases.

- Stationary sources. Fuel combustion for electric utilities and industrial processes; fuel combustion for home heating; construction and demolition; metals, minerals, petrochemical, and wood products processing; mills and elevators used in agriculture; erosion from tilled lands; food cooking; and waste disposal and recycling.

Numerous natural sources also contribute primary and secondary particles to the atmosphere:

- Primary sources. Windblown dust from undisturbed land; sea spray; and plant and insect debris.

- Secondary sources. Oxidation of naturally emitted biogenic hydrocarbons, such as terpenes, leads to formation of secondary organic $PM_{2.5}$ and can accelerate the formation of inorganic secondary $PM_{2.5}$, such as ammonium sulfate and ammonium nitrate.

- Natural and man-made sources. Ammonia gas (precursor) and wood-burning particles (fuel wood burning and forest fires) are potentially important sources of $PM_{2.5}$.

Since the precursor gases and $PM_{2.5}$ are capable of long-range transport, it is often difficult to identify individual sources of $PM_{2.5}$. It is especially difficult to distinguish contributions from natural and anthropogenic sources.

The chemical composition of $PM_{2.5}$ varies both geographically and seasonally (**Figures 2-9 and 2-10**). The composition may also vary with the magnitude of the $PM_{2.5}$ mass concentrations. There are some relatively common characteristics across much of the United States, however, including:

- The crustal component is relatively small (<10%) in $PM_{2.5}$.

- The carbonaceous aerosol component, which consists of both organic material and elemental carbon, is either the most abundant or second most abundant class of compounds (35 to 60%).

- Higher sulfate content is observed in eastern United States $PM_{2.5}$ than in the western United States $PM_{2.5}$.

- The nitrate content of the $PM_{2.5}$ in the eastern United States is consistently lower than in the western United States.

- $PM_{2.5}$ ammonium concentrations are generally present in sufficient quantities to buffer the sulfate and nitrate in the aerosol (i.e., to keep the sulfate and nitrate in the particles, and thus to keep it from volatilizing to form gaseous sulfuric or nitric acids).

- The relative amounts of secondary constituents tend to increase under air pollution episode conditions.

- Seasonal (and geographical) variations in primary emissions and secondary formation rates lead to seasonal and geographical differences in composition and concentrations. For example, wood burning for home heating is an important source of $PM_{2.5}$ OC in the winter compared to the summer and in northern states compared to southern states.
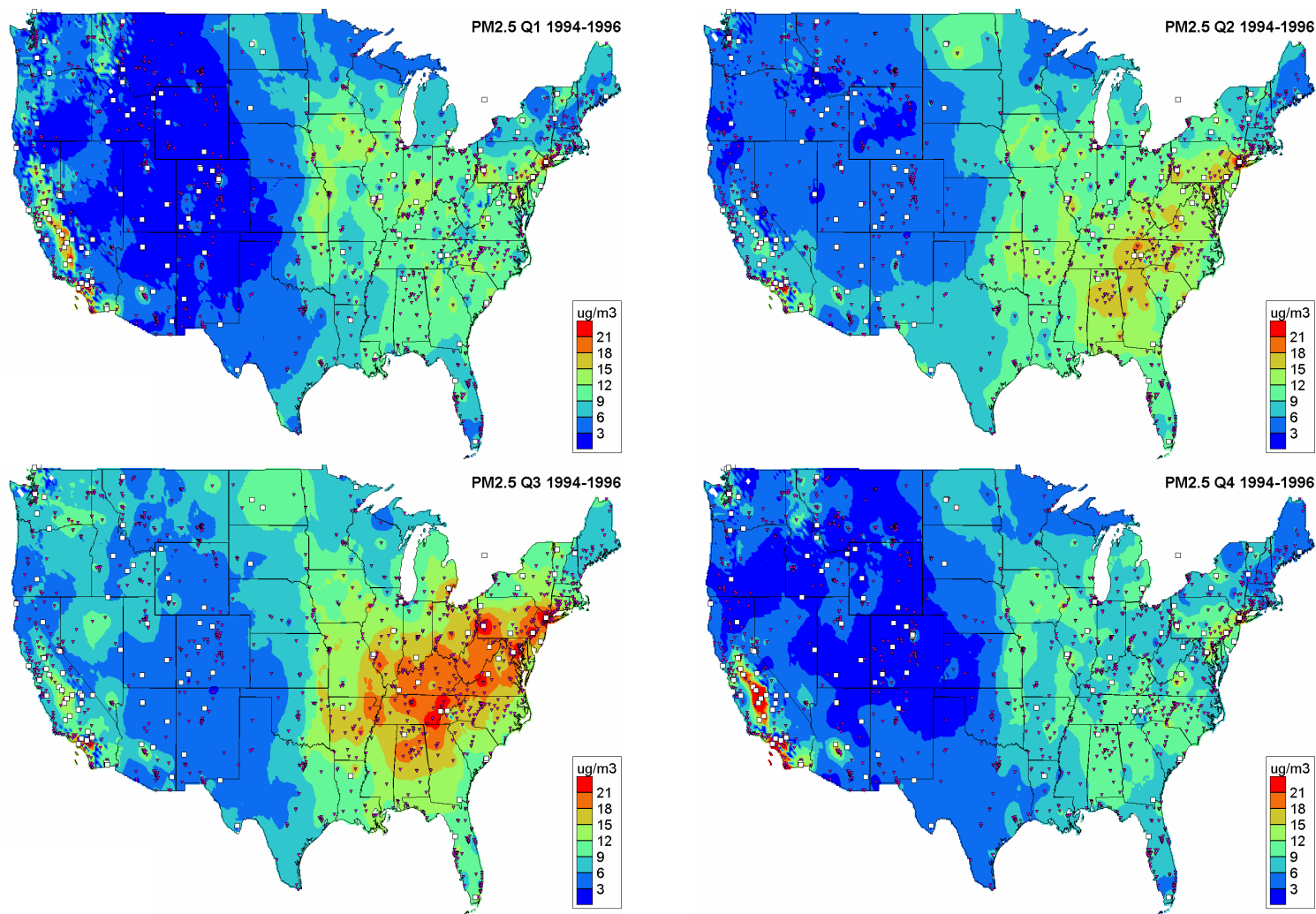
Figure 2-9. Seasonal maps of PM$_{2.5}$ mass for 1994-1996 (Falke, 1999).
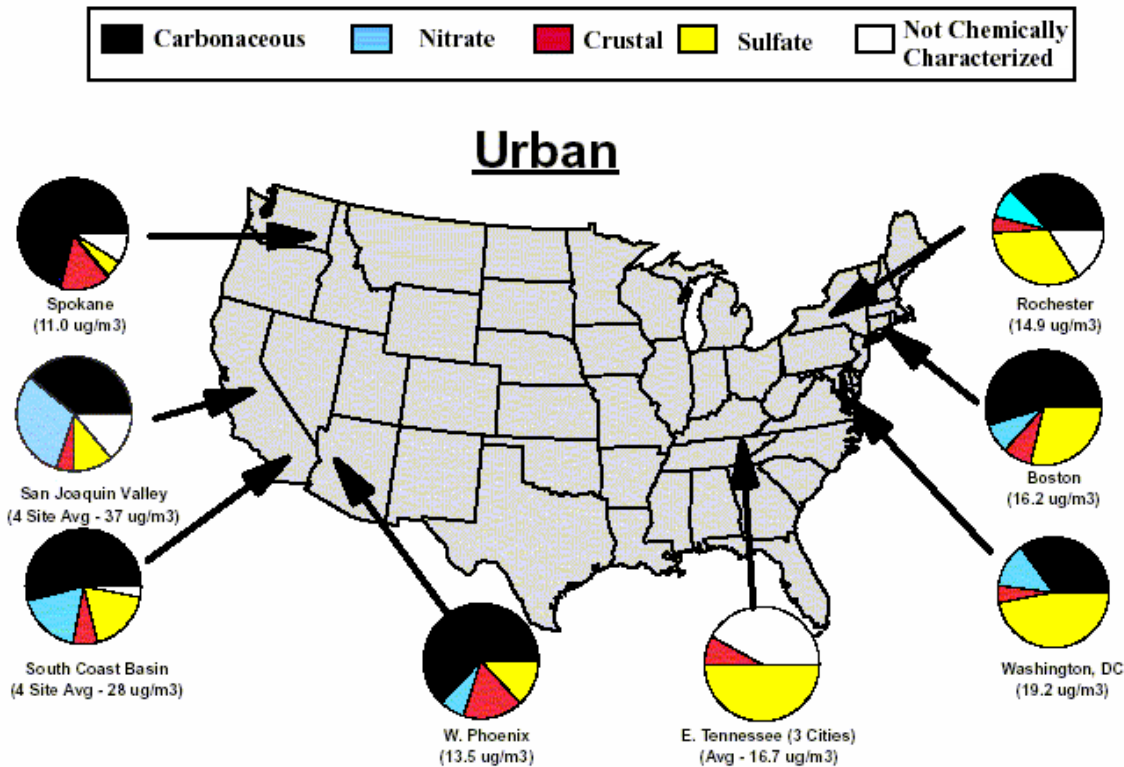
Figure 2-10.   Ambient PM$_{2.5}$ composition at urban sites in the United States (U.S. Environmental Protection Agency, 1998).

2.2.3   Monitoring Issues

Monitor types

Historically, PM mass and composition have been monitored using filter-based methods with 24-hr (or longer) averages. Inlets to the sampling systems have traditionally been set with particle diameter cut points of 2.5 or 10 microns (see **Figure 2-11**). The federal reference method (FRM) collects PM onto a 47-mm Teflon filter which is later weighed for mass.

Deployment of the continuous PM$_{2.5}$ monitoring network began in 1999 and is producing a more detailed understanding of PM$_{2.5}$ mass characteristics. Two continuous mass measurements are listed below; both can be operated with either a PM$_{10}$ or a PM$_{2.5}$ inlet.

- BAM – beta attenuation monitor; beta gauge. Beta ray transmission is measured across a clean section of filter tape and advanced to the sampling inlet. Air is then drawn into the sample inlet and PM is deposited on the filter tape. After a set interval (typically 1 hour), the filter tape is returned to its original location and the beta ray transmission is re-measured to obtain a difference which is proportional to the PM mass concentration.
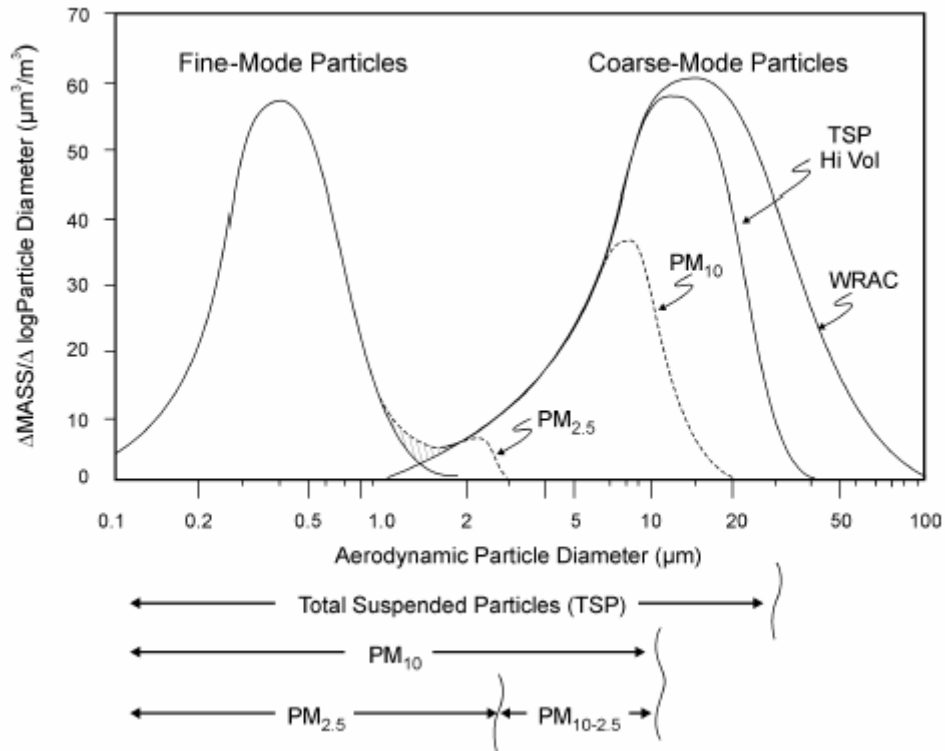
Figure 2-11. Idealized distribution of ambient PM showing fine-mode particles and coarse-mode particles and the fractions collected by size-selective samplers (Wilson and Suh, 1997). WRAC is the wide range aerosol classifier which collects the entire coarse mode.

- TEOM - tapered element oscillating microbalance. An inertial mass measurement technique for making a direct measurement of the particle mass collected on a filter in real time. The instrument works with an inertial balance that directly measures the mass collected on a filter by monitoring the frequency changes of a hollow tapered element. As more mass accumulates on the filter on top of the tapered element, the oscillating frequency changes, and that change is related to PM mass concentration.

While the above instruments provide the mass of PM in the air during a given period, chemical speciation is useful to determine the constituents that comprise the PM. Chemical speciation is still largely filter-based. In the FRM, PM is collected onto Teflon and Quartz filters which are later analyzed for mass, metals, soluble ions (e.g., sulfate, nitrate), and organic and elemental carbon. Continuous monitors are now available for black carbon (Aethalometer), organic and elemental carbon (OC/EC), sulfate, nitrate, and particle scatter (nephelometer - measures can be related to $PM_{2.5}$ mass).

Continuous real-time data serve as the cornerstone of a forecasting program and are needed to develop forecasting methods, monitor current conditions, amend forecasts, and evaluate forecasting performance. As the new instruments are being deployed and compared to historical filter samplers, uncertainties and inconsistencies with the data are emerging (see U.S.

Environmental Protection Agency, 2001, for details). Issues to consider when comparing data include the following:

- FRM – As PM is deposited on the filters, temperature, humidity, and PM compositional changes occur throughout the 24-hr sample period and thus some constituents may volatize (e.g., nitrates, organic carbon compounds, water) or adsorb (e.g., organic carbon compounds, water) with the changing weather conditions.

    o Organic gas adsorption (positive bias) comprised up to 50% of the organic carbon measured on quartz-fiber filters in southern California (Turpin et al., 1994). This study also indicated that adsorption was much more important than organic particle volatilization (negative bias).

    o Sampling losses on the order of 30% of the annual federal standard for $PM_{2.5}$ may be expected due to volatilization of ammonium nitrate in those areas of the country where nitrate is a significant contributor to the fine particle mass and where ambient temperatures tend to be warm (Hering and Cass, 1999).

- TEOM – the TEOM is typically operated at either 30°C or 50°C in an attempt to minimize the changes in PM mass with respect to humidity. But by heating the PM, other chemical components may be volatilized (e.g., nitrates, organic carbon compounds) (Allen et al., 1997) leading to the TEOM under-reporting the PM mass concentration. This bias for reporting low concentrations typically is worse during colder months when the ambient temperature is low.

Data quality

According to the EPA, "The purpose of data validation is to detect and then verify any data values that may not represent actual air quality conditions at the sampling station" (U.S. Environmental Protection Agency, 1984). Data validation is critical because serious errors in data analysis and modeling results can be caused by erroneous individual data values. The EPA's $PM_{2.5}$ speciation guidance document provides quality requirements for sampling and analysis. The guidance document also discusses data validation, including the suggested four-level data validation system. It is the monitoring agency's responsibility to prevent, identify, correct, and define the consequences of difficulties that might affect the precision and accuracy, and/or the validity, of the measurements. Once the quality-assured data are provided to data analysts, additional data validation steps need to be taken. Given the newness and complexity of the $PM_{2.5}$ mass and speciation monitoring and sample analysis methods, errors are likely to pass through the system despite rigorous application of quality assurance and validation measures by the monitoring agencies. Therefore, data analysts should also check the validity of the data before conducting their analyses.

For a forecasting program, historical data are required to investigate relationships between meteorology and $PM_{2.5}$ (or ozone) concentrations. These data need general and specific checks of consistency and validity prior to their use in analysis (e.g., Main and Roberts, 2001). To complete these checks, all relevant data need to be gathered including collocated $PM_{2.5}$ measurements, collocated gaseous (e.g., ozone, NO_x) measurements, collocated other PM (or PM-related) measurements (e.g., b_{scat}, b_{abs}, black carbon), speciated $PM_{2.5}$ and PM_{10} data, and $PM_{2.5}$ and PM_{10} mass data (by all monitoring methods).

General checks include:

- Assess data completeness. Generally, 75% of the data samples are required to make a valid average (24-hour, seasonal, or annual) to be used in a statistical analysis.

- Graphically review time series of pollutant concentrations. Inspect data spikes, dips, and outliers. Plot complementary data together (e.g., $PM_{2.5}$ and PM_{10}, $PM_{2.5}$ and light scattering). The first assumption upon finding a measurement that is inconsistent with physical expectations is that the unusual value is due to a measurement error. If, upon tracing the path of the measurement, nothing unusual is found, the value can be assumed to be a valid result of an environmental cause.

- Apply screening criteria.

  o $PM_{2.5}$ mass concentrations should be above 0 or below 200 $\mu g/m^3$ in most cases (concentrations outside this range are considered suspect and require further inspection including investigation of possible unusual events).

  o Three consecutive mass concentrations in continuous data should not be equal. If so, the data require additional inspection.

  o $PM_{2.5}$ mass concentrations should be less than or equal to PM_{10} mass concentrations.

- Compare data from collocated samplers - between the same sampler type and different sampler types. Sometimes, seasonal differences (i.e., biases) between samplers are observed because of differences in sampling methods.

  Specific checks include:

- The sum of individual chemical species concentrations should be less than the total $PM_{2.5}$ mass concentrations.

- Particle absorption (b_{abs}) should correlate well with elemental carbon.

- Cations (sodium, potassium, and ammonium) should compare well to (i.e., balance) the anions (chloride, nitrate, and sulfate) when measured by equivalents.

- The total of front and back-up filter nitrate should compare well to the front filter nitrate. A scatter plot of these two measurements will help analysts to better understand possible nitrate volatilization losses (probably higher in the summer than in winter).

- The front and back-up filter organic carbon should compare well to the front filter organic carbon. A scatter plot of these two measurements will help analysts to better understand possible organic carbon adsorption (positive filter artifact).

Once the data have been validated, the analyst can then proceed with more confidence when generating the relationships between ambient concentrations and meteorology.

2.2.4   Unusual PM Events

Unusual emission events can produce high PM concentrations and reduce visibility. Sometimes called "exceptional events" for regulatory purposes (U.S. Environmental Protection

Agency, 1986), these events are caused by uncontrollable and often infrequent activities from either natural or man-made activities, at different locations, and with different source strengths and thus are often difficult to predict. Typical unusual events include agricultural burning, wildland fires, and windblown dust. This section discusses the causes of unusual events and provides resources and suggestions for trying to understand and forecast these events.

Air quality conditions may be affected in the following ways by these unusual emission events:

- Local $PM_{2.5}$ concentrations at 1 or 2 monitors can increase due to localized fires, agricultural burning, or other activities that may not be representative of the entire forecast area.

- Transported $PM_{2.5}$ (smoke) from large wildland fires can be transported hundreds or thousands of kilometers to a forecast region and can increase the background levels of $PM_{2.5}$, thus combining transported $PM_{2.5}$ with locally-generated $PM_{2.5}$ to produce a more severe episode. Depending on concentrations, this transported $PM_{2.5}$ could produce an exceedance of the NAAQS and/or degrade visibility.

- Visibility can be significantly affected by fine particles from dust and smoke, which are efficient scatterers of light.

Table 2-4 further describes the common causes of unusual $PM_{2.5}$ events, how they affect $PM_{2.5}$ mass, and resources to help understand and forecast these events. Due to the infrequent nature of these events, they are not predicted well by air quality models or other air quality forecasting tools (e.g., statistical methods). Instead, forecasters should monitor observations (surface data and satellite images) and wildfire information for evidence of unusual emission events. Then, once an event has been detected and located, forward trajectories can be used to estimate the transport direction and potential time the smoke or dust might enter a particular forecast region (see www.arl.noaa.gov/ready/hysplit4.html for calculating forward and backward trajectories on the Internet).

To better prepare and forecast for these infrequent and unusual events, consider the following:

- Determine if these unusual events are likely to occur in a region by examining historical data and literature.

- Review the data and information sources in Table 2-4 to determine if the real-time data sources can offer early detection of an event and provide some warning.

- Develop methods and ways to anticipate when these events might occur (for example, wildland fires are more likely to occur in late summer and fall, African dust has been shown to impact the southeastern United States during June, July, and August; Prospero, 1999).

Examine satellite data to detect PM from dust storms and wildland fires (meted.ucar.edu/npoess/nrlsat). Real-time satellite data can be viewed at www.goes.noaa.gov and historical satellite data can be found at National Oceanic and Atmospheric Administration (NOAA's) satellite archive (www.saa.noaa.gov).

Table 2-4. Types of unusual events, how they affect PM concentrations, and a list of resources for acquiring data and information to forecast these events.

Event	Definition	How it affects PM	Data and Information Resources on the Internet
Agricultural burning	Burning of farm lands and farming byproducts (rice straw, orchard prunings, etc.).	Often burned at the end of growing seasons, smoke from these fires can increase local or regional $PM_{2.5}$ concentrations.	• EPA resources on agricultural burning (www.epa.gov/agriculture/tburn.html) • California Air Resources Board's smoke management plan (www.arb.ca.gov/smp/smp.htm)
Wildland fires	Large fires of 100-1000+ acres that burn all or most biomass (trees, shrubs, grasses, etc.).	Biomass burning produces substantial amounts of $PM_{2.5}$. Transport of this $PM_{2.5}$ from ten to thousands of miles can occur. Smoldering combustion releases several times more particles than flaming combustion (Ward, 1999).	• NOAA's Operational Significant Event Imagery (www.osei.noaa.gov/) • National Interagency Fire Center with real-time and historical fire data and statistics (www.nifc.gov/) • National Fire Weather Center (www.boi.noaa.gov/firewx.htm) • NOAA's Air Resources Laboratory – Wildfire/Forest Fire Smoke Forecasting (www.arl.noaa.gov/ss/transport/fires.html)
Windblown dust Local	Locally generated airborne dust from winds blowing across agricultural and barren land.	Strong winds ($\approx$6 m/s or more; Saxton et al., 2000) can cause dust to become airborne. Many factors influence the amount of $PM_{2.5}$ and PM_{10} produced by windblown dust: vegetation cover, soil moisture, soil particle size distribution, surface roughness, and changes in wind direction. Most of the PM generated during dust storms is larger than 2.5 microns; but Claiborn et al. (2000) found that $PM_{2.5}$ was about 30% of the PM_{10} mass during windblown dust events in Spokane, Washington.	• NOAA's Operational Significant Event Imagery (www.osei.noaa.gov/) • NASA site with satellite observations of dust and smoke (earthobservatory.nasa.gov/NaturalHazards/) • NASA's MODIS Land Rapid Response system for real-time satellite images (rapidfire.sci.gsfc.nasa.gov/production)
Windblown dust Global	High winds cause soil dust to become airborne. While the larger dust particles (>10 µm) have high settling velocities and fall back to earth in 100's of kilometers, fine particles have a long lifetime and low gravitational settling (and without precipitation) can remain airborne for weeks and over 1000's of kilometers	African dust can be transported by the easterly trade winds to the eastern and southeastern United States (Prospero, 1999) and can increase $PM_{2.5}$ concentrations at the surface and degrade visibility. The western United States can be affected by dust transported from Asia (Falke et al., 2001).	• NOAA's Operational Significant Event Imagery (www.osei.noaa.gov/) • SeaWIFS satellite imagery for tracking large dust storms (seawifs.gsfc.nasa.gov/SEAWIFS.html) • Naval Research Laboratory's global aerosol modeling system (www.nrlmry.navy.mil/aerosol/#currentaerosolmodeling)

The types of satellite imagery useful for detecting smoke and dust include:

o   Geostationary Operational Environmental Satellites (GOES) visible images with a 1-km resolution (www.oso.noaa.gov/goes).

o   Polar Orbiting Satellites (POES) visible images from satellites that orbit close to the earth provide higher resolution images than GOES, but less frequently (twice per day).

o   Measurements of Pollution in the Troposphere (MOPITT) sensor on NASA's Terra satellite provide global, yet coarse (22-km horizontal resolution) measurements in the lower part of the atmosphere (www.eos.ucar.edu/mopitt). Also, NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) provides an historical archive of global aerosol measurements from satellite-based instruments (modis-atmos.gsfc.nasa.gov).

2.3   METEOROLOGICAL CONDITIONS THAT INFLUENCE AIR QUALITY

This section presents the types of weather conditions that have a strong influence on $PM_{2.5}$ or ozone concentrations. Since daily weather variations best explain the day-to-day changes in air quality concentrations, understanding how weather influences air quality in a region is critical for producing accurate air quality forecasts.

Different scales of weather phenomena are important to air quality. The weather phenomena range from large storm systems that can encompass thousands of kilometers to small turbulent eddies that are a few meters in size. In general, large-scale weather phenomena are easier to characterize compared to small ones. In addition, weather forecast models typically do a better job of predicting large weather phenomena as opposed to small-scale, short lived phenomena. Therefore, to understand and predict air quality, it is usually best to use a large-scale to small-scale approach by first understanding the relationship between large-scale weather features and local air quality, and then understanding the relationship between local weather and air quality.

Meteorological conditions that strongly influence air quality include: transport by winds, recirculation of air by local wind patterns, and horizontal dispersion of pollution by wind; variations in sunlight due to clouds and season; vertical mixing and dilution of pollution within the atmospheric boundary layer; temperature; and moisture. The variability of these processes, which affects the variability in pollution, is primarily governed by the movement of large-scale high- and low-pressure systems, the diurnal heating and cooling cycle, and local and regional topography.

Figures 2-12 and 2-13 show the general relationships among meteorological phenomena and air quality. **Table 2-5** describes how specific meteorological conditions directly influence $PM_{2.5}$ and ozone concentrations. The remainder of this section discusses the key meteorological phenomena in these figures and tables. Educational resources on basic meteorology are available on the Internet (Cooperative Program for Operational Meteorology, 2002; University of Illinois Urbana-Champaign, 2002).
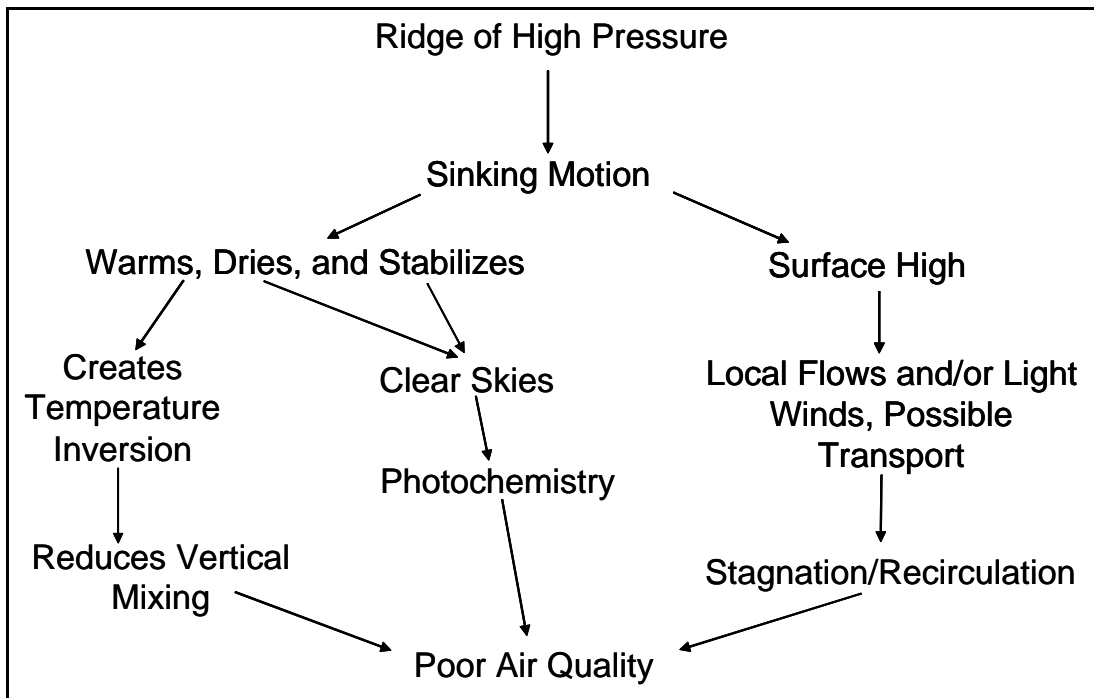
Figure 2-12.   Schematic of the typical meteorological conditions and air quality often associated with an aloft ridge of high pressure.
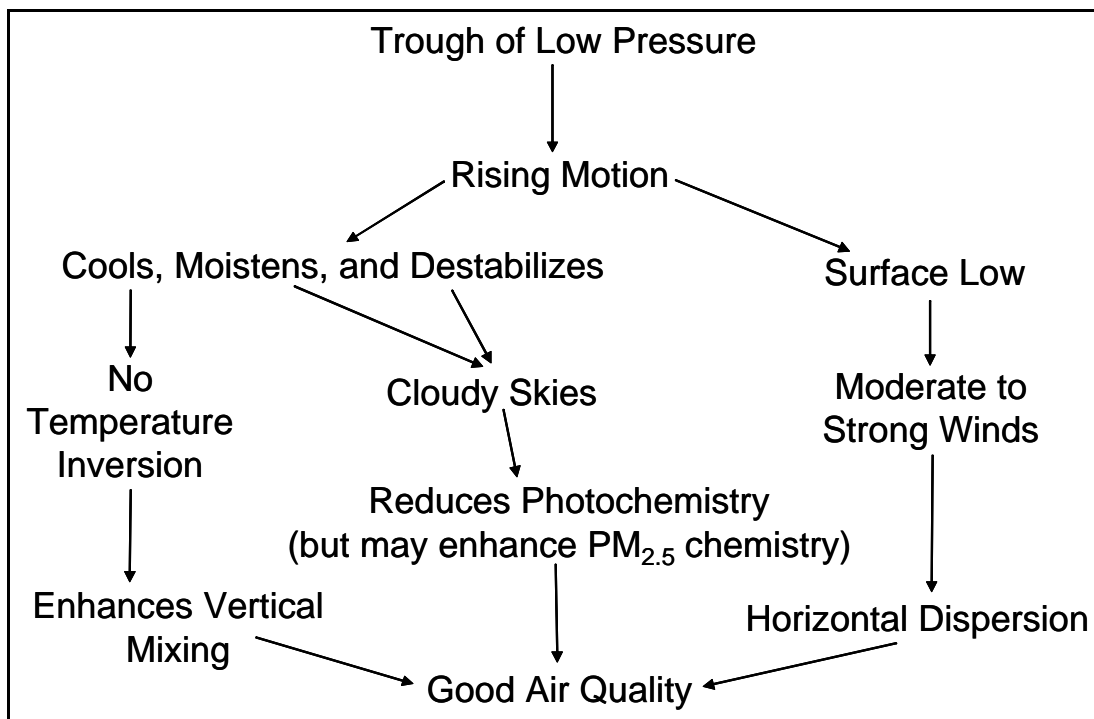
Figure 2-13.   Schematic of the typical meteorological conditions and air quality often associated with an aloft trough of low pressure.

Table 2-5. Meteorological phenomena and their influence on PM$_{2.5}$ and ozone concentrations.

Phenomena	Emissions	Chemistry	Accumulation/Dispersion/Removal
Aloft Pressure Pattern	No direct impact.	No direct impact.	Ridges tend to produce conditions conducive for accumulation of PM$_{2.5}$ and ozone. Troughs tend to produce conditions conducive for dispersion and removal of PM and ozone. In mountain-valley regions, strong wintertime inversions and high PM$_{2.5}$ levels may not be altered by weak troughs. In addition, high PM$_{2.5}$ and ozone concentrations often occur during the approach of a trough from the west.
Winds and Transport	No direct impact.	In general, stronger winds disperse pollutants, resulting in a less ideal mixture of pollutants for chemical reactions that produce ozone and PM$_{2.5}$.	Strong surface winds tend to disperse PM$_{2.5}$ and ozone regardless of season. However, strong winds can create dust which can increase PM$_{2.5}$ concentrations. In the East and Midwest, winds from a southerly direction are often associated with high PM$_{2.5}$ and ozone, due to transport from one region to another.
Temperature Inversions	No direct impact.	Inversions reduce vertical mixing and therefore increase chemical concentrations of precursors. Higher concentrations of precursors can produce faster, more efficient chemical reactions that produce ozone and PM$_{2.5}$.	A strong inversion acts to limit vertical mixing allowing for the accumulation of PM$_{2.5}$ or ozone.
Rain	No direct impact.	Rain can remove precursors of ozone and PM$_{2.5}$.	Rain can remove PM$_{2.5}$, but has little influence on existing ozone.
Moisture	No direct impact.	Moisture acts to increase the production of secondary PM$_{2.5}$ including sulfates and nitrates.	No direct impact.
Temperature	Warm temperatures are associated with increased evaporative, biogenic, and power plant emissions, which act to increase both PM$_{2.5}$ and ozone. Cold temperatures can also indirectly influence PM$_{2.5}$ concentrations (i.e., home heating on winter nights).	Photochemical reaction rates for ozone increase with temperature.	Although warm surface temperatures are generally associated with poor air quality conditions, very warm temperatures can increase vertical mixing and dispersion of pollutants.
Clouds/Fog	No direct impact.	Water droplets can enhance the formation of secondary PM$_{2.5}$. Clouds can limit photochemistry, which limits ozone production.	Convective clouds are an indication of strong vertical mixing, which disperses pollutants.
Season	Forest fires, wood burning, agriculture burning, field tilling, windblown dust, road dust, and construction vary by season.	The sun angle changes with season, which changes the amount of solar radiation available for photochemistry.	No direct impact.

2.3.1   Aloft Pressure Patterns

Aloft large-scale (1000 km or more) atmospheric circulations have a strong influence on regional and local weather conditions. Meteorologists generally focus on the so-called "500-mb level" to evaluate the aloft large-scale pressure systems. In particular, they focus on the location, size, intensity, and movement of 500-mb high-pressure ridges and low-pressure troughs (mountains of warm air and cold air, respectively). In general, poor air quality conditions are associated with high-pressure ridges and good air quality conditions are associated with low-pressure troughs. However, high $PM_{2.5}$ levels can occur without the existence of aloft ridges, from a very strong $PM_{2.5}$ emission source, such as a forest fire. **Figure 2-14** shows an example of a 500-mb ridge over the eastern United States on July 17, 1999, a day with high $PM_{2.5}$ concentrations throughout the region, and on September 21, 1999, a day with low $PM_{2.5}$ concentrations throughout the region. The existence of ridges and troughs can be diagnosed by reviewing weather charts, which are widely available as observations and forecasts on the Internet.

2.3.2   Temperature Inversions and Vertical Mixing

A temperature inversion is a layer of warm air above a layer of relatively cooler air. An inversion acts to limit the vertical mixing of pollutants, which allows concentrations to build. Several temperature inversions can exist at different altitudes in the lower part of the atmosphere. Typically, a temperature inversion can form from 25 to 300 m agl when the ground (and air near the ground) cools at night, while air above remains warmer. This type of inversion is called a nocturnal inversion. Nocturnal inversions are strongest when skies are clear at night and in the winter when nights are long. In the presence of clouds or strong winds, nocturnal inversion are often weak or do not form at all. Nocturnal inversions trap emissions, released during the overnight hours, close to the ground. As the ground warms during the day, the air near the surface warms, which erodes the nocturnal inversion. Typically, a nocturnal inversion disappears by mid-morning, allowing the trapped pollutants to mix vertically. If a nocturnal inversion is strong or if solar heating is weak, the inversion may not break until late in the day or at all. Under these circumstances, pollutants do not mix vertical and high pollutant concentrations are typical.

When there is an aloft ridge of high pressure over an area, there is often another inversion above the nocturnal inversion, called a subsidence inversion. A subsidence inversion is caused by sinking air in the mid- to low-levels of the atmosphere associated with the aloft ridge. As the air sinks, it warms due to compression. The warmest temperatures associated with the sinking air are typically found from 500 to 2000 m agl. When there is a strong subsidence inversion as indicated by aloft temperatures, the daytime heating at the surface may not be strong enough to break this inversion. Under such circumstances, vertical mixing of pollutants is weak and pollutants remain trapped near the surface for the entire day. An aloft inversion can also form when winds transport warm air at a greater rate aloft compared to the surface. This differential warming typically occurs on the west side of an upper-level ridge, ahead of an upper-level trough.
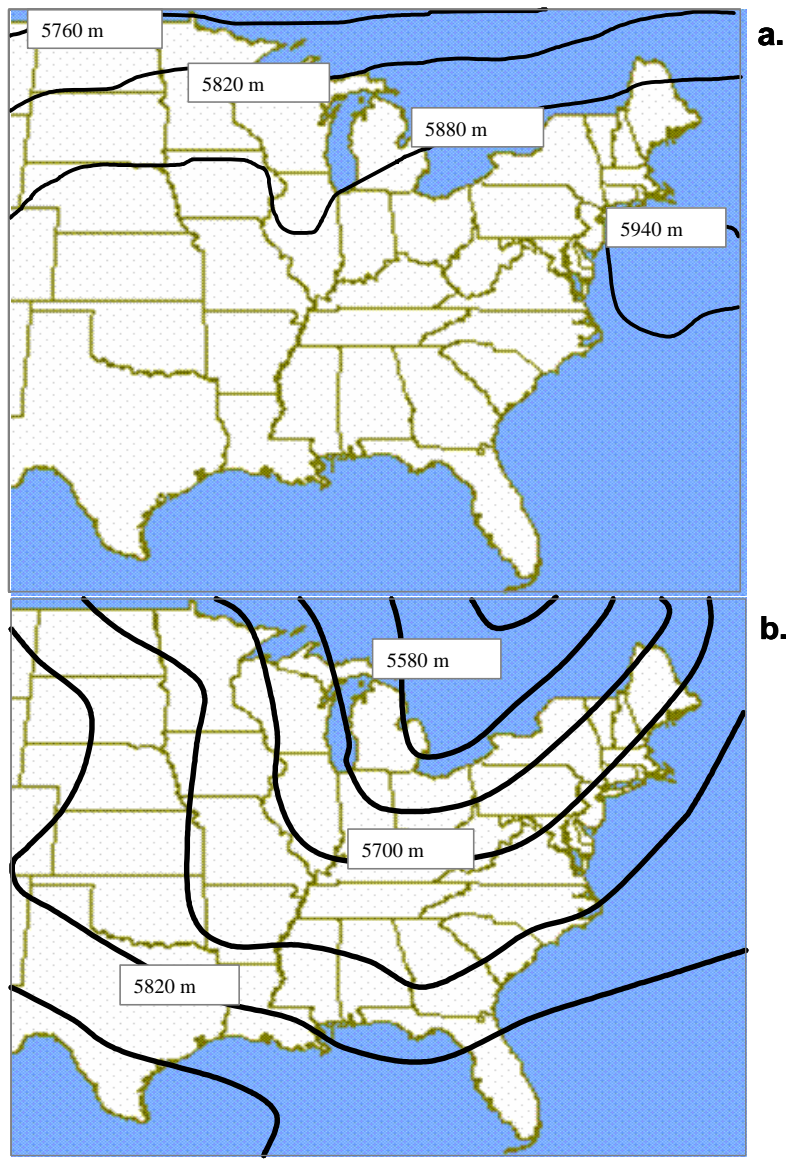
Figure 2-14.   500-mb heights (a) on the morning of July 17, 1999 (1200 UTC) and (b) on the morning of September 21, 1999 (1200 UTC).

Subsidence inversions do not form when there is an aloft trough over the region. This is because aloft troughs cause rising motion in the mid- to low-levels of the atmosphere. As the air rises, it cools due to expansion resulting in cooler air above warmer air. When there is cooler air above warmer air, the atmosphere is unstable. This instability causes vertical mixing, which dilutes pollutants whose source is near the surface.

Figure 2-15 shows the diurnal cycle of mixing, vertical temperature profiles, and boundary layer height on a day with a weak temperature inversion and on a day with a strong temperature inversion. On the day with the weak inversion, the convective boundary layer grows rapidly as the sun warms the ground during the day. The rapid growth of the convective boundary layer is associated with strong vertical mixing and the vertical dispersion of pollutants.

Temperature soundings

(a)

Weak and high inversion

Inversion Breaks

RL

CBL

NBL

NBL

Midnight   Sunrise   Sunset

(b)

Inversion Holds

Strong and low inversion

CBL

RL

NBL

NBL

Midnight   Sunrise   Sunset

RL = Residual Layer
CBL = Convective Boundary Layer
NBL = Nocturnal Boundary Layer

= Surface-based mixing depth

= Surface-based vertical mixing
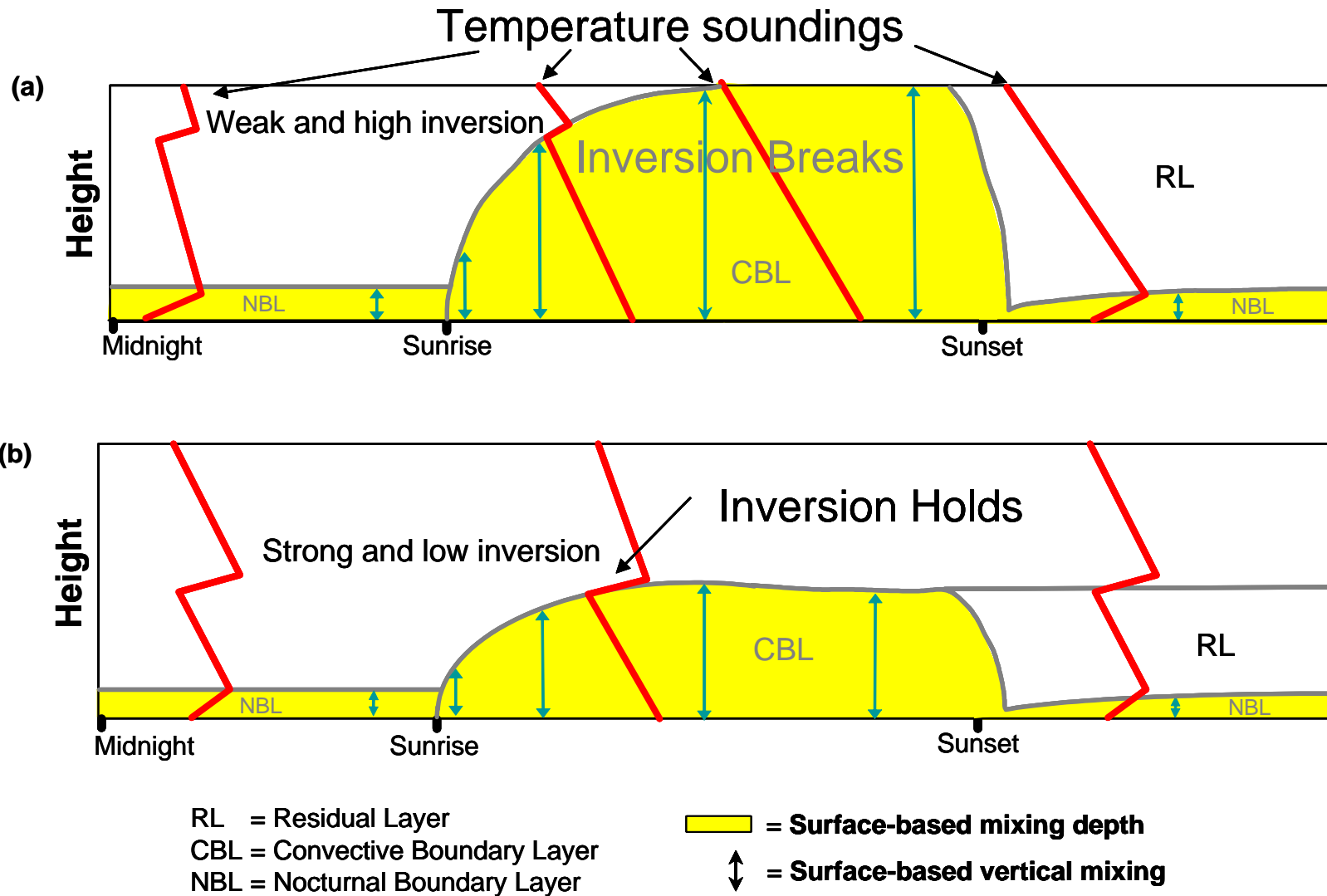
Figure 2-15.   Schematic showing diurnal cycle of mixing, vertical temperature profiles, and boundary layer height (a) on a day with a weak temperature inversion and (b) on a day with a strong temperature inversion. In (a) the pollutants mix into a large volume resulting in low pollution levels and in (b) pollutants mix into a smaller volume resulting in high pollution levels.

On the day with the strong inversion, the convective boundary layer growth is inhibited. The limited growth of the convective boundary layer is associated with weak vertical mixing and limited vertical dispersion of pollutants.

2.3.3   Winds and Transport

Winds can be described as large-, regional-, and local-scale. The large-scale winds are driven by the pressure gradients between surface high- and low-pressure systems. Light, regional, surface winds often occur near the center of the surface-high, below the ridge of high pressure, where pressure gradients are weak. Light winds are not effective at dispersing pollutants and, therefore, often occur during high pollutant concentrations. Moderate to strong winds occur between surface high and low pressure systems or near the center of low pressure systems, provided that moderate to strong pressure gradients exist. Moderate to strong surface winds act to disperse pollution and thus are typically associated with low pollutant concentrations. However, high pollutant concentrations can occur during moderate to strong wind conditions, if the winds transport pollution from one region to another.

In general, surface lows occur under the leading half of aloft troughs (typically on the eastern side), whereas, surface highs occur under the leading half of aloft ridges. **Figures 2-16 and 2-17**, respectively, show a 500-mb ridge and an associated surface high and a 500-mb trough and an associated surface low. The ridge and surface high on January 7, 2002, created conditions conducive to high $PM_{2.5}$ concentrations in Salt Lake City, Utah, including light surface winds and reduced vertical mixing. The trough and surface low on January 22, 2002, created conditions conducive to low $PM_{2.5}$ concentrations including strong surface winds, clouds, and vertical mixing.

Local winds are driven by the interaction between the large-scale pressure patterns and local forcing mechanisms. The local forcing mechanisms are driven by the diurnal temperature cycle and topography. Local winds tend to dominate over the large- and regional-scale winds when the large-scale pressure patterns are weak (i.e., at the center of a surface high pressure). The local winds may include land breezes, sea breezes, morning downslope flows, afternoon upslope flows, and terrain channeled flows, which can combine in various ways to recirculate air and cause stagnation.

2.3.4   Clouds, Fog, and Precipitation

Clouds, rain, and fog all influence pollutant concentrations through a variety of mechanisms as detailed in Table 2-5. Clouds form when air is cooled and water vapor condenses. This cooling can be caused by rising motion or contact with a cool surface such as a body of water or cool land during the night. Rising motion is generated by aloft low-pressure systems, frontal boundaries, air flowing over mountains, and convective instability (warm air below cooler air). Clouds are important because they typically reduce the amount of sunlight available for photochemical reactions that participate in the production of ozone and $PM_{2.5}$. Fog is a type of cloud that is in contact with or near the ground. Fog and clouds can dramatically increase the conversion of sulfur dioxide to sulfate (a secondary type of $PM_{2.5}$). Precipitation is a removal mechanism for fine particles.
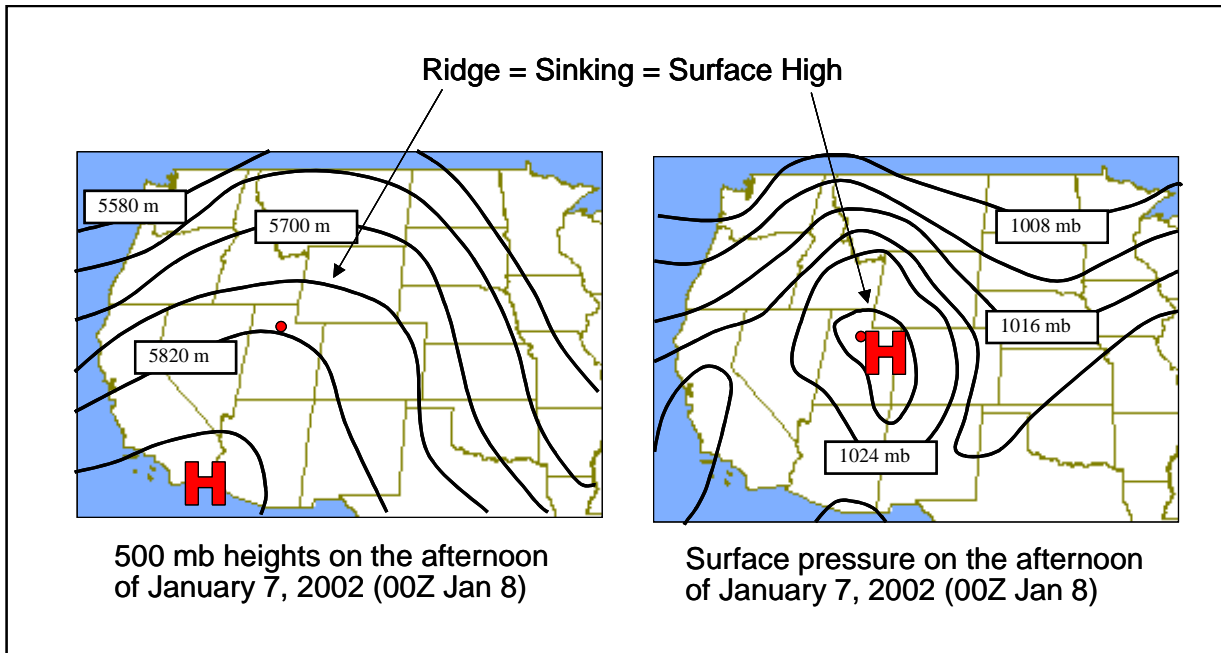
Figure 2-16.   500-mb heights (left) and surface pressure (right) on the afternoon of
January 7, 2002 (0000 UTC on January 8).
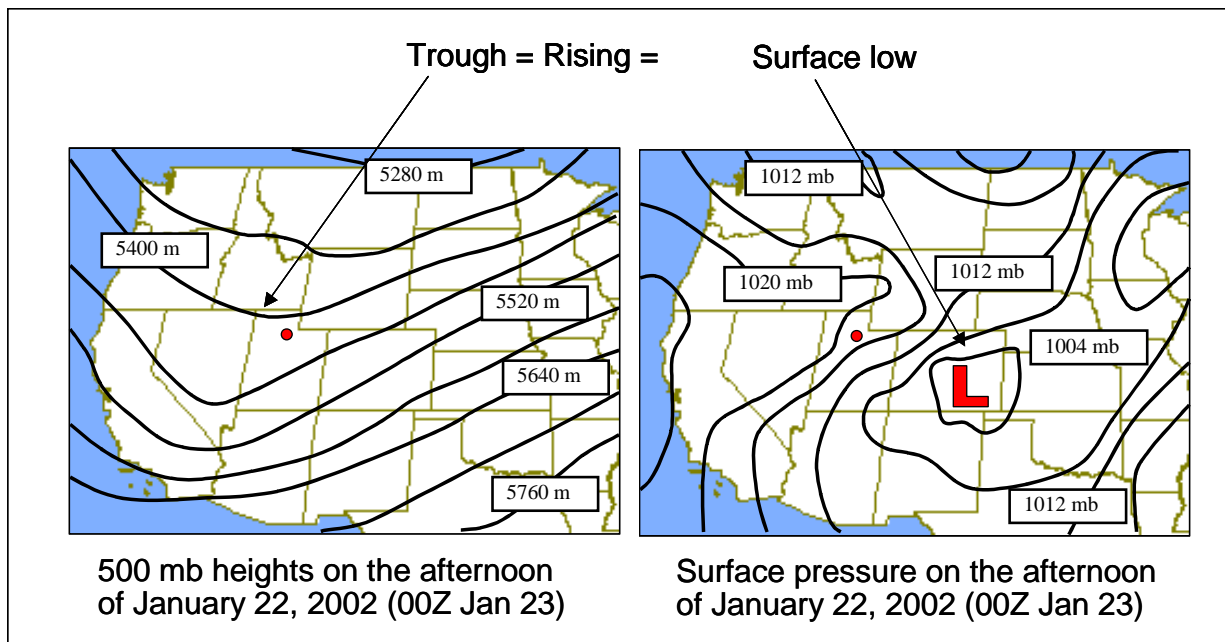
Figure 2-17.   500-mb heights (left) and surface pressure (right) on the afternoon of
January 22, 2002 (0000 UTC on January 23).

2.3.5 Weather Pattern Cycles

Typically, a region will cycle between a ridge and trough pattern every 2 to 7 days, but more stationary patterns can develop. Studying and understanding these cycles and their impact on local weather and air quality will help improve forecasting capabilities. **Figure 2-18** shows the typical life cycle of large-scale weather patterns. The following meteorological descriptions are generic and may vary from one region to another and between pollutants:

Ridge—high pressure pattern
(Figure 2-18a and b) is typically associated with poor air quality. This pattern occurs about one to two days after a cold front and trough have passed through an area. As surface high pressure develops in an area, winds become weak allowing for the accumulation of pollutants. Warming temperatures increase the biogenic and evaporative VOCs and lower humidity results in clearer skies, which are favorable for photochemistry. Sinking air (subsidence) warms and stabilizes the lower atmosphere, which suppresses cloud development and mixing. In addition, an aloft temperature inversion may form that inhibits vertical mixing and reduces dilution of pollutants. The aloft high pressure ridge typically occurs west of the surface high and can be diagnosed using 500-mb height fields.

Ridge—back side of high pattern
(Figure 2-18c and d) occurs as the surface high pressure moves east of the region and the accumulated pollutants are transported to downwind locations. In some regions, warm air is advected into the region and winds may increase from a southerly to a westerly direction depending on the orientation of the high. This pattern typically produces warm temperatures and relatively clear skies, even with a low-pressure system approaching from the west. Pollutant levels can remain high on these types of days, and the potential for longer-range transport is greater.

Trough— cold front pattern
(Figure 2-18e and f) is characterized by a low-pressure system at the surface and associated cold and warm fronts. Aloft at 500 mb, a trough of low pressure exists just upstream (west) of the surface low. This weather pattern produces clouds and precipitation that reduce photochemistry. Stronger winds and mixing also reduce pollutant concentrations.

Although aloft ridges and their associated regional and local weather conditions are generally associated with poor air quality, slight variations in the meteorological processes described above can have a dramatic affect on the spatial and temporal characteristics of air quality. It is these variations in meteorological processes that need to be analyzed and understood for different pollutants, seasons, and regions of interest to better understand the processes that produce air quality episodes.
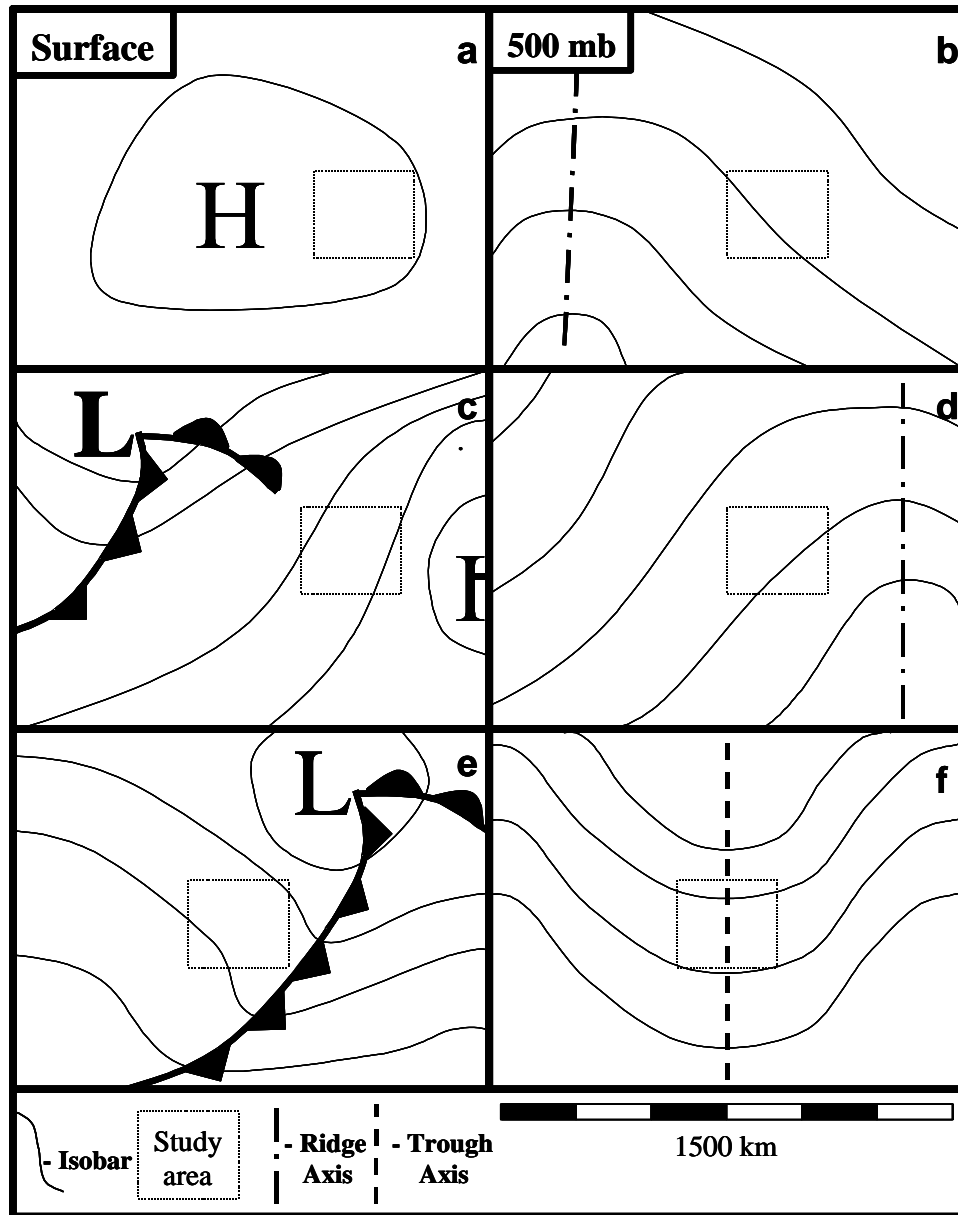
Figure 2-18. Life cycle of synoptic weather events at the surface and aloft at 500 mb for (a) and (b) Ridge—high pressure, (c) and (d) Ridge—back side of high, and (e) and (f) Trough—cold front patterns. Surface maps show isobars and frontal positions. The 500-mb maps show contours of equal height.

3. FORECASTING APPLICATIONS AND NEEDS

The success of an air quality forecasting program depends not only on accurate predictions, but also on meeting the needs and objectives of forecast recipients. For more than two decades the public has been warned of unhealthy air in several regions of the United States. Today, ozone and particulate matter forecasts are used throughout the United States for three major purposes: (1) public health notification, (2) episodic control programs (such as Ozone Action Days), and (3) scheduling specialized air monitoring programs. This section describes how air quality forecasts are used.

3.1   PUBLIC HEALTH NOTIFICATION

Pollution forecasts are typically issued by air quality agencies and communicated via television, radio, newspapers, the Internet, and fax to the public to give them adequate time to reduce or avoid exposure to unhealthy air. Forecasts are generally issued each day for the maximum ozone concentration expected for the current day and next day. For particulate matter, forecasts typically correspond to a 24-hour averaged concentration, which is standard of the Air Quality Index (AQI) for reporting PM. For example, the South Coast Air Quality Management District forecasts daily maximum ozone and average particulate matter concentrations for 40 sub-regions throughout the Los Angeles metropolitan area. For smaller cities, some agencies forecast the maximum ozone concentrations for the entire city (such as Charlotte, North Carolina). Air quality forecasts are usually formulated during the morning or early afternoon and then communicated to the public later the same day.

The needs of public health notification programs vary by region (see Section 5.1), but generally include:

- Pollution forecasting that errs on the side of public health (i.e., forecasts that tend to overpredict ozone or $PM_{2.5}$ rather than underpredict it).

- Forecasts that are as localized and specific as possible, particularly for large metropolitan regions with different geographic, emission, and source areas.

- Forecasts that provide the time of day when and location where high levels of ozone and/or particulate matter are expected.

- Forecasts that are completed as early in the day as possible, allowing sufficient time for public outreach personnel to communicate the forecast and other information to the public.

3.2   EPISODIC CONTROL PROGRAMS

Reducing air-quality violations and avoiding redesignation to nonattainment or a more severe classification is the major goal of episodic control programs (Jorquera, 1998; U.S. Environmental Protection Agency, 1997b). To accomplish this goal, episodic control programs educate the public about emission-producing activities and seek voluntary action from the public to reduce emissions on poor air quality days. More than 80 episodic control programs exist

throughout the United States and have various names, such as Ozone Action Day, Ozone Alert, Spare The Air, Don't Lite Tonight, and No Burn Days; but the underlying objectives are similar (see www.italladdsup.gov, www.epa.gov/airnow, and www.epa.gov/otaq/transp/publicat/pub_volu.htm for links to local programs).

Health officials rely on air quality forecasts (typically ozone and PM during the summer and particulate matter during the winter) to determine whether or not to call an Action Day and seek voluntary action from the public to reduce emission-producing activities (e.g., driving, mowing lawns, residential wood burning, etc.) on forecasted poor air quality days. In addition, business and industry often participate by offering services that help reduce pollution (e.g., free bus rides on high ozone days). Since these programs ask the public to reduce pollution voluntarily, the credibility of the program depends on forecast accuracy.

Typically, forecasters issue ozone forecasts midday or in the afternoon for the next-day's peak ozone or 24-hr average $PM_{2.5}$ concentration. Public outreach personnel then communicate the forecasts and emission-reduction tips for Action Days to the public so they can plan their activities for the next day (i.e., carpooling). Therefore, forecasters must issue predictions as early in the day as possible to ensure timely information dissemination.

Episodic control programs typically have the following forecasting needs:

- Minimizing the number of forecasts that falsely alert the public (i.e., minimize overpredicting). These "false alarms" may cause the public to ignore the warnings and over time would diminish the effectiveness of the program.

- Receiving forecasts as early as possible to allow sufficient time for public outreach personnel to communicate the forecast and other information to the public.

- Including a discussion of current and forecasted weather and air quality conditions in the forecast. Public outreach personnel can use this information to better communicate the forecast to the media and public.

- Providing an indication of forecast uncertainty so that public outreach personnel can plan the amount and degree of media and outreach spending necessary given the uncertainties in the forecast.

3.3   SPECIALIZED MONITORING PROGRAMS

Specialized monitoring programs are field studies run by federal, state, and private agencies to collect surface and/or aloft air quality and meteorological measurements on poor air quality days. Personnel for these programs have used ozone forecasts for decades to help schedule and plan intensive sampling efforts and more recently initiated studies using $PM_{2.5}$ forecasts. Since the 1970s, field study personnel have used ozone forecasts to plan when and where to conduct ozone sampling using expensive measurement equipment (e.g., aircraft, rawinsondes, etc.). They also use forecasts to help conserve resources by sampling only on high ozone days and to provide advanced warning to "gear up" for sampling on these days. Historically, program personnel only needed ozone forecasts for short-term projects lasting several months during selected study years. Recently, with new continuous monitoring projects

like Photochemical Assessment Monitoring Stations (PAMS), the need for accurate ozone forecasts has increased. With the PAMS program, the EPA requires some state agencies to perform more extensive ozone and ozone precursor monitoring in areas with persistently high ozone levels. Specialized carbonyl and hydrocarbon monitoring as well as aloft sampling by aircraft, are performed in many regions only on predicted high ozone days.

Specialized monitoring programs typically have the following forecasting needs:

- Forecasts that are as localized and specific as possible, particularly for large metropolitan regions so region-specific sampling can be conducted.

- Multi-day forecasts in order to allow sufficient time to prepare monitoring equipment and personnel.

- Forecast information about when an episode will begin and when it will end, including the day prior to the episode ("ramp-up" day) when sampling is often conducted to understand the air quality and meteorological conditions prior to an episode.

In summary, to make air quality forecasts as effective as possible, it is critical that forecasters understand how air quality forecasts are used in their region. The material provided in Section 5.1 will help to identify and determine these needs.

4. DEVELOPING OZONE AND PM$_{2.5}$ FORECASTING METHODS

Many methods exist for predicting air quality concentrations. Some methods are simple to develop and easy to operate, yet are not very accurate. Other methods are more difficult to develop but produce more accurate forecasts. Most forecasters use several methods—some objective, others subjective—to forecast ozone and PM$_{2.5}$. Using several methods can balance one method's strengths with another method's limitations to produce a more accurate forecast. Since PM$_{2.5}$ forecasting is new for most agencies, fewer PM$_{2.5}$ forecasting techniques have been tested and used.

Section 4.1 describes the most commonly used forecasting methods. Each subsection defines a method, explains how it works and how to develop it for a particular program, and lists its strengths and limitations.

All of the methods described here use multiple predictor variables to forecast either ozone or PM$_{2.5}$ concentrations. The process of selecting these predictor variables is described in Section 4.2.

4.1   FORECASTING METHODS

This section presents several of the most common methods used to forecast ozone concentrations; these can also be used to predict PM$_{2.5}$ concentrations, provided that long-term (ideally 3 or more years) of historical PM$_{2.5}$ data are available. For easy comparison, **Table 4-1** lists and summarizes the methods.

4.1.1 Persistence

Persistence means to continue steadily in some state. Persistence forecasting is simply saying that today's or yesterday's pollutant concentration will be the same as tomorrow's pollutant concentration. Persistence air quality forecasting is best used as a starting point and to help guide other forecasting methods. In addition, a persistence forecast can be used as a reference (or baseline) against which to compare forecasts generated from other methods. It should not be used as the only forecasting method.

How persistence forecasting works

Persistence forecasting works because atmospheric variables, including ozone and PM$_{2.5}$, exhibit a positive statistical association with their own past or future values (Wilks, 1995). That is, large values of a variable tend to be succeeded by large values; likewise, small values of a variable tend to be succeeded by small values. For example, if today's 24-hour PM$_{2.5}$ concentration was 10 ug/m^3, it is likely that tomorrow's 24-hour PM$_{2.5}$ concentration will also be relatively low. Similarly, if today's 24-hour PM$_{2.5}$ concentration is 75 ug/m^3, it is more likely that tomorrow's 24-hour PM$_{2.5}$ concentration will be high than low. Some PM events may

Table 4-1.   Comparison of forecasting methods.

	Persistence	Climatology	Criteria	CART	Regression	Neural Networks	3-D Air Quality Models	Phenomenological /Intuition
Development Effort	Low	Low/Moderate	Low/Moderate	Moderate	Moderate	Moderate/High	Very High	High
Operational Effort	Low	Low	Low	Low	Moderate	Moderate	Moderate/High	Moderate
Accuracy	Low	Low	Moderate	Moderate/High	Moderate/High	Moderate/High	Moderate/High	High
Method Description	Today's (or yesterday's) observed pollutant concentration is tomorrow's forecasted pollutant concentration.	Historical frequency of pollutant events help guide and bound pollutant forecast.	When parameters that influence pollution are forecasted to reach a pre-determined level (criteria), high pollutant concentrations are forecasted.	A decision tree predicts pollutant concentration based on values of various meteorological and air quality parameters.	A regression equation predicts pollutant concentrations using observed and forecasted meteorological and air quality variables.	A non-linear set of equations and weighting factors predicts pollutant concentrations using observed and forecasted meteorological and air quality variables.	A prognostic modeling system simulates the physical and chemical processes that lead to the formation and accumulation of air pollutants.	A person synthesizes meteorological and air quality information including pollutant concentration predictions from other methods to produce a final air quality forecast.
Development								
Expertise[1]	–	–	Ability to identify key predictor variables.	Understanding of statistics and CART.	Understanding of statistics and regression.	Understanding of statistics and neural networks.	High level understanding of meteorological and air quality relationships, and meteorological, emissions, and air quality models.	Experience in air quality forecasting and a conceptual understanding of meteorological and air quality processes.
Software /Hardware	Spreadsheet/PC	Spreadsheet/PC	Statistical Software/PC	CART Software/PC	Statistical Software/PC	Statistical and Neural Network Software/PC	Prognostic meteorological model, emissions model, and air quality models; compilers/High speed computer system with large memory and disk storage.	None
Operations								
Expertise	Ability to acquire today's and yesterday's pollutant data.	Ability to acquire pollutant data and interpret graphs and tables.	Ability to acquire observed and forecasted meteorological and air quality data.	Ability to acquire observed and forecasted meteorological and air quality data and use a decision tree.	Ability to acquire observed and forecasted meteorological and air quality data and use a computational program or spreadsheet.	Ability to acquire observed and forecasted meteorological and air quality data and use a computational program.	Basic understanding of meteorological and air quality relationships to determine reasonableness of model results.	Ability to synthesize meteorological and air quality information including pollutant predictions from other methods to produce a pollutant forecast.

[1] All methods require a basic understanding of meteorological and air quality relationships and basic data processing skills.

Table 4-1.   Comparison of forecasting methods.

	Persistence	Climatology	Criteria	CART	Regression	Neural Networks	3-D Air Quality Models	Phenomenological /Intuition
Forecast production time	<1/2hr	<1hr	<1hr	<1hr	1hr	1hr	6 to 12 hrs (90% is computational time)	1 to 3 hrs
Data needs	Yesterday's air quality data.	None	Observed and forecasted upper-air and surface meteorological and air quality data.	Observed and forecasted upper-air and surface meteorological and air quality data.	Observed and forecasted upper-air and surface meteorological and air quality data.	Observed and forecasted upper-air and surface meteorological and air quality data.	Prognostic gridded meteorological fields, gridded emissions, and boundary conditions.	Observed and forecasted meteorological data and charts and observed air quality data.
Software /Hardware	None	None	Data acquisition PC	Data acquisition PC	Computational program or spreadsheet/Data acquisition PC	Computational program/Data acquisition PC	Meteorological and air quality grid models/High-speed computer system with large memory and disk storage.	Data acquisition PC
Strengths	Works well in areas that have several continuous days of high pollutant and low pollutant concentrations.	Helps guide and bound forecasts derived from other methods.	Quick, used to get initial "idea" about forecast conditions.	Automatically differentiates between days with similar pollutant concentrations.	Commonly used and easy to operate. Produces generally good forecasts.	Allows for non-linear relationships to develop.	Predicts pollutant concentrations in areas that are not monitored. Helps in understanding pollutant processes including transport issues.	Helps temper the predictions from other methods with common sense and experience. Typically has the highest accuracy.
Potential Limitations	Doesn't predict the beginning or end of an episode; low accuracy.	Not a stand-alone method.	Is not well suited to forecast exact concentrations.	Requires a modest amount of expertise to develop.	Doesn't accurately predict extreme concentrations.	Doesn't accurately predict extreme concentrations. 50% more effort to develop than regression with only slight improvement in forecast accuracy.	Expensive and difficult to develop. Accuracy of air quality predictions depend on accuracy of meteorological and emissions predictions.	Prediction may be biased from one forecaster to another.

4-3

extend over large regions, and $PM_{2.5}$ concentrations from one day to the next may be similar and have persistence.

Air quality forecasting using the Persistence method works because air quality concentrations are highly dependent on synoptic-scale weather, which typically exhibits similar characteristics for several days, and, therefore, air quality concentrations are also typically similar for several days. For example, a high-pressure system will usually persist over an area for several days during which time weather and air quality concentrations exhibit modest day-to-day variation. Likewise, if an area is under the influence of a low-pressure system, the area will likely exhibit low air quality concentrations for several days until the synoptic pattern changes.

An analysis of the data presented in **Table 4-2** illustrates how persistence forecasting works. Table 4-2 shows peak 8-hr ozone concentrations for a sample city for 30 consecutive days. Seven days during this period had peak ozone concentrations greater than the federal 8-hr standard and five of these days occurred after an exceedance; thus, the odds of an ozone exceedance occurring on the day after an exceedance are 5 out of 7 days (71.4%). The odds of a non-exceedance occurring after a non-exceedance are 20 out of 22 days (90.9%). Therefore, in this example, if the Persistence method was used to forecast a non-exceedance or an exceedance, the forecast would be accurate 25 out of 29 days, or 86% of the time. Note that the first day of the forecast period does not count in the forecast statistics because Day 1 is not a forecast day.

As shown in Table 4-2, the Persistence method does not correctly predict the beginning or end of an episode. However, the Persistence method can be used to help guide forecasts and predictions from other methods.

Modifying a persistence forecast with forecasting experience can help improve forecast accuracy. For example, if today's weather conditions (which included clear skies) were ideal for high ozone concentrations, today's observed peak ozone concentration reached 130 ppb, and tomorrow's weather conditions are expected to be the same as today's except for partly cloudy skies, using the Persistence method the first cut at the forecast is 130 ppb, but the forecast is lowered to 100 ppb to account for the influence of cloud cover. The Persistence method provides a good starting point for the next-day ozone forecast.

Persistence forecasting development

Although the Persistence method requires no real development, forecasters should be sure that the method will work in their area. The following steps describe how to test the effectiveness of persistence forecasting in a particular area.

1. Create a data set containing at least four years of recent ozone or $PM_{2.5}$ data.

2. From this data set, use each day's maximum air quality concentration to simulate a forecast for the next day (i.e., use the Persistence method). Compare the forecast and observed pollutant concentrations for the historical data set and compute the forecast verification statistics provided in Section 5.6.

Table 4-2. Peak 8-hr ozone concentrations for a sample city for 30 consecutive days. Exceedance days are shown in bold.

Day	Ozone (ppb)	Day	Ozone (ppb)
1	80	**16**	**120**
2	50	**17**	**110**
3	50	18	80
4	70	19	80
5	80	20	70
6	**100**	21	60
7	**110**	22	50
8	**90**	23	50
9	80	24	70
10	80	25	80
11	80	26	80
12	70	27	70
13	80	28	80
14	**90**	29	60
15	**110**	30	70

3. Keep in mind the following development issues:

- Consider when the forecast will be issued to determine what ozone or $PM_{2.5}$ data are available. For example, if a forecast must be issued at 11:00 a.m. for the next day and the current day's peak ozone or $PM_{2.5}$ concentration has not yet been observed, the previous day's peak ozone concentration would be used for the next-day forecast.

- The Persistence method only works well for regions that experience several continuous days of similar air quality. This approach fails if pollutant episodes typically last only one day.

Persistence forecasting operations

Using the Persistence method to forecast pollutant concentrations requires very little expertise and is perhaps the easiest and quickest of all air quality forecasting techniques, yet its accuracy is the poorest. Effectively using the Persistence method requires forecasters to recognize when weather patterns are static and when they are changing. Persistence forecasting can be effective under static conditions, but generally ineffective under changing conditions.

Persistence forecasting strengths

- Persistence forecasting can be very accurate during several days with similar weather conditions.

- It provides a starting point for an air quality forecast that can be refined by using other forecasting methods.

- It is easy to produce and operate and requires little expertise.

Persistence forecasting limitations

- The first and last days of a pollution episode cannot be predicted using persistence forecasting.

- This method does not work well under changing weather conditions when accurate air quality predictions can be most critical.

4.1.2 Climatology

Climatology is the study of average and extreme weather conditions at a given location. Climatological techniques can be applied to air quality forecasting. Although not very accurate as a predictive tool, climatology can help forecasters bound and guide their air quality predictions.

How climatology works

Climatology works because history tends to repeat itself, especially when it comes to seasonal weather. Since pollutant concentrations are highly weather dependent, air quality climatologies can be used in the same manner as weather climatologies. For example, an initial forecast is for a maximum temperature of 105°F in downtown Boston for August 13. According to a climate table, a maximum temperature of 105°F has never occurred in Boston and the forecast is probably too high. Thus, the forecast is adjusted down to 100°F. The climate data acted as a bound and a guide to the temperature forecast. Analogously, let's say that ozone is being forecasted for April 10 for upstate New York, and the forecast techniques indicate that an exceedance may occur. A climate table (**Table 4-3**) shows that upstate New York had no exceedances in April for the 15-year period of records. Based upon the additional information provided by the climate table, a non-exceedance is forecasted for April 10. The table served as a complimentary tool to other forecast methods and helped improve forecast accuracy.

Developing climate tables

Complete the following steps to develop ozone climate tables for a particular region:

1. Create a data set containing at least four years of recent ozone or $PM_{2.5}$ data.

2. Examine the data for quality and be sure to note when emissions changed significantly due to the use of reformulated fuel in the area, for example. Changes in emissions can result in the same weather conditions producing lower pollutant concentrations. Also

Table 4-3. Annual summaries of 1-hr ozone exceedance days for New York State (1983-1997), (Taylor, 1998).

Year	April Total	April Downstate	April Upstate	May Total	May Downstate	May Upstate	June Total	June Downstate	June Upstate	July Total	July Downstate	July Upstate	August Total	August Downstate	August Upstate
1983	1	1	0	0	0	0	11	10	1	8	8	1	10	10	1
1984	0	0	0	1	0	1	5	5	0	3	3	0	7	7	0
1985	0	0	0	2	2	0	2	1	1	8	7	1	1	1	0
1986	0	0	0	2	1	1	3	3	0	4	3	1	0	0	0
1987	0	0	0	4	4	2	3	3	1	10	10	1	3	0	3
1988	0	0	0	3	3	2	9	7	6	14	12	7	7	6	2
1989	0	0	0	0	0	0	2	2	0	4	4	1	1	1	0
1990	0	0	0	0	0	0	2	1	1	3	3	0	2	2	0
1991	0	0	0	2	2	0	3	3	0	5	5	2	5	4	1
1992	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0
1993	0	0	0	0	0	0	0	0	0	6	4	2	3	3	0
1994	0	0	0	0	0	0	3	2	1	5	4	1	1	1	0
1995	0	0	0	0	0	0	1	1	1	4	2	2	3	3	0
1996	0	0	0	0	0	0	0	0	0	3	1	2	1	1	0
1997	0	0	0	0	0	0	3	3	0	4	4	0	1	1	0
Total	1	1	0	15	13	6	47	41	12	81	70	21	46	41	7
Avg/Year	0	0	0	1	1	0	3	3	1	6	5	2	3	3	1

note that changes in the monitoring network can dramatically change the maximum observed pollutant concentrations and/or the number of exceedances.

3. Create tables or charts for forecast areas containing the following types of information:

   - All-time maximum ozone or $PM_{2.5}$ concentrations (by month, by site).

   - Duration of high ozone or $PM_{2.5}$ episodes (number of consecutive days, hours of high pollutant each day).

   - Average number of days with high ozone or $PM_{2.5}$ by month and by week.

   - Day-of-week distribution of high ozone or $PM_{2.5}$ concentrations.

   - Average and peak $PM_{2.5}$ concentrations by holidays and non-holidays, weekends and weekdays.

   Examples of such charts are shown in **Figures 4-1 and 4-2** for Pittsburgh, Pennsylvania.

4. If significant changes in emissions occurred, it may be useful to divide the climate tables or charts into "before" and "after" periods.

5. Examine the tables or charts for usefulness. For example, if there are differences between weekend and weekday ozone or $PM_{2.5}$ concentrations or exceedance frequency, then a climate table showing the frequency of high ozone or $PM_{2.5}$ concentrations by day of week may be quite useful.
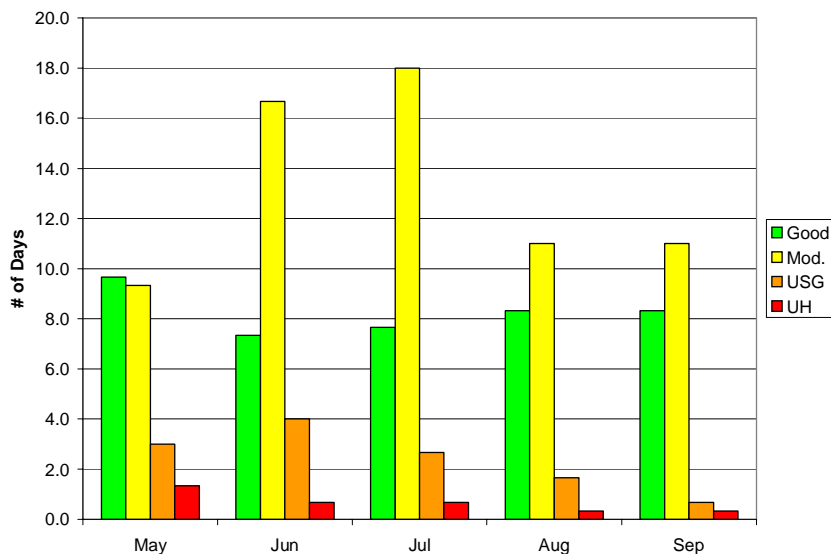
Figure 4-1.   Monthly distribution of the average number of days that $PM_{2.5}$ concentrations fell into each AQI category from 1999 to 2001 based on the peak 24-hr average $PM_{2.5}$ concentrations measured at 12 sites in the greater Pittsburgh region. Mod. = Moderate, USG = Unhealthy for Sensitive Groups, and UH = Unhealthy. Note that the totals for each month are less than 30 or 31 due to missing data (Dye et al., 2002a).
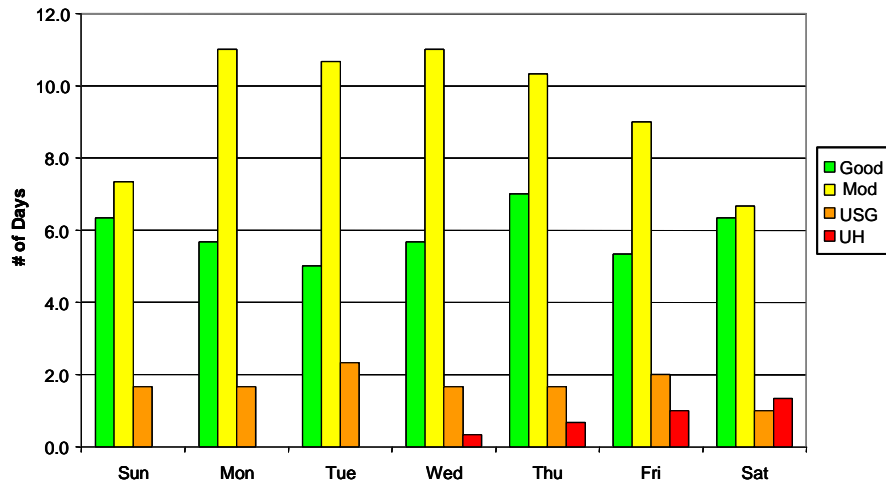
Figure 4-2. Summertime day-of-week distribution of the average number of days per year that $PM_{2.5}$ concentrations fell into each AQI category from 1999 to 2001 based on the peak 24-hr average $PM_{2.5}$ concentrations measured at 12 sites in the greater Pittsburgh region. Mod. = Moderate, USG = Unhealthy for Sensitive Groups, and UH = Unhealthy. Note that the totals for each day are different due to missing data (Dye et al., 2002a).

Climatology in operations

Using climate tables does not require much expertise. The forecaster need only understand that the tables are tools to guide and bound the air quality forecasts created using other methods. Consulting climate tables may be useful when other methods predict extreme events and they may help determine if a forecast of extreme concentrations is warranted. The forecaster can also use climatological information in the forecast discussion to provide context. For example, "Tomorrow's predicted peak ozone concentration is 150 ppb; this would be the first time in two years that ozone has reached this level."

Climatology strengths

- Climatology acts to bound and guide an air quality forecast produced by other methods.
- It is easy to develop.

Climatology limitations

- Climatology is not a stand-alone forecasting method but a tool to complement other forecast methods.

- It does not account for abrupt changes in emission patterns such as those associated with the use of reformulated fuel, a large change in population, forest fires, etc.

4.1.3 Criteria

A criterion is a principle by which something is evaluated. The Criteria method in air quality forecasting uses threshold values (criteria) of meteorological or air quality variables to forecast pollutant concentrations. Sometimes called "rules of thumb," the Criteria method is commonly used in many forecasting programs as a primary forecasting method or combined with other methods. It serves as a fundamental method on which to start an air quality forecasting program.

How the Criteria method works

This method is based on the fact that specific values of certain meteorological and air quality variables are associated with high pollutant concentrations. Once known, forecasters can look for the occurrence of the criteria in weather forecasts and predict pollutant concentrations from them. For example, high pollutant concentrations are often associated with hot temperatures and, thus, temperature can be used as one predictor of pollutant concentration. For instance, historical analysis may show that a temperature at or above 90°F is required to have an 8-hr ozone concentration greater than 85 ppb in a particular area. Thus, 90°F would be a threshold value (criterion) for an 8-hr ozone exceedance.

Since ozone and $PM_{2.5}$ formation is complex, forecasters must use several variables and associated criteria to accurately forecast ozone or $PM_{2.5}$. **Table 4-4** shows an example of multi-parameter criteria used to forecast ozone concentrations in Austin, Texas. This table indicates the conditions necessary for a 1-hr ozone exceedance. To have an exceedance in Austin in July, the predicted maximum temperature must be at least 92°F, the temperature difference between the morning low and afternoon high must be at least 20°F, the average daytime wind speed must be less than 5 knots, the afternoon wind speed must be less than 7 knots, and the prior day's peak 1-hr ozone concentration must be at least 70 ppb. Note that the meteorological criteria are predicted values for the next day. If these conditions are not met, then an exceedance is less likely and, thus, would not be forecasted.

The Criteria method is best suited to help forecast an exceedance, non-exceedance, or pollution in a particular AQI category range rather than an exact concentration.

Criteria method development

Complete the following steps to develop the Criteria method for air quality forecasting in a particular region:

1. Determine the important physical and chemical processes that influence ozone or $PM_{2.5}$ concentrations in the area. This helps with identification of variables to use for the criteria. Literature reviews, historical case studies, and climatological analysis (as discussed in Section 5.2) can help with this step.

2. Select variables that represent the important physical and chemical processes that influence ozone or $PM_{2.5}$ concentrations in the area. Useful variables may include: maximum temperature, morning and afternoon wind speed, cloud cover, relative humidity, 500-mb height, 850-mb temperature, etc. Statistical software can be used to

Table 4-4. Criteria for 1-hr ozone exceedances in Austin, Texas, used by the Texas Natural Resource Conservation Commission (Lambeth, 1998).

Month	Daily Temp Max (above °F)	Daily Temp Range (above °F)	Daily Wind Speed (below knots)	Wind Speed 15-21 UTC (below knots)	Prior Day's Ozone 1-hr Max (above ppb)
Apr	78	20	8.0	6.0	70
May	84	20	8.5	10.0	70
Jun	84	20	6.0	9.0	70
Jul	92	20	5.0	7.0	70
Aug	92	20	5.0	7.0	70
Sep	87	18	5.0	7.0	75
Oct	87	18	5.0	5.0	75

limit the number of variables by identifying the most important and significant ones. A discussion of variable selection is presented in Section 4.2.

3. Acquire at least four years of recent ozone or $PM_{2.5}$ data and surface and upper-air meteorological data.

4. Determine the threshold value for each parameter that distinguishes high and low pollutant concentrations. For example, create scatter plots of ozone or $PM_{2.5}$ vs. particular parameters to help determine the thresholds, as shown in **Figure 4-3**. In this example, the criterion of 28°C (81°F) helps distinguish higher ozone concentrations from lower concentrations in Charlotte, North Carolina. With the criterion of 28°C, only two ozone concentrations greater than or equal to 85 ppb occur when the temperature is less than the criteria. However, many low ozone concentrations (less than 85 ppb) occur when the maximum temperature is greater than or equal to 28°C, thus criteria for other variables (wind speed, cloud cover, etc.) are needed to accurately identify high ozone days.

5. Use an independent data set (i.e., a data set not used for development) to evaluate the selected criteria.

6. Keep in mind the following development issues:

- Evaluate threshold values for each month or season to understand how the values change.

- When emissions compositions change, the peak pollutant concentration associated with an established criteria may change. When this happens, the criteria method should be updated or the exceptional event should be noted.

Criteria method operations

The Criteria method is one of the easiest methods to use. Data must only be acquired and checked against the established criteria to determine the air quality forecast. Although use of this
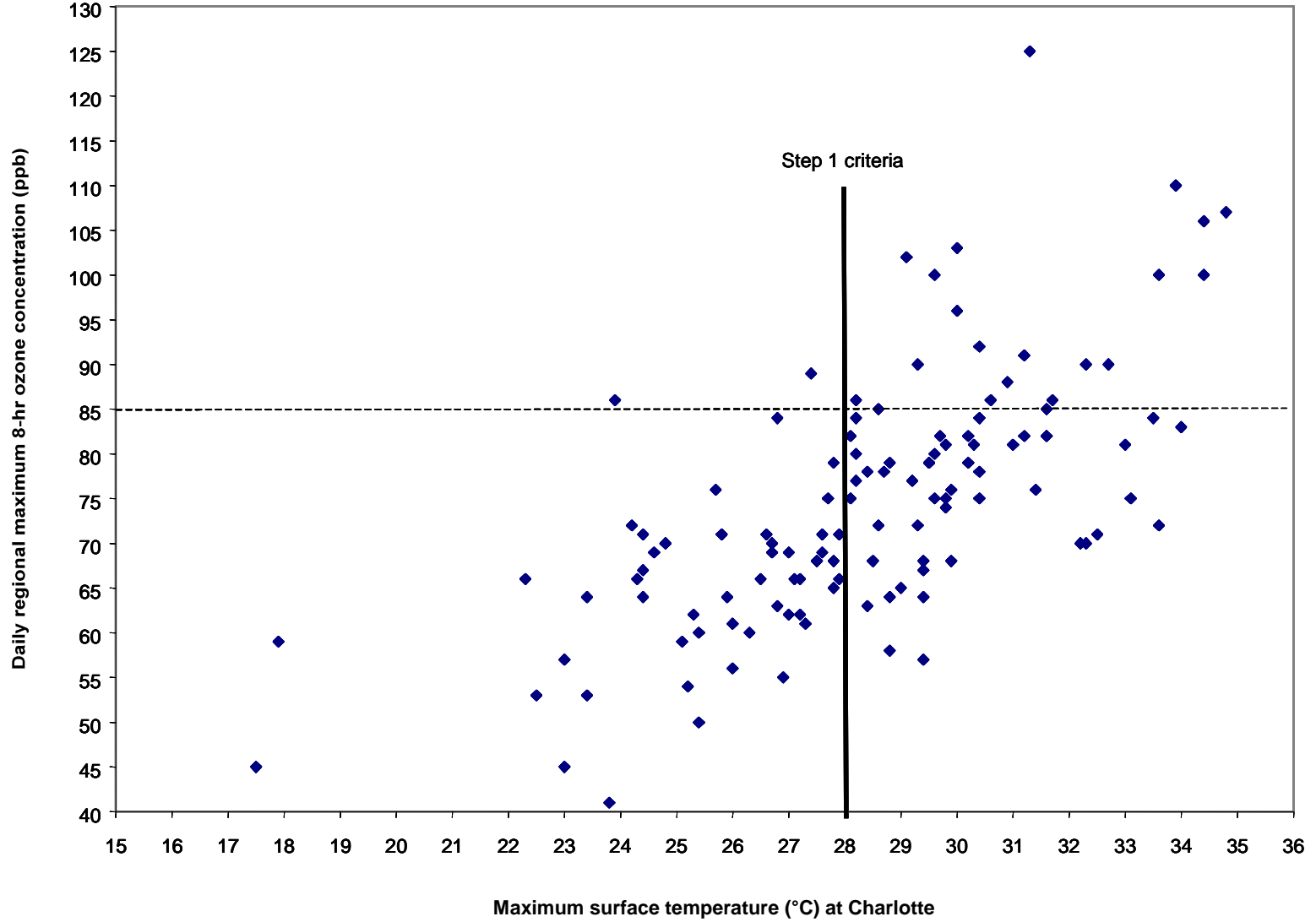
Figure 4-3.   Scatter plot of maximum surface temperature and regional maximum 8-hr ozone concentration in Charlotte, North Carolina in 1996 (MacDonald et al., 1998).

method does not require an understanding of meteorology and air quality processes, it is advisable that someone with such knowledge be involved in the development of the method and check the air quality predictions for physical reasonableness.

Criteria method strengths

- It is easy to operate.

- It is relatively easy to develop, and it can be refined each year as more knowledge is acquired.

- It is an objective method that alleviates potential biases arising from human subjectivity.

- It complements other forecasting methods. This method can easily be used first to determine whether or not the situation warrants spending more time on fine-tuning the forecast or using more sophisticated methods.

Criteria method limitations

- Selection of the variables and their associated thresholds is subjective.

- It is not well suited for predicting exact pollutant concentrations; but better suited for forecasting pollutant concentrations above or below a certain concentration or AQI category.

- This is an objective tool that can only predict pollutant concentrations based on information contained within the observed and forecasted data. Changes in the predicted weather conditions may not be reflected in the predictor variables and may cause uncertainty in the air quality predictions.

4.1.4 Classification and Regression Tree (CART)

Classification and Regression Tree (CART) is a statistical procedure designed to classify data into distinct (or dissimilar) groups. For air quality forecasting, CART enables a forecaster to develop a decision tree to predict pollutant concentrations based on the values of predictor variables that are well correlated with pollutant concentrations.

How CART works

CART uses software to develop a decision tree by continuously splitting peak pollutant concentration data into two groups based on a single value of a selected predictor variable (Horie, 1988; National Research Council, 1991; Stoeckenius, 1990). The selected predictor variable and the threshold cutoff value are determined by the CART software. The software identifies the variables with the highest correlation with the pollutant. It seeks to split the data set into the two most dissimilar groups. The splitting of the data set and tree development continues until the data in each group are sufficiently uniform. Predictor variables used in CART typically include meteorological data (i.e., temperature, wind speed, cloud cover, etc.), but may also include air quality data or other data such as the day of week or length of day. See Section 4.2 for a list of common predictor variables.

Figure 4-4 shows a decision tree for maximum ozone concentrations created using CART. This decision tree was developed by Horie (1988) for the South Coast Air Basin in California. As discussed by Horie, of the 73 variables used in the analysis, the temperature at 850 mb describes the greatest amount of the variance in maximum ozone concentration; it was used as the first data split. This split resulted in the two most dissimilar groups: Group 1 (for 850-mb temperature less than 17.1°C) had an average ozone concentration of 90 ppb, and Group 2 (for 850-mb temperature greater than 17.1°C) had an average concentration of 230 ppb. CART was then applied to each group using the same set of 73 predictor variables. The low ozone Group 1 was split again by 850-mb temperature at 9.9°C, while the high ozone Group 2 was split by 900-mb temperature at 24.3^{o}C. The tree growth continued until there were 10 distinct groups. In this example, the entire decision tree explains 80% of the variance in the daily maximum ozone concentration.

It is quite simple to forecast pollutant concentrations using the decision tree created by the CART analysis. For the example shown in Figure 4-4, if the forecasted predictor variables include an 850-mb temperature of 20°C, 900-mb temperature of 23°C, and southeast morning winds at Los Angeles International Airport, then the expected ozone concentration would be 182 ppb, as determined by the 1988 decision tree.

Note that slight differences in the predicted variables can produce significant changes in predicted pollutant levels. For example, if the predicted 900-mb temperature were 25°C instead of 23°C, the predicted ozone would have been 230 ppb instead of 182 ppb. Careful evaluation of the accuracy and quality of the predicted weather conditions is needed to ensure an accurate prediction.

Since this decision tree was developed in 1988, ozone concentrations in the South Coast Air Basin have dropped dramatically (SCAQMD, 1997) due in part to changes in fuels and automobile control technologies. To account for changes of vehicle mix and other emissions changes, a decision tree should be updated frequently.

CART development

Complete the following steps to develop a decision tree using CART:

1. Determine the important physical and chemical processes that influence pollutant concentrations in a particular area in order to identify the key variables. Literature reviews, historical case studies, and climatological analysis (as discussed in Section 5.2) can help with this step.

2. Select variables that properly represent the important physical and chemical processes that influence pollutant concentrations in the area. A discussion of variable selection is presented in Section 4.2.

3. Create a multi-year data set of the selected variables. Choose recent years that are representative of the current emission profile. Reserve a subset of the data for independent evaluation of the method.

4. Use CART software to create a decision tree on the multi-year data set.

5. Evaluate the decision tree using an independent data set.

6. When emissions compositions change, the pollutant concentrations associated with the established criteria may change. When this happens, the decision tree should be updated.

LEGEND
N = Number of Days
Mean O_3 = Average Peak Ozone Concentrations (pphm)
O_3S.D. = Standard Deviation of O_3
NZJ7D = West Morning Winds at El Toro, CA
LAX7D = Morning Winds at Los Angeles International Airport
DL = Day Length (Hrs)
850T = 850mb Temperature (°C)
900T = 900mb Temperature (°C)
TOPT = Top of two inversion temperatures (°C)

Max O_3
N = 1,096
Mean O_3 = 130
O_3 S.D. = 82

850T < 17.1
N = 760
Mean O_3 = 90
O_3 S.D. = 46

850T > 17.1
N = 336
Mean O_3 = 230
O_3 S.D. = 58

850T < 9.9
N = 386
Mean O_3 = 63
O_3 S.D. = 23

850T > 9.9
N = 374
Mean O_3 = 130
O_3 S.D. = 47

900T < 24.3
N = 234
Mean O_3 = 210
O_3 S.D. = 49

900T > 24.3
N = 102
Mean O_3 = 280
O_3 S.D. = 51

DAYL < 10.8
N = 187
Mean O_3 = 50
O_3 S.D. = 14

DAYL > 10.8
N = 199
Mean O_3 = 75
O_3 S.D. = 24

DAYL < 10.6
N = 91
Mean O_3 = 70
O_3 S.D. = 24

DAYL > 10.6
N = 283
Mean O_3 = 130
O_3 S.D. = 42

LAX7D = E to SE
N = 87
Mean O_3 = 182
O_3 S.D. = 43

LAX7D = W to NW
N = 147
Mean O_3 = 230
O_3 S.D. = 44

NZJ7D ≠ West
N = 24
Mean O_3 = 230
O_3 S.D. = 46

NZJ7D = West
N = 78
Mean O_3 = 289
O_3 S.D. = 44

850T < 12.7
N = 113
Mean O_3 = 111
O_3 S.D. = 33

850T > 12.7
N = 170
Mean O_3 = 149
O_3 S.D. = 41

TOPT < 19.9
N = 23
Mean O_3 = 183
O_3 S.D. = 40

TOPT > 19.9
N = 124
Mean O_3 = 235
O_3 S.D. = 40

Figure 4-4.   Decision tree for daily maximum ozone concentrations in the South Coast Air Basin in the Los Angeles, California, area (Horie, 1988).

CART operations

The CART method is very easy to use and requires little expertise. Data needed for the decision tree must be acquired and processed through the tree to determine the ozone forecast. Use of this method does not require an understanding of meteorology and air quality processes. However, it is advisable to have someone with meteorological experience evaluate the CART pollutant predictions for reasonableness.

CART strengths

- Requires little expertise to operate on a daily basis; runs quickly.

- Complements other subjective forecasting methods.

- Allows differentiation between days with similar pollutant concentrations if the pollutant concentrations are a result of different processes. Since PM can form through multiple pathways, this advantage of CART can be particularly important to PM forecasting.

CART limitations

- Requires a modest amount of expertise and effort to develop.

- Slight changes in predicted variables may produce large changes in the predicted ozone or $PM_{2.5}$ concentrations.

- This is an objective tool that can only predict pollutant concentrations based on information contained within the observed and forecasted data. Changes in the predicted weather conditions may not be reflected in the predictor variables and may cause uncertainty in the pollutant predictions.

- CART may not predict pollutant concentrations during periods of unusual emissions patterns due to holidays or other events; however, human forecasters can account for these changes and their potential impact on pollutant concentrations.

4.1.5   Regression Equations

Regression is a statistical method for describing the relationship among variables. For air quality forecasting, regression equations are developed to describe the relationship between pollutant concentration (referred to as the predictand, what is being predicted) and other predictor variables (e.g., temperature, wind speed, etc.). Regression equations have been successfully used to forecast pollutant concentrations in many areas of the country (Cassmassi, 1987; Dye et al., 1996; Hubbard and Cobourn, 1997; Ryan, 1994).

How regression equations work

If two variables are correlated, a line or a curve can describe the relationship between those variables using a mathematical equation. With this equation, pollutant concentrations can be predicted from other variables. Multi-linear regression is most commonly used to forecast ozone (Equation 4-1). However, curvilinear regression (Equation 4-2) is useful in ozone

forecasting because it captures the non-linear relationships of ozone and predictor variables. The same approach can be applied to predicting $PM_{2.5}$ concentrations.

$$O_3 = c_1 V_1 + c_2 V_2 \ldots\ldots c_n V_n + \text{constant} \qquad (4\text{-}1)$$

$$O_3 = c_1 V_1 + c_2 V_2^2 + c_3 V_3^3 \ldots\ldots, c_n V_n^n + \text{constant} \qquad (4\text{-}2)$$

where:
O_3 = predictand (could also be PM_{10}, $PM_{2.5}$, or other pollutants)
c = coefficients (weighting factors)
V = predictor variables

An example of a multi-linear regression equation is shown in Equation 4-3. This model was developed for forecasting wintertime $PM_{2.5}$ concentrations for Salt Lake City, Utah (Dye et al., 2002b). This equation predicts the next-day's average 24-hr $PM_{2.5}$ concentration. **Table 4-5** describes the variables used.

$$PM_{2.5} = 53.429 + 3.382*\text{Holiday} - 0.189*\text{Precip} - 0.31*\text{Tmax}$$
$$- 0.541*\text{SurfaceWS} + 1.008*(\text{T@700mb} - \text{Tmin})$$
$$+ 0.838*(\text{Stability}) + 0.183*\text{Td@700mb00Z} - 0.292*\text{WS@850mb00Z} \qquad (4\text{-}3)$$

Table 4-5. Variables used in regression Equation 4-3.

Variable	Description
Holiday	1 for Valentine's Day, Martin Luther King, Jr. Day, Presidents' Day, Veterans' Day, and Super Bowl Sunday. 2 for Thanksgiving weekend and Christmas Eve through New Year's Day. 1 for weekends immediately preceding or following any of the above holidays. 0 for all other days.
Precip	Forecasted precipitation in inches during the 24-hr forecast period.
Tmax	Forecasted daytime maximum temperature (ºF)
SurfaceWS	Average resultant wind speed from 12Z to 00Z (0500 to 1700 MST)
T@700mb	Temperature at 700 mb at 12Z (0500 MST) (ºC)
Tmin	Forecasted or observed minimum temperature (ºC)
Stability	Temperature at 700 mb at 00Z (1700 MST) (ºC) minus the forecasted daytime maximum temperature (ºC) at the surface
Td@700mb00Z	Dew-point temperature at 700 mb at 00Z (1700 MST) (ºC)
WS@850mb00Z	Wind speed at 850 mb at 00Z (1700 MST) (m/s)

To use the equation, a forecaster simply inputs the forecasted values into the equation. Notice that the model uses only weather variables. Thus, a forecaster can use input values from the 24-hr weather forecasts to make the next-day $PM_{2.5}$ forecasts.

Regression equation development

Complete the following steps to develop a regression model for ozone or $PM_{2.5}$ concentrations:

1. Determine the important physical and chemical processes that influence ozone or $PM_{2.5}$ concentrations in a particular area. Literature reviews, historical case studies, and climatological analysis (as described in Section 5.2) can help with this step.

2. Select variables that represent the important physical and chemical processes that influence ozone or $PM_{2.5}$ concentrations in the area. Statistical software can be used to limit the number of variables by identifying the most important ones. A discussion of variable selection is presented in Section 4.2.

3. Create a data set of ozone or $PM_{2.5}$ values and selected predictor variables. Choose a minimum of three recent years that are representative of the current emissions profile. Randomly select about 25% of the data and set them aside for independent evaluation (Step 5).

4. Use statistical software to calculate the coefficients and a constant for the regression equation. The process is straightforward and is likely described in the statistical software manual.

5. Perform an independent evaluation of the regression model using the verification statistics listed in Section 5.6. Evaluate the performance of the regression equations using a data set other than the developmental data set.

6. Other development issues to consider include:

   - Ozone and $PM_{2.5}$ are often log-normally distributed; yet regression is best suited for predicting data that are normally distributed.

   - Use the natural log of ozone or $PM_{2.5}$ concentrations as the predictand to improve performance.

   - Regression tends to predict the mean better than the tails (i.e., high pollutant concentrations) of the distribution. Creating secondary regression equations to predict only the high pollutant concentrations may improve forecast accuracy. These secondary equations can be used when the primary equation reaches a specified concentration level.

   - Be careful not to "over fit" the model by using too many prediction variables. An "over-fit" model will decrease the forecast accuracy. A reasonable number of variables to use in predicting ozone or $PM_{2.5}$ is 5 to 10.

   - One variable can likely represent a whole subset of variables. Unique (i.e., dissimilar) variables should be used to avoid redundancy and co-linearity.

   - Stratifying the data set may improve regression performance. Consider dividing the data set by seasons, weather type, or other meteorological variables. For example, separate equations might be developed for spring, summer, and fall.

Regression equation operations

Compared to the relatively extensive effort required to develop regression equations, operation of the model requires modest expertise. Running the forecast equation only requires data input into a simple computational program or spreadsheet that contains the regression equation(s). Although use of the equation(s) does not require an understanding of meteorology and air quality processes, it is advisable that someone with meteorological experience check the pollutant prediction for physical reasonableness.

Because the predictor variables are forecasted, they have inherent uncertainty, which results in an air quality forecast that has a degree of uncertainty. To help quantify this uncertainty the input values can be altered slightly and the effect this has on the air quality forecasted can be evaluated.

Regression equation strengths

- Regression analysis is well documented and widely used in a variety of disciplines. It has been successfully used in ozone forecasting is many areas of the country (Cassmassi, 1987; Dye et al., 1996; Hubbard and Cobourn, 1997; Ryan, 1994).

- Regression software is widely available and runs on a personal computer. It is generally easy to use.

- Regression is an objective forecasting method that reduces potential biases arising from human subjectivity.

- Regression can properly weight relationships that are difficult to subjectively quantify.

- Regression analysis can be used in combination with other forecasting methods, or it can be used as the primary forecasting method.

Regression equation limitations

- Regression equations require a modest amount of expertise and effort to develop.

- Regression equations tend to predict the mean better than the tails (i.e., the highest pollutant concentrations) of the distribution. They will likely underpredict the high concentrations and overpredict the low concentrations.

4.1.6 Artificial Neural Networks

Artificial neural networks are computer algorithms designed to simulate biological neural networks (e.g., the human brain) in terms of learning and pattern recognition. Artificial neural networks have been under development for many years in a variety of disciplines to derive meaning from complicated data and to make predictions. In recent years, neural networks have been investigated for use in pollution forecasting (Comrie, 1997; Gardner and Dorling, 1998; Ruiz-Suarez et al., 1995). Artificial neural networks can be "trained" to identify patterns and extract trends in imprecise and complicated non-linear data. Because ozone and $PM_{2.5}$ formation are complex non-linear processes, neural networks are well suited for ozone and $PM_{2.5}$

forecasting. Note that neural networks require about 50% more effort to develop than regression equations and provide only a modest improvement in forecast accuracy (Comrie, 1997).

How artificial neural networks work

     Artificial neural networks use a complex combination of weights and functions to convert input variables (such as wind speed and temperature) into an output prediction (such as ozone or PM$_{2.5}$ concentration). **Figure 4-5** shows the artificial neural network components. A forecaster supplies the neural network software with meteorological and air quality input data. The software then weights each datum and sums these values with other weighted datum at each hidden node. The software then modifies the node data by a non-linear equation (transfer function). The modified data are again weighted and summed as they pass to the output node. At the output node, the software modifies the summed data using another transfer function and then outputs an ozone or PM$_{2.5}$ prediction. The neural network software offers several choices for transfer functions.
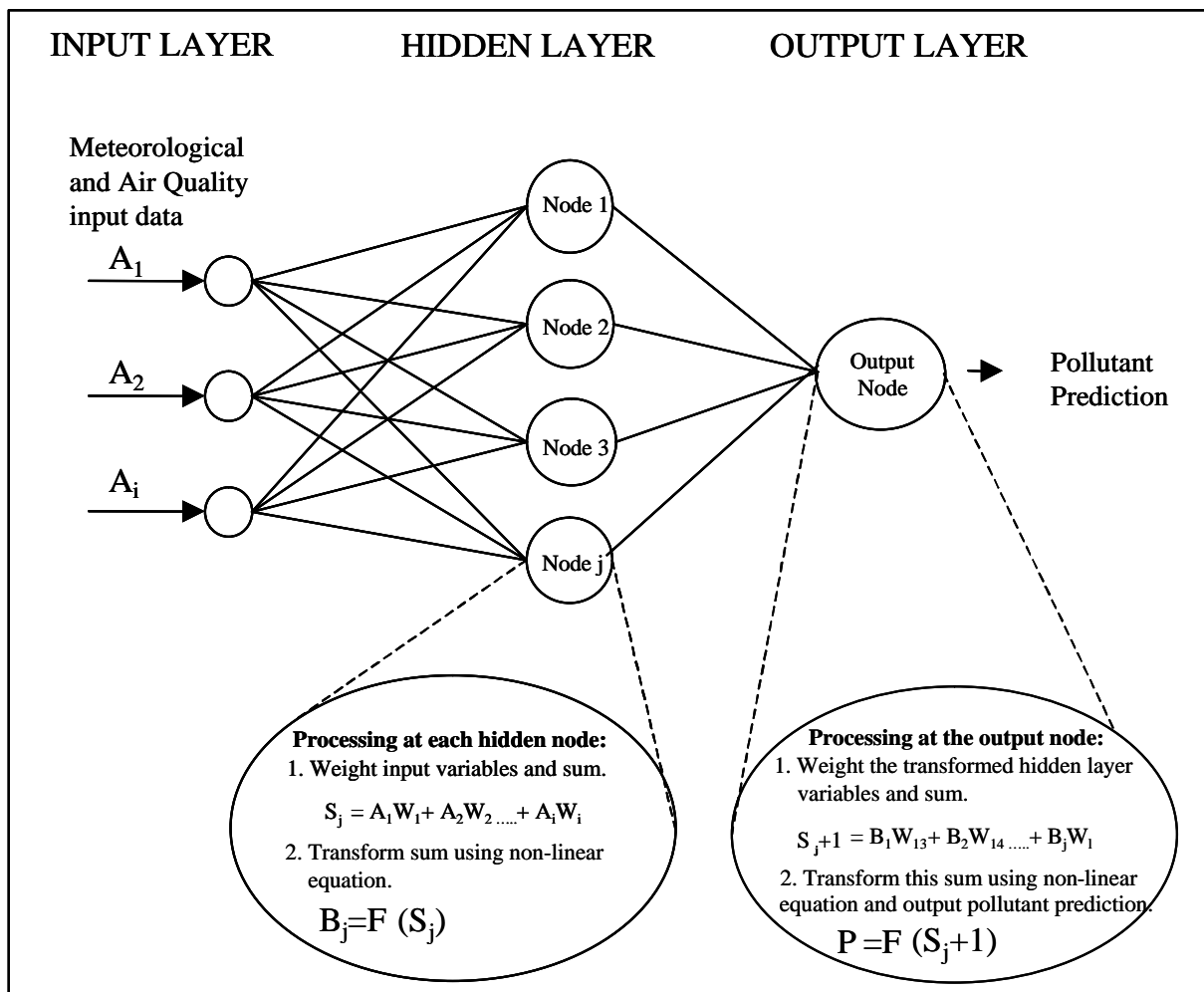
Figure 4-5.   A schematic of an artificial neural network (Comrie, 1997).

Commercial software can be purchased to help forecasters develop and operate a neural network. Before a prediction can be made, the network software must be "trained" and developed. Complete the following steps to train artificial neural networks:

1. Supply the software with historical meteorological and air quality data for the input layer.

2. Supply the software with the historical ozone or $PM_{2.5}$ data.

3. The software establishes nodes within the hidden layer. It then iteratively adjusts the weights until the error between the output data and the actual data (observed) is minimized.

4. Neural networks typically use a backpropagation algorithm to adjust the weights to minimize the error. The error information propagates back through the network. The software first adjusts the weights between the output layer and the hidden layer and then adjusts the weights between the hidden layer and the input layer. With each iteration, the software adjusts the weights to produce the least amount of error in the output data. This process "trains" the network.

Once the network has been trained (i.e., developed) it can be used to operationally forecast ozone or $PM_{2.5}$ concentrations.

Three data sets are required to train a neural network to achieve good generalization on new data: a developmental set, a validation set, and a test set. The developmental set is used to develop the neural network. The validation set is used to determine when the network's general performance is maximized. And the test data set is used to evaluate the trained network.

It is important not to over train the neural network on the developmental data set because an over trained network will predict ozone or $PM_{2.5}$ concentrations based on random noise associated with the developmental data set (Gardner and Dorling, 1998). When presented with a new data set the network will likely give incorrect output since the new data's random noise will be different than the random noise of the developmental data set.

Developing artificial neural networks

Complete the following steps to develop neural networks to forecast ozone or $PM_{2.5}$.

1. Complete historical data analysis and/or literature reviews to establish the air quality and meteorological phenomena that influence ozone or $PM_{2.5}$ concentrations in a particular area. A detailed discussion of this process is contained in Section 5.2.

2. Select parameters that accurately represent these phenomena. Be sure to select parameters that are readily available on a forecast basis. A detailed discussion of variable selection can be found in Section 4.2.

3. Confirm the importance of each meteorological and air quality parameter using forward stepwise regression, for example. See Section 4.2 for more details.

4. Create three data sets: (1) a data set to train the network, (2) a data set to validate the network's general performance without over fitting the data, and (3) a data set to evaluate the trained network. The developmental data set should contain at least three years of data. The validation and evaluation data sets should each contain about one year of data.

However, with today's changing emissions, a five-year-old data set may have significantly different characteristics than a current data set.

5. Train the data using neural network software. Be sure not to over train the network as it must be general enough to work well on new data sets. As the network is trained, use the validation data set to determine when the network's general performance is maximized. See Gardner and Dorling (1998) for details.

6. Test the generally trained network on a test data set to evaluate the performance. If the results are satisfactory, the network is ready to use for forecasting. A discussion of forecast accuracy and performance is presented in Section 5.6.

Artificial neural networks operations

Compared to the development of the network, the operation of the network is straightforward and requires little expertise. Data only needs to be acquired and input into the input layer of the neural network. Although use of the network does not require an understanding of meteorology and air quality processes, it is advisable that someone with meteorological experience be involved in the development of the method and evaluate the pollutant prediction for reasonableness.

Strengths of artificial neural networks

- Ozone and $PM_{2.5}$ formation are non-linear processes. This method can weight relationships that are difficult to subjectively quantify and neural networks allow for non-linear relationships between variables.

- Neural networks should predict extreme values more effectively than regression equations, provided that the network developmental set contains such outliers.

- Once a neural network is developed, a forecaster does not need specific expertise to operate it.

- Neural networks can be used in combination with other forecasting methods, or it can be used as the primary forecasting method.

Limitations of artificial neural networks

- Neural networks are complex and not commonly understood; thus, the method can be inappropriately applied and difficult to develop.

- Neural networks do not extrapolate data well. Thus, extreme pollutant concentrations not included in the developmental data set will not be taken into consideration in the formulation of the neural network prediction.

4.1.7 Deterministic Air Quality Modeling

Deterministic air quality modeling attempts to mathematically represent the important processes that affect ambient air quality. Air quality modeling actually requires a system of models that work together to simulate the emission, transport, diffusion, transformation, and removal of air pollution. These models include:

- Meteorological models – These models forecast meteorological conditions that determine transport and mixing; and influence chemistry, emissions, and deposition.

- Emissions models – These models simulate the temporal, spatial, and chemical distribution of emissions of the pollutant in question, and/or its precursors, from both anthropogenic and natural sources.

- Air quality models – These models use the forecasts from meteorological and emissions models to simulate the transport, diffusion, transformation by chemical reaction, and removal of air pollution. A more detailed discussion of air quality models can be found in Seinfeld and Pandis (1998).

Historically, deterministic air quality models have been used in air quality planning to estimate the impact of population growth and emission controls on future air quality. For air quality planning, these models have been and continue to be used in case study analyses to understand air pollution processes and to estimate the effects of emissions changes on pollutant concentrations during episodic conditions. These modeling analyses may take years to set up, evaluate, and complete. In recent years, high performance computing at low cost has become available and air quality models have been used with forecasts from prognostic meteorological models to produce daily air quality forecasts. While it is possible to forecast pollutant concentrations with simple one-dimensional air quality models, three-dimensional (3-D) air quality models that simulate the complex interaction of physical and chemical processes are more suitable for forecasting air quality because they handle multiple processes and allow for future improvements to the model components.

How 3-D air quality modeling systems work

To predict pollutant concentrations with a 3-D air quality model, meteorological factors, and emissions must be predicted first. An air quality modeling system links the meteorological, emissions, and air quality models together to make air pollution forecasts. **Figure 4-6** shows how these components are related in an air quality model system.

Meteorological models

Prognostic meteorological models solve sets of equations that represent fundamental atmospheric behavior. During the past 10 years, prognostic mesoscale modeling has become an increasingly common method of developing inputs for air quality modeling. Many newer air quality models use mesoscale meteorological models as their preferred meteorological driver and are well suited for air quality forecasting. Currently, the two most widely used meteorological models for air quality applications are:

- The Penn State/NCAR Mesoscale Model version 5 - MM5 (Grell et al., 1994)
- The Regional Atmospheric Modeling System - RAMS (Pielke et al., 1992)

The most common approach to prognostic meteorological modeling for air quality applications is to use National Center for Environmental Prediction (NCEP) operational forecast model (Eta, AVN, NGM, etc.) analysis fields to provide initial and boundary conditions for MM5. When used for air quality forecasting, the NCEP forecast meteorological fields can be
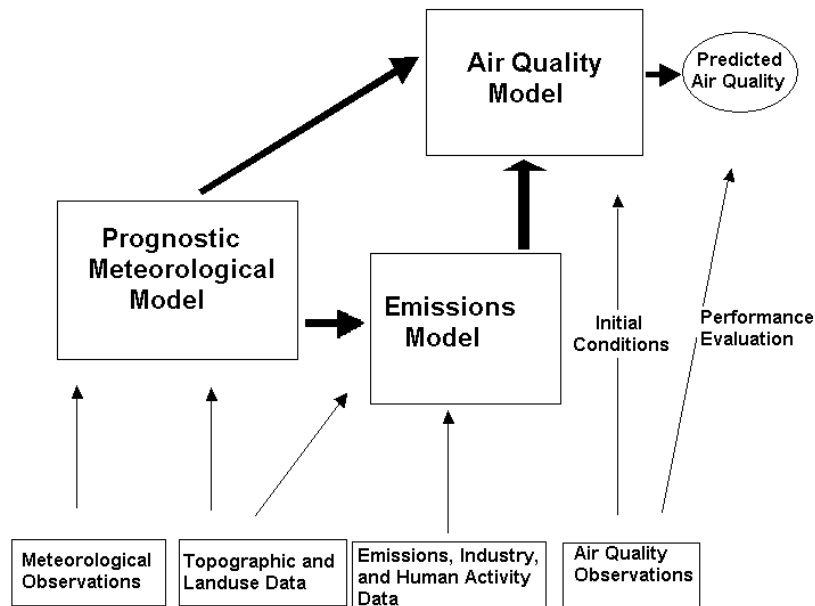
Figure 4-6.   Schematic showing the component models of an air quality modeling
              system.

used for MM5's boundary conditions. In the forecast mode, the MM5 modeling systems can
make weather predictions for use in an air quality model.

Emissions models

Emissions modeling is the process of estimating emissions with the spatial, temporal, and
chemical resolution needed for air quality modeling. The emission inventory includes data for
mobile sources, stationary point sources, area sources, and natural sources. Mobile, biogenic,
and some point/area source emissions can vary substantially with temperature. Mobile source
and some industrial/commercial emission sources also exhibit significant variations by day of the
week. Currently, there are three emissions modeling systems commonly used to provide
Eulerian air quality models with emissions input data.

- Emission Processing System (EPS 2.0) (U.S. Environmental Protection Agency, 1992)
- Emissions Modeling System (EMS-95 – EMS-2002) (Bruckman, 1993)
- Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system (Coats, 1996)

When performing emissions modeling to support air quality forecasting, environmental
(i.e., temperature and solar radiation) and day-of-week effects need to be taken into account.
These effects may be included in pre-computed, model-ready emissions inputs for various cases,
or may be calculated at run-time based on the predictions of the prognostic meteorological
model. Of the currently available emissions models, only SMOKE has been used to calculate
emissions at run-time in a real-time forecasting system (McHenry et al., 1999).

Air quality models

Three-dimensional air quality models are classified as being either Lagrangian or Eulerian depending on the method used to simulate the time-varying distribution of pollution concentrations. Lagrangian (trajectory) models follow individual air parcels over time using the meteorological data to transport and diffuse the pollutants; they may also include chemical transformations. This approach is computationally efficient when treating a limited number of emission sources. However, it is difficult to properly characterize the interaction of a large number of individual sources when nonlinear chemistry is involved and these models have limited usefulness in forecasting secondary pollutants. Eulerian models use a grid of cells (vertical and horizontal) where the chemical transformation equations are solved in each cell and pollutants are exchanged between cells. These models can produce three-dimensional concentration fields for several pollutants but require significant computational power. Typically, the computational requirements are reduced through the use of nested grids, with a coarse grid used over rural areas and a finer grid used over urban areas where concentration gradients tend to be more pronounced.

The Hybrid Single-Particle Lagrangian Integrated Trajectories with a generalized non-linear Chemistry Module (HY-SPLIT CheM) model is an example of a Lagrangian model used to forecast air quality on a regional scale (Stein et al., 2000). However, because of the limited ability of Lagrangian models to handle the large number of emission sources needed for urban-scale forecasting, Eulerian models are often used to forecast air quality.

The typical Eulerian air quality model is bounded on the bottom by the ground, on the top at some specified height, and at some distance on all four sides, depending on the size of the meteorological modeling domain and the area of interest. The volume of the modeling domain is divided into grid cells. For regional air quality models, the grid cells are typically on the order of tens of kilometers in length and width (4 to 36 km are common) with 5 to 50 vertical layers. The grid cell size is chosen to maximize the resolution for a given computational budget. Smaller grid cells will result in higher resolution and generally greater model accuracy, but at a higher computational cost. Within each grid cell a series of physical and chemical processes are simulated as shown in **Figure 4-7.**

Currently available Eulerian air quality models include.

- The California Grid Model: CALGRID (Yamartino et al., 1992).
- The Comprehensive Air Quality Model with Extensions: CAMx (Environ, 1998).
- The Community Multiscale Air Quality model: CMAQ (U.S. Environmental Protection Agency, 1998).
- The Multiscale Air Quality Simulation Platform: MAQSIP (Odman and Ingram, 1996).
- The SARMAP Air Quality Model: SAQM (Chang et al., 1996).
- The Urban Airshed Model with Aerosols: UAM-AERO (Lurmann, 2000; Lurmann et al., 1997).
- The Urban Airshed Model with Carbon Bond IV Chemistry: UAM-IV (Morris et al., 1990).
- The Variable Grid Urban Airshed Model: UAM-V (Systems Applications International, 1999).
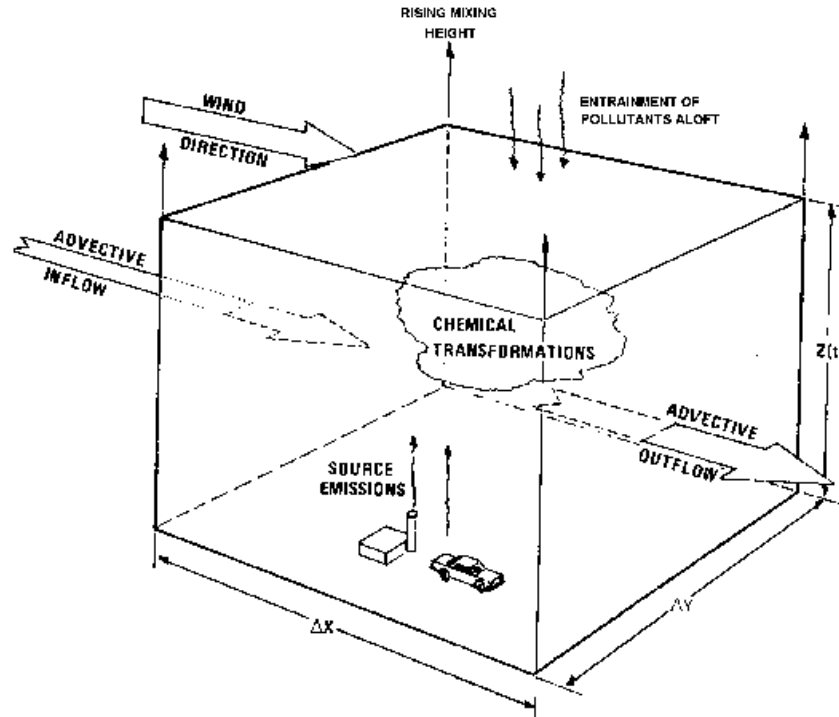
Figure 4-7.   Schematic illustration of the processes in an Eulerian photochemical model cell (Schere and Demerjian, 1984).

Setting up a 3-D air quality model

Substantial staff and computer resources are needed to establish a scientifically sound and automated air quality forecast system based on a 3-D air quality model. Even when using existing meteorological, emissions, and air quality models, the effort to integrate and refine the entire system enough to produce reliable forecasts may be large.

To implement a 3-D air quality modeling system for predicting air quality in a region, the following steps are suggested.

1. Design and plan the system

   - Decide on which pollutants to forecast.

   - Define modeling domains (meteorology and air quality) considering geography and emission sources. The definition should include horizontal and vertical resolution of the models.

   - Select component models considering forecast pollutants, domains, component model compatibility, availability of interface programs, and available resources.

   - Determine hardware and software requirements for the system.

   - Identify sources of needed meteorological, emissions, and air quality data, and methods for acquiring these data.

4-26

- Prepare a detailed plan for acquiring and integrating data acquisition, modeling, and analysis software.

- Identify what the final products will be.

- Plan for continuous real-time evaluation of the modeling system.

- Prepare a reasonable implementation schedule that plans for problems.

2. Identify and allocate the resources needed

- Staff for system implementation and operations.

- Computing and storage consistent with the selection of domains and models.

- Communications for data transfer into and out of the modeling system. Sufficient network bandwidth must be provided for downloading external data from the Internet and transferring data within the local network to ensure other operations are not affected.

3. Acquire required geophysical data

- Topographical data

- Land use data

4. Implement the data acquisition and processing tools, component models (emissions, meteorological, and air quality), and analysis programs.

- Implement each program individually.

- Use standard test cases to verify correct implementation.

5. Develop the emission inventory

- Acquire needed emission inventory related data.

- Review the emissions data for accuracy. Errors in the emissions data can result in large errors in the air quality model output.

- Be sure that the emissions data reflect the most recent emissions data available unless there are known problems that indicate earlier data is more suitable.

- Project emissions to the current year using growth and control factors.

- Update the base emission inventory annually unless annual projections are to be included in the run-time processing of emissions. If projections are to be included in the run-time processing, growth and control factors should be reviewed and updated annually.

6. Test the data acquisition and processing tools, component models, and analysis programs

- Test the operation of all data acquisition programs, preprocessor programs, component models, and analysis programs as a system using real data sources.

- Review the prognostic meteorological forecast data for accuracy over several weeks under various weather patterns. Errors in the meteorological input field can result in large errors in the air quality output. If the meteorological model shows persistent and significant errors (particularly winds) it may not be reasonable to continue system implementation without first resolving the source of errors.

- Run the combined meteorological/emissions/air quality modeling system in a prognostic mode using a variety of meteorological and air quality conditions. Evaluate the performance of the modeling system by comparing it with observations. Refine the model application procedures (i.e., the methods of selecting boundary conditions or initial concentration fields, the number of spin-up days, the grid boundaries, etc.) to improve performance. It may be necessary to refine the options or inputs for any one of the component models that make up the system.

7. Integrate data acquisition and processing tools, component models, and analysis programs into an operational system

- After achieving satisfactory results in the testing phase, implement automated processes for data acquisition, the daily data exchange from the prognostic meteorological model and the emissions model to the 3-D air quality model and analysis programs, and forecast product production.

- Implement automated processes by using scripting and scheduling tools. Since the entire modeling process will take a significant amount of time to complete, there should be some method of tracking the progress of the modeling.

- Verify that the forecast products reflect the actual model predictions.

8. Test, evaluate, and improve the integrated system

- Run the model in real-time test mode for an extended period. Compare output to observed data and note when there are model failures.

- After obtaining satisfactory results on a consistent basis, use the modeling system to forecast pollutant concentrations.

- Document the modeling system.

- Continuously evaluate the system's performance by comparing observations and predictions.

- Implement improvements as needed based on performance evaluations and new information.

Three-dimensional air quality modeling system operations

Operation of the 3-D air quality modeling system should be completely automated. Ideally, the forecaster should only need to review the model forecast for reasonableness. However, in reality there will be times when parts of the system will fail. Therefore, operational procedures should be established for monitoring of the models as they execute and for recovery from failures.

The forecaster should continuously monitor model performance statistics and graphics to identify persistent errors or biases. Problems with the model predictions that are significant should be discussed with the system developer.

Strengths of 3-D air quality models

- Three-dimensional air quality forecast models are phenomenological based, simulating the physical and chemical processes that result in the formation and destruction of air pollutants.

- They can forecast for a large geographic area.

- They can predict air pollution in areas where there are no air quality measurements.

- The model forecasts can be presented as maps of air quality to show how predicted air quality varies over a region hour by hour. The maps can be animated to show where air pollution is expected to form and how it will evolve over the course of a day.

- Three-dimensional air quality forecast models can be used to further understand the processes that control air pollution in a specific area. For example, they can be used to assess the importance of local emission sources or long-range transport.

Limitations of 3-D air quality models

- Inaccuracies in the prognostic model forecasts of wind speeds, wind directions, extent of vertical mixing, and solar insulation may limit 3-D air quality model performance. Small discrepancies in winds over 24-hr to 48-hr periods can produce significant shifts in the spatial pattern of predicted ozone concentrations over a region.

- Emission inventories used in current models are often out of date and based on uncertain emission factors and activity levels. Three-dimensional air quality forecast model accuracy depends on accurate emission inventory modeling.

- Site-by-site ozone concentrations predicted by 3-D air quality forecast models may not be accurate due to small-scale weather and emission features that are not captured in the model.

4.1.8 The Phenomenological/Intuition Method

Phenomenological/intuition forecasting involves analyzing and conceptually processing air quality and meteorological information to formulate an air quality prediction. Phenomenological/intuition forecasting can be used alone or with other forecasting methods such as regression or criteria. Although intuition is commonly defined as "the perception of truth or fact, independent of any reasoning," for air quality forecasting intuition is the perception of truth or fact (the prediction) derived from reason (the conceptual processing of meteorological and air quality data).

This method is heavily based on the experience provided by a meteorologist or air quality scientist who understands the phenomena that influence ozone or $PM_{2.5}$. This method balances

some of the limitations of objective prediction methods (i.e., criteria, regression, CART, and neural networks).

How the Phenomenological/Intuition method works

This method depends on an individual's capabilities and/or experience in three major areas:

1. Understanding the processes that influence ozone or $PM_{2.5}$. The basic component to phenomenological/intuition forecasting is developing a robust and accurate conceptual understanding of the important phenomena that control pollutant concentrations. This conceptual understanding should include information on synoptic, regional, and local meteorological conditions, plus air quality characteristics in the forecast area.

2. Synthesizing information. Vast amounts of data are needed to forecast air pollution. Forecasters will analyze both observed and forecasted weather charts, satellite information, air quality observations, and pollutant predictions from other methods. Each piece of information or prediction from other methods must be evaluated and given a relative weight.

3. Developing a consensus. Some information or data will likely be contradictory and should be dismissed. For example, available weather data may be conducive to high ozone concentrations (light winds, clear skies, and high temperatures) and forecasting criteria may suggest high ozone concentrations, yet the regression equation may predict only modest ozone concentrations. The forecaster must take into account the historical performance of each method/data source, accept some, reject others, and issue the ozone forecast based on a general agreement of the forecasts.

Phenomenological/Intuition method development

The fundamental step in developing a Phenomenological/Intuition forecasting method is acquiring a conceptual understanding of the important physical and chemical processes that influence ozone or $PM_{2.5}$ concentrations in an area. Literature reviews, historical case studies, and climatological analysis (as discussed in Section 5.2) can help with this. Although much knowledge can be gleaned from these sources, the greatest benefit to the method is the development of intuition, which only comes from forecasting experience.

Phenomenological/Intuition method operations

Compared to other forecasting methods, phenomenological/intuition forecasting requires a high level of expertise. The forecaster needs to have a strong understanding of the processes that influence ozone or $PM_{2.5}$ concentrations and needs to apply this understanding on a daily basis. Typically, the forecaster will evaluate meteorological forecast models and use pattern recognition that equates the meteorological fields to ozone or $PM_{2.5}$ concentrations. For example, the forecaster may observe a high-pressure ridge building into the forecast area and equate this with high $PM_{2.5}$ concentrations. The forecaster will repeat this process for several other meteorological and air quality data fields, weigh the combined influence of these fields, and output a forecast. For example, some predictor variables may indicate high pollutant

concentrations, while others indicate moderate and low pollutant concentrations. By processing all of this information in the conceptual model, the forecaster develops a pollutant prediction.

Strengths of the Phenomenological/Intuition method

- The Phenomenological/Intuition method allows for easy integration of new data sources. For example, if a new wind monitor is installed, the forecaster can quickly make use of this additional data. Whereas, other objective methods, such as regression, may require re-creation of forecasting algorithms.

- The Phenomenological/Intuition method allows for the integration and selective processing of large amounts of data in a relatively short period of time.

- This method can be immediately adjusted as new truths are learned about the processes that influence ozone or $PM_{2.5}$.

- The effect of unusual emissions patterns associated with holidays and other events can easily be taken into account.

- Extreme or rare events may be more accurately forecasted. Generally, objective methods such as regression or neural networks do not capture extreme or rare events.

- The Phenomenological/Intuition method is a good complement to other more objective forecasting methods because it tempers their results with common sense and experience.

Limitations of the Phenomenological/Intuition method

- The Phenomenological/Intuition method requires a high level of expertise. The forecaster needs to have a strong understanding of the processes that influence ozone or $PM_{2.5}$ concentration and needs to apply this understanding in both the developmental and operational processes of this method.

- Since the Phenomenological/Intuition method is subjective, forecaster bias is likely to occur. Using an objective method as a complement to this method can alleviate these biases.

4.2   SELECTING PREDICTOR VARIABLES

Many of the methods discussed in Section 4.1 use predictor variables to forecast ozone or $PM_{2.5}$. This section provides guidelines for selecting the candidate predictor variables to use in ozone or PM forecasting efforts. **Tables 4-6 and 4-7** list common predictor variables that influence ozone and $PM_{2.5}$ concentrations. Consider the following issues when selecting predictor variables.

- Understand the phenomena. Before selecting particular variables it is important that forecasters understand the phenomena that affect ozone or $PM_{2.5}$ concentrations in their region. This understanding can be gained through review of past air quality studies in the area, conducting a historical analysis of meteorology and ozone or $PM_{2.5}$, and/or doing a literature review as described in Section 5.2.

Table 4-6. Common predictor variables used to forecast ozone.

Variable	Usefulness	Condition for High Ozone[a]
Maximum temperature	Highly correlated with ozone and ozone formation	High
Morning wind speed	Associated with dispersion and dilution of ozone precursor pollutants	Low
Afternoon wind speed	Associated with transport of ozone	-
Cloud cover	Controls solar radiation, which influences photochemistry	Few
Relative humidity	Surrogate for cloud cover	Low
500-mb height	Indicator of the synoptic-scale weather pattern	High
850-mb temperature	Surrogate for vertical mixing	High
Pressure gradients	Causes winds/ventilation	Low
Length of day	Amount of solar radiation	Longer
Day of week	Emissions differences	-
Morning NO_x concentration	Ozone precursor levels	High
Previous day's peak ozone concentration	Persistence, carry-over	High
Aloft wind speed and direction	Transport from upwind region	-

[a] Relative condition is location- and season-dependent.

Table 4-7. Common predictor variables used to forecast $PM_{2.5}$.

Variable	Usefulness	Condition for High $PM_{2.5}$[a]
500-mb height	Indicator of the synoptic-scale weather pattern	High
Surface wind speed	Associated with dispersion and dilution of pollutants	Low
Surface wind direction	Associated with transport of pollutants	-
Pressure gradient	Causes wind/ventilation	Low
Previous day's peak $PM_{2.5}$ concentration	Persistence, carry-over	High
850-mb temperature	Surrogate for vertical mixing	High
Precipitation	Associated with clean-out	None or light
Relative Humidity	Affects secondary reactions	High
Holiday	Additional emissions	-
Day of week	Emissions differences	-

[a] Relative condition is location- and season-dependent.

- Capture the important phenomena. The variables selected should capture the important phenomena that affect pollutant concentrations in the region. For example, research may show that high background $PM_{2.5}$ concentrations are needed to produce high $PM_{2.5}$ concentrations in the forecast area. Thus, using yesterday's $PM_{2.5}$ concentration as a surrogate for background $PM_{2.5}$ concentration may improve forecast accuracy.

- Select observed and forecasted variables. Predictor variables can consist of observed variables (e.g., yesterday's ozone or $PM_{2.5}$ concentration) and forecasted variables (e.g., tomorrow's maximum temperature). Using forecasted predictor variables is critical since tomorrow's pollutant concentrations are more strongly related to tomorrow's weather conditions than to today's or yesterday's pollutant concentrations.

- Ensure data availability and reliability. Make sure that data is easily obtainable from reliable source(s). Ensure that data will be available by a specified time every day, so that forecasts can be issued in a timely manner. For example, if a forecast needs to be issued for tomorrow's ozone or $PM_{2.5}$ concentration by 1100 LST, all predictor variables and data must be available before 1100 LST.

Using the above guidelines and understanding the mechanisms and phenomena that influence ozone or $PM_{2.5}$ concentrations, a forecaster might select as many as 50 to 100 variables for consideration. These variables are the starting point for the statistical analysis, but will need to be reduced to a smaller number of the most useful variables.

Statistical analysis techniques can be used to identify the most significant variables. Following is a list of the types of statistical analyses that can be performed. For further details on statistical methods, see Wilks (1995).

- Cluster analysis is a method used to partition data into similar and dissimilar subsets. Many of the variables may be somewhat similar (e.g., maximum surface temperature and 900-mb temperature), and cluster analysis can be used to identify these similarities. One variable can likely represent a whole set of similar variables. Unique (i.e., dissimilar) variables should be used to avoid redundancy. Statistical software can be purchased and used for performing cluster analysis.

- Correlation analysis is used to evaluate the relationship between the predictand (i.e., pollutant levels) and various predictor variables. Correlations range from +1 (high-positive relationship) to 0 (no relationship) to -1 (high-negative relationship). Variables used for this type of analysis should have a high-positive or high-negative correlation. A high-positive correlation indicates that increases in the variable are associated with increases in the next day's pollutant concentration. Some of the variables may be both similar and highly correlated, for example, maximum surface and 850-mb temperatures. In this case, one or two variables would suffice. Spreadsheet programs (Excel, Lotus, etc.) or statistical software can be used to calculate correlation.

- Step-wise regression is an automatic procedure that allows the statistical software (SAS, Statgraphics, Systat, etc.) to select the most important variables and generate the best regression equation. When using this approach, it is important to question and evaluate the results. A common problem with this technique is that the resulting

regression equations may contain too many variables that cause them to over fit the data, producing inaccurate predictions.

- Human selection is another means of selecting the most important predictor variables. A forecaster can visually evaluate the relationship among variables using scatter plot matrices, for example.

This selection process results in a series of key variables that can be used with the forecast methods described in Section 4.1 to predict pollutant concentrations in a forecast region.

5. STEPS FOR DEVELOPING AN AIR QUALITY FORECASTING PROGRAM

This section describes the major steps for setting up and operating an air quality forecasting program. For each step, major issues are identified and suggestions are provided for resolving them. Understanding the users' needs (Section 5.1) and understanding the processes that control air quality (Section 5.2) are the first steps to developing a forecasting program. Information to help forecasters choose one or more forecasting methods is presented in Section 5.3. Section 5.4 identifies the types and sources of air quality and meteorological data. Section 5.5 explains the importance of having a forecasting protocol. Section 5.6 explains how to evaluate the quality of a pollutant forecast.

5.1   UNDERSTANDING FORECAST USERS' NEEDS

The success of an air quality forecasting program depends partly on accurate predictions, but also on meeting the needs and objectives of forecast users. As discussed in Section 3, air quality forecasts are used for three major purposes: public health notification, episodic control programs (Action Days), and scheduling specialized monitoring programs. The questions provided below are designed to help identify forecast users' needs.

- *Who will use the forecast?* Forecasters who understand their audience will have more insight into potential ways to improve the forecast.

- *For how many months are forecasts needed?* Understanding how long the ozone or $PM_{2.5}$ season lasts helps forecasters plan the resources (labor and data) needed for air quality forecasting. The analysis techniques described in Section 5.2 can help determine the length of the season for each pollutant.

- *What periods should a forecast cover?* Typically, air quality forecasts are made for the current- and next-day periods; however, they can be extended to include two- to five-day predictions. Note that longer-range predictions will likely be less accurate.

- *Are three-day forecasts needed for weekend/holiday periods?* During weekends and holidays staff may be unavailable to produce daily forecasts. In this case, two- and/or three-day forecasts may be needed to cover these periods. A plan should be in place to handle the situation if conditions change appreciably from initial forecasts.

- *When should forecasts be issued to ensure meeting public outreach deadlines?* Preparatory work is needed to communicate forecast information to the public, particularly during air quality episodes. Issuing forecasts as early in the day as possible helps ensure that they can be effectively communicated to the public.

- *Should forecasts be re-issued? If so, under what conditions?* Sometimes weather conditions change rapidly after a forecast has been issued. Re-issuing an ozone or $PM_{2.5}$ forecast may improve the forecast accuracy, but could lead to public confusion and questioning of the forecast and possibly jeopardize the credibility of the outreach program.

- *What are the accuracy requirements? For example, is an error of ±20 ppb acceptable? What about missing a forecast by two AQI categories?* It is important to understand the error tolerance of forecast users. Exceeding this threshold can lead to reduced credibility.

- *Are air quality forecasts issued for maximum regional concentrations or for site-specific maximums?* Forecasting difficulty and uncertainty is greater for smaller forecasting regions. It is more difficult to make a site-specific forecast than a regional one. The user's forecast needs and tolerance for accuracy must be balanced with the resources available to produce the forecast. Most air quality agencies issue regional ozone and $PM_{2.5}$ forecasts.

- *Are $PM_{2.5}$ forecasts issued for 24-hr average $PM_{2.5}$ concentrations or for sub-daily averages?* Since the AQI for $PM_{2.5}$ and PM_{10} is based on a 24-hr average, forecasts are generally made for the AQI standard. However, when conditions are rapidly changing, forecasting $PM_{2.5}$ on a shorter averaging interval may be more desirable for the end-users of the forecast.

- *Should forecasts be made for specific concentrations or concentration ranges (e.g., AQI categories)?* Generally, forecasts used for public health notification are provided in concentration ranges or AQI categories, which allow for easier forecast interpretation. However, agency personnel who make decisions about specialized sampling (e.g., collecting VOC measurements) may benefit from specific concentration forecasts.

- *Should a forecaster hedge high or low?* Ideally all forecasts would be accurate. In reality, all forecasts contain uncertainty. Hedging a forecast either high or low allows the forecaster to account for conditions that could influence pollutant concentrations such as fronts or winds. It is important to identify whether forecast users want to minimize "false negatives" (forecast for low ozone, but actually observe high ozone) or "false positives" (forecast for high ozone that doesn't occur).

- *What types of interactions with forecast users are needed?* In addition to receiving forecasts, forecast users may benefit from a brief discussion with the forecaster. This discussion allows the forecaster to pass on verbally any details or uncertainty about the forecast and any possible scenarios that might result in changes to the forecasted values.

- *Are written forecast discussions of predicted weather and air quality conditions required?* These discussions provide additional information to help users interpret the predicted pollutant values. Written explanations can convey fine points and uncertainties about the forecast.

- *How should forecasts be disseminated?* Many methods exist for disseminating forecasts (fax, phone, e-mail, Internet, pager, etc.). Identify appropriate primary and secondary (backup) methods to disseminate forecasts to users.

- *How should missed forecasts be handled?* Missed forecasts, particularly large misses, should be examined and discussed with forecast users. By identifying and explaining the causes of error, forecasters can learn from past mistakes, and users can better understand the forecast process and its limitations.

5.2   UNDERSTANDING THE PROCESSES THAT CONTROL AIR QUALITY

The next step in developing an air quality forecasting program is understanding how and why pollution forms in an area. Section 2 provides a general discussion of the chemical processes and weather phenomena that influence ozone and $PM_{2.5}$ concentrations. This section presents methods and examples to help with identification and understanding of the processes and phenomena that influence ozone and $PM_{2.5}$. Understanding these processes and phenomena will improve a forecaster's capabilities. Common methods for developing this understanding include reviewing literature from past research and conducting data analyses.

5.2.1 Literature Reviews

The most efficient and generally the easiest way to start understanding pollution in a particular area is by reviewing existing literature on the topic. Ozone pollution has been studied for three decades, and scientists have produced a plethora of papers and reports for most areas of the country. Although $PM_{2.5}$ pollution has also been studied for many years, few areas are forecasting for $PM_{2.5}$.

Articles published in the *Journal of Applied Meteorology* and *Atmospheric Environment* often contain very pertinent information. Other literature sources, such as reports from local/regional ozone studies that may be available through government agencies, should also be considered. Broadening the literature review to include nearby regions may provide important information that is directly applicable to air quality processes in the forecast area.

Some good general reference sources include:

- National Research Council (1991) – Explains how tropospheric ozone forms and provides details about ozone chemistry.

- Seinfeld and Pandis (1998) – Provides a basic overview of atmospheric chemistry (ozone and $PM_{2.5}$) and describes how meteorology affects atmospheric chemistry.

- Wallace and Hobbs (1977) – Provides general meteorological information about weather maps, atmospheric stability, and atmospheric motions from the synoptic-scale to the local-scale.

- Wilks (1995) – Describes statistical techniques and how these can be applied to meteorological data. Many of the techniques discussed can also be applied to air quality.

5.2.2 Data Analyses

Once a literature search is completed, data analysis can help forecasters learn more about the processes that control ozone or $PM_{2.5}$ concentrations in their area. Data analysis is the process of exploring data to answer questions. It can be performed in three steps: developing questions (i.e., hypotheses), acquiring data, and using analytical methods to answer the questions. Depending on the air quality agency's resources, data analysis efforts can range from simple statistical analyses to large field studies with subsequent research and computer modeling. Following is a discussion of some basic analysis procedures to help explain the processes that control ozone or $PM_{2.5}$ concentrations in a forecast area.

The first step in performing data analysis is having clearly defined questions; this will increase the effectiveness of the research. Types of questions to ask include:

Temporal distribution of ozone or PM$_{2.5}$

- *During what weeks/months do ozone or PM$_{2.5}$ episodes occur?*

- *At what time of day do the highest ozone or PM$_{2.5}$ concentrations occur? For how many hours do high ozone or PM$_{2.5}$ concentrations typically last?*

- *For how many consecutive days do high ozone or PM$_{2.5}$ episodes typically last?*

- *Do maximum ozone or PM$_{2.5}$ concentrations vary by day of week?*

- *During what time of year do PM$_{2.5}$ episodes occur?*

- *Do ozone and PM$_{2.5}$ episodes coincide with one another? When do PM$_{2.5}$-only episodes occur?*

Spatial distribution of ozone or PM$_{2.5}$

- *Where do the highest ozone or PM$_{2.5}$ concentrations occur? Do the highest concentrations occur at different times for different sites?*

- *Have emissions patterns changed in recent years?*

Monitoring issues

- *Has the monitoring network changed recently?*

- *What are the different PM$_{2.5}$ monitoring methods and how do they compare to one another?*

- *Are adjustments made to the continuous PM$_{2.5}$ data to make them better match the FRM standards? If so, how accurate are these adjustments?*

Meteorological and air quality processes

- *What types of synoptic weather patterns are associated with high ozone or PM$_{2.5}$ concentrations?*

- *Does local carryover contribute to ozone or PM$_{2.5}$ concentrations?*

- *Do surface or aloft transport pollutants from other areas contribute to ozone or PM$_{2.5}$ in the forecast area?*

- *How do local flow patterns influence ozone or PM$_{2.5}$ concentrations?*

- *How does the aloft temperature structure influence peak ozone or PM$_{2.5}$ concentrations?*

- *What types of weather patterns are associated with cloud cover?*

- *Are there rare emission events such as forest fires, agricultural burning and tilling, windblown dust, etc. that cause PM$_{2.5}$ episodes?*

The remainder of this section discusses these questions in more detail and explains why each question is important. Also included is an example analysis technique. These examples are intended as a starting place for an understanding of the important processes that produce ozone or $PM_{2.5}$ in a forecast area.

Temporal distribution of ozone or $PM_{2.5}$

    Question: *During what weeks/months do ozone or $PM_{2.5}$ episodes occur?*
          Why: Helps define the forecasting period.
    Technique: Create frequency plots of the number of exceedances by month (or week) for several years. **Figure 5-1** shows that in the New Jersey and New York City metropolitan region, a forecasting season for ozone would last from May through September, since most of the 1-hr and 8-hr exceedances are confined to these months.
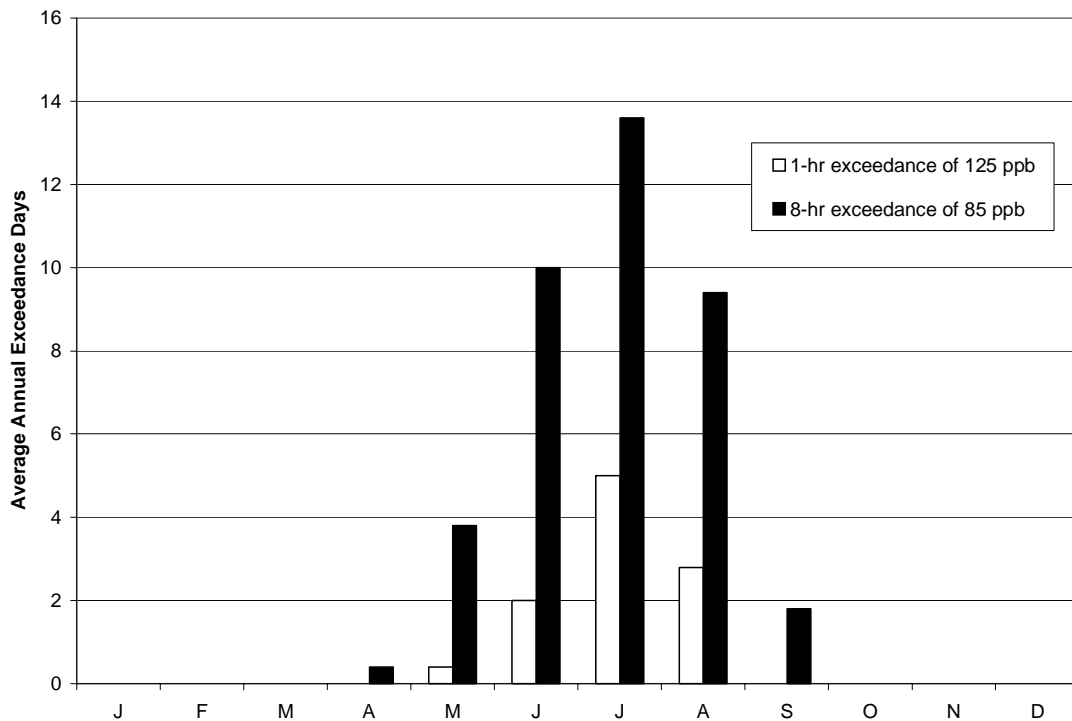
Figure 5-1. Distribution of the average number of days with 8-hr and 1-hr exceedances by month for the New Jersey and New York City region from 1993-1997 (NESCAUM, 1998).

Question: *At what time of day do the highest ozone or PM$_{2.5}$ concentrations occur? For how many hours do high concentrations typically last?*

Why: Knowing the typical time and duration of high ozone and PM$_{2.5}$ concentrations can help public outreach personnel properly notify the public so they can take appropriate action to minimize exposure. Understanding the diurnal cycle of ozone or PM$_{2.5}$ can help in producing a more accurate forecast.

Technique: Create frequency plots of the time of peak ozone or PM$_{2.5}$ concentrations. For example, **Figure 5-2** shows that the highest occurrence of 1-hr ozone exceedances is at 1400 EST in the New Jersey and New York City region, but ranges from 1000 to 1700 EST.
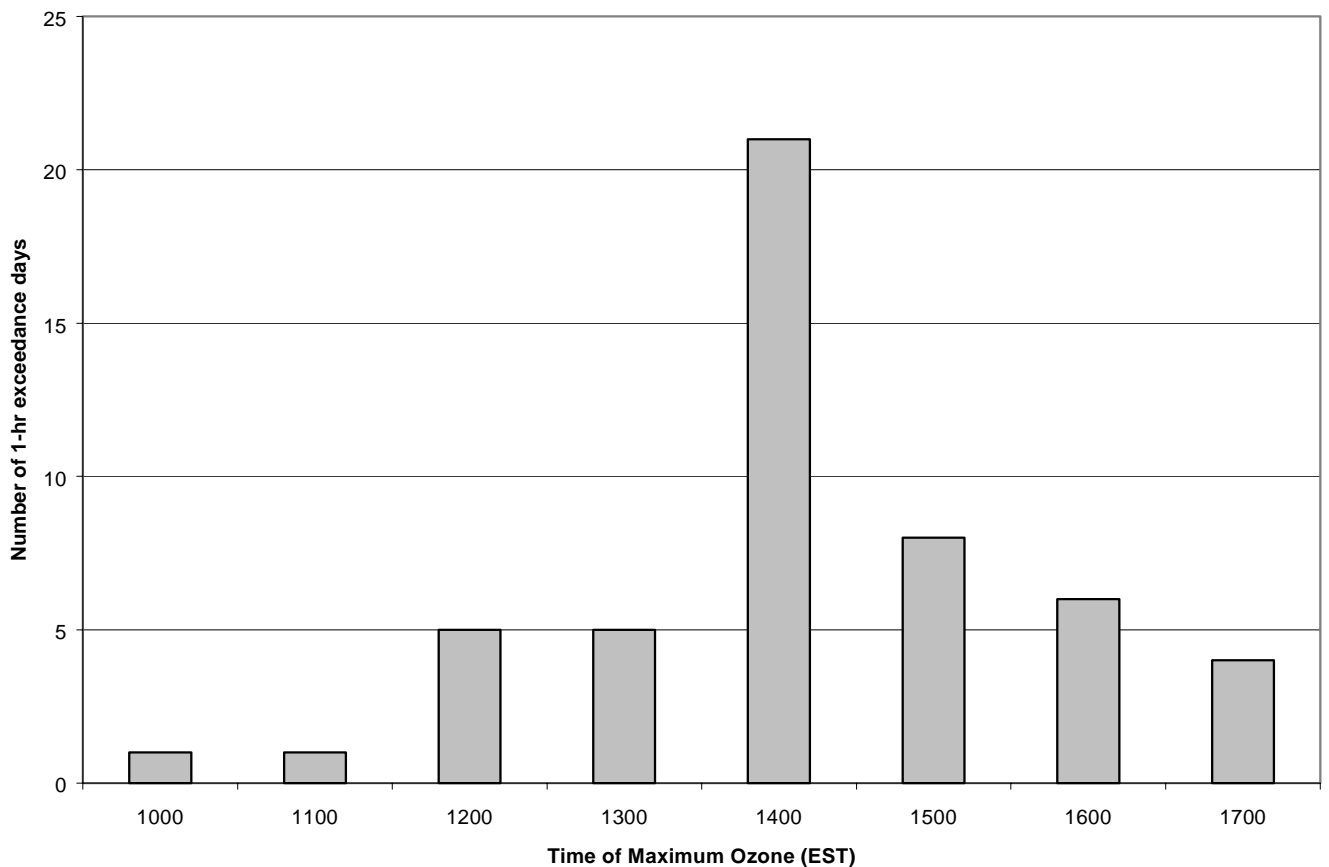
Figure 5-2.   Distribution by hour of daily maximum 1-hr ozone concentration on days that exceeded 125 ppb in the New Jersey and New York City region from 1993-1997 (NESCAUM, 1998).

Question: *For how many consecutive days do high ozone or* PM$_{2.5}$ *episodes typically last?*

Why: Knowing the typical duration of episodes can help guide the forecast. For example, if episodes never last more than two days in a particular area, the occurrence of a three-day episode in the future is unlikely; therefore, the forecaster would be cautious to forecast high pollution for three straight days.

Technique: Create a frequency plot (such as the one shown in **Figure 5-3**) of the number of continuous days with high ozone or PM$_{2.5}$ concentrations. Figure 5-3 indicates that a typical ozone episode of 125 ppb exceedances lasts one to two days and is never longer than four days in the New Jersey and New York City region.
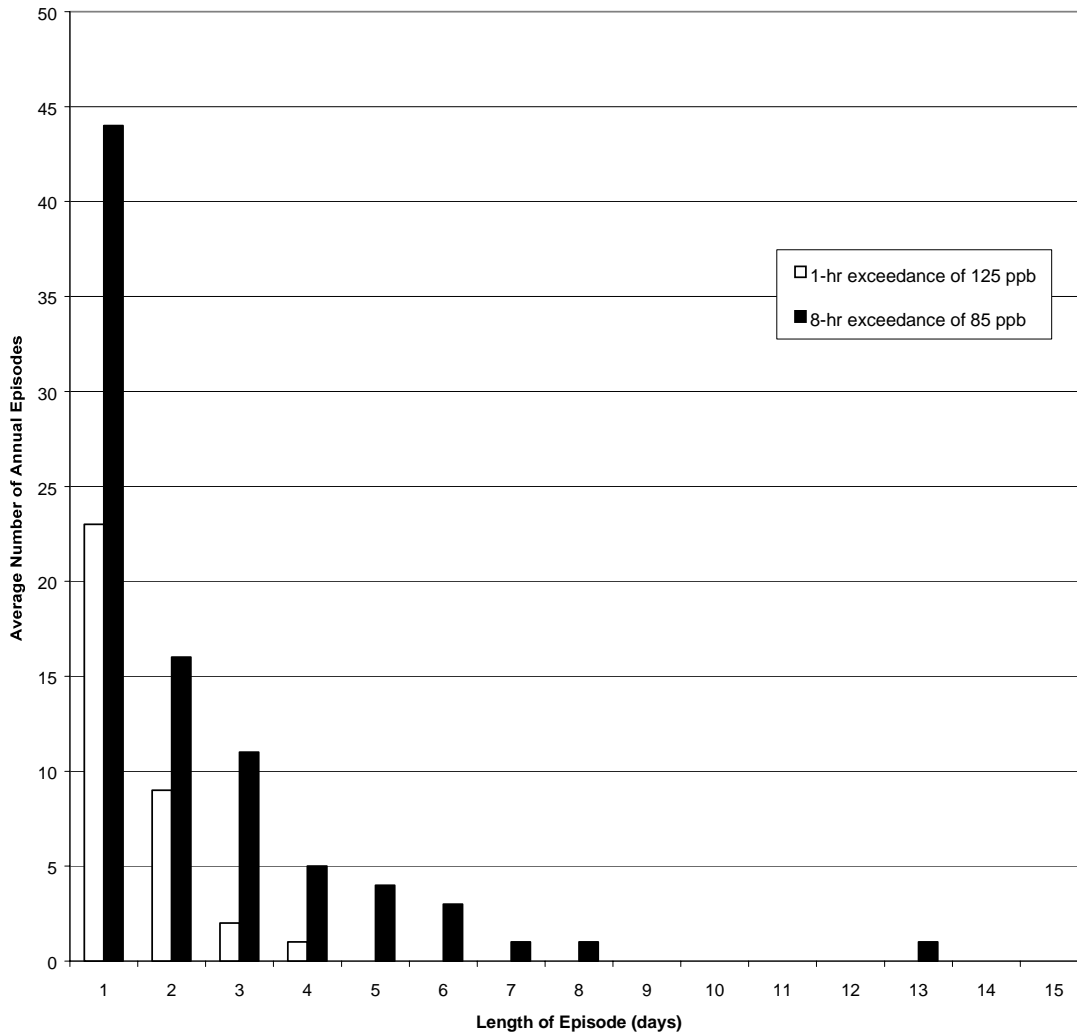
Figure 5-3. Average annual frequency of ozone episode length for the 8-hr and 1-hr standards in the New Jersey and New York City region from 1993-1997 (NESCAUM, 1998).

Question: *Do maximum ozone or PM$_{2.5}$ concentrations vary by day of week?*
Why: Weekday and weekend differences in commute traffic and some industrial processes can lead to a variation in ozone and PM$_{2.5}$ concentrations given similar weather conditions.
Technique: Create frequency plots of the number of exceedances by day of week. For example, **Figure 5-4** shows that in the New Jersey and New York City region 1-hr ozone exceedances are more likely to occur on Tuesdays and Wednesdays and somewhat less likely to occur on the weekends. Notice that the 8-hr exceedances show no day-of-week dependence. Thus, given similar meteorological conditions, a 1-hr ozone forecast on Saturday through Monday should be lower than one for Tuesday through Friday in this region. Notice that weekend 8-hr exceedance frequency is higher than all days except Tuesday.
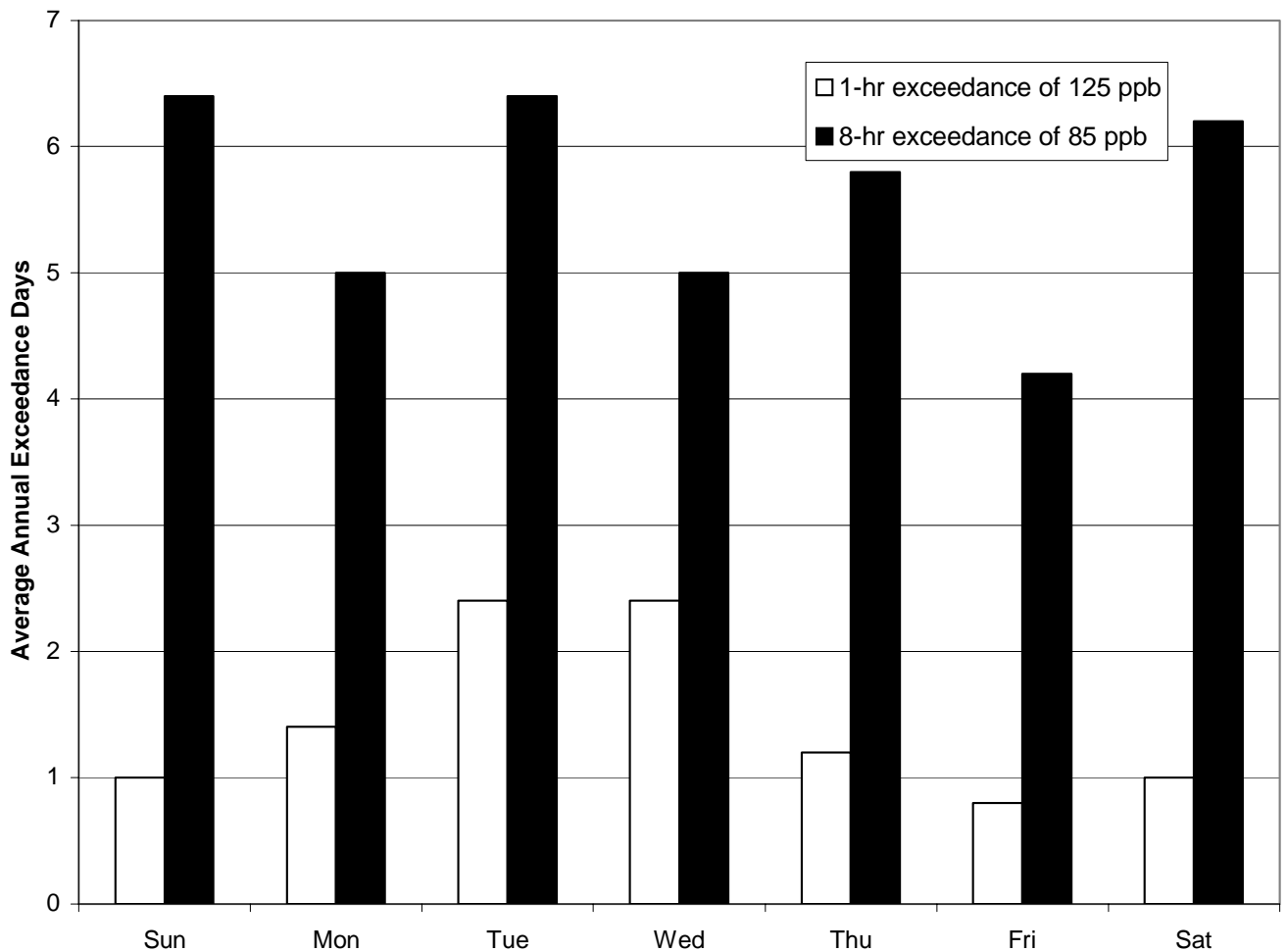
Figure 5-4. Distribution of the average number of 8-hr and 1-hr exceedances by day of week for the New Jersey and New York City region (NESCAUM, 1998).

Question: *During what time of year do PM$_{2.5}$ episodes occur?*

Why: Unlike ozone, for many areas of the United States PM$_{2.5}$ episodes can occur at any time of the year. Understanding what time of year PM$_{2.5}$ episodes occur in the forecast area will help define the forecast season or seasons, so resources can be effectively allocated. Also, understanding when PM$_{2.5}$ episodes occur helps the forecaster determine what meteorological phenomena may be influencing air quality.

Technique: Create histograms of the number of days with PM$_{2.5}$ concentration above a certain threshold by month. The threshold may be the NAAQS 24-hr PM$_{2.5}$ standard.

Question: *Do ozone and PM$_{2.5}$ episodes coincide with one another? When do PM$_{2.5}$-only episodes occur?*

Why: In many areas of the country, PM$_{2.5}$ and ozone episodes coincide during the "summertime" (May through September). If this is the case for the forecast area, then understanding the processes that influence ozone may help forecasters understand the processes that influence PM$_{2.5}$ during the summertime. PM$_{2.5}$-only episodes occur in the wintertime under weather conditions conducive to high pollutant concentrations and when there is not enough sunlight (due to the low sun angle) to drive the photochemical reactions that produce ozone. PM$_{2.5}$-only episodes can also occur in the summertime, for example, when there are unusual emissions that produce PM$_{2.5}$ or if conditions are conducive to pollution buildup, but clouds block sunlight thus keeping ozone concentrations low while PM$_{2.5}$ concentrations are high.

Technique: Create a scatter plot of the daily AQI based on ozone versus the daily AQI based on PM$_{2.5}$. For example, **Figure 5-5** shows a scatter plot of the summertime daily AQI based on ozone and PM$_{2.5}$ for Washington D.C. for 1999 through 2001. As noted on this plot the AQIs are not well correlated meaning that ozone and PM$_{2.5}$ episodes don't coincide very often in Washington D.C. However, there are days on which high AQIs do agree. These days should be investigated to understand what makes them different from the others.

Spatial distribution of ozone or PM$_{2.5}$

Question: *Where do the highest ozone or PM$_{2.5}$ concentrations occur? Do the highest concentrations occur at different times for different sites?*

Why: Different areas in the forecast region may have very different ozone or PM$_{2.5}$ characteristics due to spatial variation in emissions and meteorology. When forecasting for large areas it may be necessary to sub-divide the region to account for differences in ozone or PM$_{2.5}$ concentrations based on differences in emissions and weather across the region.

Technique: Plot a map of the average peak ozone or PM$_{2.5}$ concentration and the time of the peak for each site on exceedance days.
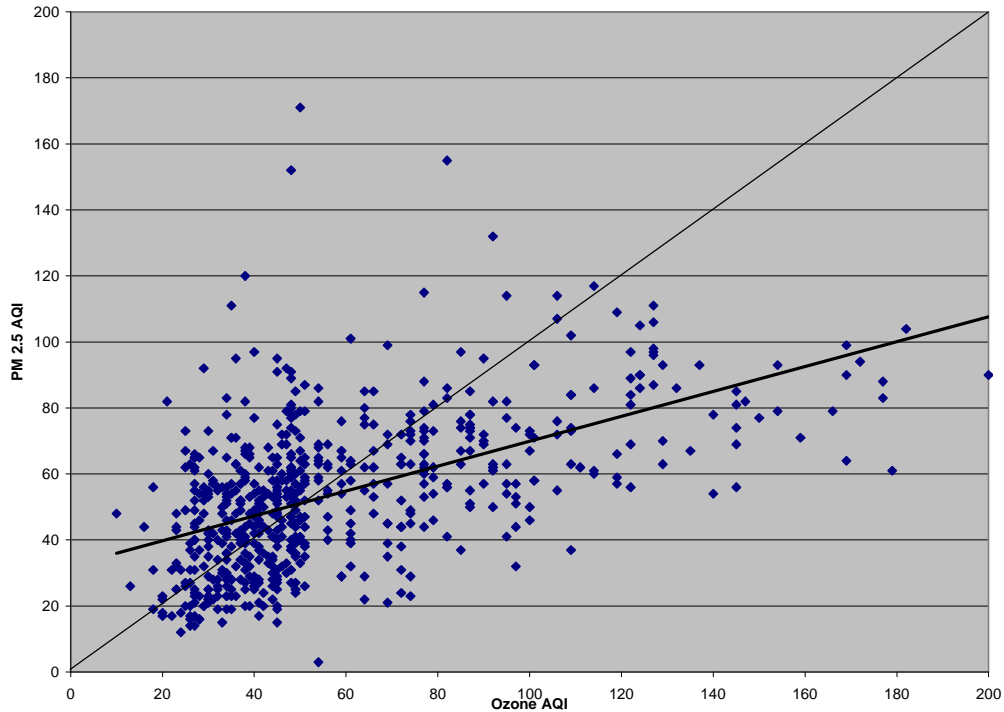
Figure 5-5. Scatter plot of the summertime daily AQI based on peak 8-hr ozone concentrations versus the AQI based on 24-hr average $PM_{2.5}$ concentrations for Washington D.C. for 1999 through 2001. The one-to-one line (thin line) and linear best fit (thick line) are also shown.

Question: *Have emissions patterns changed in recent years?*
Why: If emissions patterns have changed, weather conditions that have historically produced ozone or $PM_{2.5}$ exceedances, may now result in lower concentrations.
Technique: Determine if significant emissions changes (e.g., the use of reformulated fuel) or shifts in population have occurred in the forecast region. For example, less reactive emissions may result in peak ozone concentrations occurring farther downwind and/or in lower ozone concentrations. Changes in SO_2, NO_x, or ammonia emissions can influence the production of secondary $PM_{2.5}$.

Monitoring issues

Question: *Has the monitoring network changed recently?*
Why: Changes in a monitoring network can cause significant differences between historic and currently observed ozone or $PM_{2.5}$ concentrations. If a new monitor was recently installed downwind of a major emission source area, then the observed time and peak ozone or $PM_{2.5}$ concentrations for the entire area may change significantly due to this new site. These types of monitoring network changes must be taken into account when analyzing historic and current ozone or $PM_{2.5}$ concentration data.

Technique: Create a plot of the historic monitoring network and compare it to a plot of the current network. If new sites have been added in recent years, determine if the new sites have caused an increase in the number of exceedance days in the forecast region.

Question: *What are the different PM$_{2.5}$ monitoring methods and how do they compare to one another?*

Why: Different monitoring methods (FRM, BAM, TEOM) can report very different concentrations, even when sites are collocated. The differences are often weather condition dependent (temperature and humidity). If forecasting tools are developed using historic FRM data and then real-time TEOM or BAM data are used for daily forecasting, then these differences need to be considered.

Technique: Create a scatter plot of daily 24-hr average PM$_{2.5}$ concentrations from collocated FRM and continuous monitors. For example, **Figure 5-6** shows a comparison of 1999-2001 FRM and TEOM 24-hr average PM$_{2.5}$ data collected at a site in Forsyth County, North Carolina. The blue line is the 1:1 line. The regression equation shows a slope near 1 and the r^2 is high (.96), which indicates a good relationship among the data.
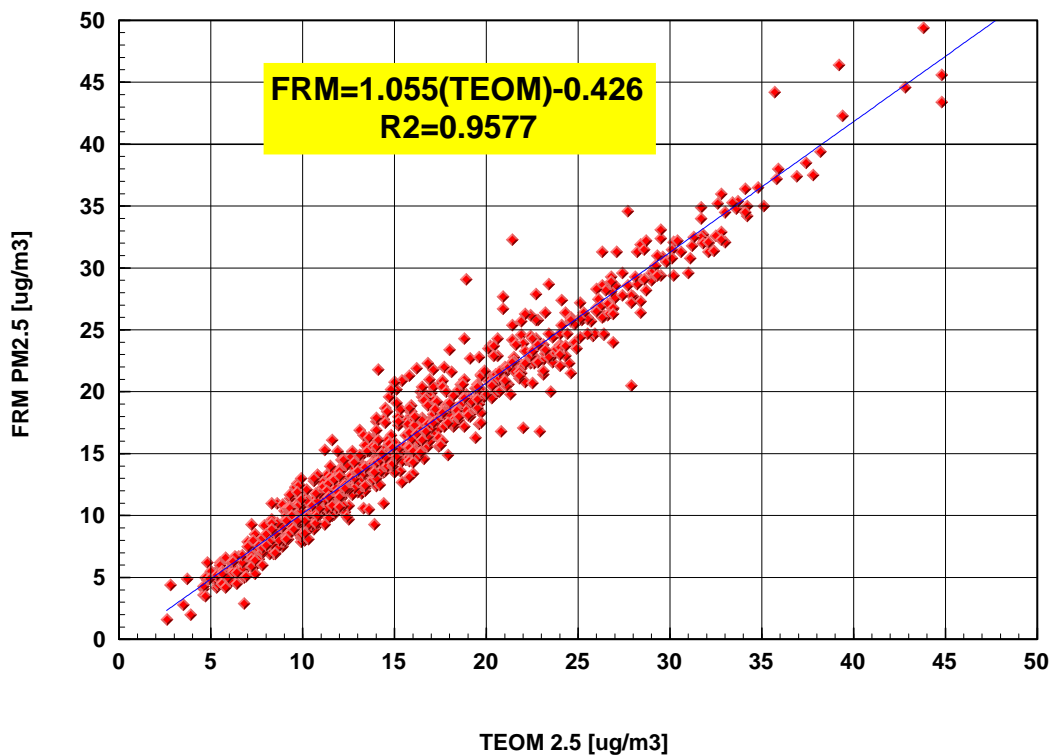

FRM=1.055(TEOM)-0.426
R2=0.9577

Figure 5-6. Comparison of FRM and TEOM 24-hr average PM$_{2.5}$ data collected in Forsyth County, North Carolina, from 1999 to 2001 (Courtesy of Lewis Weinstock).

Question: *Are adjustments made to the continuous $PM_{2.5}$ data to make them better match the FRM standards? If so, how accurate are these adjustments?*

Why: Forecast tools are usually developed from FRM data. In real-time, only continuous data are available to verify the forecast. Since collocated FRM and continuous monitors often report different $PM_{2.5}$ concentrations, adjustments to the continuous data need to be made to properly verify the forecast.

Technique: Create scatter plots of collocated FRM and adjusted continuous monitor data (similar to that shown in Figure 5-6). If the data points fall within roughly 15% of the 1:1 line, then the adjusted data represent the FRM data; otherwise, evaluate why the data are different and consider developing new techniques to adjust the continuous data to match the FRM data. For example, developing equations for each season may produce better results.

Meteorological and air quality processes

Question: *What types of synoptic weather patterns are associated with high ozone or $PM_{2.5}$ concentrations?*

Why: Synoptic-scale weather features are large (1000 km or more) weather circulations that influence regional weather conditions that, in turn, strongly influence the production and transport of ozone and $PM_{2.5}$ and their precursors. By reviewing weather forecast charts, forecasters can identify historical weather patterns associated with particular pollutant concentrations in the forecast region.

Technique: Analyze historical weather charts that depict synoptic features. Classify the surface and aloft synoptic weather patterns and create frequency plots showing the synoptic pattern versus the number of high ozone or $PM_{2.5}$ concentration days, moderate ozone or $PM_{2.5}$ concentration days, and low ozone or $PM_{2.5}$ concentration days. EPA's AIRNow program provides historical ozone maps since 1998 (www.epa.gov/airnow/).

For example, **Figure 5-7** shows a surface synoptic pattern associated with high ozone in Pittsburgh, Pennsylvania (Comrie and Yarnal, 1992). Historic daily weather map sources include the Daily Weather Map series issued daily by the National Oceanic and Atmospheric Administration (NOAA)[3] and the archive analysis of the National Center for Environmental Prediction weather models available on the Internet at www.arl.noaa.gov/ready/arlplota.html.

Question: *Does local carryover contribute to ozone or $PM_{2.5}$ concentrations?*

Why: When ozone or $PM_{2.5}$ episodes occur over several days, day-to-day pollution buildup can contribute to the daily ozone and $PM_{2.5}$ concentrations. That is, today's pollution (if not dispersed, deposited, or permanently reacted away) will contribute to tomorrow's pollution.

---

[3] Daily Weather Maps, Climate Prediction Center, Room 811, World Weather Building, Washington, DC 20233
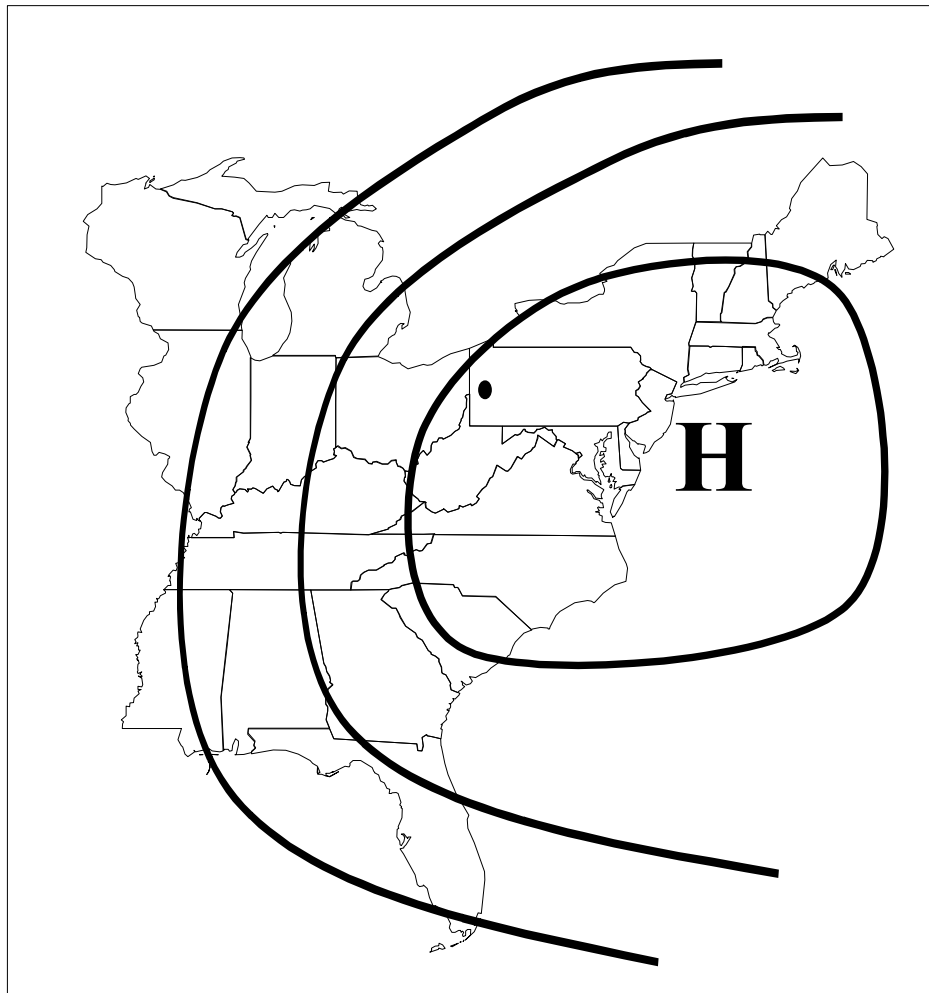
Figure 5-7. A surface synoptic pattern associated with high ozone in Pittsburgh, Pennsylvania (Comrie and Yarnal, 1992).

Technique: Carryover ozone can be investigated by examining ozone data from surface sites at which the ozone data show no overnight titration by NO. Since there is no titration for $PM_{2.5}$ most sites that are not significantly influenced by local sources can be analyzed. Create scatter plots of overnight ozone or $PM_{2.5}$ concentrations vs. the next-day concentrations. Examine the plots to see if there is a relationship between overnight concentrations and the next-day concentrations.

For example, **Figure 5-8** shows the relationship between 0200 EST ozone concentrations at a mountainous site in western North Carolina (a site that is representative of regional carryover) and North Carolina daytime peak ozone concentrations. The influence of background ozone or $PM_{2.5}$ can also be accessed by analyzing aloft data collected by aircraft, on a tower, or on a nearby mountain instead of or in addition to the surface data.
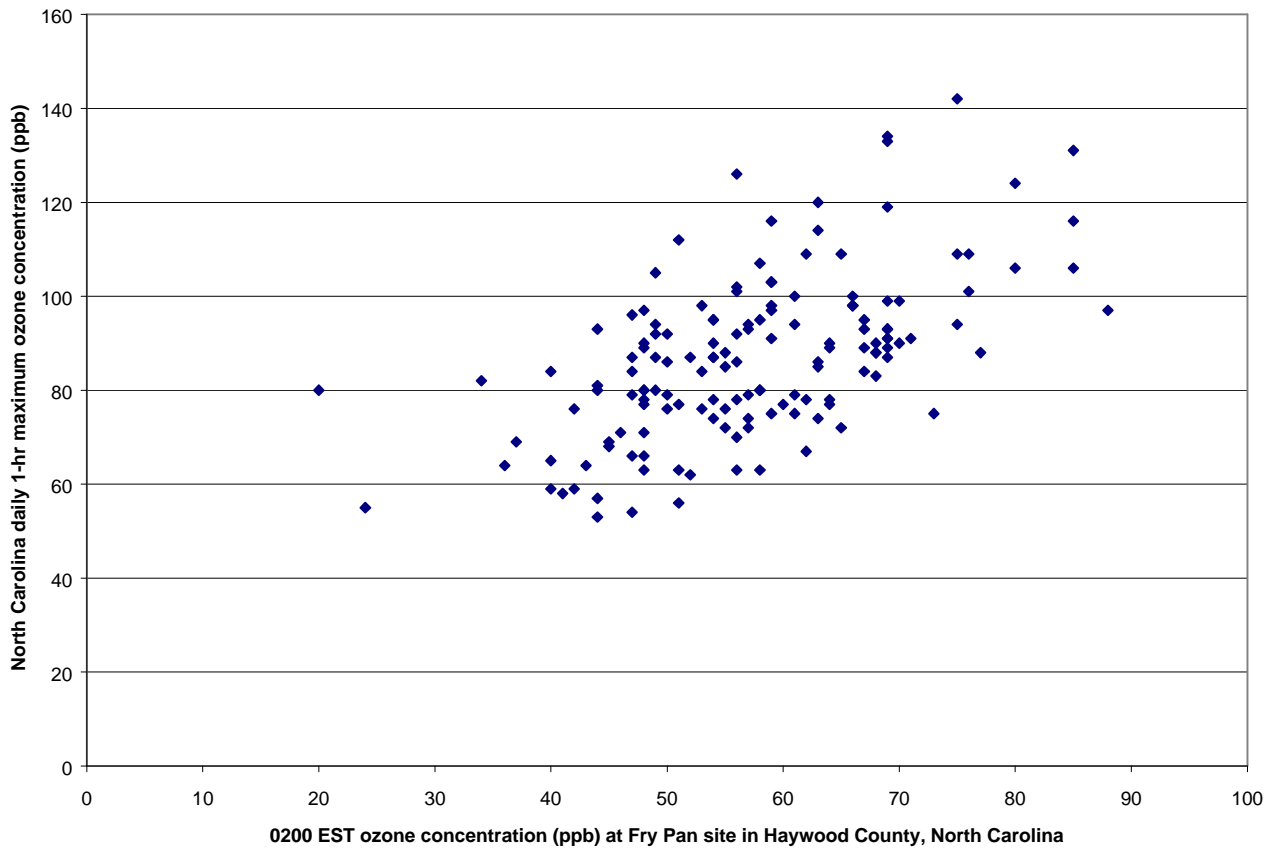
Figure 5-8. Scatter plot of 0200 EST ozone concentrations at a mountainous site (Fry Pan) in Haywood County, North Carolina, versus North Carolina daily regional maximum ozone concentrations for June to September, 1996 (MacDonald et al., 1998).

Question: *Do surface or aloft transport of pollutants from other areas contribute to ozone or $PM_{2.5}$ in the forecast area?*

Why: Long-range transport of pollutants can contribute significantly to local ozone or $PM_{2.5}$ concentrations. It is important for a forecaster to understand if and when this occurs in order to accurately forecast ozone or $PM_{2.5}$.

Technique: Computing back trajectories is a useful way to examine the potential for long-range transport **(Figure 5-9)**. For selected days, create 12-hr, one-day, and two-day back trajectories at several levels. Determine if these trajectories originate in areas with high ozone or $PM_{2.5}$ concentrations. An excellent tool for computing back trajectories is interactively available on the Internet at www.arl.noaa.gov/ready/hysplit4.html (Draxler and Hess, 1997).
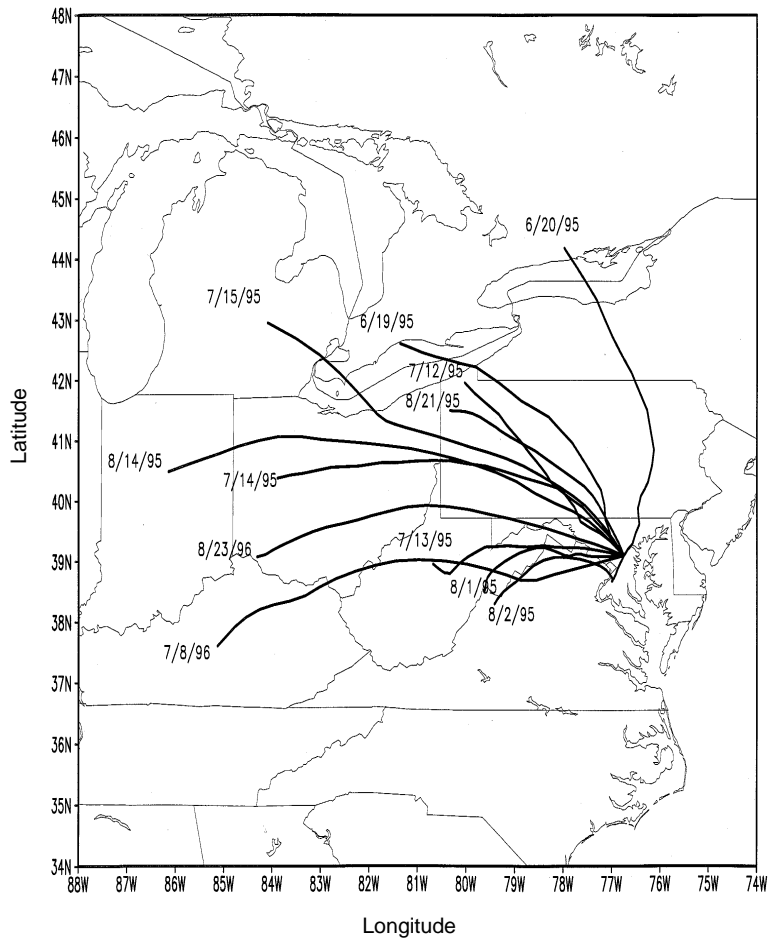
Figure 5-9. Back trajectories during an ozone episode in the northeastern
            United States showing possible transport of pollutants from regions to
            the west (Ryan et al., 1998).

Question: *How do local flow patterns influence ozone or PM$_{2.5}$ concentrations?*
Why: Local flow patterns such as land-sea breezes, up/down slope flows, and terrain
         guided flows can play a large role in transporting ozone and PM$_{2.5}$. Such flows
         may locally transport pollutants from upwind sources to downwind cities or
         recirculate pollutants within metropolitan areas. Understanding the flow
         processes in the forecast area will greatly improve the forecasts.
Technique: Compute back trajectories on high, moderate, and low ozone or
         PM$_{2.5}$ concentration days. A good resource for computed trajectories is
         www.arl.noaa.gov/ready/hysplit4.html.

         An example of a 24-hr back trajectory for a monitoring site in Crittenden County,
         Arkansas, (near Memphis, Tennessee) on a high ozone day is shown in
         Figure 5-10. This simple trajectory shows both surface and aloft flow from the
         northeast portion of the domain.
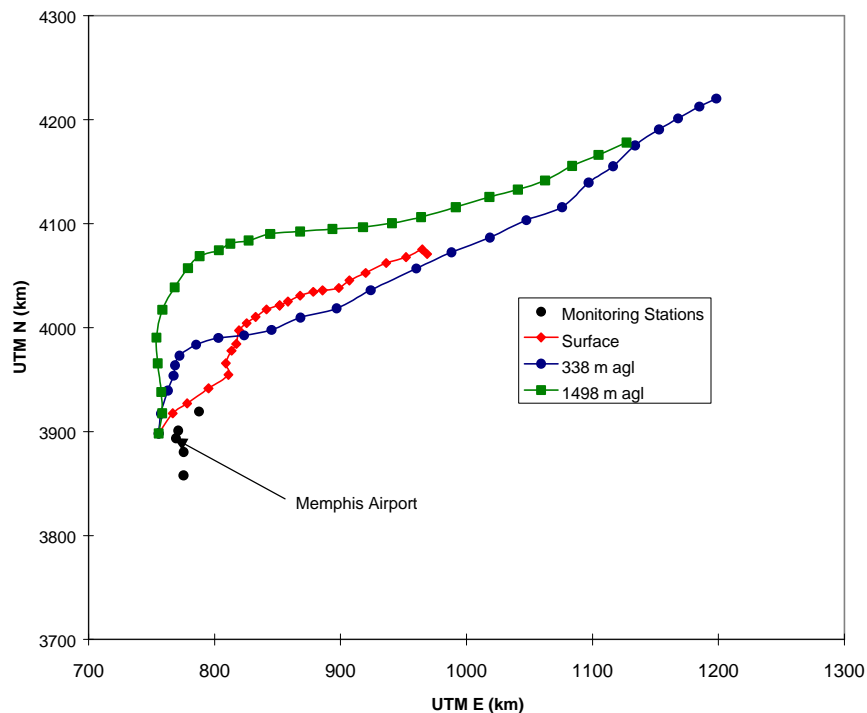
Figure 5-10. A 24-hr back trajectory from Crittenden County, Arkansas, starting at 1400 EST on August 25, 1995, and ending at 1300 EST on August 26, 1995. Trajectories were computed using surface wind data from the National Weather Service's site at the Memphis airport and upper-air data from a radar wind profiler located at the airport (Chinkin et al., 1998).

Question: *How does the aloft temperature structure influence peak ozone or PM$_{2.5}$ concentrations?*

Why: Aloft temperature structure strongly influences vertical mixing and dilution of pollutants. A stable atmosphere produces less vertical mixing and dilution of pollutants which leads to higher ozone or PM$_{2.5}$ concentrations.

Technique: In many areas of the country 850-mb temperature is a good indicator of aloft stability and inversion strength. Forecasted 850-mb temperatures can therefore be used to estimate the amount of mixing and dilution of ozone and PM$_{2.5}$. For example, New York State ozone forecasters use an 850-mb temperature greater than 15°C as one criterion for forecasting high ozone concentrations (Taylor, 1998), while Sacramento, California, forecasters set the 850-mb temperature criterion at a temperature greater than 18°C (Dye et al., 1996).

Question: *What types of weather patterns are associated with cloud cover?*

Why: Cloud coverage limits the photodissociation of NO$_2$; this is a key step in ozone formation. Accurately predicting cloud coverage will improve ozone forecast accuracy. Clouds can also increase the conversion of SO$_2$ to sulfate.

Technique: The National Weather Service' (NWS') computer forecast models predict relative humidity at several altitudes with reasonable accuracy. Analyzing these predictions along with satellite images can help to forecast cloud cover. Model

output statistics (MOS) predict the amount of cloud cover but are not always accurate. For example, anvils from distant thunderstorms are often not accurately predicted. Performing case studies of days when a model's cloud predictions are wrong and understanding the types of weather patterns that cause inaccurate cloud forecasts, allow forecasters to identify such conditions in the future.

Question: *Are there exceptional emission events such as forest fires, agricultural burning and tilling, windblown dust, etc. that cause $PM_{2.5}$ episodes?*

Why: Rare emission events often cause $PM_{2.5}$ episodes. Since these events are rare, they are difficult to predict using objective forecasting tools or techniques alone. By understanding when such events are likely to occur (for example, wildland fires occur most often in the summer and fall), forecasters can watch for such events and account for them in the daily forecasting when needed.

Technique: Perform case study analysis of past $PM_{2.5}$ episodes to determine the cause of the high concentrations. See Section 2.2.5 for more information on rare events.

Once forecasters understand the chemical and meteorological processes that influence ozone and $PM_{2.5}$ concentrations, they can start selecting methods to forecast these pollutants, as discussed in the next section.

5.3   CHOOSING FORECASTING METHODS

Once the needs of the forecasting program have been identified, forecasters will need to choose a forecasting method or combination of methods to predict ozone or $PM_{2.5}$. The method(s) chosen will primarily depend on the available resources and experience. Listed below are a number of issues to consider when selecting a forecasting method.

Resources Cost may be the major factor guiding method selection. When determining the overall cost of a particular method, the costs associated with both developing and operating the method should be considered. Development costs versus operating costs can vary greatly between methods. For example, the development of a regression model may be fairly expensive compared to the development of a criteria method; however, the operational costs may only be slightly different.

Severity of problem The severity of the air quality problem and the frequency of high concentrations in the forecast region will also guide the choice of method. For example, a region with very few ozone episodes may only need a simple and inexpensive method to forecast a few high ozone days. On the other hand, a region that experiences many days with high ozone during the summer and numerous days with high $PM_{2.5}$ in the winter may benefit from developing several forecasting methods that can be utilized year-round.

Balancing methods	Balancing resources between multiple forecasting methods may minimize the limitations of the methods while compounding their strengths. Also, balancing objective and subjective methods may increase forecast accuracy.
Adding methods	Once a forecasting method has been selected, the program is not limited to retaining this single method. Building a program from one simple method in the first year to multiple methods in future years is a cost-effective approach to increase the accuracy of a forecasting program.
Expertise	Some methods require a high level of meteorological experience and forecasting expertise. Working with a university or other agency to develop a forecasting method may be beneficial if in-house resources are not available.

5.4   DATA TYPES, SOURCES, AND ISSUES

Once a forecasting method or methods has been selected, data needs should be addressed. Air quality and meteorological data are needed for both developing the method(s) to predict air quality and for operationally forecasting. This section identifies the types and sources of meteorological and air quality data as well as issues to consider when acquiring and using data.

A variety of data types, both meteorological and air quality, are available for developing prediction methods and forecasting air quality. The general data requirements of each method are listed in Table 4-1. **Table 5-1** summarizes data types and typical parameters. These data types include surface and upper-air meteorological data, both observed and forecasted. Data needs for a particular forecast region will depend on the specific meteorological and air quality phenomena to be predicted.

Locating a data source is often a major part of developing a forecasting method. **Table 5-2** lists many of the major sources of air quality and meteorological data. Historical data can also be found in air quality studies that were conducted in the forecast region.

Most forecasting programs require many types of data from difference sources to fulfill all of the forecaster's needs. Each of these data sources provide data at different costs, in different file formats, and with varying degrees of reliability and quality. When acquiring data, consider the following issues:

Table 5-1. Data products for developing forecasting methods and for forecasting weather and ozone.

Data Type	Variables	Forecasted/ Observed	Frequency
Surface Meteorological	WS, WD, T, RH, Solar Rad., Cloud Cover, Vis, P	Observed	Hourly
Surface Air Quality	Ozone, PM_{10}, $PM_{2.5}$, Oxides of Nitrogen, Carbon Monoxide, VOCs	Forecasted and Observed	Hourly
Upper-air Meteorological *(rawinsondes, radar profilers, and sodars)*	Vertical Profiles of WS, WD, T, RH	Observed	Twice per day to hourly
Aloft Air Quality Observations *(towers, mountains, and aircraft)*	Ozone, PM_{10}, $PM_{2.5}$, Oxides of Nitrogen, Carbon Monoxide, VOCs	Observed	Variable
Weather Charts	Surface (WS, WD, T, RH, P) 850 mb (WS, WD, T, Height) 700 mb (WS, WD, T, Height) 500 mb (WS, WD, T, Height), Others	Forecasted and Observed	Typically twice per day
Weather Radar	Precipitation	Observed	Sub-hourly
Satellite	Cloud Cover (visible and infrared)	Observed	Hourly and/or Sub-hourly
Meteorological Model Forecasts	T, RH, WS, WD, Cloud Cover, Vis, P, and others at many levels	Forecasted	Twice per day
Text Weather Forecasts	Discussions	Forecasted and Observed	Four or more times per day

WS = wind speed   T= temperature   Vis = visibility
WD = wind direction   RH = relative humidity   P = pressure

Table 5-2. Major sources of air quality and meteorological data.

Data Source	Type of Data Source	Types of Data	Phone Number	Web Site
U.S. EPA Aerometric Information Retrieval System (AIRS)	Historical	Surface air quality	(916) 541-5586	www.epa.gov/ttn/airs
National Climate Data Center (NCDC)	Historical	Surface Meteorological, Upper-air Meteorological, Weather Charts, Radar, Satellite	(828) 271-4800	www.ncdc.noaa.gov
NOAA National Data Centers (NNDC)	Historical	Surface Meteorological, Upper-Air Meteorological, Weather Charts, Satellite, Radar, Climate	(828) 271-4800	www.nndc.noaa.gov
Regional Climate Centers	Historical	Surface Meteorological, Upper-air Meteorological, Climate Information	Western Regional Climate Center (775) 674-7010 High Plains Climate Center (402) 472-6706 Midwestern Climate Center (217) 244-8226 Northeast Regional Climate Center (607) 255-1751 Southeast Regional Climate Center (803) 734-9560 Southern Regional Climate Center (225) 388-5021	www.wrcc.dri.edu www.hpccsun.unl.edu mcc.sws.uiuc.edu met-www.nrcc.cornell.edu www.sercc.com www.srcc.lsu.edu
Purdue University	Historical and Real-time	Surface Meteorological, Upper-air Meteorological, Satellite, Radar, Model Forecast, Text Weather Forecast	-	www.arl.noaa.gov/ready/arlplota.html
Commercial Weather Service Providers (WSP)	Real-time	Surface Meteorological, Upper-Air Meteorological, Weather Charts, Satellite, Radar, Model Forecast, Text Weather Forecast	-	Comprehensive list of WSPs at: www.ugems.psu.edu/~owens/WWW_Virtual_Library/commercial.html

Cost	A significant amount of data is available for free on the Internet from the NWS, Government Laboratories, and Universities. The reliability of these data is reasonable for forecasting, but access to it may suffer from Internet outages. Weather Service Providers (WSPs) supply weather data to television stations, private industry, and government agencies. WSPs typically charge a startup fee for display and data acquisition software. Most charge a monthly/yearly data subscription fee and automatically send the data to a subscriber's computer. Reliability for this type of service is generally very high.
Reliability	Knowing that data will always be available when needed it is critical to a forecasting program's success. Unreliable data will reduce forecast effectiveness and may lower forecast accuracy.
Quality control	Receiving higher quality data (data with fewer errors and inconsistencies) decreases the necessity for personnel to thoroughly review the data before it is used. All historical data should be reviewed for quality prior to developing a forecasting method.
Data formats	The number of data formats used should be minimized in order for data to be decoded and processed more efficiently. Different time standards, reporting units, quality control codes, etc., can produce additional decoding/processing effort.
Ingest methods	Determining how data will flow from the source to the forecasting location is important. Ingest methods are typically the Internet, telephone, and satellite. Internet and telephone telemetry are cost effective. Satellite delivery systems are very reliable, yet are typically more expensive. Seek to automate as many of the data ingest tasks as possible, so the forecaster can spend more time on the nuances of the prediction.
Hardware requirements	Hardware needs are a function of the amount of data and data processing required. The greatest convenience of WSPs is that they can supply multiple data types through one software package and computer. Combining this service with Internet use on the same computer not only improves resource efficiency, but also provides additional data types and redundancy at little extra cost.
Redundancy	Having a backup or secondary data source is a prudent practice. Consider the risks of not having data available for forecasting. For key information and data, identify several sources from which the data can be obtained.

Using high quality meteorological and air quality data is important to develop accurate forecasting methods. In addition, obtaining reliable real-time data is a key component for operationally forecasting.

5.5   FORECASTING PROTOCOL

A forecasting protocol describes the daily operating procedures from data acquisition to forecast production and dissemination. A protocol helps guide personnel through the forecasting process. It ensures that all activities are performed on time without the need for last minute decisions and helps maintain consistency from one forecaster to the next. This section explains what to include in a forecasting protocol.

Preparing a forecast often requires that various personnel complete numerous steps. To standardize this process, written procedures should be prepared and tested that the forecast team can follow on a regular basis. A forecasting protocol will likely include:

- Descriptions of the meteorological conditions that produce high ozone and $PM_{2.5}$ concentrations in the area.

- A schedule of daily tasks and personnel responsibilities. The easiest way to create this schedule is to work backwards from the time that the forecast is due to the time initial procedures need to begin. It is likely that the schedule will differ for high, moderate, and low ozone days. An example of a basic schedule is shown in **Table 5-3**.

- Steps to take to arrive at a forecast, including key decision points that help forecasters to quickly identify low ozone and $PM_{2.5}$ days, thus allowing time for the more difficult forecasts.

- Forms and worksheets for documenting data, forecast information, forecast rationale, and comments that forecasters can analyze and evaluate later.

- Phone and fax numbers and e-mail addresses of key personnel.

- Names, fax and phone numbers, and e-mail addresses of forecast recipients.

- Troubleshooting and backup procedures for the key components necessary to produce and issue the pollutant forecasts such as: backup forecasters, redundant data acquisition methods, and forecast dissemination.

These written procedures save time and effort and should be an integral part of any forecasting program.

5.6   FORECAST VERIFICATION

Verification is the process of evaluating the quality of a forecast by comparing the predicted air quality to the observed air quality. As part of a forecasting program, forecasters should regularly evaluate the forecast quality. The benefits of verifying ozone or $PM_{2.5}$ forecasts include:

- Quantifying the performance of forecasters and/or the forecast program,

- Identifying trends in forecast performance over time,

- Quantifying improvements from new (or changes in) forecasting methods/tools,

Table 5-3. Example of a forecasting protocol schedule.

Time	Activity
0900 – 0930	Run air quality data acquisition programs. Review data for completeness and accuracy.
0930 – 0945	Acquire observed and forecasted meteorological data from the Internet.
0945 – 1015	Review forecast weather maps.
1015 – 1030	Run regression model to forecast ozone.
1030 – 1100	Evaluate forecasted weather conditions and air quality using the Phenomenological/Intuition method.
1100 - 1125	Produce the final forecast; write forecast discussion.
1125 – 1130	Fax forecast to air district officials and place forecast on the Internet.

- Comparing verification statistics to those from other agencies that forecast air pollution.

- Demonstrating the performance of forecasts to program participants, stakeholders, and the media.

The verification process can be complex since there are many ways to evaluate a forecast including accuracy, bias, and skill. Since no one statistic can fully reflect the performance of a program, many verification statistics must be completed in order to completely evaluate the quality of a forecast program.

Two basic types of forecasts exist: discrete forecasts of specific concentrations and AQI category forecasts (e.g., good, moderate, etc.). Verification statistics differ for these two types of forecasts. This section explains how verification statistics can be computed and interpreted for both types of forecasts and provides a schedule for verifying forecasts (Section 5.6.1). Section 5.6.2 describes verification statistics for discrete forecasts. Section 5.6.3 describes verification statistics for category forecasts. An on-line glossary of verification terms can be found at www.sec.noaa.gov/forecast_verification.

5.6.1 Forecast Verification Schedule

Forecasts should be evaluated frequently to identify any problems or downward performance trends. A schedule of verification tasks follows:

Daily   If forecasts are significantly missed (off by more than 30 ppb or two AQI categories), the causes of the missed forecasts should be examined. Write a forecast retrospective, which is a several page document that details what went wrong and includes recommended changes to forecast methods or procedures. **Figure 5-11** shows an example outline for a forecast retrospective.

Monthly	Compute the forecast verification statistics described in this section. Compare these with statistics from previous months and review statistics with forecasters.
Annually	Compute the forecast verification statistics described in this section. Compare these statistics from previous years, review statistics with forecasters, and compare them to other forecasting programs.

Forecast Retrospective
Date

1. Summary of event
  Provide a brief synopsis of what happened.
2. Forecast rationale
  Explain the steps and thought processes used to make the forecast.
3. Actual weather and air quality conditions
  Discuss all aspects of the weather that occurred. Use weather maps, satellite images, observations. Review the relevant air quality conditions.
4. Revision to forecasting guidelines
  Recommend any changes to forecasting procedures.

Figure 5-11.   Example outline of a forecast retrospective.

5.6.2   Verification Statistics for Discrete Forecasts

Several verification statistics can be computed for forecasting programs that predict discrete pollutant concentration values. **Table 5-4** lists four statistics commonly used to verify discrete forecasts and explains how to compute and interpret these statistics. The four statistics are:

Accuracy	Average "closeness" between the forecast and observed values.
Bias	Indicates, on average, if the forecasts are underpredicted or overpredicted.
Skill score	Percentage improvement of a forecast with respect to a reference forecast (typically a climatology or persistence forecast).
Correlation	A measure of the relationship between forecasts and observations and if the two sets of data change together.

The first step in the process is to pair the forecast and observation data for each forecast issued. Then use the equations listed in Table 5-4 to calculate the verification statistics.

Table 5-4.   Verification statistics computed on discrete concentration forecasts.

Statistic Name	What it Measures	How to Compute it	Equation	Units	How to Interpret it
Accuracy (mean absolute error)	Average "closeness" between the forecast and observed values. Summarizes the overall quality of the forecasts.	1. Take the absolute difference between forecast (*f*) and observation (*o*) for all forecasts (*N*). 2. Sum the differences and divide by N.	$A = \dfrac{1}{N}\left(\sum_{1}^{N}\lvert f - o\rvert\right)$	Units of variable (ppm, ppb, $\mu g/m^3$, etc.)	• Lower numbers are best. • Values indicate the uncertainty in any single forecast.
Bias (mean error)	It indicates, on average, if the forecasts are underpredicted or overpredicted.	1. Take the difference between forecast (*f*) and observation (*o*) for all forecasts (*N*). 2. Sum the differences and divide by N.	$B = \dfrac{1}{N}\left(\sum_{1}^{N}(f - o)\right)$	Units of variable (ppm, ppb, $\mu g/m^3$ etc.)	• Values near 0 are best. • Values <0 indicate underpredicting (i.e., forecasts are too low). • Values >0 indicate overpredicting.
Skill score	Percentage improvement of a forecast with respect to a reference forecast (typically a climatology or persistence forecast).	1. Compute accuracy for a reference forecast (A_{ref}), either climatology or persistence (see Section 4.1 for details). 2. Compute accuracy (A) for the forecast. 3. Divide accuracy by the reference accuracy and subtract from 1.	$SS = \left(\dfrac{A}{A_{Ref}} - 1\right)*100$	%	• 0% indicates no improvement (or skill) over the reference forecast. • 50% or more indicates a significant improvement in skill.
Correlation	A measure of the relationship between forecasts and observations. It measures if two sets of data change together.	Use correlation functions in spreadsheet or statistics software (Excel, Lotus) or compute using the following: 1. Compute the co-variance (Cov(f,o)) using the equation. 2. Compute the standard deviation for the forecasts (sf) and observations (so). 3. Divide the co-variance by the product of the standard deviations to compute the correlation (C_{fo}).	$C_{fo} = \left(\dfrac{Cov(f,o)}{sf \cdot so}\right)$ where: $-1 > C_{fo} < 1$ $Cov(f,o) = \dfrac{1}{N}\left(\sum_{j=1}^{N}(f_j - \mu_f)(o_j - \mu_o)\right)$ where μ_0 = mean observed value μ_f = mean forecasted value $sf = \sqrt{\dfrac{N \cdot \sum_{j=1}^{N} f_j^2 - \left(\sum_{j=1}^{N} f_j\right)^2}{N(N-1)}}$ $so = \sqrt{\dfrac{N \cdot \sum_{j=1}^{N} o_j^2 - \left(\sum_{j=1}^{N} o_j\right)^2}{N(N-1)}}$	-	• Values close to 1 are best. • Positive correlation indicates that large forecast values are associated with large observed values. • Negative correlation occurs when small values of one set are associated with large values of the other. • No correlation occurs when values in both sets are unrelated. • High correlation does not necessarily denote high accuracy.

To illustrate how to compute and interpret these statistics, **Table 5-5** shows hypothetical forecasts and verification statistics. In this case, a forecaster made hypothetical forecasts (F) for 11 days. For reference purposes, forecasts were also made using the Persistence method (F_{Pers}) discussed in Section 4.1. A reference forecast is a baseline against which to compare a forecast. Any other forecast method can be used, but typically persistence, climatology, and random chance are used as reference forecasts.

Table 5-5. Hypothetical forecasts for an 11-day period showing a human forecast (F), observed values (O), and forecasts using the Persistence method (F_{Pers}). Accuracy (A), accuracy for persistence forecast (A_{Pers}), bias (B), skill-score (SS), and correlation (C) were computed using the equations provided in Table 5-4.

Date	F (ppb)	O (ppb)	F_{Pers} (ppb)
1-Jul	80	80	70
2-Jul	90	100	80
3-Jul	70	80	100
4-Jul	80	80	80
5-Jul	90	90	80
6-Jul	120	90	90
7-Jul	120	120	90
8-Jul	100	110	120
9-Jul	130	100	110
10-Jul	100	100	100
11-Jul	60	60	100

A = 8 ppb
A_{Pers} = 14 ppb
B = 3 ppb
SS = 40%
C = 0.77

For the 11-day period, the forecaster's accuracy was 8 ppb, meaning that, on average, the forecasts were within ±8 ppb of the observed maximum. A slight positive bias of 3 ppb indicates that the forecasts are slightly higher than the observed values. The skill score is 40%, which indicates that the forecasts are 40% better than the reference forecast using the Persistence method. The last statistic, correlation, had a value of 0.77 indicating that most day-to-day changes in the forecasts are also reflected in the observations.

More detailed information on forecast verification can be found in Murphy (1991; 1993), Murphy and Winkler (1987), and Wilks (1995).

5.6.3   Verification Statistics for Category Forecasts

This section describes verification statistics that can be computed for category forecasts and provides several examples of how to interpret the verification statistics. Verification methods differ slightly based on the number of forecast categories. At the simplest level, a forecast can be issued for two categories (e.g., forecast an Ozone Action Day to occur or not to occur). A two-category forecast is discussed below. Three- or more-category forecasts are discussed later in this section.

Two-category forecast

Creating a contingency table (also called a frequency table) is the first step in evaluating a category forecast. **Figure 5-12** shows a contingency table of forecasted and observed events. This table is the basis for calculating all verification statistics for category forecasts. It is constructed by counting the frequency of occurrence of each event and assigning it to the appropriate cell.

Figure 5-12.   Contingency table for a two-category forecast.

Using the contingency table shown in Figure 5-12, a perfect forecasting program would have values in cells "a" and "d" indicating correct "hits". While both "a" and "d" represent correct forecasts, events forecasted in cell "a" are generally more frequent and easier to predict than those in cell "d". In the real world, imperfect forecasts result in values in cells "b" (false alarms) and "c" (misses). The verification statistics listed in **Table 5-6** are used to evaluate the quality of two-event categorical forecasts. The statistics include:

Accuracy	Percent of forecasts that correctly predicted the event or non-event.
Bias	Indicates, on average, if the forecasts are underpredicted (false negatives) or overpredicted (false positives).
False alarm rate	Percent of times a forecast of high pollution did not actually occur.
Critical success index	How well the high-pollution events were predicted; it is unaffected by a large number of correctly forecasted, low-pollution events.
Probability of detection	Ability to predict high-pollution events.
Skill score	Percentage improvement of a forecast with respect to a reference forecast, typically a climatology or persistence forecast.

Statistic name	What it measures	How to compute it	Equation	Units	How to interpret
Accuracy (A) (percent correct)	Percent of forecasts that correctly predicted the event or non-event.	Divide the number of "hits" (cells a plus d) by the total number of forecasts issued.	$A=(a+d)/N *100$	%	• Higher numbers are better. • For example, 65 means that 65% of the forecasts were correct in predicting ozone or $PM_{2.5}$ above or below a given threshold and 35% of the forecasts missed.
Bias (B)	Indicates, on average, if the forecasts are underpredicted (false negatives) or overpredicted (false positives).	Divide the number of forecasted high ozone events (cells b plus d) by observed high ozone events (cells c plus d).	$B = \dfrac{b + d}{c + d}$	-	• Values closer to 1 are best. • Values <1 indicate underpredicting (i.e., the event occurred more often than it was forecasted). • Values >1 indicate overpredicting.
False Alarm Rate (FAR)	The percent of times a forecast of high pollution did not actually occur.	Divide the high ozone forecasts that were missed (cell b) by the total number of high ozone forecasts (cells b plus d).	$FAR=(b)/(b+d) *100$	%	• Smaller values are best. • 0=no false alarms (perfect forecast of high events). • 50 means that half of the forecasts for high events did not materialize.
Critical Success Index (CSI), also called Threat Score	How well the high-pollution events were predicted. Useful for evaluating rarer events like high-pollution days. It is not affected by a large number of correctly forecasted, low-pollution events.	Divide the number of high-pollution "hits" (cell d) by the total number of forecasts plus the number of misses (cells b, c, and d).	$CSI = d/(b+c+d)*100$	%	• Higher numbers are best. • For example, 66% indicates that two-thirds of the forecasts, or actual exceedances, were correctly predicted.
Probability of Detection (POD)	Ability to predict high-pollution events (i.e., the percentage of forecasted high-pollution events that actually occurred).	Divide the correct forecasts of high pollution (cell d) by the total number of observed high-pollution events (cells c plus d).	$POD=d/(c+d)*100$	%	• Higher numbers are best. • For example, 70% indicates that, of the 10 high-pollution events, 7 were correctly predicted.
Skill Score (SS)	Percentage improvement of a forecast with respect to a reference forecast, typically a climatology or persistence forecast.	1. Compute accuracy for a reference forecast (A_{ref}), such as climatology or persistence (see Section 4.1 for details). 2. Compute accuracy (A) for the forecast. 3. Divide accuracy by the reference accuracy and subtract from 1.	$SS = (A/A_{ref} - 1) *100$	%	• 0% indicates no improvement (or skill) over the reference forecast. • 50% indicates a significant improvement in skill.

The most important statistics for evaluating the success of a program are Accuracy, False Alarm Rate, and Critical Success Index.

To help understand these verification measures, statistics were computed for two hypothetical forecasting programs as shown in **Figure 5-13**. This example evaluates the forecast performance for a two-category forecast: a prediction of 8-hr ozone concentrations at or above 85 ppb and below 85 ppb. "Program LM" is typical of a large metropolitan area with many 8-hr exceedances, whereas "Program SC" represents a smaller city with few exceedances. Both programs were evaluated for a 180-day period.

Program LM

Forecasted

Observed		No	Yes
	No	130	8
	Yes	20	22

A	84
B	0.71
FAR	27
CSI	44
POD	52
SS	78
A_{ref}	50

Program SC

Forecasted

Observed		No	Yes
	No	160	11
	Yes	6	3

A	91
B	1.56
FAR	79
CSI	15
POD	33
SS	87
A_{ref}	50

Figure 5-13. Hypothetical verification statistics for a two-category forecast for Program LM that has many ozone exceedances and Program SC with fewer exceedances.

The accuracy (A) of Program SC is slightly higher than that for Program LM mostly due to correctly predicting the non-event (i.e., below 85 ppb), which can skew accuracy to be higher in areas with few high ozone days. Accuracy by itself does not fully describe the performance differences between the two programs. The second statistic, bias (B), measures the tendency to underpredict an event (false-negative) or overpredict an event (false-positive). The two programs exhibit opposite values of bias with Program LM forecasting over twice as many false-negatives as false-positives and Program SC forecasting nearly twice as many false-positives as false-negatives.

False alarm rates (FAR) for the two programs are significantly different; Program SC has nearly three times the FAR of Program LM, 79% versus 27%, respectively. This means that 79% of the high ozone forecasts from Program SC missed. "Crying wolf" almost eight out of ten times may decrease credibility of the ozone-forecasting program. Program LM's FAR is typical of many ozone forecasting programs. The critical success index (CSI) measures the forecaster's

ability to predict the high ozone events, while excluding the large occurrence of correctly forecasted low ozone days. Program SC has a very low CSI of 15%, meaning that only 15% of the high ozone events were forecasted correctly even though the accuracy was higher for Program SC. Program LM does a much better job of predicting the high ozone events and has a CSI of 44%. Program SC also has a lower probability of detection (POD) than Program LM, meaning that forecasters in Program SC have a difficult time predicting the high ozone event when it actually does happen.

The last of the statistics is the skill score (SS), which measures the forecaster's performance relative to mere chance or another reference method. In the example, the accuracy, for the reference method (i.e., random chance) is 50%. As with accuracy, Program SC has the higher skill score; it represents an 87% improvement over chance. But again, Program SC's results are skewed by a few high ozone events and the frequent (easier to forecast) low ozone events. Overall, Program SC has a higher accuracy and skill score due to the high number of correctly forecast low ozone events. Yet, Program LM does a much better job at predicting the high ozone events as measured by the FAR, CSI, and POD statistics.

A contingency table can be constructed with random forecasts and can be used as a reference forecast in the skill score computation. **Figure 5-14** shows the expected number of events for each cell. The marginal totals of both the forecasts and the observations are the same as the marginal totals in the contingency table shown in Figure 5-12 (Stephenson, 2000). Contingency tables and statistics were computed for Program LM and its random forecast as shown in **Figure 5-15**.

Figure 5-14.   Contingency table for random forecast.

The accuracy, critical success index (CSI), and probability of detection (POD) are all higher for Program LM than those of its random forecast's. This is reflected in the skill score which represents a 25% improvement over the random forecast.

Program LM
Forecasted

Observed

	No	Yes
No	130	8
Yes	20	22

A	84
B	0.71
FAR	27
CSI	44
POD	52
SS	25
A_{ref}	68

Random
Forecasted

Observed

	No	Yes
No	115	23
Yes	35	7

A	68
B	0.71
FAR	77
CSI	11
POD	17

Figure 5-15.   Hypothetical verification statistics for a two-category forecast for Program LM and its random forecast.

Three- or more-category forecasts

Categorical forecasts can have three or more categories. In this case, computing the verification statistics becomes more complicated. For the four-category table shown in **Figure 5-16**, the accuracy (A) of the entire forecasting program is computed using the following equation:

$$A = [(k+p+u+z)/N]*100 \qquad (5\text{-}2)$$

where:
    N = total number of events in the table

Computing the verification statistics listed in Table 5-6 first involves collapsing a four-category table to a two-category table. To collapse the table, complete the following steps:

1. Pick an ozone or $PM_{2.5}$ concentration that separates two categories (such as 85 ppb, which separates the Good and Moderate categories from the Unhealthy for Sensitive Groups and Unhealthy categories in Figure 5-16). The ability to forecast above or below this value will be evaluated, not the performance of each category in the table.

2. Each cell of the four-category table must be assigned to one of the four cells (a, b, c, or d) of the two-category table shown in Figure 5-12. Each cell should be assigned according to the following criteria for the four possible scenarios:

    Cell a = event not forecasted and not observed.

    Cell b = event forecasted but not observed.

Figure 5-16.   Contingency table for a four-category forecast.

Cell c = event not forecasted but observed.

Cell d = event forecasted and observed.

3. Once the assignments are made, total all of the values corresponding to each letter and place them in the respective cell of the two-category table.

4. Calculate the verification statistics described in Table 5-6.

These forecast evaluation measures enable a forecaster to objectively quantify how well pollution is forecasted. They should become a regular part of any forecasting program.

5.6.4   Methods to Further Evaluate Forecast Performance

Statistics provide useful values but do not indicate the reasons why forecast performance may be poor. The following questions can help determine the cause of performance problems.

1. Are there any modifications in the monitoring network that could possibly skew the observations, for example, outage problems and new or decommissioned monitoring sites? Have the emissions patterns changed around the monitoring site (e.g., is a rural site now surrounded by buildings and roadways?) that might make the data collected at that site unrepresentative?

2. Are there any systematic forecast biases that are not evident in annual summary statistics, but are shown by evaluating the forecast performance on a finer time scale? For example, **Figure 5-17** shows the forecast bias each day and with a 2-week running average that reveals a positive bias starting around July 1 that persists throughout the season.

3. Are there any personnel changes in the forecasting team? Should there be more cross-training to transfer knowledge and experiences to the new forecasters?

4. Did the forecasters interpret meteorological information incorrectly? Were all forecast and observation data available to make an accurate forecast?

5. How reliable is the forecast guidance, including meteorological observations, forecast model output, and other meteorological forecast products? How accurate were the key weather forecasts (e.g., maximum temperature, wind speed, cloud cover, etc.) used to formulate the air quality forecast?
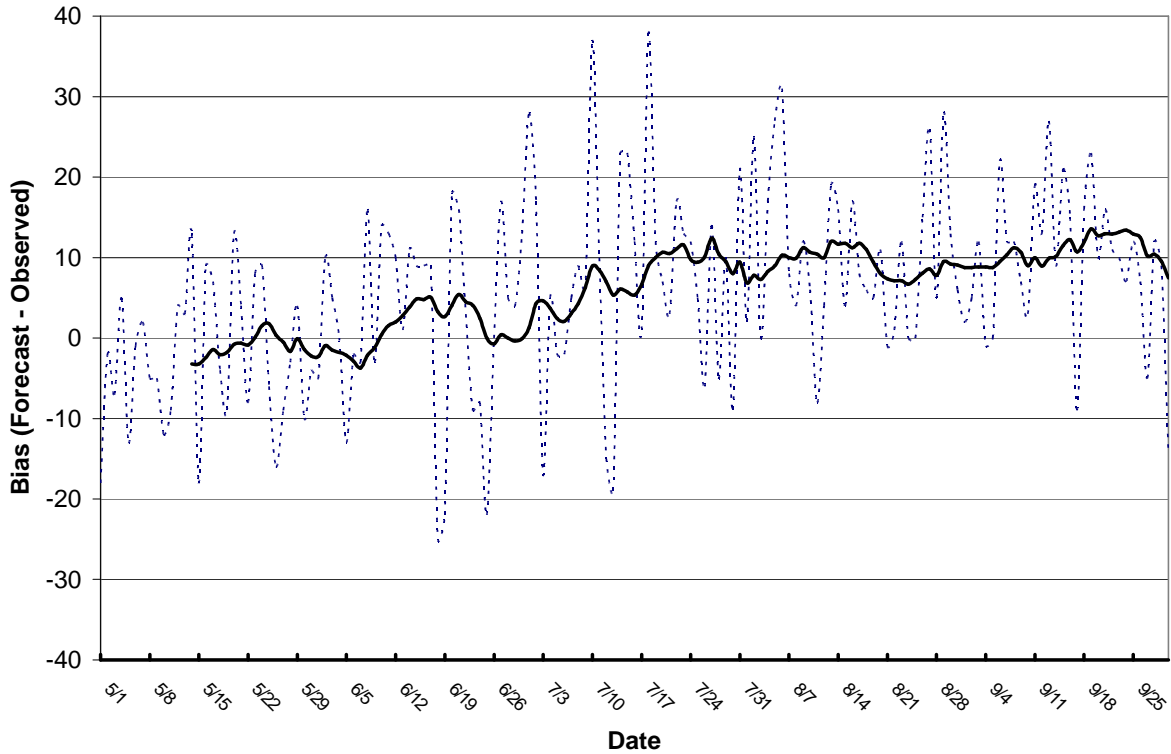
Figure 5-17. An example of forecast bias for a 24-hr ozone forecast. The dashed line shows the bias each day, and the solid line shows the 2-week moving average of the bias.

6. REFERENCES

Allen G., Sioutas C., Koutrakis P., Reiss R., Lurmann F.W., Roberts P.T., and Burton R.M. (1997) Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas. *J. Air & Waste Manag. Assoc.* **47**, pp. 682-689.

Altshuller A.P. and Lefohn A.S. (1996) Background ozone in the planetary boundary layer over the United States. *J. Air & Waste Manag. Assoc.* **46**, pp. 134-141.

Bruckman L. (1993) Overview of the Enhanced Geocoded Emissions Modeling and Projection (Enhanced GEMAP) System. In *proceeding of the Air & Waste Management Association's Regional Photochemical Measurements and Modeling Studies Conference*, p 562, San Diego, CA.

Cassmassi J.C. (1987) Development of an objective ozone forecast model for the South Coast Air Basin. Presented at the *Air Pollution Control Association 80th Annual Meeting, New York, NY, June 21-26*.

Chang J.S., Jin S., Li Y., Beauharnois M., Chang K.H., Huang H.C., Lu C.H., and Wojcik G. (1996) The SARMAP Air Quality Model, SARMAP Final Report Part 1.

Chinkin L.R., Main H.H., Anderson C.B., Coe D.L., Haste T.L., Hurwitt S.B., and Kumar N. (1998) Study of air quality conditions including ozone formation, emission inventory evaluation, and mitigation measures for Crittenden County, Arkansas. Report prepared for the Arkansas Department of Pollution Control and Ecology, Little Rock, AR by Sonoma Technology, Inc., Petaluma, CA, STI-998310-1837-DFR, November.

Claiborn C.S., Finn D., Larson T.V., and Koeing J.Q. (2000) Windblown dust contributes to high $PM_{2.5}$ concentrations. *J. Air & Waste Manag. Assoc.* **50**, pp. 1440-1455.

Coats C.J. (1996) High performance algorithms in the sparse matrix operator kernel emissions modeling system. Proceedings of the *Ninth Joint Conference on Applications of Air Pollution Meteorology of the American Meteorological Society and the Air and Waste Management Association, Atlanta, GA*.

Comrie A.C. (1997) Comparing neural networks and regression models for ozone forecasting. *J. Air & Waste Manag. Assoc.* **47**, pp. 653-663.

Comrie A.C. and Yarnal B. (1992) Relationships between synoptic-scale atmospheric circulation and ozone concentrations in metropolitan Pittsburgh, Pennsylvania. *Atmos. Environ.* **26B**, pp. 301-312.

Cooperative Program for Operational Meteorology, Education and Training (COMET) (2002). Available on the Internet at http://www.comet.ucar.edu.

Dockery D.W. and Pope C.A., III (1994) Acute respiratory effects of particulate air pollution. *Annu Rev Public Health* **15**, pp. 107-132.

Draxler R.R. and Hess G.D. (1997) Description of the Hysplit 4 modeling system. Technical memorandum by NOAA, ERL ARL-224, December 24.

Dye T.S., MacDonald C.P., and Miller D.S. (2002a) Summary of $PM_{2.5}$ forecasting tool development for Pittsburgh, Pennsylvania. Technical memorandum prepared for the

Pennsylvania Department of Environmental Protection and Allegheny County Health Department, Pittsburgh, PA, by Sonoma Technology, Inc., Petaluma, CA, STI-901491-2259-TM, September.

Dye T.S., Miller D.S., and MacDonald C.P. (2002b) Summary of $PM_{2.5}$ forecasting program development and operations for Salt Lake City, Utah during winter 2002. Technical memorandum prepared for Ms. Cheryl Heying, Utah Department of Environmental Quality by Sonoma Technology, Inc., Petaluma, CA 94954, 901491-2243-TM, August.

Dye T.S., Ray S.E., Lindsey C.G., Arthur M., and Chinkin L.R. (1996) Summary of ozone forecasting and equation development for the air districts of Sacramento, Yolo-Solano, and Placer. Vol. I: ozone forecasting. Vol. II: equation development. Final report prepared for Sacramento Metropolitan Air Quality Management District, Sacramento, CA by Sonoma Technology, Inc., Santa Rosa, CA, STI-996210-1701-FR, December.

Environ (1998) User's Guide - Comprehensive Air Quality Model with Extensions (CAMx). Version 2.0. By Environ International Corporation, Novato, CA, December.

Falke S. (1999) Draft $PM_{2.5}$ topic summary. Available on the Internet at http://capita.wustl.edu/PMFine/Workgroup/Status&Trends/Reports/In-progress/PM25Maps/PM25Maps/sld001.htm.

Falke S.R., Husar R.B., and Schichtel B.A. (2001) Fusion of SeaWiFS and TOMS satellite data with surface observations and topographic data during extreme aerosol events. *J. Air & Waste Manag. Assoc.* **51**, pp. 1579-1585.

Gardner M.W. and Dorling S.R. (1998) Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmos. Environ.* **32**, pp. 2627-2636.

Grell G.A., Dudhia J., and Stauffer D.R. (1994) A description of the fifth-generation Penn State/NCAR mesocale model (MM5). Prepared by National Center for Atmospheric Research, Boulder, CO, NCAR Technical Note-398.

Health Effects Institute (2002) Prespectives: understanding the health effects of components of the particulate matter mix: progress and next steps. Available on the Internet at www.healtheffects.org.

Hering S. and Cass G. (1999) The magnitude of bias in the measurement of $PM_{2.5}$ arising from volatilization of particulate nitrate from Teflon filters. *J. Air & Waste Manag. Assoc.* **50**, pp. 613-632.

Horie Y. (1988) Air Quality Management Plan 1988 Revision, Appendix V-P: Ozone episode representativeness study for the South Coast Air Basin. Report prepared for the South Coast Air Quality Management District, El Monte, CA by Valley Research Corporation, Van Nuys, CA, VRC Project Number 057, March.

Hubbard M.C. and Cobourn W.G. (1997) Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. *Atmos. Environ.* (submitted).

Husar R.B. (1998) Spatial pattern of 1-hour and 8-hour daily maximum ozone over the OTAG region. Presented at the *Air & Waste Management Association's 91st Annual Meeting & Exhibition, San Diego, CA, June 14-18.*

Jorquera M.E. (1998) The use of episodic controls to reduce the frequency and severity of air pollution events. Submitted to the Transportation Research Board, Transportation and Air Quality Committee, Federal Highway Administration, Baltimore, MD, February.

Lambeth B. (1998) Texas Natural Resource Conservation Commission, Austin, TX. Personal communication.

Lurmann F.W. (2000) Simplification of the UAMAERO Model for seasonal and annual modeling: the UAMAERO-LT Model. Report prepared for South Coast Air Quality Management District, Diamond Bar, CA by Sonoma Technology, Inc., Petaluma, CA, STI-999420-1996-FR, August.

Lurmann F.W., Wexler A.S., Pandis S., Musarra S., Kumar N., Seinfeld J.H., and Hering S.V. (1997) Development of an acid deposition model (UAM-AERO) for the South Coast Air Basin. Final report prepared for California Air Resources Board, Sacramento, CA by California Institute of Technology, Pasadena, CA, Sonoma Technology, Inc., Santa Rosa, CA, and Aerosol Dynamics, Inc., Berkeley, CA, ARB Contract No. 92-311.

MacDonald C.P., Roberts P.T., Main H.H., Kumar N., Haste T.L., Chinkin L.R., and Lurmann F.W. (1998) Analysis of meteorological and air quality data for North Carolina in support of modeling. Report prepared for North Carolina Department of Environment and Natural Resources, Division of Air Quality, Raleigh, NC by Sonoma Technology, Inc., Petaluma, CA, STI-997420-1818-DFR, October.

Main H.H. and Roberts P.T. (2001) PM$_{2.5}$ Data Analysis Workbook. Draft workbook prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-900242-1988-DWB, February.

McHenry J., Seaman N.L., Coats C.J., Lario-Gibbs A., Vukovich J., Wheeler N., and Hayes E. (1999) Real-time nested mesoscale forecast of lower tropospheric ozone using a highly optimized coupled model numerical prediction system. Preprints in *AMS Symposium on Interdisciplinary Issues in Atmospheric Chemistry, Dallas, TX, January 10-15*, American Meteorological Society.

Morris R.E., Myers T.C., Carr E.L., Causley M.C., and Douglas S.G. (1990) User's Guide for the Urban Airshed Model. Volume II: User's Guide for the UAM (CB-IV) Modeling System. Prepared for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-450/4-90-007B.

Murphy A.H. (1991) Forecast verification: its complexity and dimensionality. *Mon. Wea. Rev.* **119**, pp. 1590-1601.

Murphy A.H. (1993) What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* **8**, pp. 281-293.

Murphy A.H. and Winkler R.L. (1987) A general framework for forecast verification. *Mon. Wea. Rev.* **115**, pp. 1330-1338.

National Research Council (1991) *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Academy of Sciences/National Research Council, National Academy Press, Washington, DC.

NESCAUM (1998) 8-hr and 1-hr ozone exceedances in the NESCAUM Region (1993-1997). Report prepared by the Northeast States for Coordinated Air Use Management, Boston, MA.

Odman T. and Ingram C.L. (1996) Multiscale Air Quality Simulation Platform (MAQSIP): source code documentation and validation. MCNC Technical Report, ENV-96TR002-v1.0.

Paul R.A., Biller W.F., and McCurdy T. (1987) National estimates of population exposure to ozone. Paper no. 87-42.7 presented at the *Air Pollution Control Association 80th Annual Meeting and Exhibition, Pittsburgh, PA*.

Pielke R.A., Cotton W.R., Walko R.L., Tremback C.J., Lyons W.A., Grasso L., Nicholls M.E., Moran M.D., Wesley D.A., Lee T.J., and Copeland J.H. (1992) A comprehensive meteorological modeling system - RAMS. *Meteor. Atmos. Phys.* **49**, pp. 69-91.

Prospero J.M. (1999) Long-range transport of mineral dust in the global atmosphere: Impact of African dust on the environment of the southeastern United States. *Proc. Natl. Acad. Sci.* **96**, pp. 3396-3404.

Ruiz-Suarez J.C., Mayora-Ibarra O.A., Torres-Jimenez J., and Ruiz-Suarez L.G. (1995) Short-term ozone forecasting by artificial neural networks. *Advances in Engineering Software* **23**, pp. 143-149.

Ryan W.F. (1994) Forecasting severe ozone episodes in the Baltimore metropolitan area. *Atmos. Environ.* **29**, pp. 2387-2398.

Ryan W.F., Dickerson R.R., Doddridge B.G., Morales R.M., and Piety C.A. (1998) Transport and meteorological regimes during high ozone episodes in the mid-Atlantic region: observations and regional modeling. Preprints of the *10 Joint Conference of the Applications of Air Pollution Meteorology with the Air and Waste Management Association, January 11-16, Phoenix, AZ*, pp. 168-172, American Meteorological Society, Boston, MA.

Saxton K., Chandler D., Stetler L., Lamb B., Claiborn C., and Lee B.H. (2000) Wind erosion and fugitive dust fluxes on agricultural lands in the Pacific northwest. *Transactions of the ASAE* **43**, pp. 623-630.

SCAQMD (1997) 1997 revision to the Air Quality Management Plan - Appendix II: current air quality. South Coast Air Quality Management District, Diamond Bar, CA.

Schauer J.J., Rogge W.F., Hildemann L.M., Mazurek M.A., Cass G.R., and Simoneit B.R.T. (1996) Source apportionment of airborne particulate matter using organic compounds as tracers. *Atmos. Environ.* **30**, pp. 3837-3855.

Schere K.L. and Demerjian K.L. (1984) User's guide for the photochemical box model (PBM). U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC 27711, EPA-600/8-84-022a.

Schwartz J. (1994) What are people dying of on high air pollution days? *Environ Res* **64** (19), pp. 26-35.

Seinfeld J.H. and Pandis S.N. (1998) *Atmospheric chemistry and physics: from air pollution to global change*, J. Wiley, New York.

Stein A.F., Lamb D., and Draxler R.R. (2000) Incorporation of detailed chemistry into a three-dimensional Lagrangian-Eulerian hybrid model: Application to regional tropospheric ozone. *Atmos. Environ.* **34**, pp. 4361-4372.

Stephenson D.B. (2000) Use of the "odds ratio" for diagnosing forecast skill. *Weather and Forecasting* **15** (2), pp. 221-232.

Stoeckenius T. (1990) Adjustment of ozone trends for meteorological variation. Presented at the *Air and Waste Management Association's Specialty Conference, Tropospheric Ozone and the Environment, Los Angeles, CA, March 19-22*.

Systems Applications International (1999) User's Guide to the Variable-Grid Urban Airshed Model (UAM-V). By Systems Applications International, Inc./ICF Consulting, 101 Lucas Valley Road, Suite 160, San Rafael, California 94903.

Taylor R. (1998) New York State Department of Environmental Conservation, Albany, NY. Personal communication.

Turpin B.J., Huntzicker J.J., and Hering S.V. (1994) Investigation of organic aerosol sampling artifacts in the Los Angeles basin. *Atmos. Environ.* **28**, pp. 3061-3071.

Turpin B.J., Huntzicker J.J., Larson S.M., and Cass G.R. (1991) Los Angeles summer midday particulate carbon-primary and secondary aerosol. *Environ. Sci. Technol.* **25**, pp. 1788-1793.

U.S. Environmental Protection Agency (1984) Quality assurance handbook for air pollution measurement systems: Volume II. Ambient air specific methods. Sections 2.1, 2.2, 2.6, and 2.9. Report prepared by the Environmental Monitoring Systems Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-600/4-77-027a, July.

U.S. Environmental Protection Agency (1986) Guideline on the identification and use of air quality data affected by exceptional events. EPA-450/4-86-007, July. Available on the Internet at www.epa.gov/ttn/amtic/files/ambient/criteria/reldocs/4-86-007.pdf.

U.S. Environmental Protection Agency (1992) User's guide for the urban airshed model. Volume IV: User's manual for the emissions preprocessor system 2.0. Part A: Core FORTRAN system. Report prepared by U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-450/4-90-007D(R), June.

U.S. Environmental Protection Agency (1996) Clearinghouse for inventories and emission factors. On U.S. Environmental Protection Agency electronic bulletin board.

U.S. Environmental Protection Agency (1997a) National air pollutant emission trends, 1990-1996. Report prepared by the Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/R-97-011, December.

U.S. Environmental Protection Agency (1997b) Survey and review of episodic control programs in the United States. EPA 420-R-97-003, September.

U.S. Environmental Protection Agency (1998) EPA third-generation air quality modeling system, Models-3, Volume 9B: user manual. Report prepared by the National Exposure

Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-600/R-98/069(a), June.

U.S. Environmental Protection Agency (2001) Data quality objectives (DQOs) and model development for relating federal reference method (FRM) and continuous $PM_{2.5}$ measurements to report an air quality index (AQI). Available on the Internet at www.epa.gov/ttn/amtic/files/ambient/monitorstrat/aqidqorept.pdf. EPA-454/R-01-002, February.

University of Illinois Urbana-Champaign, Department of Atmospheric Sciences (2002) Weather World 2010 educational web site. Available on the Internet at http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/home.rxml.

Wallace J.M. and Hobbs P.V. (1977) *Atmospheric Science*, Academic Press, New York.

Ward D.E. (1999) Smoke from wildland fires. *Health Guidelines for Vegetation Fire Events, Lima, Peru, 6-9 October 1998,* Background papers, WHO.

Wilks D.S. (1995) *Statistical methods in the atmospheric sciences*, Academic Press, San Diego, CA, p 467.

Wilson W.E. and Suh H.H. (1997) Fine particles and coarse particles: concentration relationships relevant to epidemiologic studies. *J. Air & Waste Manag. Assoc.* **47** (33), pp. 1238-1249.

Yamartino R.J., Scire J.S., Carmichael G.R., and Chang Y.S. (1992) The CALGRID mesoscale photochemical grid model - I. Model formulation. *Atmos. Environ.* **26A**, pp. 1493-1512.

This page is intentionally blank.

TECHNICAL REPORT DATA
(Please read Instructions on reverse before completing)

1. REPORT NO. EPA-456/R-03-002	2.	3. RECIPIENT'S ACCESSION NO.

4. TITLE AND SUBTITLE Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program	5. REPORT DATE June 2003
	6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S) Timothy S. Dye, Clinton P. MacDonald, Craig B. Anderson, Hilary R. Hafner Neil J. M. Wheeler, Alan C. Chan (Sonoma Technology, Inc.)	8. PERFORMING ORGANIZATION REPORT NO. STI-902461-2295-FR

9. PERFORMING ORGANIZATION NAME AND ADDRESS Information Transfer Group-AIRNow Program (E143-03) U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Research Triangle Park, NC 27711	10. PROGRAM ELEMENT NO. 27/53/E
	11. CONTRACT/GRANT NO. Contract No. GS-10F-0181K Task Order No. 1303

12. SPONSORING AGENCY NAME AND ADDRESS Office of Air Quality Planning and Standards Information Transfer and Program Integration Division U.S. Environmental Protection Agency Research Triangle Park, NC 27711	13. TYPE OF REPORT AND PERIOD COVERED Final
	14. SPONSORING AGENCY CODE EPA/200/04

15. SUPPLEMENTARY NOTES
Document is available electronically at "http://www.epa.gov/airnow/"

16. ABSTRACT

This report provides technical guidance to help air quality agencies develop, operate, and evaluate ozone and PM2.5 forecasting programs. This document provides:

1) Background information about ozone and PM2.5 and the weather's effect on these pollutants.
2) A list of how air quality forecasts are currently used.
3) A summary and evaluation of methods currently used to forecast ozone and PM2.5.
4) Steps to develop and operate an air quality forecasting program.
5) Information on the level of effort needed to set up and operate a forecasting program.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Ozone, Air Quality Forecasting, AQI PM2.5, Ambient Monitoring	Air Quality Data and Forecasting, AQI Reporting	

18. DISTRIBUTION STATEMENT Release Unlimited	19. SECURITY CLASS *(Report)* Unclassified	21. NO. OF PAGES 125
	20. SECURITY CLASS *(Page)* Unclassified	22. PRICE

EPA Form 2220-1 (Rev. 4-77)   PREVIOUS EDITION IS OBSOLETE