

United States  
Environmental Protection  
Agency

Office of Air Quality  
Planning and Standards  
Research Triangle Park, NC 27711

EPA-454/R-99-009  
July 1999

---

Air

---



# GUIDELINE FOR DEVELOPING AN OZONE FORECASTING PROGRAM



EPA-454/R-99-009  
July 1999

# **GUIDELINE FOR DEVELOPING AN OZONE FORECASTING PROGRAM**

**U.S. Environmental Protection Agency  
Office of Air Quality Planning and Standards  
Research Triangle Park, North Carolina 27711**

## **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

During the past several months, we have discussed issues concerning ozone forecasting techniques and how ozone forecasts are used with numerous individuals knowledgeable in these areas. We have spoken with colleagues at state and federal agencies and universities and in the private sector who either forecast ozone or use ozone forecasts. Their ideas and suggestions have been instrumental in producing these guidelines.

The authors wish to especially thank the following individuals for giving us their time and experience: Mr. Lee Alter, Mr. Rafael Ballagas, Mr. Mark Bishop, Mr. Chris Carlson, Mr. Joe Casmassi, Mr. Joe Chang, Mr. Aaron Childs, Dr. Geoffrey Cobourn, Dr. Andrew Comrie, Ms. Lillie Cox, Ms. Laura DeGuire, Ms. Beth Gorman, Ms. Sheila Holman, Mr. Michael Koerber, Mr. Larry Kolczak, Mr. Bryan Lambeth, Mr. Erich Linse, Ms. April Linton, Mr. Michael Majewski, Mr. Cliff Michaelson, Ms. Eve Pidgeon, Ms. Katherine Pruitt, Mr. Chris Roberie, Mr. Bill Ryan, Mr. Kerry Shearer, Mr. Till Stoeckenius, Ms. Susan Stone, Mr. Troy Stuckey, Mr. Bob Swinford, Mr. Richard Taylor, Mr. Brian Timan, Mr. Alan VanArsdale, Mr. Chet Wayland, Ms. Leah Weiss, Mr. Neil Wheeler, and Mr. Robert Wilson. We also wish to thank our colleagues at STI for their comments and contributions: Dr. Donald Blumenthal, Dr. Paul Roberts, Mr. Lyle Chinkin, Ms. Hilary Main, Mr. Fred Lurmann, and Mr. Joe Kwiatkowski.

## TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
ACKNOWLEDGMENTS .....	iii
LIST OF FIGURES .....	vii
LIST OF TABLES.....	ix
1. INTRODUCTION AND GUIDE TO DOCUMENT .....	1-1
1.1 INTRODUCTION.....	1-1
1.2 GUIDE TO THIS DOCUMENT.....	1-2
1.3 FORECASTING OTHER POLLUTANTS.....	1-2
2. PROCESSES AFFECTING OZONE CONCENTRATIONS.....	2-1
2.1 BASIC OZONE CHEMISTRY .....	2-1
2.2 OZONE PRECURSOR EMISSIONS.....	2-2
2.3 METEOROLOGICAL CONDITIONS THAT INFLUENCE OZONE LEVELS .....	2-6
2.4 RELATIONSHIP BETWEEN THE 1-HR AND 8-HR OZONE STANDARDS .....	2-10
3. FORECASTING APPLICATIONS AND NEEDS.....	3-1
3.1 PUBLIC HEALTH NOTIFICATION .....	3-1
3.2 EPISODIC CONTROL PROGRAMS.....	3-1
3.3 SPECIALIZED MONITORING PROGRAMS .....	3-2
4. DEVELOPING OZONE FORECASTING METHODS .....	4-1
4.1 FORECASTING METHODS.....	4-1
4.1.1 Persistence.....	4-1
4.1.2 Climatology .....	4-6
4.1.3 Criteria.....	4-11
4.1.4 Classification and Regression Tree (CART) .....	4-14
4.1.5 Regression Equations.....	4-18
4.1.6 Artificial Neural Networks.....	4-21
4.1.7 Three-dimensional (3-D) Air Quality Models.....	4-24
4.1.8 The Phenomenological/Intuition Method .....	4-27
4.2 SELECTING PREDICTOR VARIABLES .....	4-29
5. STEPS FOR DEVELOPING AN OZONE FORECASTING PROGRAM .....	5-1
5.1 UNDERSTANDING FORECAST USERS' NEEDS.....	5-1
5.2 UNDERSTANDING THE PROCESSES THAT CONTROL OZONE.....	5-2
5.2.1 Literature Reviews .....	5-3
5.2.2 Data Analyses.....	5-3

## TABLE OF CONTENTS (Concluded)

<b><u>Section</u></b>		<b><u>Page</u></b>
5.3	CHOOSING OZONE FORECASTING METHODS.....	5-14
5.4	DATA TYPES, SOURCES, AND ISSUES .....	5-15
5.5	FORECASTING PROTOCOL .....	5-19
5.6	FORECAST VERIFICATION .....	5-20
	5.6.1 Forecast Verification Schedule .....	5-21
	5.6.2 Verification Statistics for Discrete Forecasts.....	5-22
	5.6.3 Verification Statistics for Category Forecasts.....	5-24
6.	REFERENCES.....	6-1

## LIST OF FIGURES

<b><u>Figure</u></b>	<b><u>Page</u></b>
2-1. Average diurnal profile of ozone, NO, and VOC concentrations for August 1995 in Lynn, Massachusetts (an urban site).....	2-2
2-2. 1996 Volatile Organic Compounds (VOC) emissions from anthropogenic sources by county .....	2-5
2-3. 1996 Nitrogen Oxide (NO) emissions from anthropogenic sources by county.....	2-5
2-4. 1996 Volatile Organic Compounds (VOC) emissions from biogenic sources by county.....	2-6
2-5. 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.....	2-9
2-6. Scatter plot showing the relationship between 1-hr and 8-hr daily maximum ozone concentrations for a site in Hancock County, Indiana.....	2-11
4-1. Scatter plot of maximum surface temperature and regional maximum 8-hr ozone concentration in Charlotte, North Carolina.....	4-13
4-2. Decision tree for daily basin maximum ozone concentrations in the South Coast Air Basin in the Los Angeles, California area.....	4-16
4-3. A schematic of an artificial neural network.....	4-22
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.....	5-5
5-2. Distribution of 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.....	5-6
5-3. Average annual frequency of episode length for the 8-hr and 1-hr standards in the New Jersey and New York City region from 1993-1997.....	5-7
5-4. Distribution of the average number of 8-hr and 1-hr exceedance by day of week for the New Jersey and New York City region.....	5-8
5-5. A surface synoptic pattern associated with high ozone in Pittsburgh, Pennsylvania .....	5-10

## LIST OF FIGURES (Concluded)

<b><u>Figure</u></b>		<b><u>Page</u></b>
5-6.	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.....	5-11
5-7.	Back trajectories at 1500 m msl during ozone episodes in Baltimore, Maryland showing possible transport of pollutants from regions to the west .....	5-12
5-8.	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.....	5-13
5-9.	Example outline of a forecast retrospective .....	5-21
5-10.	Contingency table for a two-category forecast .....	5-25
5-11.	Hypothetical verification statistics for a two-category forecast for Program LM that has many ozone exceedances and Program SC with fewer exceedances.....	5-27
5-12.	Contingency table for a four-category forecast .....	5-29



## LIST OF TABLES

<b><u>Table</u></b>	<b><u>Page</u></b>
2-1. Summary of total anthropogenic VOC and NO <sub>x</sub> emissions in the United States during 1994 .....	2-4
4-1. Comparison of forecasting methods.....	4-2
4-2. Peak 8-hr ozone concentrations for a sample city for 30 consecutive days.....	4-4
4-3. Annual summaries of 1-hr ozone exceedance days for New York State (1983-1997).....	4-7
4-4. Historical maximum ozone concentrations for three air districts in the Sacramento, California region (1990-1995).....	4-9
4-5. Duration of high ozone episodes for three air districts in the Sacramento, California region (1990-1995). .....	4-9
4-6. Information on health advisory days ( $\geq 150$ ppb) from 1990 through 1995 for three air districts in the Sacramento, California region.....	4-9
4-7. Average number of days with high ozone for three air districts in the Sacramento, California region (1990-1995).....	4-10
4-8. Distribution of high ozone concentrations by day of week for three air districts in the Sacramento, California region (1990-1995).....	4-10
4-9. Criteria for 1-hr ozone exceedances in Austin, Texas used by the Texas Natural Resource Conservation Commission.....	4-11
4-10. Common predictor variables used to forecast ozone .....	4-30
5-1. Data products for developing forecasting methods and for forecasting weather and ozone.....	5-16
5-2. Major data sources for air quality and meteorological data.....	5-17
5-3. Example of a forecasting protocol schedule.....	5-20

## LIST OF TABLES (Concluded)

<b><u>Table</u></b>		<b><u>Page</u></b>
5-4.	Verification statistics computed on discrete concentration forecasts .....	5-23
5-5.	Hypothetical forecasts for an 11-day period showing a human forecast (F), observed values (O), and forecasts using the Persistence method ( $F_{\text{pers}}$ ).....	5-24
5-6.	Verification statistics used to evaluate two-category forecasts.....	5-26

# 1. INTRODUCTION AND GUIDE TO DOCUMENT

## 1.1 INTRODUCTION

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 plant-life 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).

In light of the health effects of ground-level ozone, a few air quality agencies have been forecasting ozone concentrations for many years to warn the public of unhealthy air and to encourage people to voluntarily reduce emissions-producing activities. From 1978 to 1997, forecasts were based on the 1-hr National Ambient Air Quality Standard (NAAQS) for ozone, which was 0.12 parts per million (ppm). 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.<sup>1</sup> Under the revised standard, regions exceed the NAAQS when the three-year average of the annual fourth highest 8-hour average ozone concentrations is above 0.08 ppm. More regions will have daily maximum 8-hour ozone concentrations that exceed the level of the revised NAAQS than the old standard, and more agencies may need to forecast ozone to alert the public. The purpose of this document is to provide guidance to help air quality agencies develop, operate, and evaluate ozone forecasting programs. This guidance document provides:

- Background information about ozone and the weather's effect on ozone.
- A list of how ozone forecasts are currently used.
- A summary and evaluation of methods currently used to forecast ozone.
- Steps you can follow to develop and operate an ozone forecasting program.

The intended audience of this document is project managers, meteorologists, air quality analysts, and data analysts. Project managers can learn about the level of effort needed to set up and operate a forecasting program. Meteorologists can learn about the various methods to predict ozone and the steps needed to create a program.

The information presented in this document is based on literature reviews and on telephone interviews with ozone forecasters throughout the country.

## 1.2 GUIDE TO THIS DOCUMENT

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

---

<sup>1</sup> This revision was challenged in the U.S. Court of Appeals for the District of Columbia Circuit, and on May 14, 1999, the Court remanded it to the Agency for further consideration, principally in light of constitutional concerns regarding section 109 of the Act as interpreted by EPA. American Trucking Associations v. EPA, Nos. 97-1440, 97-1441 (D.C. Cir. May 14, 1999). On June 28, 1999, the EPA filed a petition for rehearing seeking review of the Court's decision by the entire Court of Appeals.

**Section 2: Processes Affecting Ozone Concentrations** describes the principal chemical and meteorological factors that produce ozone and its precursor emissions. It also describes how atmospheric phenomena affect ozone concentrations.

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

**Section 4: Developing Ozone Forecasting Methods** explains the different approaches used to forecast ozone. This section describes each method and compares its strengths and limitations, thus allowing you to select the methods that meet your agency's needs and resources.

**Section 5: Steps for Developing an Ozone Forecasting Program** identifies the steps that you can follow to develop, operate, and evaluate an ozone forecasting program.

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

### 1.3 FORECASTING OTHER POLLUTANTS

This guidance document is focused on forecasting ozone concentrations. However, the methods discussed in Section 4 and the procedures to setting up a program in Section 5 may be applied to other pollutants.

In order to accurately forecast other pollutants you must be knowledgeable about the atmospheric and chemical processes that affect pollutant formation, transport, and dispersion. Once you have a physical understanding of how these processes affect a particular pollutant, follow these major steps to develop methods for forecasting the pollutant:

1. Understand the nature of the pollutant by determining:
  - How it forms by identifying the physical and chemical processes that produce the pollutant.
  - When it forms by analyzing data to develop a climatological record when a particular pollutant's concentrations are high.
  - How weather affects the pollutant by understanding the key meteorological and air quality interactions that create and transport it.
2. Apply the forecasting methods described in Section 4.1 to the particular pollutant. The techniques described in that section are generally valid for all pollutants, however, the weather parameters (Section 4.2) used to predict pollutant levels may differ for each type of pollutant.
3. Follow the steps outlined in Sections 5.1 through 5.6 to develop a forecasting program for the new pollutant.

## 2. PROCESSES AFFECTING OZONE CONCENTRATIONS

Ozone concentrations are strongly affected by weather. Developing a basic understanding of how ozone forms and where emissions originate will help you forecast the effects of weather on ozone and emissions.

This section summarizes the basic chemical reactions that generate ozone in the troposphere (Section 2.1); describes the sources of precursor emissions that create ozone (Section 2.2); explains how weather affects ozone formation, transport, and dispersion (Section 2.3); and discusses relationships between the 1-hr and 8-hr ozone standards (Section 2.4). A discussion of how to develop a more detailed understanding of the processes that control ozone in a specific area is presented in Section 5.2.

### 2.1 BASIC OZONE CHEMISTRY

Understanding basic ozone chemistry is important because weather influences many aspects of ozone. Ozone (O<sub>3</sub>) 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<sub>x</sub>) 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<sub>x</sub> and hydrocarbons. Many species of vegetation including trees and plants emit hydrocarbons; and fertilized soils release NO<sub>x</sub>.

In the presence of ultraviolet radiation ( $h\nu$ ), oxygen (O<sub>2</sub>) and nitrogen dioxide (NO<sub>2</sub>) react in the atmosphere to form ozone and nitric oxide (NO) through the reactions given in Equations 2-1 and 2-2.



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 parts per billion (ppb) (Altshuller and Lefohn, 1996).



Ozone cannot accumulate further unless volatile organic compounds (VOCs), which include hydrocarbons, are present to consume or convert NO back to NO<sub>2</sub> as shown by

Equation 2-4. This equation is a simplified 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<sub>2</sub>, resulting in greater ozone formation. Additionally, anthropogenic sources of NO result in greater levels of NO<sub>2</sub> in the atmosphere. This NO<sub>2</sub> is then available for photolysis to NO and O (Equation 2-1) and, ultimately, for conversion to NO<sub>2</sub> (Equation 2-4) and ozone (Equation 2-2).



The formation and increase in ozone concentrations occurs 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.

## 2.2 OZONE PRECURSOR EMISSIONS

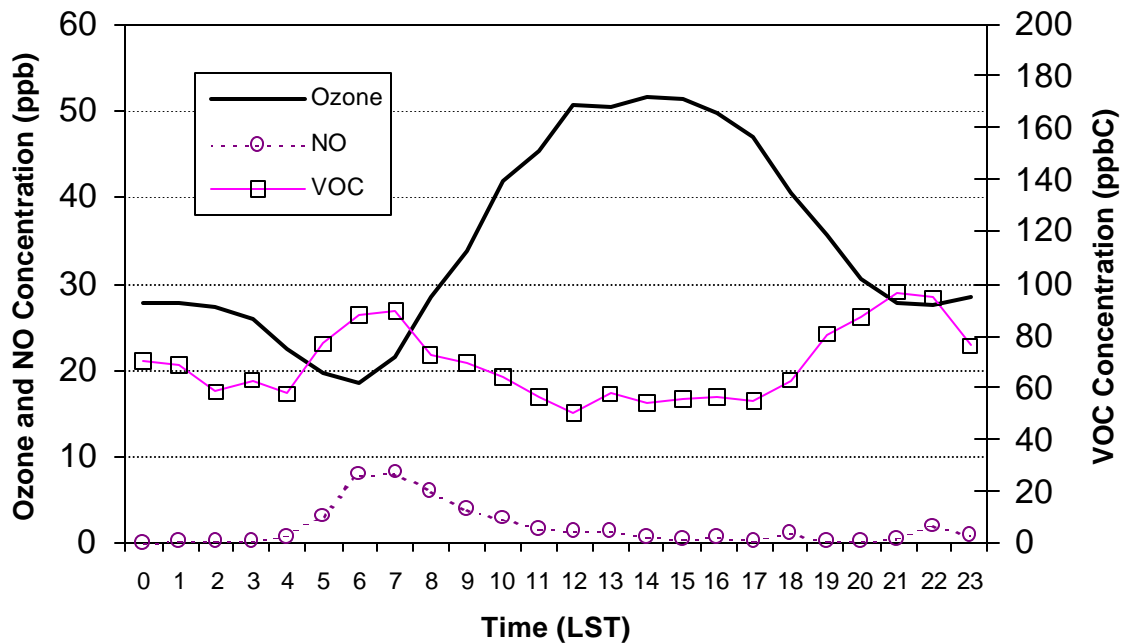


Figure 2-1. Average diurnal profile of ozone, NO, and VOC concentrations for August 1995 in Lynn, Massachusetts (an urban site).

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 you factor day-to-day emissions changes into your forecast. 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 compute patterns. This section provides a brief overview of the sources and spatial distribution of VOC and NO<sub>x</sub> (NO and NO<sub>2</sub>) emissions.

**Table 2-1** summarizes the total anthropogenic VOC and NO<sub>x</sub> emissions in the United States for 1994. The dominant NO<sub>x</sub> 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.

Anthropogenic VOC and NO<sub>x</sub> 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<sub>x</sub> emissions levels are typically much lower than anthropogenic NO<sub>x</sub> emissions levels.

Table 2-1. Summary of total anthropogenic VOC and NO<sub>x</sub> emissions in the United States during 1994 (U.S. Environmental Protection Agency, 1996). Note that 1 short ton equals 2000 pounds.

Source Type	NO <sub>x</sub> 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
<b>Total Emissions</b>	<b>23,666</b>	<b>—</b>	<b>23,175</b>	<b>—</b>



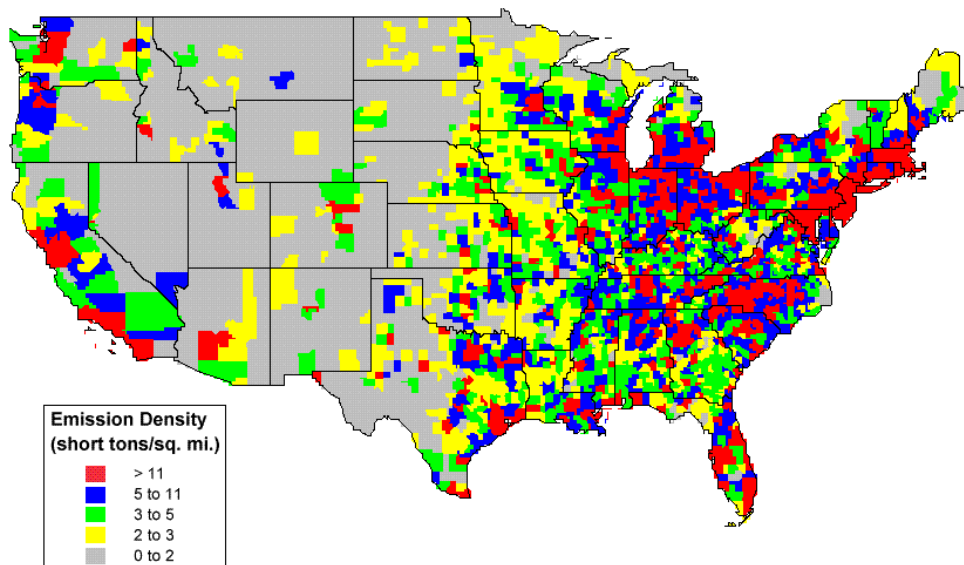
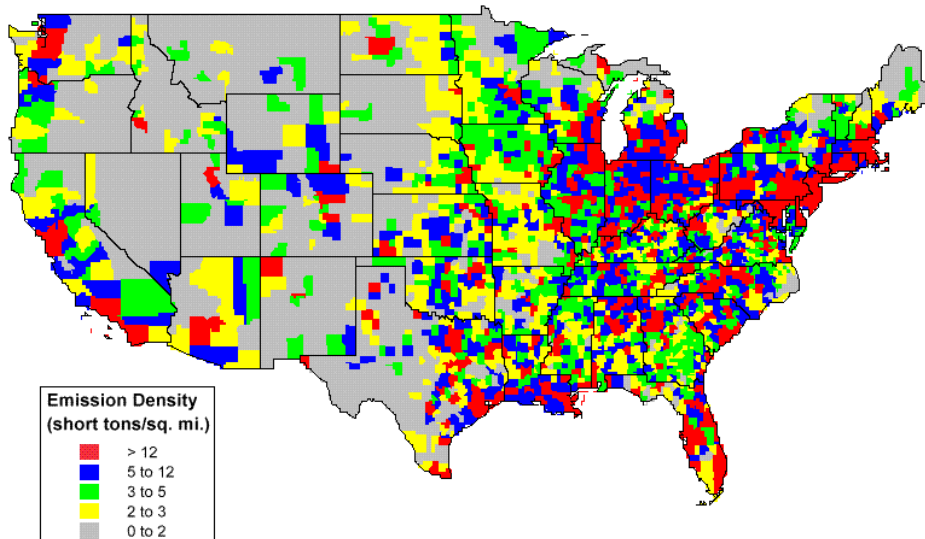


Figure 2-2. 1996 Volatile Organic Compounds (VOC) emissions from anthropogenic sources by county (U.S. Environmental Protection Agency, 1997a).

Figure 2-3. 1996 Nitrogen Oxide (NO) emissions from anthropogenic sources by county



(U.S. Environmental Protection Agency, 1997a).

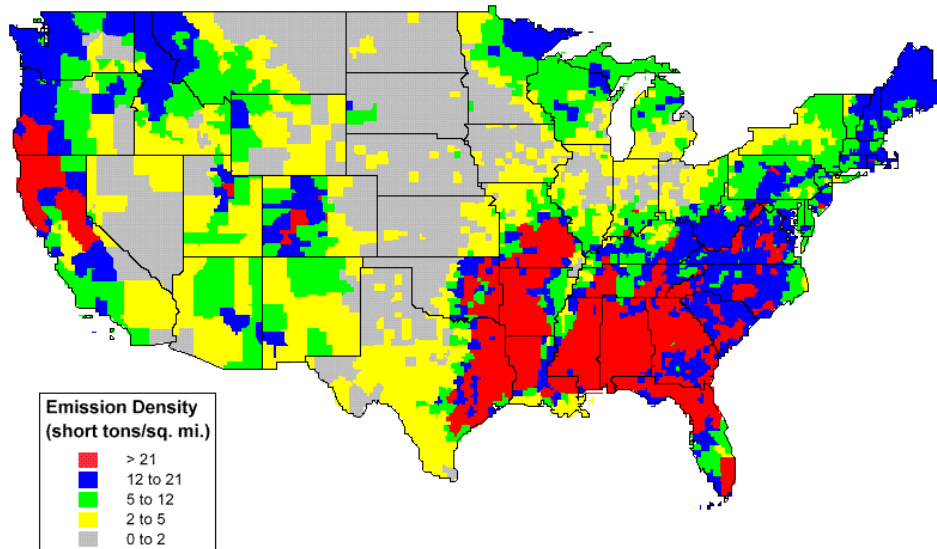


Figure 2-4. 1996 Volatile Organic Compounds (VOC) emissions from biogenic sources by county (U.S. Environmental Protection Agency, 1997a).

### 2.3 METEOROLOGICAL CONDITIONS THAT INFLUENCE OZONE LEVELS

In addition to the chemical and emission variables, a variety of meteorological variables also influence ozone concentrations. Although changes in daily emissions can affect daily ozone concentrations, it is the daily weather variations that best explain the day-to-day changes in ozone concentrations.

Understanding the types of weather that affect ozone is important for selecting variables to help predict ozone (Section 4.2) and for letting you relate forecasted weather parameters and patterns to future ozone concentrations. This section first examines the types of weather parameters that are important for controlling ozone concentrations. It then examines the types of synoptic weather patterns that produce conditions conducive for high ozone.

The types of weather parameters and how they influence ozone concentrations are as follows:

- *Sunlight*
  - *Temperature*
  - *Vertical temperature structure*
  - *Surface winds*
  - *Aloft winds*
- Ultraviolet radiation from clear to scattered skies is needed for ozone photochemistry. Clouds can also influence maximum temperatures.
  - Photochemical reaction rate increases as temperatures rise. In addition, temperature can affect emissions (e.g., evaporative emissions of VOC's increase with high temperatures). Biogenic emissions also increase in certain high temperature ranges. Demand for power may also increase during high temperatures.
  - Atmospheric lapse rate or stability (temperature change by height) controls the amount of vertical mixing that takes place. Strong stability tends to reduce mixing (i.e., reduce dilution) and confine emissions and ozone closer to the ground. This is important because, as discussed in Section 2.1, higher concentrations of precursors are needed to form higher ozone concentrations. In addition, aloft temperature inversions can act to trap pollutants below the inversion and inhibit vertical mixing.
  - Wind speeds control the degree of ventilation. Calm or light winds produce weak ventilation and allow more emissions to accumulate in a given volume of air, resulting in higher precursor concentrations.
  - Upper-air winds are important because they transport ozone and precursors into a region overnight and in the early morning hours or transport locally formed ozone out of a region during the afternoon. For example, low-level jets with winds of 10 to 20 m/s form throughout the United States shortly after sunset and remain through the night (Blackadar, 1957). Low-level jets are efficient at transporting ozone and its precursors several hundred kilometers during the night (Clark, 1997; Samson, 1978; Ray et al., 1998).

Synoptic meteorological patterns affect the mixing, ventilation, sunlight, and temperature in an area (Pagnotti, 1987; Chu, 1987; Comrie and Yarnal, 1992; Chu and Doll, 1991). **Figure 2-5** shows the typical life cycle of synoptic-scale weather patterns. The following meteorological descriptions are generic and may vary from one region to another:

- Ridge—high pressure pattern* (Figure 2-5a and b) is typically associated with the highest ozone concentrations. This pattern occurs about 1 to 2 days after a cold front and trough have passed through the area. As the surface high pressure develops in an area, winds become weak allowing for the accumulation of ozone and its precursor emissions. 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 ozone and ozone precursors. The aloft high pressure ridge typically occurs west of the surface high and can be diagnosed with 500-mb height fields.
- Ridge—back side of high pattern* (Figure 2-5c and d) occurs as the surface high pressure moves east of the region and the accumulated ozone can be 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 continues to produce warm temperatures and relatively clear skies even with a low-pressure system approaching from the west. Ozone levels can remain high on these types of days, and the potential for longer-range ozone and precursor transport is greater.
- Trough—cold front pattern* (Figure 2-5e 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 act to reduce ozone concentrations by diluting ozone and its precursors.

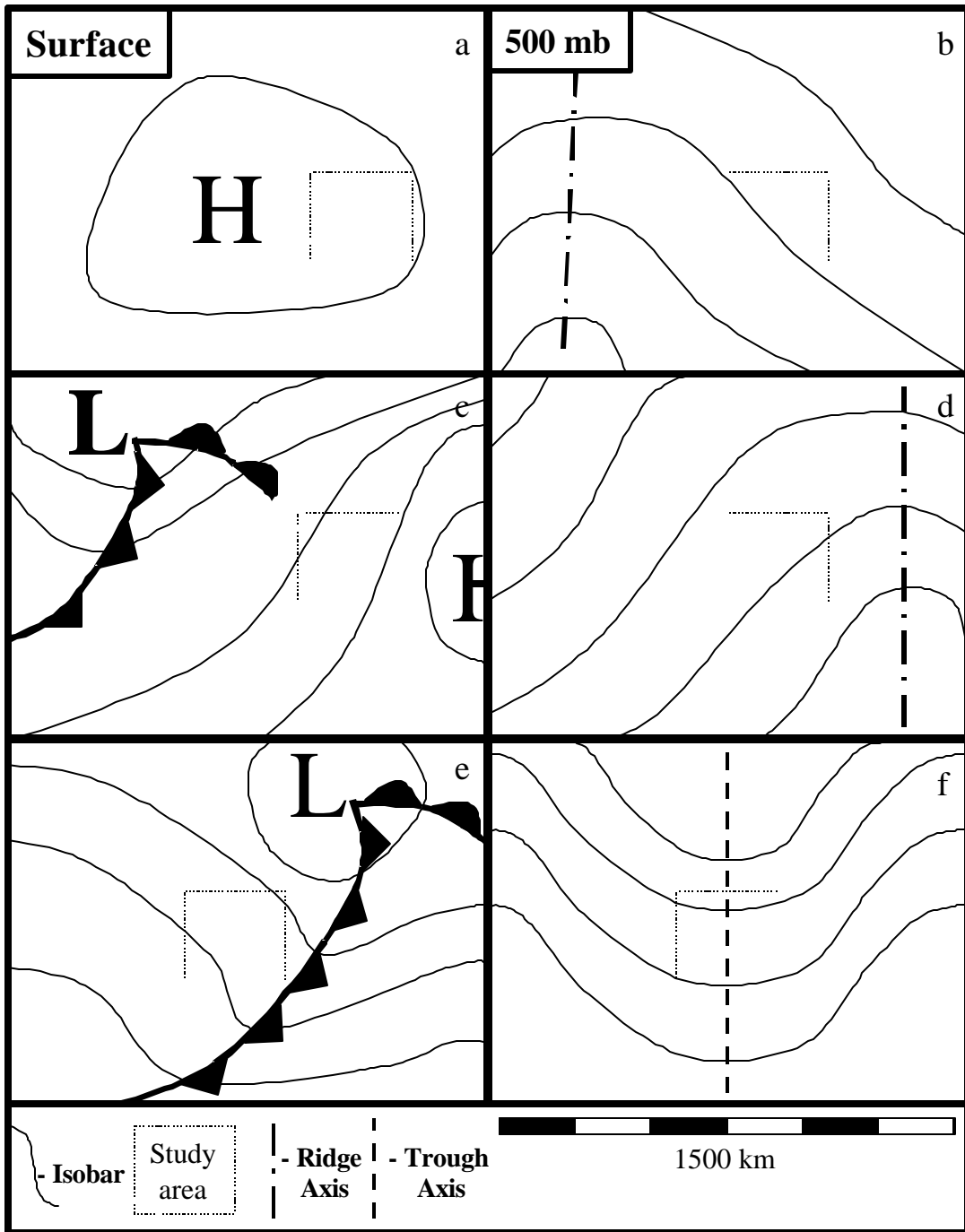


Figure 2-5. Life cycle of synoptic weather events at the surface and aloft at 500 mb for a and b) Ridge—high pressure, c and d) Ridge—back side of high, and e and f) Trough—cold front patterns. Surface maps show isobars and frontal positions. The 500-mb maps show contours of equal height.

The relative influences of emissions, mixing, ventilation, temperature, sunlight, and transport from upwind areas control the local ozone concentrations in a region. Forecasting the affects of meteorology on these influences is the key to forecasting ozone.

## 2.4 RELATIONSHIP BETWEEN THE 1-HR AND 8-HR OZONE STANDARDS

On July 18, 1997, the EPA promulgated a revision to the National Ambient Air Quality Standard (NAAQS) for ozone. Previously, the level of the standard was exceeded when the ozone concentration was greater than 0.12 ppm averaged over 1 hr. Under the revised NAAQS, the level of the standard is exceeded when the 8-hour average ozone concentration is above 0.08 ppm.

The 1-hr standard has been the basis for ozone-forecasting techniques developed over the past decade. Since agencies now need to forecast for 8-hour ozone concentrations, it is important to develop an understanding of how these two standards differ. The potential impacts of the 8-hour standard on ozone forecasting are as follows:

Increase in number of exceedances with the 8-hour standard. Agencies can expect to see a twofold to fourfold (or more) increase in the number of days with 8-hour ozone concentrations at or above 85 ppb (Hyde and Barnett, 1998; Dye et al., 1998; Husar, 1998). This increase in the number of days and the lengthening of the ozone season can be attributed to the lower threshold of 85 ppb for the 8-hour standard.

Broader range of weather conditions contributing to 8-hour exceedances. Since the threshold for 8-hour exceedances is lower, a wider range of weather conditions may produce exceedances of 85 ppb. For example, with the new 8-hour standard, exceedances might occur under clear to partly cloudy skies, whereas with the 1-hour standard, exceedances might have only occurred under ideal, clear sky conditions. Forecasters must now predict the broader range of weather conditions that produce 8-hour exceedances and not only the extreme conditions (hot temperatures, light winds, clear skies) that produce 1-hour exceedances. Thus, the difference in weather conditions between 8-hour exceedance and non-exceedance days will be slight and likely more difficult to forecast. It is important to conduct additional analyses to better understand the range of weather conditions that produce 8-hour exceedances in each region.

Larger number of regions affected by 8-hr exceedances. Due to the lower threshold of the 8-hour standard, more sites will experience exceedances. These new exceedance sites may differ from the traditional 1-hour peak sites and may peak on different days. In addition, regions that did not experience 1-hour exceedances may begin experiencing 8-hour exceedances. Forecasting the new 8-hour ozone exceedances over a broader region may mean that more local/regional weather conditions or terrain can influence ozone transport and dispersion.

8-hour and 1-hour ozone correlations. Many researchers have shown (Hyde and Barnett, 1998; Dye et al., 1998; Conroy, 1998) that the daily peak ozone concentrations for 1-hour and 8-hour standards are highly correlated. This means that you can convert predictions of 1-hour ozone concentrations using previously proven methods to 8-hour predictions. Use historical 1-hour and 8-hour data and statistical software to determine the correlation. Generally, correlations range from .75 to .98 for most of the monitoring sites in the country.

Several researchers have developed linear regression equations that use this high correlation to convert a forecasted 1-hour ozone concentration into a forecasted 8-hour concentration. Equation 2-5 shows an example of such a method.

$$\text{Forecasted 8-hr ozone} = \text{Slope} * (\text{Forecasted 1-hr ozone}) + \text{Constant} \quad (2-5)$$

**Figure 2-6** shows a scatter plot of daily 1-hour and 8-hour maximum ozone concentrations for a three-year period (total of 539 days) for a site in Hancock County, Indiana. In this example, the correlation between the 1-hour and 8-hour concentrations is 0.95. Therefore, using Equation 2-5 and the slope and constant from Figure 2-6, a 1-hour forecast of 130 ppb would be converted to an 8-hour forecast of 115 ppb.

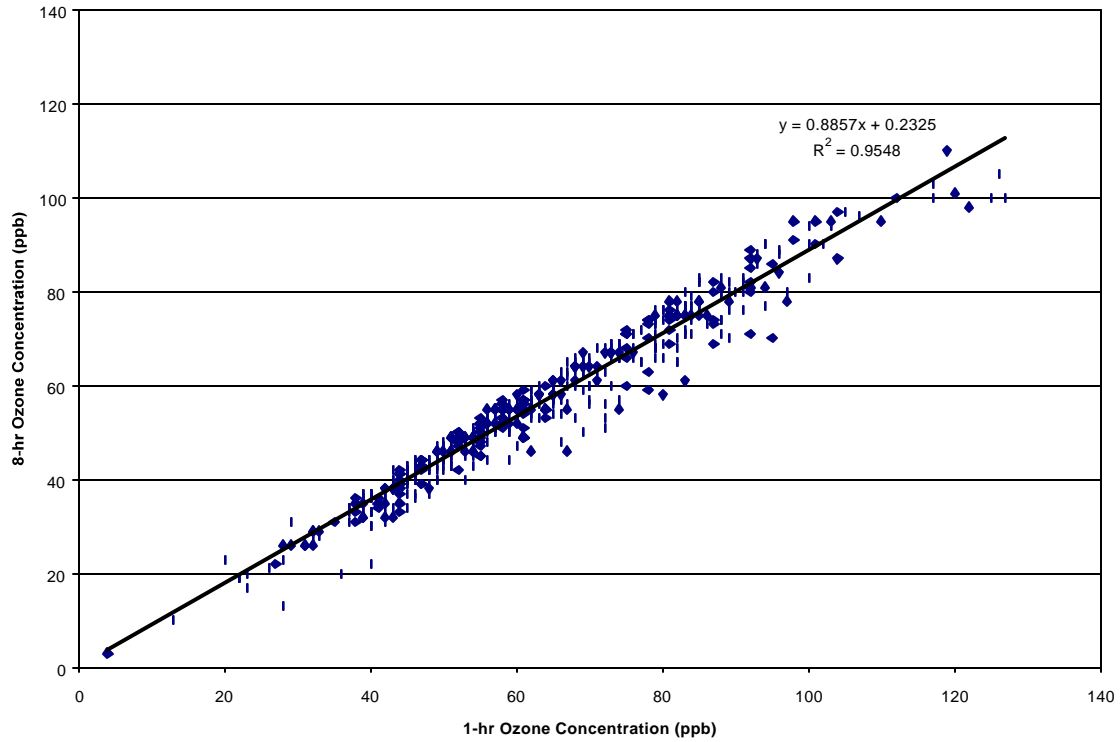


Figure 2-6. Scatter plot showing the relationship between 1-hour and 8-hour daily maximum ozone concentrations for a site in Hancock County, Indiana.

### **3. FORECASTING APPLICATIONS AND NEEDS**

The success of an ozone-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 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) for scheduling specialized air monitoring programs. This section describes these forecast applications and lists some of their needs as they relate to ozone forecasts.

#### **3.1 PUBLIC HEALTH NOTIFICATION**

Ozone 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 ozone. Forecasts are generally issued each day during the ozone season for maximum concentrations expected for the current day and next day. For example, the South Coast Air Quality Management District forecasts maximum ozone 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). The ozone forecast is usually formulated during the morning or early afternoon and then communicated to the public later that day.

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

- Ozone forecasting that errs on the side of public health (i.e., that tends to over-predict ozone rather than under-predict it).
- Forecasts that are as localized and specific as possible, particularly for large metropolitan regions.
- 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 (U.S. Environmental Protection Agency, 1997b; Jorquera, 1998). 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 30 episodic control programs exist throughout the United States and have various names, such as Ozone Action Day, Ozone Alert, and Spare The Air; but the underlying objectives are similar.

Health officials rely on ozone forecasts to determine whether or not to call an Ozone Action Day and seek voluntary action from the public to reduce emission-producing activities (e.g., driving, mowing lawns, etc.) on high ozone 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 concentration. Public outreach personnel then communicate the forecasts and plans for Ozone 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 forecast dissemination.

Episodic control programs typically have the following ozone forecasting needs:



- Minimizing the number of forecasts that falsely alert the public (i.e., minimize over-predicting). 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

### **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 high ozone days. Personnel for these programs have used ozone forecasts for decades to help schedule and plan intensive sampling efforts. 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 ozone 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 your forecasts as effective as possible, it is critical that you understand how ozone forecasts are used in your region. The material provided in Section 5.1 will help you to identify and determine these needs.

## 4. DEVELOPING OZONE FORECASTING METHODS

Many methods exist for predicting ozone 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 ozone forecasters use several methods—some objective, others subjective—to forecast ozone. Using several methods can balance one method's strengths with another method's limitations to produce a more accurate forecast.

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 your program, and lists its strengths and limitations.

All of the methods described here use multiple variables to predict ozone. 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. Each method presentation contains a definition, a discussion of how the method works, how you can develop it for your area, and its strengths and limitations. For easy comparison, **Table 4-1** lists and summarizes the methods.

#### 4.1.1 Persistence

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

#### How persistence forecasting works

Persistence forecasting works because atmospheric variables, including ozone, 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 peak ozone concentration was 50 ppb, it is likely that tomorrow's peak ozone concentration will also be relatively low. Similarly, if today's peak ozone concentration is 120 ppb, it is more likely that tomorrow's peak ozone concentration will be high (say over 100 ppb) than low (say less than 50 ppb).

Table 4-1. Comparison of forecasting methods.

	Persistence	Climatology	Criteria	CART	Regression	Neural Networks	Phenomenological /Intuition	3-D Air Quality Models
<b>Development Effort</b>	Low	Low/Moderate	Low/Moderate	Moderate	Moderate	Moderate/High	High	Very High
<b>Operational Effort</b>	Low	Low	Low	Low	Moderate	Moderate	Moderate	Very High
<b>Accuracy</b>	Low	Low	Moderate	Moderate/High	Moderate/High	Moderate/High	High	Moderate/High
<b>Method Description</b>	Today's (or yesterday's) observed ozone concentration is tomorrow's forecasted ozone concentration.	Historical frequency of ozone events help guide and bound ozone forecast.	When parameters that influence ozone are forecasted to reach a pre-determined level (criteria), high ozone concentrations are forecasted.	A decision tree predicts ozone based on values of various meteorological and air quality parameters.	A regression equation predicts ozone concentrations using observed and forecasted meteorological and air quality variables.	A non-linear set of equations and weighting factors predicts ozone concentrations using observed and forecasted meteorological and air quality variables.	A person synthesizes meteorological and air quality information including ozone predictions from other methods to produce a final ozone forecast.	A three-dimensional prognostic model replicates the meteorological and air quality processes that create ozone.
<b>DEVELOPMENT</b>								
Expertise <sup>1</sup>	—	—	Ability to identify key predictor variables.	Understanding of statistics and CART.	Understanding of statistics and regression.	Understanding of statistics and neural networks.	Experience in ozone forecasting and a conceptual understanding of meteorological and air quality processes.	High level understanding of meteorological and air quality relationships, and meteorological, emissions, and air quality models.
Software /Hardware	Spreadsheet /PC	Spreadsheet /PC	Statistical Software/PC	CART Software/PC	Statistical Software/PC	Statistical and Neural Network Software /PC	None	Prognostic meteorological model, emissions model, and air quality grid model /Cray or other high-speed computer system.
<b>OPERATIONS</b>								
Expertise	Ability to acquire today's and yesterday's ozone data.	Ability to acquire ozone 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 spreadsheet.	Ability to acquire observed and forecasted meteorological and air quality data and use a computational program.	Ability to synthesize meteorological and air quality information including ozone predictions from other methods to produce an ozone forecast.	Basic understanding of meteorological and air quality relationships to determine reasonableness of model results.

4-2

<sup>1</sup> 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	Phenomenological /Intuition	3-D Air Quality Models
Forecast production time	<1/2hr	<1hr	<1hr	<1hr	1hr	1hr	1 to 3 hrs	6 to 12 hrs (90 percent is computational time)
Data needs	Yesterday's ozone data.	No operational needs.	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.	Observed and forecasted meteorological data and charts, and observed air quality data.	Prognostic gridded meteorological fields, gridded emissions, and boundary conditions.
Software /Hardware	None	None	Data acquisition PC	Data acquisition PC	Computational program or spreadsheet /Data acquisition PC	Computational program/Data acquisition PC	Data acquisition PC	3-D meteorological and air quality grid model/Cray or other high-speed computer.
<b>STRENGTHS</b>	Works well in areas that have several continuous days of high ozone and low ozone concentrations.	Helps guide and bound forecasts derived from other methods.	Quick, use it to get initial "idea" about forecast conditions.	Automatically differentiates between days with similar ozone concentrations.	Commonly used and easy to operate. Produces generally good forecasts.	Allows for non-linear relationships to develop.	Helps temper the predictions from other methods with common sense and experience. Typically has the highest accuracy.	Predicts ozone concentrations in areas that are not monitored. Helps in understanding ozone processes including transport issues.
<b>POTENTIAL LIMITATIONS</b>	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 percent more effort to develop than regression with only slight improvement in forecast accuracy.	Prediction may be biased from one forecaster to another.	Expensive and difficult to develop and operate.

Ozone forecasting using the Persistence method works because ozone concentrations are highly dependent on synoptic-scale weather, which typically exhibits similar characteristics for several days, and, therefore, ozone 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 ozone 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 ozone 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 percent). The odds of a non-exceedance occurring after a non-exceedance are 19 out of 22 days (86.3 percent). Therefore, in this example, if you used the Persistence method to forecast a non-exceedance or an exceedance, your forecast would be accurate 24 out of 29 days, or 83 percent 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.

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

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

As shown in Table 4-2, you cannot use the Persistence method to correctly predict the beginning or end of an episode. However, you can use the Persistence method to help guide your forecasts and predictions from other methods.

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

### **Persistence forecasting development**

Although the Persistence method requires no real development, you need to be sure that the method will work in your area. The following steps describe how to test the effectiveness of persistence forecasting in your area.

1. Create a data set containing at least four years of recent ozone data.
2. From this data set, use each day's maximum ozone concentration to simulate a forecast for the next day (i.e., use the Persistence method). Compare the forecast and observed ozone concentrations for the historical data set and compute the forecast verification statistics provided in Section 5.6
3. Keep in mind the following development issues:
  - Consider when the forecast will be issued to determine what ozone data are available. For example, if you must issue a forecast at 11:00 a.m. for the next day and the current day's peak ozone concentration has not yet been observed, you would use the previous day's peak ozone concentration for your next-day forecast.
  - The Persistence method only works well for regions that experience several continuous days of high or low ozone. This approach fails if ozone episodes typically last only one day.

### **Persistence forecasting operations**

Using the Persistence method to forecast ozone concentrations requires very little expertise and is perhaps the easiest and quickest of all ozone forecasting techniques, yet its accuracy is the poorest. However, 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 ozone forecast that can be refined by using other forecasting methods.
- It is easy to produce and operate and requires little expertise.

### **Persistence forecasting limitations**

- Using persistence forecasting, you are unable to predict the first day and end of an episode.
- It does not work well under changing weather conditions when accurate ozone 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 ozone forecasting. Although not very accurate as a predictive tool, climatology can help forecasters bound and guide their ozone predictions.

### **How climatology works**

Climatology works because history tends to repeat itself, especially when it comes to seasonal weather. Since ozone concentrations are highly weather dependent, ozone climatology can be used in the same manner as weather climatology. For example, let's say your initial forecast is for a maximum temperature of 105°F in downtown Boston for August 13. After consulting a climate table, you learn that a maximum temperature of 105°F has never occurred in Boston and your forecast is probably too high.

Thus, you adjust your forecast down to 100°F. The climate data acted as a bound and a guide to your temperature forecast. Analogously, let's say that you are forecasting ozone for April 10 for upstate New York, and your forecast techniques indicate that an exceedance may occur. Consulting a climate table (**Table 4-3**), you learn that upstate New York had no exceedance in April for the 15-year period of records. Based upon the additional information provided by the climate table, you forecast a non-exceedance for April 10. The table has served as a complimentary tool to other forecast methods and helps improve your forecast accuracy.

### **Developing climate tables**

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

1. Create a data set containing at least four years of recent ozone data.
2. Examine the data for quality and be sure to note if emissions changed significantly over the time period of interest. Emissions tend to change slowly over time, but certain changes can occur quickly such as implementation of reformulated fuels. Changes in emissions can result in the same weather conditions producing lower ozone concentrations. Also note that changes in the monitoring network can dramatically change the maximum observed ozone concentrations and/or the number of exceedances.



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

Year	April			May			June			July			August		
	Total	Downstate	Upstate	Total	Downstate	Upstate	Total	Downstate	Upstate	Total	Downstate	Upstate	Total	Downstate	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

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

- All-time maximum ozone concentrations (by month, by site).
- Duration of high ozone episodes.
- Average number of days with high ozone by month and by week.
- Day-of-week distribution of high ozone concentrations.

Examples of such tables are shown in **Tables 4-4 through 4-8** for three air districts in the Sacramento, California region.

4. If significant changes in emission occurred, you may wish to divide the climate tables into “before” and “after” periods.

5. Examine your tables for usefulness. For example, if there are differences between weekend and weekday ozone concentrations or exceedance frequency, then a climate table showing the frequency of high ozone concentrations by day of week may be quite useful.

### **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 ozone forecasts you create using other methods. Consulting climate tables may be useful when other methods predict extreme events. Such events may include multiple days of high ozone concentrations or an extreme 1-day high ozone concentration. You can also use climatological information in the forecast discussion to provide context. For example, “Tomorrow’s predicted peak ozone concentration of 150 ppb would be the first time in two years that ozone has reached this level.”

### **Climatology strengths**

- Climatology acts to bound and guide an ozone 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 or large changes in population.

Table 4-4. Historical maximum ozone concentrations for three air districts in the Sacramento, California region (1990-1995). (Courtesy of Sacramento Metropolitan Air Quality Management District.)

District	Maximum Ozone Concentration (ppb)	Date
Yolo-Solano	130	8/2/93
District	160	7/2/91 & 9/17/91
Placer	170	5/4/92

Table 4-5. Duration of high ozone episodes for three air districts in the Sacramento, California region (1990-1995). (Courtesy of Sacramento Metropolitan Air Quality Management District.)

Concentration (ppb)	District	Maximum No. Days	Average No. Days	Median
≥ 100	Yolo-Solano	4	1.63	1
≥ 120	Sacramento	4	1.58	1
≥ 120	Placer	4	1.48	1

Table 4-6. Information on health advisory days (≥150 ppb) from 1990 through 1995 for three air districts in the Sacramento, California region. (Courtesy of Sacramento Metropolitan Air Quality Management District.)

District	Average # per year	Concentration Range (ppb)	Duration		Sites
			Average (hours)	Range (hours)	
Yolo-Solano	None				
Sacramento	2.7	150-180	1.5	1 to 3	FOL, DPM
Placer	1.5	150-160	2.2	1 to 4	ROC

Table 4-7. Average number of days with high ozone for three air districts in the Sacramento, California region (1990-1995). (Courtesy of Sacramento Metropolitan Air Quality Management District.)

Maximum Concentration (ppb)	District	Month					
		May	June	July	Aug.	Sept.	Oct.
≥ 100	Yolo-Solano	0	1	2	3	1	0
≥ 110	Yolo-Solano	0	0	1	1	0	0
≥ 120	Yolo-Solano	0	0	0	1	0	0
≥ 130	Yolo-Solano	0	0	0	0	0	0
≥ 140	Yolo-Solano	0	0	0	0	0	0
≥ 150	Yolo-Solano	0	0	0	0	0	0
≥ 160	Yolo-Solano	0	0	0	0	0	0
≥ 100	Sacramento	3	5	9	8	8	4
≥ 110	Sacramento	1	4	6	6	5	3
≥ 120	Sacramento	1	2	4	3	2	2
≥ 130	Sacramento	1	1	3	2	1	1
≥ 140	Sacramento	0	1	2	1	1	0
≥ 150	Sacramento	0	0	1	1	0	0
≥ 160	Sacramento	0	0	0	0	0	0
≥ 100	Placer	3	4	10	10	6	2
≥ 110	Placer	1	2	6	5	3	1
≥ 120	Placer	0	2	3	3	2	0
≥ 130	Placer	0	1	2	2	1	0
≥ 140	Placer	0	0	1	1	0	0
≥ 150	Placer	0	0	0	1	0	0
≥ 160	Placer	0	0	0	0	0	0

Table 4-8. Distribution of high ozone concentrations by day of week for three air districts in the Sacramento, California region (1990-1995). (Courtesy of Sacramento Metropolitan Air Quality Management District.)

Concentration (ppb)	District	Day						
		Sun	Mon	Tue	Wed	Thu	Fri	Sat
≥ 100	Yolo-Solano	8%	23%	27%	19%	15%	0%	8%
≥ 120	Yolo-Solano	0%	50%	0%	50%	0%	0%	0%
≥ 120	Sacramento	9%	16%	18%	15%	19%	15%	9%
≥ 120	Placer	7	11%	13%	24%	20%	17%	9%

### 4.1.3 Criteria

A criterion is a principle by which something is evaluated. The Criteria method in ozone forecasting uses threshold values (criteria) of meteorological or air quality variables to forecast ozone concentrations. 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 ozone 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 ozone concentrations. Once known, forecasters can look for the occurrence of the criteria in weather forecasts and predict ozone concentrations from them. For example, high ozone concentrations are often associated with hot temperatures and, thus, temperature can be used as one predictor of ozone 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 your area. Thus, 90°F would be a threshold value (criterion) for an 8-hr ozone exceedance.

Since ozone formation is complex, forecasters must use several variables and associated criteria to accurately forecast ozone. **Table 4-9** 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 yesterday'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.

Table 4-9. 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 kt)	Wind Speed 15-21 UTC (below kt)	Yesterday's Ozone Max (above 1-hr 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

The Criteria method is better suited to help you forecast an exceedance or non-exceedance rather than a particular ozone concentration. If you wished to forecast a particular ozone concentration using the Criteria method, you would need to establish threshold values for each parameter for each ozone concentration level.

## **Criteria method development**

Complete the following steps to develop the Criteria method for ozone forecasting in your region:

1. Determine the important physical and chemical processes that influence ozone concentrations in your area. This helps you identify which variables to use for the criteria. You can do this with literature reviews, historical case studies, and climatological analysis as discussed in Section 5.2
2. Select variables that represent the important physical and chemical processes that influence ozone concentrations in your area. Useful variables include: maximum temperature, morning and afternoon wind speed, cloud cover, relative humidity, 500-mb height, 850-mb temperature, etc. You can use statistical software to 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 data and surface and upper-air meteorological data.
4. Determine the threshold value for each parameter that distinguishes high and low ozone concentrations. For example, create scatter plots of ozone vs. particular parameters to help you determine the thresholds, as shown in **Figure 4-1**. The criterion of 28°C (81°F) helps distinguish higher ozone concentrations from lower concentrations. 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 differentiate high ozone days.
5. Use an independent data set (i.e., a data set not used for development) to evaluate the selected criteria (for example, data from a different time period).
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 ozone concentration associated with your established criteria may change. When this happens, you should update your criteria method.

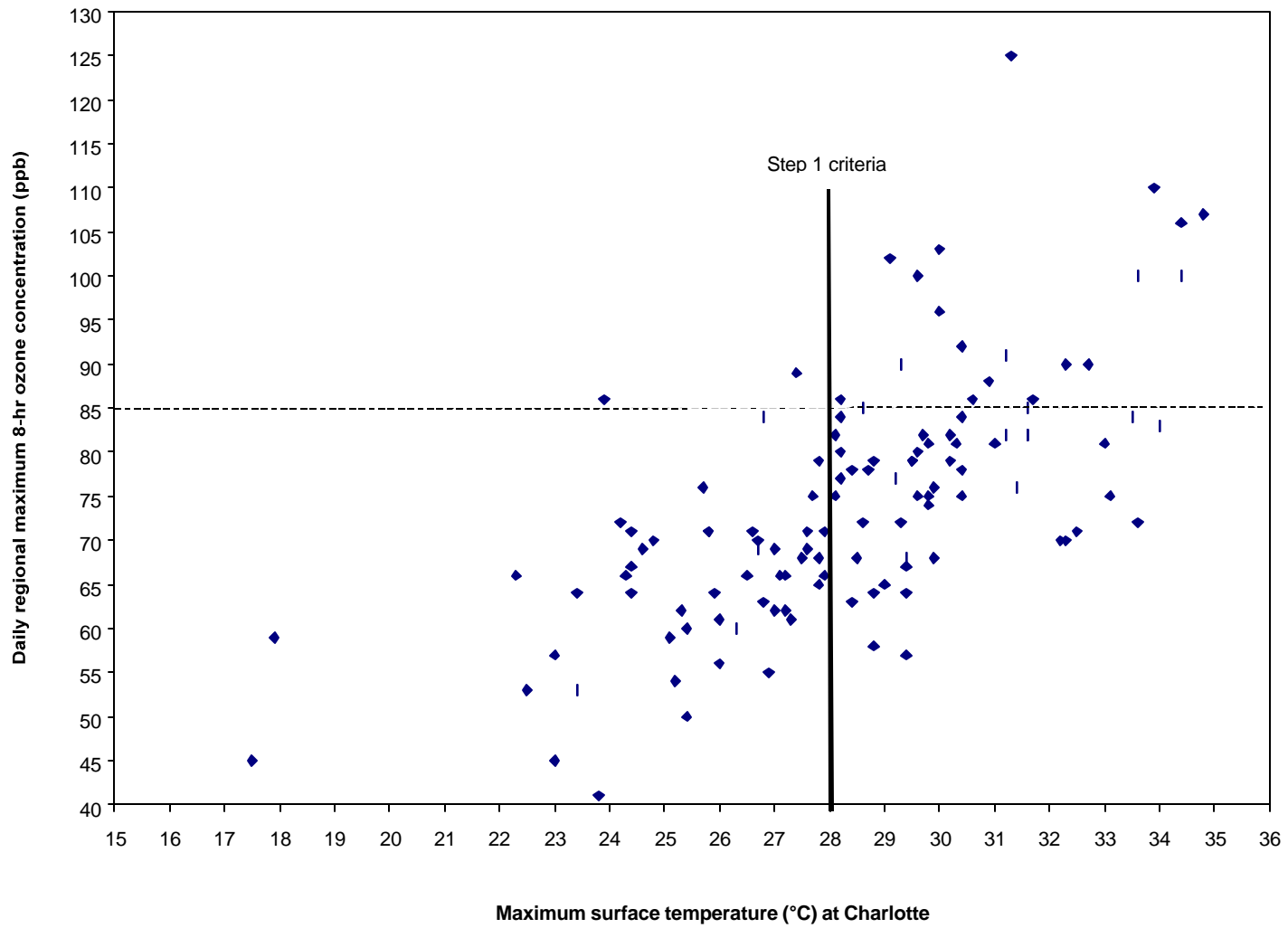


Figure 4-1. Scatter plot of maximum surface temperature and regional maximum 8-hr ozone concentration in Charlotte, North Carolina (MacDonald et al., 1998).

### **Criteria method operations**

The Criteria method is one of the easiest methods to use. You need only acquire data and check the data against the established criteria to determine the ozone forecast. Although use of this 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 ozone predictions for physical reasonableness.

### **Criteria method strengths**

- Easy to operate.
- Relatively easy to develop, and it can be refined each year as more knowledge is acquired.
- Objective method that alleviates potential biases arising from human subjectivity.
- Complement to other forecasting methods. You can easily use this method 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.
- Not well suited for predicting exact ozone concentrations; better suited for forecasting ozone concentrations above or below a certain concentration.
- Objective tool that can only predict ozone 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 ozone 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 ozone forecasting, CART enables you to develop a decision tree to predict ozone concentrations based on the values of predictor variables that are well correlated with ozone concentrations.

#### **How CART works**

CART uses software to develop a decision tree by continuously splitting peak ozone concentration data into two groups based on a single value of a selected predictor variable (Stoeckenius, 1990; Horie, 1988; National Research Council, 1991). 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 ozone. 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 4-2** 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°C. The tree growth continued until there were 10 distinct groups. In this example, the entire decision tree explains 80 percent of the variance in the daily maximum ozone concentration.

It is quite simple to forecast ozone concentrations using the decision tree created by the CART analysis. For the example shown in Figure 4-2, 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 the predicted ozone. 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 ozone 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, the decision tree should be updated frequently.

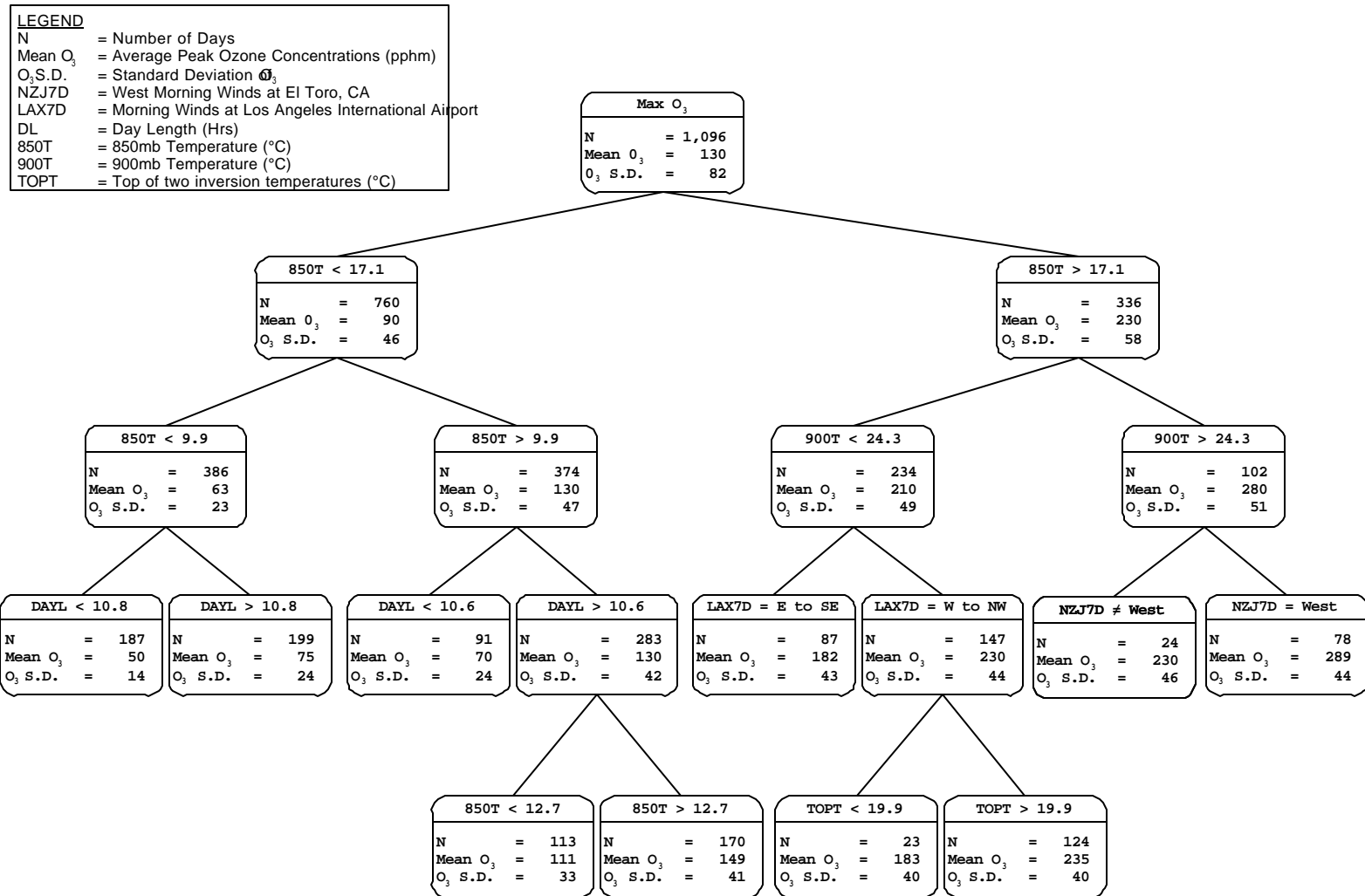


Figure 4-2. Decision tree for daily basin maximum ozone concentrations in the South Coast Air Basin in the Los Angeles, California area (Horie, 1988).

## **CART development**

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

1. Determine the important physical and chemical processes that influence ozone concentrations in your area in order to identify the key variables. You can do this through literature reviews, historical case studies, and climatological analysis as discussed in Section 5.2.
2. Select variables that properly represent the important physical and chemical processes that influence ozone concentrations in your 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 peak ozone associated with your established criteria may change. When this happens, you should update the decision tree.

## **CART operations**

The CART method is very easy to use and requires little expertise. You need only acquire data that is in the decision tree and process those data 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 ozone predictions for reasonableness.

## **CART strengths**

- Requires little expertise to operate; runs quickly.
- Complements other subjective forecasting methods.
- Allows you to differentiate between days with similar ozone concentrations if the ozone concentrations are a result of different processes.

## **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.
- Objective tool that can only predict ozone 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 ozone predictions.
- May not predict ozone 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 ozone concentrations.

### 4.1.5 Regression Equations

Regression is a statistical method for describing the relationship among variables. For ozone forecasting, regression equations are developed to describe the relationship between ozone 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 peak ozone concentrations in many areas of the country (Cassmassi, 1987; Hubbard and Cobourn, 1997; Ryan, 1994; Dye et al., 1996).

#### 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, you can predict ozone concentration 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.

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

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

where:

- $O_3$  = predictand
- $c$  = 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 peak 1-hr ozone concentrations in the Baltimore, Maryland metropolitan area (Ryan, 1994).

$$\text{Ozone} = 1.671 * T_{\text{max}} - 1.163 * T_{\text{min}} - 1.750 * \text{TSKC} - 0.786 * \text{WS} + 3.048 * T_{950} - 1.457 * \text{WS}_{850} - 1.075 * \text{SZ} + 16.15 \quad (4-3)$$

where:

- $T_{\text{max}}$  = Maximum surface temperature ( $^{\circ}\text{F}$ )
- $T_{\text{min}}$  = Minimum surface temperature ( $^{\circ}\text{F}$ )
- $\text{TSKC}$  = Fraction of cloud coverage 1600-1800 UTC
- $\text{WS}$  = Surface wind speed (kts) at 0900 UTC
- $T_{950}$  = 950-mb temperature ( $^{\circ}\text{C}$ ) at 1200 UTC
- $\text{WS}_{850}$  = 850-mb wind speed (m/s) at 1200 UTC
- $\text{SZ}$  = Daily solar zenith angle (degrees)

To use the equation, a forecaster simply inputs the forecasted values into the equation. For example, if the values of the input variables are  $94^{\circ}\text{F}$ ,  $55^{\circ}\text{F}$ , 0, 1 kt,  $25^{\circ}\text{C}$ , 3 m/s, and  $60^{\circ}$ , respectively, then the model would forecast a peak ozone concentration of 115 ppb. Notice that the model uses only weather variables. Thus you can use input values from the 24- and 48-hr weather forecasts to make one- and two-day ozone forecasts.

#### Regression equation development

Complete the following steps to develop a regression model for ozone concentrations in your area:

1. Determine the important physical and chemical processes that influence ozone concentrations in your area. You can do this with literature reviews, historical case studies, and climatological analysis as described in Section 5.2.
2. Select variables that represent the important physical and chemical processes that influence ozone concentrations in your area. You can use statistical software 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 and selected predictor variables. Choose a minimum of four recent years that are representative of the current emissions profile. Randomly select about 25 percent 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 is log-normally distributed; yet regression is best suited for predicting data that are normally distributed.
  - Use the natural log of ozone concentrations as the predictand instead of just ozone concentrations to improve performance.
  - Regression tends to predict the mean better than the tails (i.e., high ozone concentrations) of the distribution. Creating secondary regression equations to predict only the high ozone concentrations may improve your 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 is 5 to 10.
  - One variable can likely represent a whole subset of variables. You should attempt to use variables that are unique (i.e., dissimilar) to avoid redundancy and co-linearity.
  - Stratifying your data set may improve regression performance. Consider dividing your data set by seasons, weather type, or other meteorological variables. For example, you might develop separate equations for spring, summer, and fall.

### **Regression equation operations**

Compared to the development of the regression equations, operation of the model requires modest expertise. You need only acquire data and input the data into a simple computational program or spreadsheet that contains the regression equations. Although use of the equation does not require an understanding of meteorology and air quality processes, it is advisable that someone with meteorological experience check the ozone prediction for physical reasonableness.

Because the predictor variables are forecasted, they have inherent uncertainty, which results in an ozone forecast that has a degree of uncertainty. To help quantify this uncertainty you can slightly alter the input values and evaluate the effect this has on the forecasted ozone.

### **Regression analysis strengths**

- Regression analysis is well documented and widely used in a variety of disciplines. It has been successfully used in ozone forecasting in many areas of the country (Cassmassi, 1987; Hubbard and Cobourn, 1997; Ryan, 1994; Dye et al., 1996).
- 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.
- You can use regression analysis to complement other forecasting methods, or you can use it as your primary forecasting method.

### **Regression analysis 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 ozone concentrations) of the distribution. They will likely under predict the high concentrations and over predict the low concentrations.

#### **4.1.6 Artificial Neural Networks**

Artificial neural networks (ANN) 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 formation is a complex non-linear process, neural networks are well suited for ozone forecasting. Note that neural networks require about 50 percent more effort to develop than regression equations and provide only a modest improvement in forecast accuracy (Comrie, 1997).

### **How artificial neural networks work**

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 concentration). **Figure 4-3** is a schematic showing the neural network components. You supply 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 prediction. The neural network software offers several choices for transfer functions.

You can purchase commercial software to help you develop and operate a neural network. Before you can make a prediction, you must train and develop the network software. Complete the following steps to train your 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 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.

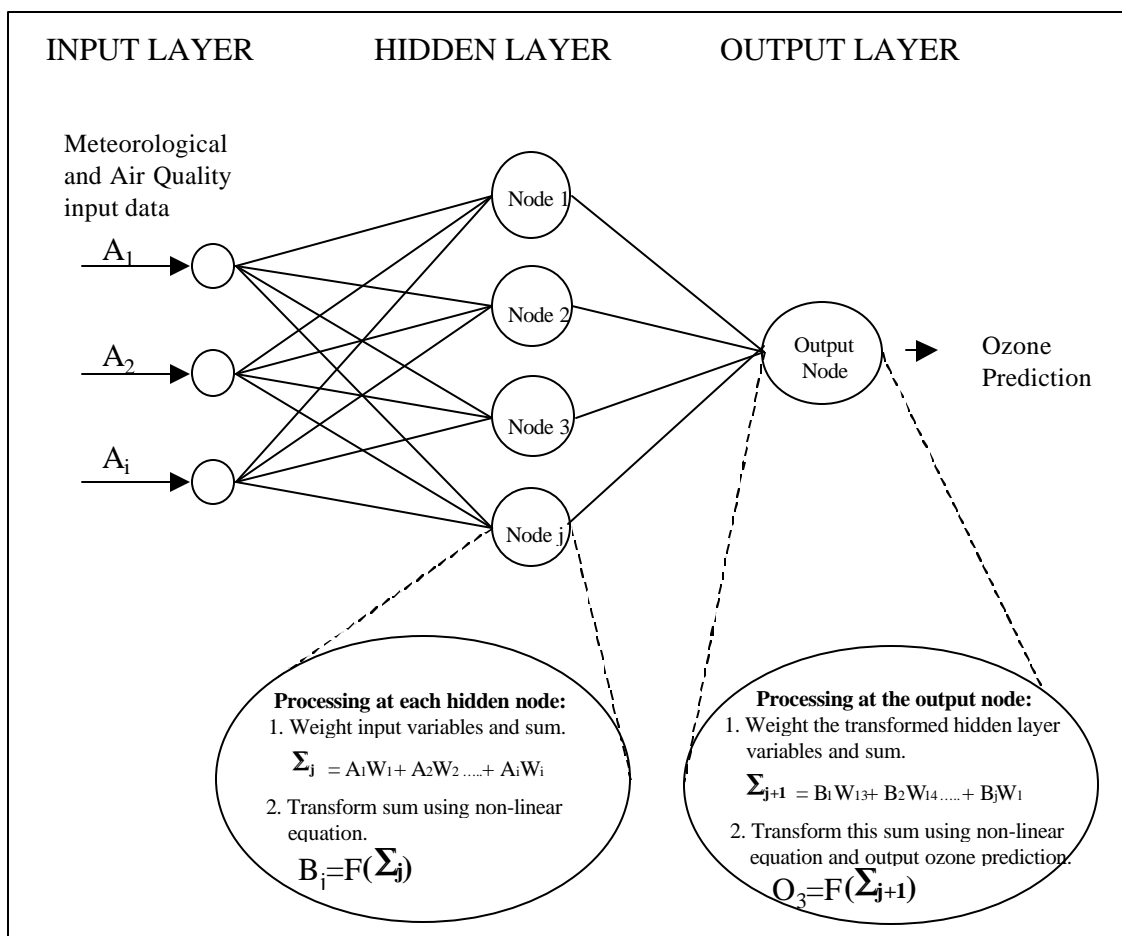


Figure 4-3. A schematic of an artificial neural network (Comrie, 1997).

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.
5. Once the network has been trained (i.e., developed) you can use it operationally to forecast ozone.

To train a neural network to achieve good generalization on new data, you need three data sets: a developmental set, a validation set, and a test set. You use the developmental set to develop the neural network. You use the validation set to determine when the network’s general performance is maximized. And you use the test data set 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 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:



1. Complete historical data analysis and/or literature reviews to establish the air quality and meteorological phenomena that influence ozone concentrations in your 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 four 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 your 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 you train the network, 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. You need only acquire data and input the data 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 ozone prediction for reasonableness.

### **Strengths of artificial neural networks**

- Ozone formation is a non-linear process. 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, 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.
- You can use neural networks to complement other forecasting methods, or you can use it as your primary forecasting method.

### **Limitations of artificial networks**

- Neural networks are complex and not commonly understood; thus, the method can be inappropriately applied and more difficult to develop.
- Neural networks do not extrapolate data well. Thus, extreme ozone concentrations not included in the developmental data will not be taken into consideration in the formulation of the neural network prediction.

#### **4.1.7 Three-dimensional (3-D) Air Quality Models**

Three-dimensional (3-D) air quality simulation models are mathematical descriptions designed to mimic the atmospheric processes that influence pollutant concentrations. Historically, 3-D air quality models have been used extensively in case study analyses to understand ozone processes and to estimate the effects of emissions changes on ozone concentrations during episodic conditions. Recently, air quality models have been applied using prognostic meteorological inputs to produce daily ozone forecasts. A sample meteorological and ozone air quality forecasting system is described on the Internet at <http://envpro.ncsc.org/NAQP/>.

#### **How 3-D air quality models work**

Three-dimensional air quality models use computer algorithms that are designed to simulate the atmospheric processes that influence ozone including transport, dispersion, and chemistry. The air quality model integrates and processes meteorological, emissions, and chemistry information to estimate the state of the atmosphere at some future time. To do this it uses equations that capture the current state of knowledge of atmospheric pollutant dynamics. The meteorology and emissions input data are derived from prognostic meteorological and emissions models.

The common 3-D regional 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 modeling regime. 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 15 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 greater model accuracy, but also higher computational cost. Modern 3-D air quality models used nested grids that have coarse resolution in the outlying areas and fine resolution in the areas of greatest interest.

The prognostic meteorological models solve an approximation of the equations that govern atmospheric behavior. During the past ten years, prognostic mesoscale modeling has become an increasingly common method of developing inputs for air quality modeling. Several air quality modeling

systems—CAMx (Environ, 1998), MAQSIP (Odman and Ingram, 1996), SAQM (Chang et al., 1996), Models-3 and UAM-V (U.S. Environmental Protection Agency, 1998)—use mesoscale meteorological models as their preferred meteorological driver. In a current program that uses 3-D air quality models to forecast ozone in the northeastern United States, forecasters are using a combination of the National Weather Service’s operational forecast Eta model and the mesoscale MM5 model (Grell et al., 1994; Dudhia, 1993; Steenburgh and Onton, 1996) to supply meteorological inputs to the MAQSIP air quality model (see <http://envpro.ncsc.org/NAQP/> for more information).

Emissions modeling is the process of estimating emissions with the spatial, temporal, and chemical resolution needed for air quality modeling. The emissions inventory includes data for mobile sources, stationary area and point sources, and natural and agricultural sources. Mobile, biogenic, and selective point/area source emissions can vary substantially with temperature. Mobile source and selective industrial/commercial source emissions also exhibit significant variations by day of the week. When you use emissions modeling to support ozone forecasting you must take temperature and day-of-week effects into account. These effects may be included in precomputed, model-ready emissions inputs for various cases. Currently three main emission processing tools provide gridded air quality models with emissions data:

1. Emission Processing System (EPS 2.0) (U.S. Environmental Protection Agency, 1992)
2. Emissions Modeling System – 1995 (EMS-95) (Bruckman, 1993)
3. Sparse Matrix Operator Kernel Emissions (SMOKE) (Coats, 1996)

### **Setting up a 3-D air quality model for your region**

You need substantial personnel and computer resources to establish credible and automated meteorological, emission, and air quality model forecast systems. Even using existing models, you may still have to undertake a large effort to refine the application methodologies enough to produce reliable ozone forecasts.

Complete the following steps to develop a 3-D forecasting model for your region. A more detailed discussion of air quality models can be found in Seinfeld and Pandis (1998).

1.
  1. Review the gridded 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.
  2. Review the emissions data for accuracy. Errors in the emission field can also result in large errors in the air quality output. Be sure that the emissions data you use reflect the most recent emission profiles available. It does not make sense to use 1990 emissions data if 1998 emissions data are available.
  3. Run the combined meteorological/emissions/air quality modeling system in a prognostic mode using a wide 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 in your region. You may need to refine any one of the three prognostic models that make up the system: meteorological, emissions, or air quality models.
  4. Once you achieve satisfactory results in the testing phase, establish (and automate) the mechanisms for the daily data exchange from the prognostic meteorological model and the emissions model to the 3-D air quality model, also develop ways to display model output.
  5. Run the model in real-time test mode for an extended period. Compare output to observed data and note when the model fails.

6. After obtaining satisfactory results on a consistent basis, you can use the modeling system to forecast ozone concentrations.

### **Three-dimensional air quality model operations**

Operation of the 3-D air quality forecast model should be completely automated. The forecaster need only review the model output forecast for physical reasonableness.

### **Strengths of 3-D air quality models**

- Three-dimensional air quality forecast models are phenomenological based, simulating the physical and chemical processes that influence ozone.
- They forecast a large geographic area.
- They can predict ozone in areas that are not monitored.
- You can use 3-D air quality forecast models to further understand the processes that control ozone in a specific area. For example, you can use them to assess the importance of long-range transport.

### **Limitations of 3-D air quality models**

- 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.
- Inaccuracies in the prognostic model forecasts of wind speeds, wind directions, extent of vertical mixing, and solar insolation 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.
- 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 ozone forecasting involves analyzing and conceptually processing air quality and meteorological information to formulate an ozone 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 ozone forecasting intuition is the perception of truth or fact (the ozone 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. This method balances some of the limitations of objective prediction methods (i.e., criteria, regression, CART, and neural networks).

### **How the Phenomenological/Intuition methods works**

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

1. **Understanding the processes that influence ozone.** The basic component to phenomenological/intuition forecasting is developing a robust and accurate conceptual understanding of the important phenomena that control ozone concentrations. This conceptual understanding should include information on synoptic, regional, and local meteorological conditions, plus air quality characteristics in your area.
2. **Synthesizing information.** Vast amounts of data are needed to forecast ozone. Forecasters will analyze both observed and forecasted weather charts, satellite information, air quality observations, and ozone 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, weather data are conducive for high ozone (light winds, clear skies, and hot temperatures) and forecasting criteria suggest high ozone, yet the regression equation predicts only modest ozone concentrations. A 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 how ozone forms in your forecast area. This task requires you to determine the important physical and chemical processes that influence ozone concentrations in your area. You can do this through literature reviews, historical case studies, and climatological analysis as discussed in Section 5.2. Although you can gain much knowledge 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 concentrations and needs to apply this understanding on a daily basis. Typically, the ozone forecaster will evaluate meteorological forecast models and use pattern recognition that equates the meteorological fields to ozone concentrations. For example, the forecaster may observe a high-pressure ridge building into the forecast area and equate this with high ozone. 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 ozone concentrations, while others indicate moderate and low ozone concentrations. By processing all of this information in the conceptual model, the forecaster develops an ozone 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, require re-creation of the forecasting algorithm.
- The Phenomenological/Intuition method allows for the integration and selective processing of large amounts of data in a relatively short period of time.
- You can immediately adjust this method as new truths are learned about ozone formation.

- You can easily take into account the effect of unusual emissions patterns associated with holidays and other events on the ozone forecast.
- You may be able to more accurately forecast extreme or rare events. 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 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. This section provides guidelines to help you select the candidate predictor variables to use in your ozone forecasting efforts. **Table 4-10** contains a list of common predictor variables to get you started. Consider the following issues when selecting predictor variables:

- **Understand the phenomena.** Before selecting particular variables it is important that you understand the phenomena that affect ozone concentrations in your region. You can gain this understanding through review of past air quality studies in your area, conducting a historical analysis of meteorology and ozone, and/or doing a literature review as described in Section 5.2.

Table 4-10. Common predictor variables used to forecast ozone.

Variable	Usefulness	Condition for high ozone
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 <sub>x</sub> concentration	Ozone precursor levels	High
Previous day's peak ozone	Persistence, carry-over	High
Aloft wind speed and direction	Transport from upwind region	-

- **Capture the important phenomena.** The variables you select should capture the important phenomena that affect ozone concentrations in your region. For example, research may show that high background ozone concentrations are needed to produce high ozone concentrations in your area. Thus, using yesterday's maximum ozone concentrations as a surrogate for background ozone concentration may improve forecast accuracy.
- **Select observed and forecasted variables.** Predictor variables can consist of observed variables that have been measured (e.g., yesterday's peak ozone concentration) and forecasted variables (e.g., tomorrow's maximum temperature). Using forecasted predictor variables is critical since tomorrow's ozone concentrations are more strongly related to tomorrow's weather conditions than to today's or yesterday's ozone concentrations.
- **Ensure data availability and reliability.** Make sure that you can easily obtain data from reliable source(s). Ensure that data will be available by a specified time every day, so that you can issue a timely forecast. For example, if you need to issue a forecast for tomorrow's maximum ozone concentration by 1100 LST, all predictor variables and data must be available before 1100 LST.

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

You can use statistical analysis techniques to identify the most significant variables. Following is a list of the types of statistical analyses you can perform. 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 you can use cluster analysis to identify these similarities. One variable can likely represent a whole set of similar variables. You should use variables that are unique (i.e., dissimilar) to avoid redundancy. You may find it beneficial to purchase and use statistical software.

- **Correlation analysis** is used to evaluate the relationship between the predictand (i.e., peak ozone levels) and various predictor variables. Correlations range from +1 (high-positive relationship) to 0 (no relationship) to -1 (high-negative relationship). Select variables for this type of analysis that 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 ozone 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. You can calculate correlation with spreadsheet programs (Excel, Lotus, etc.) or statistical software.
- **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. You can visually evaluate the relationship among variables using scatter plot matrices, for example.

This selection process results in a series of key variables you can use with the forecast methods described in Section 4.1 to predict ozone concentrations in your forecast region.



## 5. STEPS FOR DEVELOPING AN OZONE-FORECASTING PROGRAM

This section describes the major steps you can follow to set up and operate an ozone-forecasting program. For each step, we identify the major issues that you might face and provide suggestions for tackling them. Understanding the users' needs (Section 5.1) and how/why ozone forms in your area (Section 5.2) are the key first steps to developing a forecasting program. Information to help you choose one or more forecasting methods is presented in Section 5.3. Section 5.4 can help you to identify the types and sources of air quality and meteorological data you will need to run your program. Section 5.5 explains the importance of having a forecasting protocol. To evaluate the quality of an ozone forecast, follow the verification procedures described in Section 5.6.

### 5.1 UNDERSTANDING FORECAST USERS' NEEDS

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

- ***Who will use the forecast?*** Understanding who will use your forecasts will give you insight into potential ways to improve the forecast.
- ***For how many months are forecasts needed?*** Understanding how long the ozone season lasts will help you plan the resources (labor and data) needed to forecast ozone. The analysis techniques described in Section 5.2 can help you determine the length of the ozone season.
- ***What periods should your forecast cover?*** Typically, ozone forecasts are made for the current- and next-day periods; however, they can be extended to include two- to three-day predictions. Keep in mind that longer-range predictions will likely be less accurate.
- ***Do you need three-day forecasts for weekend/holiday periods?*** During weekends and holidays, staff may be unavailable to produce daily forecasts. In this case, you may need two- and/or three-day forecasts to cover this period. Have a plan 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 high ozone events. 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 forecast may improve the forecast accuracy, but could lead to public confusion and jeopardize credibility.
- ***What are the accuracy requirements? For example, is an error of  $\pm 20$  ppb acceptable?*** It is important to understand the error tolerance of your forecast users. Exceeding this threshold can lead to reduced credibility.
- ***Are forecasts issued for maximum regional ozone 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. Balance the user's forecast needs and tolerance for accuracy with the resources you have available to produce the forecast. Most air quality agencies issue regional ozone forecasts.
- ***Should forecasts be made for specific concentrations or concentration ranges (e.g., Air Quality Index (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 ozone 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 ozone concentrations such as fronts or winds. It is important to identify whether your 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). This may depend on where you are forecasting and the goals of the users.
- ***What types of interactions with ozone forecast users are needed?*** In addition to receiving your forecasts, your forecast users may benefit from a brief discussion with the forecaster. This discussion allows you, the forecaster, to pass on verbally any details or uncertainty about the forecast.
- ***Do you need to provide written forecast discussions of predicted weather and air quality conditions?*** These discussions provide additional information to help users interpret the predicted ozone values. Written explanations can convey fine points and uncertainties about the forecast.
- ***How should forecasts be disseminated?*** Many methods exist for disseminating your forecast (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, you can learn from past mistakes, and users can better understand the forecast process and its limitations.

Once you have determined how to meet the needs of ozone forecast users, the next step is understanding how and why ozone forms in your region.

## 5.2 UNDERSTANDING THE PROCESSES THAT CONTROL OZONE

The next step in developing an ozone-forecasting program is understanding how and why ozone forms in your area. Section 2 provides a general discussion of the chemical processes and weather phenomena that influence ozone concentrations. This section presents methods and examples to help you identify and understand the processes and phenomena that influence ozone in your area. Understanding these processes and phenomena will improve your ozone forecasting capabilities. Common methods for developing this understanding include reviewing literature from past ozone research and conducting data analyses.

### 5.2.1 Literature Reviews

The most efficient and generally the easiest way to start understanding ozone in your 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.

Articles published in the *Journal of Applied Meteorology* and *Atmospheric Environment* are good places to start your research. Consider other literature sources such as reports from local/regional ozone studies that may be available through government agencies. Broadening your literature review to include other regions may provide important information that is directly applicable to ozone processes in your 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 other pollutants) in both the troposphere and stratosphere and describes how meteorology affects atmospheric chemistry.
- Wallace and Hobbs (1977) – Provides general meteorological information about weather maps, atmospheric stability, and atmospheric motions from synoptic-scale to local-scale.
- Wilks (1995) – Describes statistical techniques and how you can apply them to meteorological data.

### 5.2.2 Data Analyses

Once you have completed a literature search, data analysis can help you learn more about the processes that control ozone concentrations in your area. Data analysis is the process of exploring data to answer questions. You can perform data analysis in three steps: developing questions (i.e., hypotheses), acquiring data, and using analytical methods to answer the questions. Depending on your resources, data analysis efforts can range from simple statistical analyses to large field studies with subsequent research and computer modeling. What follows is a discussion of some basic analysis procedures that will help you understand the processes that control ozone concentrations in your forecast area.

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

#### Temporal distribution of ozone

- *During what weeks/months are exceedances of the 8-hr and 1-hr ozone standard likely to occur?*
- *At what time of day do the highest ozone concentrations occur? How many hours do high ozone concentrations typically last?*
- *How many consecutive days do high ozone episodes typically last?*
- *Do maximum ozone concentrations vary by day of week?*

#### Spatial distribution of ozone

- *Where do the highest ozone concentrations occur? Do peak ozone concentrations occur at different times for different sites?*
- *Have emissions patterns changed in recent years?*
- *Has your monitoring network changed recently?*

#### Meteorological and air quality processes

- *What types of synoptic weather patterns are associated with high ozone concentrations?*
- *Does local carryover contribute to peak surface ozone concentrations?*
- *Does surface or aloft transport of ozone or ozone precursors from other areas contribute to ozone in your forecast area?*
- *How do local flow patterns influence ozone concentrations?*
- *How does the aloft temperature structure influence peak ozone concentration?*
- *What types of weather patterns are associated with cloud cover?*

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 to get you started. These examples are intended as a starting place for your understanding of the important processes that produce ozone in your area.

**Temporal distribution of ozone**

**Question:** *During what weeks/months are exceedances of the 8-hr and 1-hr ozone standard likely to occur?*

**Why:** Helps define your ozone-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 would last from May through September, since most of the 1-hr and 8-hr exceedances are confined to these months.

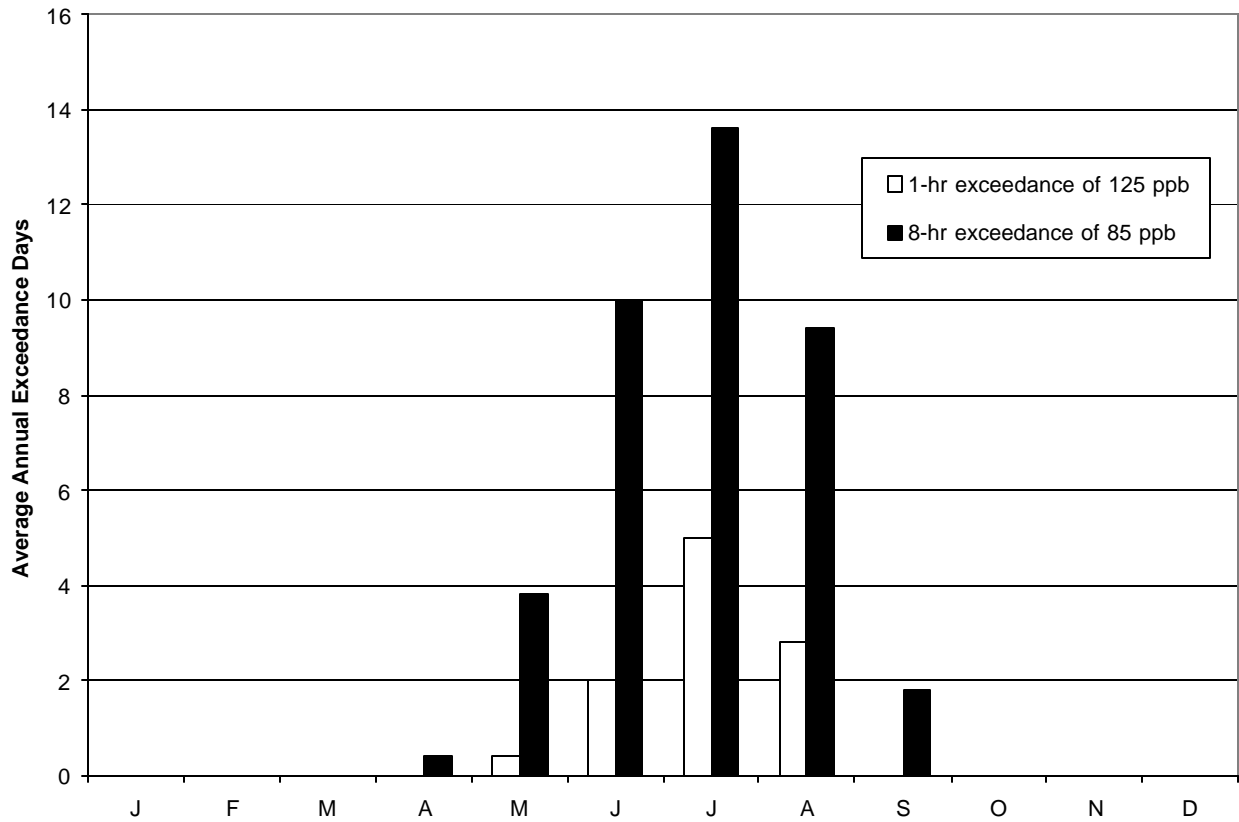


Figure 5-1. Distribution of the average number of days with 8-hr and 1-hr exceedances by month for the New Jersey and New York City region from 1993-1997 (NESCAUM, 1998).

**Question:** *At what time of day do the highest ozone concentrations occur? How many hours do high ozone concentrations typically last?*

**Why:** Knowing the typical time and duration of high ozone concentrations can help public outreach personnel properly notify the public so they can take appropriate action to minimize exposure.

**Technique:** Create frequency plots of the time of peak ozone 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 1100 to 1700 EST.

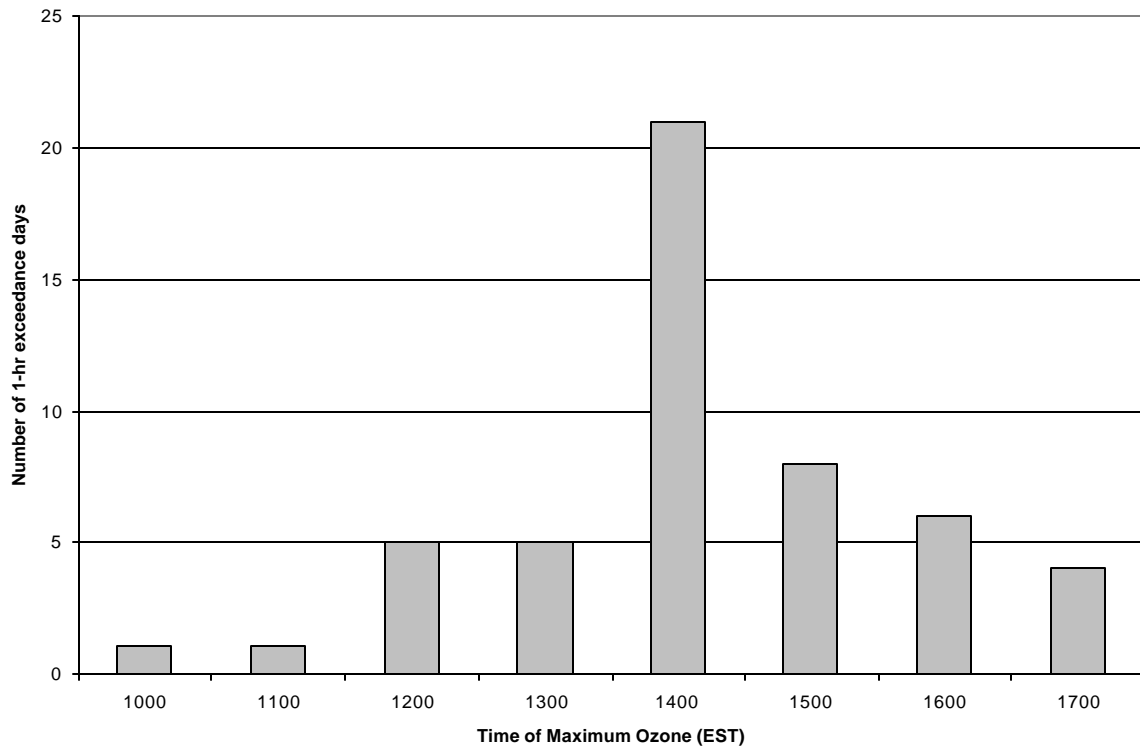


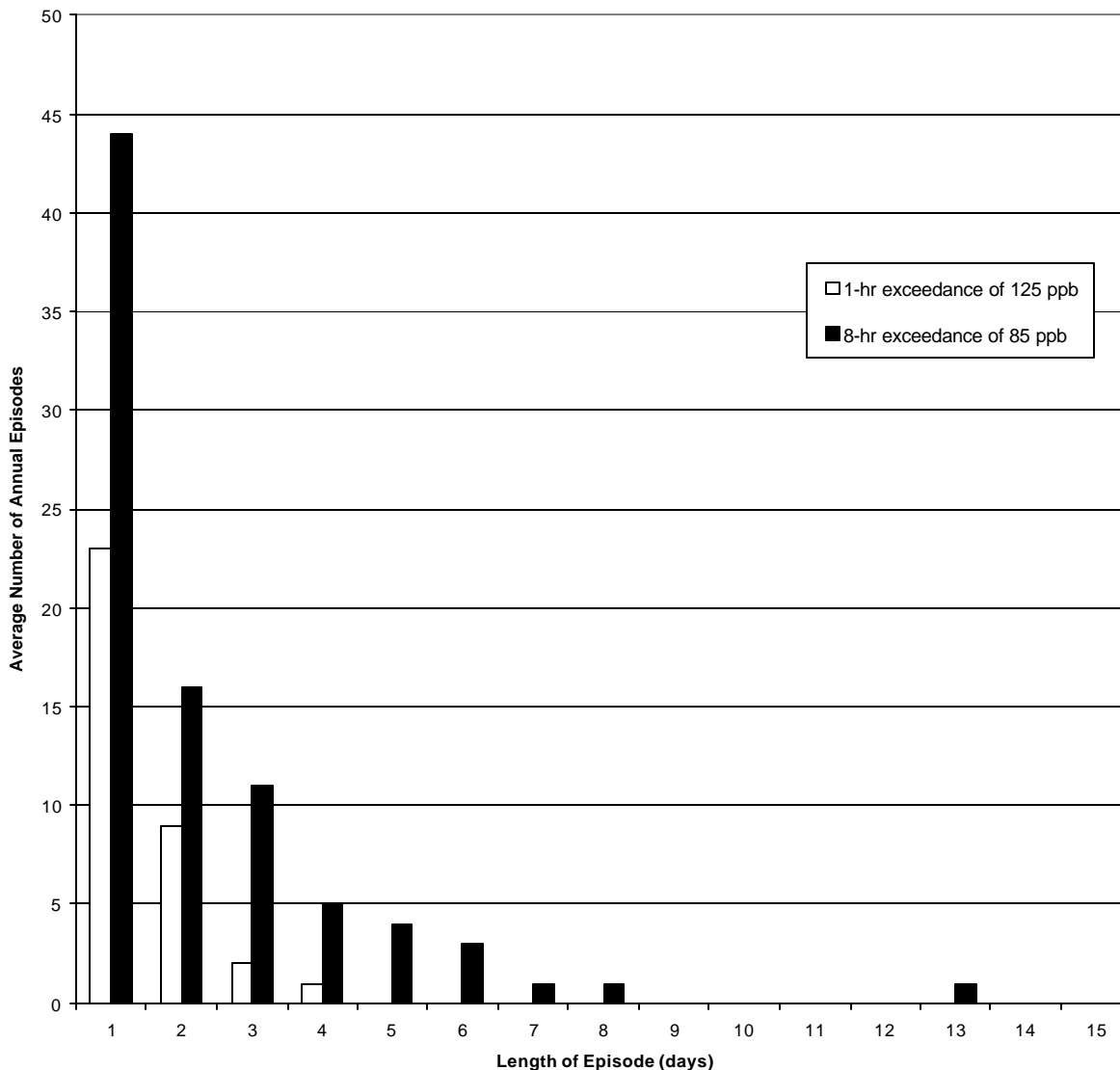
Figure 5-2. Distribution of hour of daily maximum 1-hr ozone concentration on days that exceeded 125 ppb in the New Jersey and New York City region from 1993-1997 (NESCAUM, 1998).

**Question:** *How many consecutive days do high ozone episodes typically last?*

**Why:** Knowing the typical duration of high ozone episodes can help guide your forecast. For example, if ozone episodes never last more than two days in your area, the occurrence of a three-day episode in the future is unlikely; therefore, you would be cautious to forecast high ozone 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 concentrations. Figure 5-3 indicates that a typical 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 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 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 concentrations given similar weather conditions.

**Technique:** Create frequency plots of the number of ozone 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, an 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.

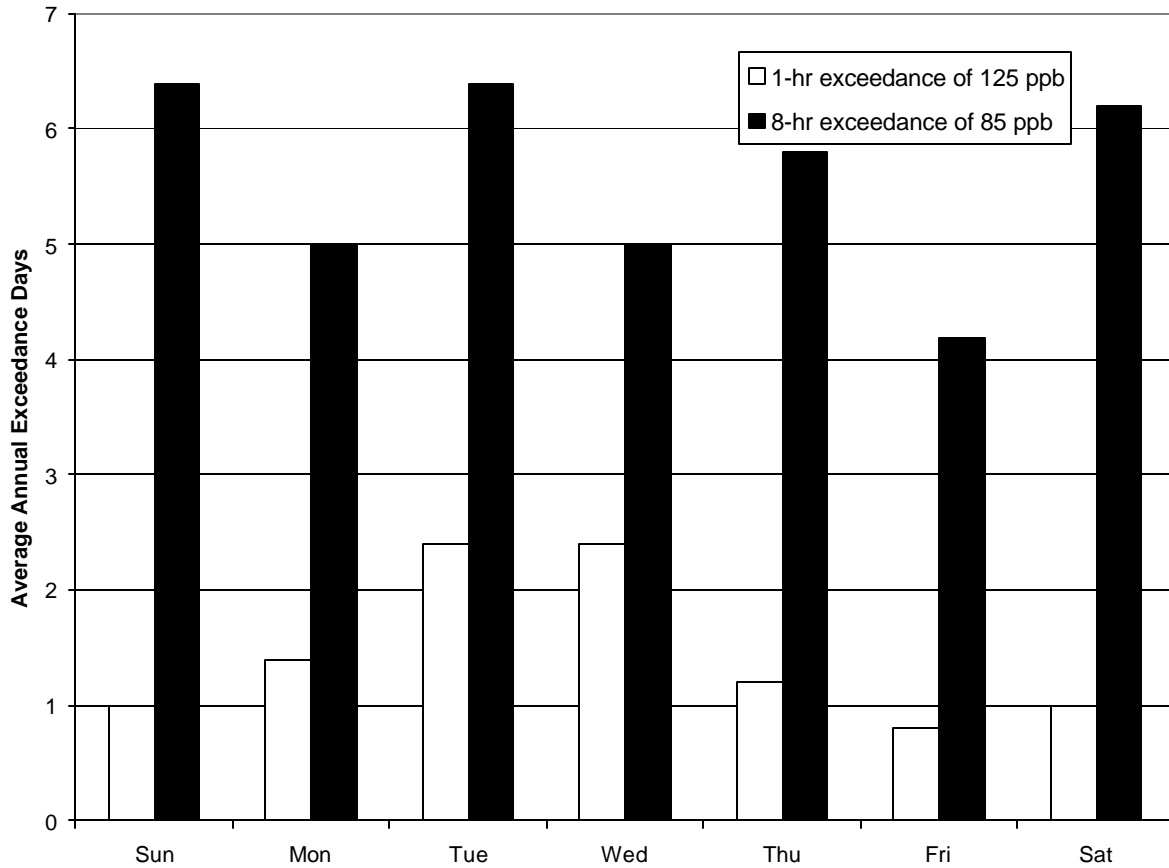


Figure 5-4. Distribution of the average number of 8-hr and 1-hr exceedances by day of week for the New Jersey and New York City region (NESCAUM, 1998).

## Spatial distribution of ozone

**Question:** *Where do the highest ozone concentrations occur? Do peak ozone concentrations occur at different times for different sites?*

**Why:** Different areas in your forecast region may have very different ozone characteristics due to spatial variation in emissions and meteorology. When forecasting for large areas it may be necessary to sub-divide your region to account for differences in ozone concentrations based on differences in emissions and weather across the region. This may also depend on the goals of the forecasting program

**Technique:** Plot a map of the average peak ozone concentration and the time of peak ozone concentration for each site on exceedance days.

**Question:** *Have emissions patterns changed in recent years?*

**Why:** If emissions patterns have changed, weather conditions that have historically produced ozone 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 your region. Less reactive emissions may result in peak ozone concentrations occurring farther downwind and/or in lower ozone concentrations.

**Question:** *Has your monitoring network changed recently?*

**Why:** Changes in a monitoring network can cause significant differences between historic and currently observed ozone concentrations. If a new monitor was recently installed downwind of a major emission source area, then the observed time and peak ozone concentrations for the entire area may change significantly due to this new site. You must take these types of monitoring network changes into account when analyzing historic and current ozone 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 your region.

## Meteorological and air quality processes

**Question:** *What types of synoptic weather patterns are associated with high ozone concentrations?*

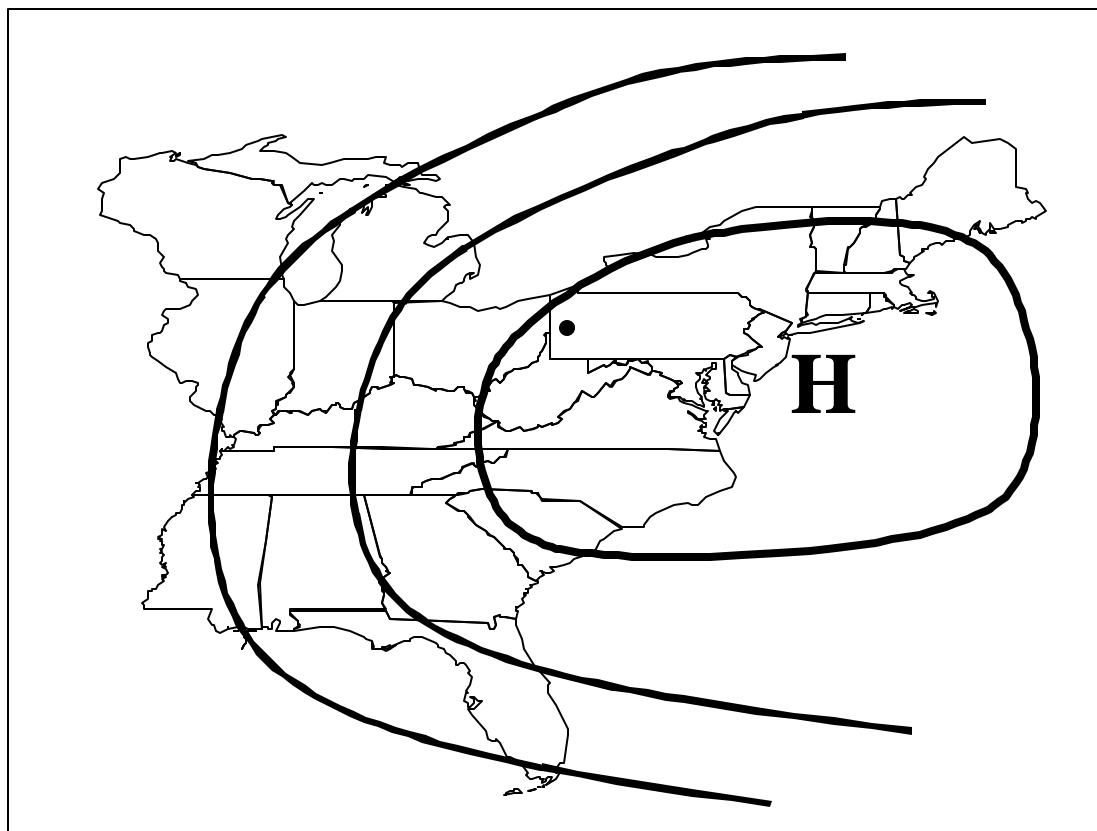
**Why:** Synoptic-scale weather features are large (1000 km or more) weather circulations that influence regional weather conditions that, in turn, strongly influence the production and transport of ozone and its precursors. By reviewing weather forecast charts, you can identify historical weather patterns associated with particular ozone concentrations in your 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 concentration days, moderate ozone concentration days, and low ozone concentration days.



For example, **Figure 5-5** 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)<sup>1</sup> and the archive analysis of the National Center for Environmental Prediction Eta model available on the Internet at <http://wxp.eas.purdue.edu/archive/index.html>.

Figure 5-5. A surface synoptic pattern associated with high ozone in Pittsburgh, Pennsylvania (Comrie and Yarnal, 1992).



<sup>1</sup> Daily Weather Maps, Climate Prediction Center, Room 811, World Weather Building, Washington, DC 20233

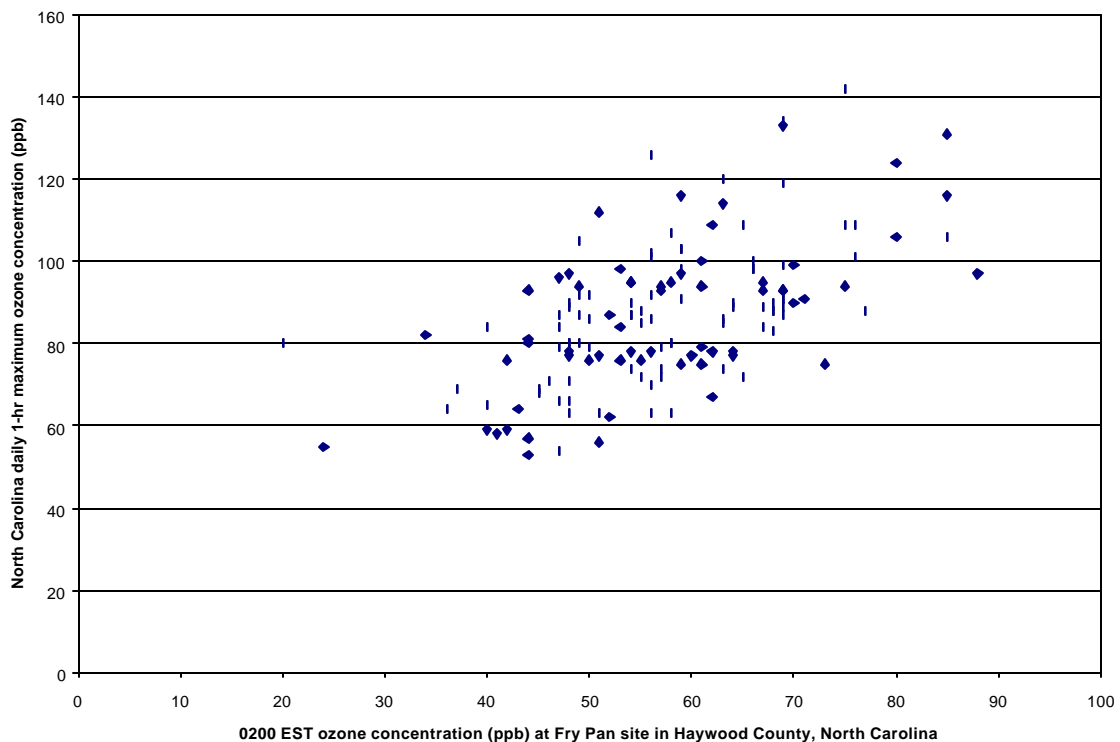
**Question:** *Does local carryover contribute to peak surface ozone concentrations?*

**Why:** When ozone episodes occur over several days, day-to-day pollution buildup can contribute to peak ozone concentrations. That is, today's ozone and ozone precursors (if not dispersed, deposited, or permanently reacted away) will contribute to tomorrow's ozone.

**Technique:** You can investigate carryover by examining ozone data from surface sites at which the ozone data show no overnight titration by NO. Create scatter plots of overnight ozone concentrations at non-titrated sites vs. peak daytime ozone concentrations. Examine the plots to see if there is a relationship between overnight ozone levels and peak ozone concentrations.

For example, **Figure 5-6** 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. You can also assess the influence of background ozone 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 5-6. 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:** *Does surface or aloft transport of ozone or ozone precursors from other areas contribute to ozone in your forecast area?*

**Why:** Long-range transport of ozone and ozone precursors can contribute significantly to local ozone concentrations. It is important for a forecaster to understand if and when this occurs in order to accurately forecast ozone.

**Technique:** Computing back trajectories is a useful way to examine the potential for long-range transport. 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 concentrations. An excellent tool for computing back trajectories is interactively available on the Internet at <http://www.arl.noaa.gov/ready/hysplit4.html> (Draxler and Hess, 1997).

**Figure 5-7** depicts 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).

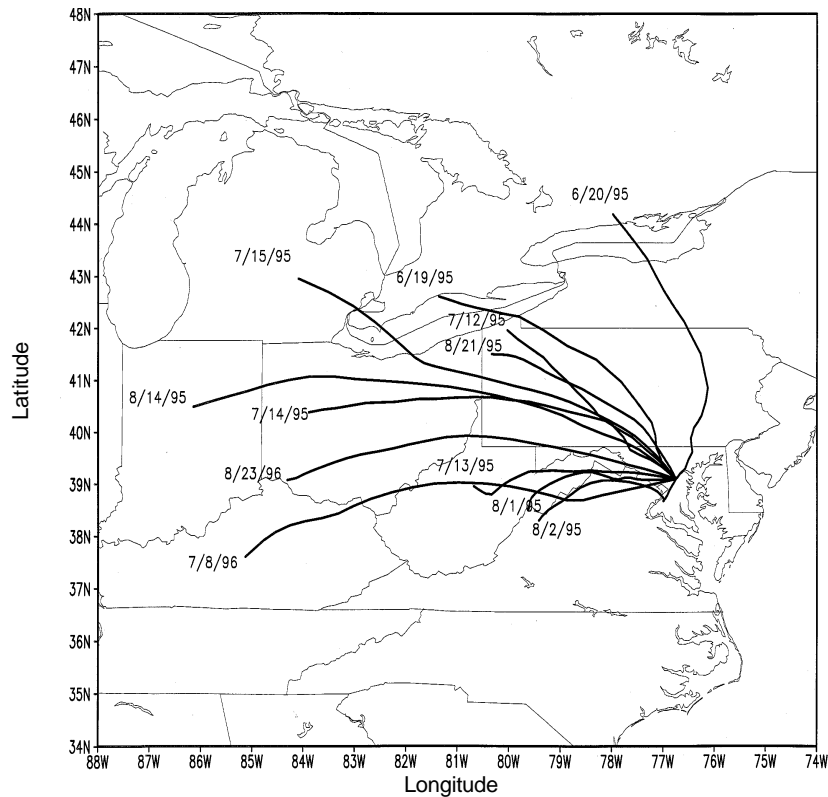


Figure 5-7. Back trajectories at 1500 m msl during ozone episodes in Baltimore, Maryland showing possible transport of pollutants from regions to the west (Ryan et al., 1998).

**Question:** *How do local flow patterns influence ozone 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. Such flows may locally transport pollutants from upwind sources to downwind cities or recirculate pollutants within metropolitan areas. Whatever the flow processes are in your area, understanding them will greatly improve your ozone forecasts.

**Technique:** Compute back trajectories on high, moderate, and low ozone days.

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 5-8**. This simple trajectory shows both surface and aloft flow from the northeast portion of the domain.

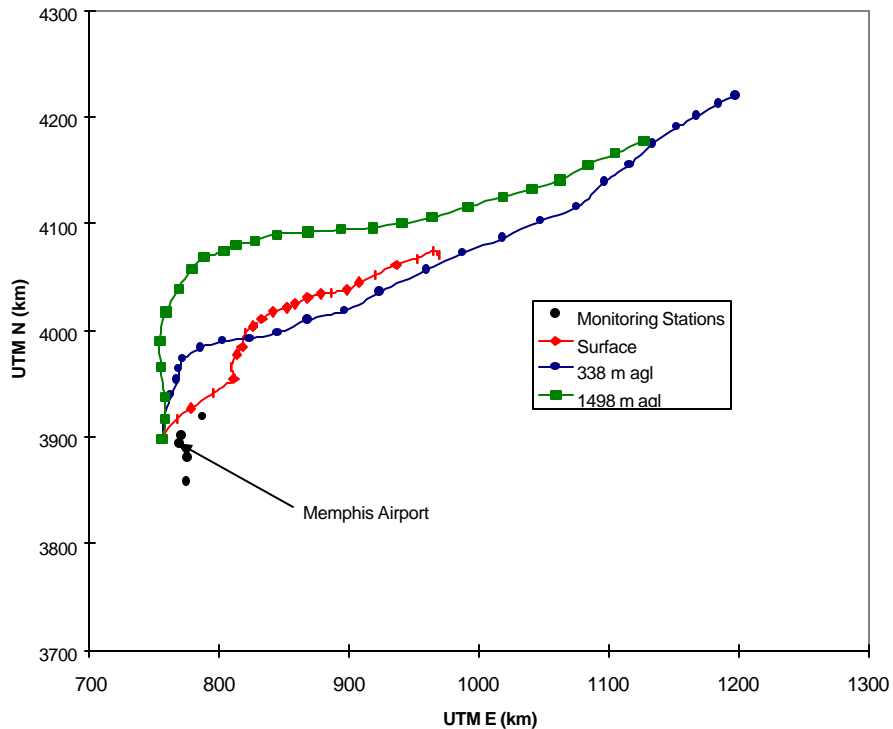


Figure 5-8. A 24-hr back trajectory from Crittenden County, Arkansas starting at 1400 EST on August 25, 1995 and ending at 1300 EST on August 26, 1995. Trajectories were computed using surface wind data from the National Weather Service's (NWS) site at the Memphis airport and upper-air data from a radar wind profiler located at the airport (Chinkin et al., 1998).

**Question:** *How does the aloft temperature structure influence peak ozone concentration?*

**Why:** Aloft temperature structure strongly influences vertical mixing and dilution of pollutants. A stable atmosphere produces less vertical mixing and dilution of ozone and ozone precursors which leads to higher ozone 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.

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<sub>2</sub>; this is a key step in ozone formation. Accurately predicting cloud coverage will improve your forecast accuracy.

**Technique:** The NWS' computer forecast models predict relative humidity at several altitudes with reasonable accuracy. Analyzing these predictions along with satellite images can help you 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, will allow you to identify such conditions in the future.

Once you understand the chemical and meteorological processes that influence ozone, you can start selecting methods to forecast ozone, as discussed in the next section.

### 5.3 CHOOSING OZONE FORECASTING METHODS

Once you understand the needs of your forecasting program, you will need to choose a forecasting method or combination of methods to predict ozone. The method(s) that you choose will primarily depend on the available resources and experience. This section presents a number of issues that you need to consider when selecting an ozone forecasting method.

<i>Resources</i>	Cost may be the major factor that will guide your method selection. When determining the overall cost of a particular method, consider the cost associated with both developing and operating the method. 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.
<i>Severity of problem</i>	The severity of your ozone problem and the frequency of high ozone concentrations in your region will also guide your method choice. 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, if a region experiences many exceedances, several methods may be needed to accurately predict ozone concentrations.
<i>Balancing methods</i>	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.
<i>Adding methods</i>	Once you have selected a forecasting method, your 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 your forecasting program.
<i>Expertise</i>	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

After you select a forecasting method, or methods, you need to address your data needs. Air quality and meteorological data are needed for both developing the method(s) to predict ozone and for operationally forecasting ozone. 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 ozone. 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. Your data needs will depend on the specific meteorological and air quality phenomena to be predicted in your area.

Locating a data source is often a major part of developing a forecasting method. **Table 5-2** lists many of the major sources for obtaining data. Another source of historical data includes past air quality studies that were conducted in your region.

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, Oxides of Nitrogen, Carbon Monoxide, VOCs	Observed	Hourly
Upper-air Meteorology <i>(rawinsondes, radar profilers, and sodars)</i>	Vertical Profiles of WS, WD, T, RH	Observed	Twice-per-day to hourly
Aloft Air Quality Observations <i>(towers, mountains, and aircraft)</i>	Ozone, Oxides of Nitrogen	Observed	Variable
Weather Charts	Surface (WS, WD, T, RH, P) 850 mb (WS, WD, T, Hgt) 700 mb (WS, WD, T, Hgt) 500 mb (WS, WD, T, Hgt), Others	Forecasted and Observed	Twice-per-day
Weather Radar	Precip	Observed	Hourly
Satellite	Cloud Cover (visible and infrared)	Observed	Hourly and/or Sub-hourly
Meteorological Model Forecasts	T, RH, WS, WD, Cloud Cover, Vis, P, others at many levels	Forecasted	Twice-per-day
Text Weather Forecasts	Discussions	Forecasted and Observed	Four or more times per day

WS = wind speed  
WD = wind direction

T = temperature  
RH = relative humidity

Vis = visibility  
P = pressure

Precip = precipitation  
Hgt = height

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

Data Source	Type of Data Source	Types of Data	Phone	Web Site
-------------	---------------------	---------------	-------	----------

U.S. EPA Aerometric Information Retrieval System (AIRS)	Historical	Surface air quality	(703)487-4146	<a href="http://www.epa.gov/ttn/airs">www.epa.gov/ttn/airs</a>
National Climate Data Center (NCDC)	Historical	Surface Meteorology, Upper-air Meteorology, Weather Charts, Radar, Satellite	(828)271-4800	<a href="http://www.ncdc.noaa.gov">www.ncdc.noaa.gov</a>
NOAA National Data Centers (NNDC)	Historical	Surface Meteorology, Upper-Air Meteorology, Weather Charts, Satellite, Radar, Climate	(828)271-4800	<a href="http://www.nndc.noaa.gov">www.nndc.noaa.gov</a>
Regional Climate Centers	Historical	Surface Meteorology, Upper-air Meteorology, Climate Information	Western Regional Climate Center (775)677-3106  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)737-0849  Southern Regional Climate Center (225)388-5021	<a href="http://climate.sage.dri.edu">climate.sage.dri.edu</a>  <a href="http://hpccsun.unl.edu">hpccsun.unl.edu</a>  <a href="http://mcc.sws.uiuc.edu">mcc.sws.uiuc.edu</a>  <a href="http://sercc.dnr.state.sc.us/sercc.html">sercc.dnr.state.sc.us/sercc.html</a>  <a href="http://met-www.cit.cornell.edu/nrcc_home.html">met-www.cit.cornell.edu/nrcc_home.html</a>  <a href="http://maestro.srcc.lsu.edu/srcc.html">maestro.srcc.lsu.edu/srcc.html</a>
Purdue University	Historical and Real-time	Surface Meteorology, Upper-air Meteorology, Satellite, Radar, Model Forecast, Text Weather Forecast	-	<a href="http://wxp.eas.purdue.edu/archive/index.html">wxp.eas.purdue.edu/archive/index.html</a>
Commercial Weather Service Providers (WSP)	Real-time	Surface Meteorology, Upper-Air Meteorology, Weather Charts, Satellite, Radar, Model Forecast, Text Weather Forecast	-	Comprehensive list of WSPs at: <a href="http://www.ugems.psu.edu/~owens/WWW_Virtual_Library/commercial.html">www.ugems.psu.edu/~owens/WWW_Virtual_Library/commercial.html</a>



Most forecasting programs require many types of data from different 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:

<i>Cost</i>	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 your access to it may suffer from Internet outages. Weather Service Providers (WSPs) supply weather data to TV 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 your computer. Reliability for this type of service is generally very high.
<i>Reliability</i>	Knowing that your data will always be available when you need it is critical to your program's success. Unreliable data will reduce forecast effectiveness and may lower your accuracy.
<i>Quality control</i>	Receiving higher quality data (data with fewer errors and inconsistencies) decreases the necessity for personnel to thoroughly review the data before it is used. We recommend that all historical data be reviewed for quality prior to developing a forecasting method.
<i>Data formats</i>	Seek to limit the types of data formats you use. This will help decode and process the data more efficiently. Different time standards, reporting units, quality control codes, etc., can produce additional decoding/processing effort.
<i>Ingest methods</i>	Determining how data will flow from the source to your 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.
<i>Hardware requirements</i>	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.
<i>Redundancy</i>	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 ozone.

## **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, you should prepare and test written procedures that the forecast team can follow on a regular basis. Your forecasting protocol will likely include:

- Descriptions of the meteorological conditions that produce high ozone concentrations in your 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 your 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 you to quickly identify low ozone days, thus allowing time for the high ozone, more difficult forecasts.

Forms and worksheets for documenting data, forecast information, forecast rationale, and comments, which 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 your forecast recipients.

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

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.

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

By this point, you should understand the needs of forecast users (Section 5.1), know how and why ozone forms and how weather affects it (Section 5.2), have chosen methods to predict ozone (Section 5.3 and Section 4), determined your data needs (Section 5.4), and documented the steps necessary to produce the forecast.

Next, it is important to evaluate how well you forecast ozone, which is the focus of the next section.

## 5.6 FORECAST VERIFICATION

Verification is the process of evaluating the quality of a forecast by comparing the predicted ozone to the observed ozone. As part of a forecasting program, forecasters should regularly evaluate the forecast quality. The benefits of verifying your ozone 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,
- Comparing your verification statistics to those from other agencies that forecast ozone.

The verification process can be complex since there are many ways to evaluate a forecast including, accuracy, bias, and skill. No one statistic can fully reflect the performance of a program so you need to compute many verification statistics in order to evaluate completely the quality of your forecast program.

Two basic types of forecasts exist: discrete forecasts of specific concentrations and category forecasts (e.g., good, moderate, etc.). Verification statistics differ for these two types of forecasts. This section explains how you can compute and interpret verification statistics for both types of forecasts and

provides a schedule for verifying your forecasts (Section 5.6.1). If you make discrete forecasts, read Section 5.6.2 to understand the verification statistics. If you make category forecasts, read Section 5.6.3.

### 5.6.1 Forecast Verification Schedule

Evaluate your forecasts frequently to identify any problems or downward performance trends. A schedule of verification tasks follows:

<i>Daily</i>	If forecasts were significantly missed (off by more than 30 ppb or two categories), then examine what caused the missed forecast. Write a forecast retrospective, which is a several page document that details what went wrong and includes recommended changes to forecast methods or procedures. <b>Figure 5-9</b> shows an example outline for a forecast retrospective.
<i>Monthly</i>	Compute the forecast verification statistics described in this section. Compare these with statistics from previous months and review statistics with forecasters.
<i>Annually</i>	Compute the forecast verification statistics described in this section. Compare these statistics from previous years and review statistics with forecasters.

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

Figure 5-9. Example outline of a forecast retrospective.

### 5.6.2 Verification Statistics for Discrete Forecasts

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

<i>Accuracy</i>	Average "closeness" between the forecast and observed values.
<i>Bias</i>	Indicates, on average, if the forecasts are under- or over- predicted.
<i>Skill score</i>	Percentage improvement of a forecast with respect to a reference forecast (typically a climatology or persistence forecast).
<i>Correlation</i>	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.

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 ( $F_{\text{Pers}}$ ) discussed in Section 4.1. A reference forecast is a baseline against which to compare your forecast. Any other forecast method can be used, but typically persistence, climatology, and random chance are used as reference forecasts.

For the 11-day period, the forecaster's accuracy was 8 ppb, meaning that, on average, the forecasts were within  $\pm 8$  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 percent, which indicates that the forecasts are a 40 percent improvement over 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.

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

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

Statistic Name	What it measures	How to compute it	Equation	Units	How to interpret
Accuracy (mean absolute error)	Average "closeness" between the forecast and observed values. Summarizes the overall quality of the forecasts.	<ol style="list-style-type: none"> <li>1. Take the <u>absolute</u> difference between forecast (<math>f</math>) and observation (<math>o</math>) for all forecasts (<math>N</math>).</li> <li>2. Sum the differences and divide by <math>N</math>.</li> </ol>	$A = \frac{1}{N} \left( \sum_1^N  f - o  \right)$	Units of variable (ppm, ppb, etc.)	<ul style="list-style-type: none"> <li>• Lower numbers are best.</li> <li>• Values indicate the uncertainty in any single forecast.</li> </ul>
Bias (mean error)	It indicates, on average, if the forecasts are under predicted or over predicted.	<ol style="list-style-type: none"> <li>1. Take the difference between forecast (<math>f</math>) and observation (<math>o</math>) for all forecasts (<math>N</math>).</li> <li>2. Sum the differences and divide by <math>N</math>.</li> </ol>	$B = \frac{1}{N} \left( \sum_1^N (f - o) \right)$	Units of variable (ppm, ppb, etc.)	<ul style="list-style-type: none"> <li>• Values near 0 are best.</li> <li>• Values &lt;0 indicate under-forecasting (i.e., forecasts are too low).</li> <li>• Values &gt;0 indicate over-forecasting.</li> </ul>
Skill score	Percentage improvement of a forecast with respect to a reference forecast (typically a climatology or persistence forecast).	<ol style="list-style-type: none"> <li>1. Compute accuracy for a reference forecast (<math>A_{ref}</math>), either climatology or persistence (see Section 4.1 for details).</li> <li>2. Compute accuracy (<math>A</math>) for your forecast.</li> <li>3. Divide accuracy by the reference accuracy and subtract from 1.</li> </ol>	$SS = \left( 1 - \frac{A}{A_{ref}} \right) * 100$	%	<ul style="list-style-type: none"> <li>• 0% indicates no improvement (or skill) over the reference forecast.</li> <li>• 50% or more indicates a significant improvement in skill.</li> </ul>
Correlation	A measure of the relationship between forecasts and observations. It measures if two sets of data change together.	<p>Use correlation functions in spreadsheet or statistics software (Excel, Lotus) or compute using the following:</p> <ol style="list-style-type: none"> <li>1. Compute the co-variance (Cov(<math>f, o</math>)) using the equation.</li> <li>2. Compute the standard deviation for the forecasts (<math>sf</math>) and observations (<math>so</math>).</li> <li>3. Divide the co-variance by the product of the standard deviations to compute the correlation (<math>C_{fo}</math>).</li> </ol>	$C_{fo} = \left( \frac{Cov(f, o)}{sf \cdot so} \right)$ <p>where: <math>-1 &gt; C_{fo} &lt; 1</math></p> $Cov(f, o) = \frac{1}{N} \left( \sum_{j=1}^N (f_j - \mu_f)(o_j - \mu_o) \right)$ <p>where <math>\mu_o</math> = mean observed value</p> <p><math>\mu_f</math> = mean forecasted value</p> $sf = \sqrt{\frac{N \cdot \sum_{j=1}^N f_j^2 - \left( \sum_{j=1}^N f_j \right)^2}{N(N-1)}}$ $so = \sqrt{\frac{N \cdot \sum_{j=1}^N o_j^2 - \left( \sum_{j=1}^N o_j \right)^2}{N(N-1)}}$	-	<ul style="list-style-type: none"> <li>• Values close to 1 are best.</li> <li>• Positive correlation indicates that large forecast values are associated with large observed values.</li> <li>• Negative correlation occurs when small values of one set are associated with large values of the other.</li> <li>• No correlation occurs when values in both sets are unrelated.</li> <li>• High correlation does not necessarily denote high accuracy.</li> </ul>

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<sub>Pers</sub>). Accuracy (A), accuracy for persistence forecast (A<sub>Pers</sub>), bias (B), skill-score (SS), and correlation (C) were computed using the equations provided in Table 5-4.

Date	F (ppb)	O (ppb)	F <sub>Pers</sub> (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<sub>Pers</sub>=14 ppb  
 B=3 ppb  
 SS=40%  
 C=0.77

### 5.6.3 Verification Statistics for Category Forecasts

This section describes numerous verification statistics you can compute for category forecasts and provides several illustrative 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 frequency table (also called a contingency table) is the first step in evaluating a category forecast. **Figure 5-10** shows a frequency table of the 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.

		Forecasted Exceedance	
		no	yes
Observed Exceedance	no	a	b
	yes	c	d

Figure 5-10. Contingency table for a two-category forecast.

Using the contingency table shown in Figure 5-10, a perfect forecast program would have values in cells "a" and "d" only. In the real world, imperfect forecasts result in values in cells "b" and "c." Thus, the verification statistics listed in **Table 5-6** are used to evaluate the quality of two-event categorical forecasts. The statistics include:

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

The most important statistics for evaluating the success of the 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 5-11**. 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.



Table 5-6. Verification statistics used to evaluate two-category forecasts. Lower case letters in the equations correspond to those in Figure 5-10.

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.	$A=(a+d)/N *100$	%	<ul style="list-style-type: none"> <li>Higher numbers are better.</li> <li>For example, 65 means that 65% of the forecasts were correct in predicting ozone above or below a given threshold and 35% of the forecasts missed.</li> </ul>
Bias (B)	Indicates, on average, if the forecasts are under predicted (false negatives) or over predicted (false positives).	Divide the number of forecasted high ozone events (cells b plus d) by observed high ozone events (cells c plus d).	$B = \frac{b + d}{c + d}$	-	<ul style="list-style-type: none"> <li>Values closer to 1 are best.</li> <li>Values &lt;1 indicate under-forecasting (i.e., the event occurred more often than it was forecasted).</li> <li>Values &gt;1 indicate over-forecasting.</li> </ul>
False Alarm Rate (FAR)	The percent of times a forecast of high ozone 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).	$FAR=(b)/(b+d) *100$	%	<ul style="list-style-type: none"> <li>Smaller values are best.</li> <li>0=no false alarms (perfect forecast of high events).</li> <li>50 means that half of the forecasts for high ozone did not materialize.</li> </ul>
Critical Success Index (CSI), also called Threat Score	How well the high ozone events were predicted. Useful for evaluating rarer events like high ozone days. It is not affected by a large number of correctly forecasted, low-ozone events.	Divide the number of high ozone "hits" (cell d) by the total number of forecasts plus the number of misses (cells b, c, and d).	$CSI = d/(b+c+d)*100$	%	<ul style="list-style-type: none"> <li>Higher numbers are best.</li> <li>For example, 66% indicates that two-thirds of the forecasts for high ozone were correct.</li> </ul>
Probability of Detection (POD)	Ability to predict high ozone events (i.e., the percentage of forecasted high ozone events that actually occurred).	Divide the correct forecasts of high ozone (cell d) by the total number of observed high ozone events (cells c plus d).	$POD=d/(c+d)*100$	%	<ul style="list-style-type: none"> <li>Higher numbers are best.</li> <li>For example, 70% indicates that 7 in 10 forecasts of high ozone actually occurred.</li> </ul>
Skill Score (SS)	Percentage improvement of a forecast with respect to a reference forecast, typically a climatology, chance, or persistence forecast.	<ol style="list-style-type: none"> <li>1. Compute accuracy for a reference forecast (<math>A_{ref}</math>), such as chance, climatology, or persistence (see Section 4.1 for details).</li> <li>2. Compute accuracy (A) for your forecast.</li> <li>3. Divide accuracy by the reference accuracy and subtract from 1.</li> </ol>	$SS = (1-A/A_{ref}) *100$	%	<ul style="list-style-type: none"> <li>0% indicates no improvement (or skill) over the reference forecast.</li> <li>50% indicates a significant improvement in skill.</li> </ul>

Program LM			Program SC		
Forecasted			Forecasted		
	No	Yes		No	Yes
No	130	8	No	160	11
Yes	20	22	Yes	6	3
A	84		A	91	
B	0.71		B	1.56	
FAR	27		FAR	79	
CSI	44		CSI	15	
POD	52		POD	33	
SS	78		SS	87	
A <sub>ref</sub>	50		A <sub>ref</sub>	50	

Figure 5-11. 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 under-forecast an event (false-negative) or over-forecast 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 percent versus 27 percent, respectively. This means that 79 percent 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 percent, meaning that only 15 percent 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 percent. 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., chance) is 50 percent. As with accuracy, Program SC has the higher skill score; it represents an 87 percent 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.

### 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 5-12**, the accuracy (A) of the entire forecasting program is computed using the following equation:

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

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 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 5-12, a standard four-event table). You will then be evaluating the ability to forecast above or below this value, 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 5-10. To do this, assign each cell 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.
  - 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 forecast statistics described in Table 5-6.

These forecast evaluation measures let you objectively quantify how well you forecast ozone. They should become a regular part of your forecasting program.

		Forecasted			
		Good	Moderate	Unhealthy for Sensitive Groups	Unhealthy
Observed	Good	k	l	m	n
	Moderate	o	p	q	r
	Unhealthy for Sensitive Groups	s	t	u	v
	Unhealthy	w	x	y	z

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

## 6. REFERENCES

- 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, 134-141.
- Blackadar A.K. (1957) Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bull. Am. Meteorol. Soc.* 38, 283-290.
- 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, San Diego, CA, p. 562.
- 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., Hurwitz 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.
- Chu S.H. (1987) Coupling high pressure systems and outbreaks of high surface ozone concentration. Proceedings of the 80th APCA Annual Meeting, pp. 87-113.
- Chu S.H. and Doll D.C. (1991) Summer blocking highs and regional ozone episodes. Preprints of the Seventh Joint Conference on Applications of Air Pollution Meteorology with AWMA American Meteorological Society, pp. 274-277.
- Clark R.D. (1997) Vertical profiles of meteorological variables and ozone concentrations in the nocturnal boundary layer at Gettysburg, PA. Proceedings of the 12th Symposium on Boundary Layers and Turbulence, Vancouver, BC, August, 1997, pp. 417-418.
- 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, 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, 301-312.
- Conroy D. (1998) U.S. Environmental Protection Agency, Region 1. Personal communication.
- Draxler R.R. and Hess G.D. (1997) Description of the Hysplit4 modeling system. NOAA Technical Memorandum ERL ARL-224, December 24.
- Dudhia J. (1993) A non-hydrostatic version of the Penn State/NCAR mesoscale model: validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.* 121, 1493-1513.

- 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.
- Dye T.S., Reiss R., Kwiatkowski J.J., MacDonald C.P., Main H.H., and Roberts P.T. (1998) 8-Hr and 1-Hr ozone exceedances in the NESCAUM region (1993-1997). Report prepared for the Northeast States for Coordinated Air Use Management, Boston, MA by Sonoma Technology, Inc., Petaluma, CA, STI-998100-1810-FR, September.
- Environ (1998) User's Guide - Comprehensive Air Quality Model with Extensions (CAMx). Version 2.0. Environ International Corporation, Novato, CA, December.
- 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, June, 2627-2636.
- Grell G.A., Dudhia J., and Stauffer D.R. (1994) A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). Prepared by National Center for Atmospheric Research, Boulder, CO, NCAR Technical Note-398.
- 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.
- Hyde P. and Barnett B. (1998) Compliance with the 1-hour and 8-hour ozone standards. Paper no. 98-TA32.05 presented at Air & Waste Management Association's 91st Annual Meeting and 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, <http://www.ozoneaction.com/new/org/orgarchive/episctrl.htm>, February.
- Lambeth B. (1998) Texas Natural Resource Conservation Commission, Austin, TX. Personal communication.
- 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.
- Murphy A.H. (1991) Forecast verification: its complexity and dimensionality. *Mon. Wea. Rev.* 119, 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, 281-293.

- Murphy A.H. and Winkler R.L. (1987) A general framework for forecast verification. *Mon. Wea. Rev.* 115, 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.
- Pagnotti V. (1987) A meso-meteorological feature associated with high ozone concentrations in the Northeastern United States. *J. Air Pollut. Control Assoc.* 37, 720-722.
- 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.
- Ray S.E., Dye T.S., Roberts P.T., and Blumenthal D.L. (1998) Analysis of nocturnal low-level jets in the Northeastern United States during the summer of 1995. Paper No. 5A.4 presented at the 10th Joint Conference on the Applications of Air Pollution Meteorology, Phoenix, AZ, January 11-16, (STI 1750).
- 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, 143-149.
- Ryan W.F. (1994) Forecasting severe ozone episodes in the Baltimore metropolitan area. *Atmos. Environ.* 29, 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, American Meteorological Society, Boston, MA, pp. 168-172.
- Samson P.J. (1978) Nocturnal ozone maxima. *Atmos. Environ.* 12, 951-955.
- SCAQMD (1997) 1997 revision to the Air Quality Management Plan - Appendix II: current air quality. South Coast Air Quality Management District, Diamond Bar, CA.
- Seinfeld J.H. and Pandis S.N. (1998) Atmospheric chemistry and physics: from air pollution to global change. J. Wiley, New York.
- Steenburgh W.J. and Onton D.J. (1996) An evaluation of a real-time mesoscale prediction system based on the Penn State/NCAR mesoscale model. In Preprints. Fifteenth Conference on Weather Analysis and Forecasting, Norfolk, VA, August 19-23, American Meteorological Society.
- Stoeckenius T. (1990) Adjustment of ozone trends for meteorological variation. Presented at the Air and Waste Management Specialty Conference, Tropospheric Ozone and the Environment, Los Angeles, CA, March.
- Taylor R. (1998) New York State Department of Environmental Conservation, Albany, NY. Personal communication.

- 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, Research Triangle Park, NC EPA-600/R-98/069(a), June.
- Wallace J.M. and Hobbs P.V. (1977) Atmospheric Science. Academic Press, New York.
- Wilks D.S. (1995) Statistical methods in the atmospheric sciences. Academic Press, San Diego, CA, 467 pp.