NATIONAL COOPERATIVE
HIGHWAY RESEARCH PROGRAM REPORT **238**

ESTIMATING EXCEEDANCES AND DESIGN VALUES FROM URBAN OZONE MONITORING NETWORK DATA

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

TRANSPORTATION RESEARCH BOARD EXECUTIVE COMMITTEE 1981

Officers

Chairman
THOMAS D. LARSON, *Secretary, Pennsylvania Department of Transportation*

Vice Chairman
DARRELL V MANNING, *Director, Idaho Transportation Department*

Secretary
THOMAS B. DEEN, *Executive Director, Transportation Research Board*

Members
RAY A. BARNHART, *Federal Highway Administrator, U.S. Department of Transportation* (ex officio)
ROBERT W. BLANCHETTE, *Federal Railroad Administrator, U.S. Department of Transportation* (ex officio)
FRANCIS B. FRANCOIS, *Executive Director, American Association of State Highway and Transportation Officials* (ex officio)
WILLIAM J. HARRIS, JR., *Vice President, Research and Test Department, Association of American Railroads* (ex officio)
J. LYNN HELMS, *Federal Aviation Administrator, U.S. Department of Transportation* (ex officio)
PETER G. KOLTNOW, *President, Highway Users Federation for Safety and Mobility* (ex officio, Past Chairman, 1979)
ELLIOTT W. MONTROLL, *Chairman, Commission on Sociotechnical Systems, National Research Council* (ex officio)
RAYMOND A. PECK, JR., *National Highway Traffic Safety Administrator, U.S. Department of Transportation* (ex officio)
ARTHUR E. TEELE, JR., *Urban Mass Transportation Administrator, U.S. Department of Transportation* (ex officio)
JOHN F. WING, *Senior Vice President, Booz, Allen & Hamilton, Inc.* (ex officio, MTRB liaison)
CHARLEY V. WOOTAN, *Director, Texas Transportation Institute, Texas A&M University* (ex officio, Past Chairman 1980)
GEORGE J. BEAN, *Director of Aviation, Hillsborough County (Florida) Aviation Authority*
THOMAS W. BRADSHAW, JR., *Secretary, North Carolina Department of Transportation*
RICHARD P. BRAUN, *Commissioner, Minnesota Department of Transportation*
ARTHUR J. BRUEN, JR., *Vice President, Continental Illinois National Bank and Trust Company of Chicago*
LAWRENCE D. DAHMS, *Executive Director, Metropolitan Transportation Commission, San Francisco Bay Area*
ADRIANA GIANTURCO, *Director, California Department of Transportation*
JACK R. GILSTRAP, *Executive Vice President, American Public Transit Association*
MARK G. GOODE, *Engineer-Director, Texas State Department of Highways and Public Transportation*
WILLIAM C. HENNESSY, *Commissioner, New York State Department of Transportation*
ARTHUR J. HOLLAND, *Mayor, City of Trenton, New Jersey*
JACK KINSTLINGER, *Executive Director, Colorado Department of Highways*
MARVIN L. MANHEIM, *Professor, Department of Civil Engineering, Massachusetts Institute of Technology*
DANIEL T. MURPHY, *County Executive, Oakland County Courthouse, Michigan*
RICHARD S. PAGE, *General Manager, Washington (D.C.) Metropolitan Area Transit Authority*
PHILIP J. RINGO, *Chairman of the Board, ATE Management and Service Co., Inc.*
MARK D. ROBESON, *Chairman, Finance Committee, Yellow Freight Systems, Inc.*
GUERDON S. SINES, *Vice President, Information and Control Systems, Missouri Pacific Railroad*
JOHN E. STEINER, *Vice President, Corporate Product Development, The Boeing Company*

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Transportation Research Board Executive Committee Subcommittee for NCHRP
THOMAS D. LARSON, *Pennsylvania Dept. of Transp.* (Chairman)
DARRELL V MANNING, *Idaho Transp. Dept.*
FRANCIS B. FRANCOIS, *Amer. Assn. State Hwy. & Transp. Officials*
RAY A. BARNHART, *U.S. Dept. of Transp.*
ELLIOTT W. MONTROLL, *National Research Council*
CHARLEY V. WOOTAN, *Texas A&M University*
THOMAS B. DEEN, *Transportation Research Board*

Field of Special Projects
Project Panel, SP20-14A

EARL C. SHIRLEY, *California Dept. of Transportation* (Chairman)
DENIS E. DONNELLY, *Colorado Dept. of Highways*
AL HAMMACK, *Bessent, Hammack & Ruckman, Inc.*
RODERICK D. MOE, *Texas State Dept. of Hwys. and Public Transp.*
RONALD J. PIRACCI, *New York State Dept. of Transportation*
STEPHEN C. SNYDER, *Santa Fe Industries, Inc.*
JOHN SUMMERHAYS, *U.S. Environmental Protection Agency*
HOWARD A. JONGEDYK, *Federal Highway Administration*
STEPHEN E. BLAKE, *Transportation Research Board*

Program Staff

KRIEGER W. HENDERSON, JR., *Director, Cooperative Research Programs*
LOUIS M. MacGREGOR, *Administrative Engineer*
CRAWFORD F. JENCKS, *Projects Engineer*
R. IAN KINGHAM, *Projects Engineer*
ROBERT J. REILLY, *Projects Engineer*
HARRY A. SMITH, *Projects Engineer*
ROBERT E. SPICHER, *Projects Engineer*
HELEN MACK, *Editor*

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT 238

ESTIMATING EXCEEDANCES AND DESIGN VALUES FROM URBAN OZONE MONITORING NETWORK DATA

**F. L. LUDWIG, H. S. JAVITZ, A. VALDES,
C. MAXWELL, R. PATTERSON, J. A. LEE,
E. SHEBR, and R. L. MANCUSO
SRI International
Menlo Park, California**

RESEARCH SPONSORED BY THE AMERICAN ASSOCIATION OF STATE HIGHWAY AND TRANSPORTATION OFFICIALS IN COOPERATION WITH THE FEDERAL HIGHWAY ADMINISTRATION

AREAS OF INTEREST:
ENERGY AND ENVIRONMENT
(HIGHWAY TRANSPORTATION)

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. NOVEMBER 1981

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as: it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to its parent organization, the National Academy of Sciences, a private, nonprofit institution, is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the Academy and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the Academy and its Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

NCHRP REPORT 238

Project 20-14A FY '79
ISSN 0077-5614
ISBN 0-309-03170-2
L. C. Catalog Card No. 81-85313

Price: $9.60

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council, acting in behalf of the National Academy of Sciences. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the National Academy of Sciences, or the program sponsors.

Each report is reviewed and processed according to procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

The Transportation Research Board evolved from the 54-year-old Highway Research Board. The TRB incorporates all former HRB activities and also performs additional functions under a broader scope involving all modes of transportation and the interactions of transportation with society.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America.

FOREWORD

By Staff
Transportation
Research Board

This report will be of principal interest to transportation administrators, planners, and technologists in state and local levels of government. State highway program managers and local transportation planning directors will find the summary to describe concisely the capabilities that technologists have to determine areawide base-line ozone conditions against which the impacts of proposed transportation actions should be evaluated. Air quality technologists will find a new methodology for calculating ozone exceedances and design values from observed data for a network of sites throughout an urban area.

Federal and state regulations mandate air quality studies to provide the basis for evaluating transportation strategies to assist in achieving compliance with ozone standards. These air quality studies must frequently include monitoring programs to determine ozone concentrations and the degree of compliance with the air quality standards. However, use of historical data, wherever possible, is a more efficient and practical means to quantify ozone problems and minimize requirements for special monitoring. The major objective of the study described here has been to develop the methods needed to analyze existing data and obtain as much information as possible from those data. Corollary objectives have been to provide information about additional data needs and ways in which those needs can be met with the least additional monitoring.

This report presents the concepts of design value and expected number of exceedances (of the 120 ppb standard) as defined by the U.S. Environmental Protection Agency (EPA) for a single monitoring site and extends the concepts to a network of sites describing conditions throughout an extended urban area. Using those definitions, methods are described for calculating values from observed data. The methods apply nonlinear interpolation and Monte Carlo simulations (based on empirically derived conditional-probability distributions) to locate areas where the greatest numbers of exceedances and highest design values are expected and to calculate the values in those areas. The methods were applied to an extensive data base collected during special monitoring programs in four different urban areas: Houston, Los Angeles, Philadelphia, and St. Louis.

The appendixes to the report provide all the necessary instructions (including computer programs) for applying the methods. The instructions encompass:

- Acquisition, screening, and preprocessing of data.
- Application of methodologies and computer programs for estimating design values and number of exceedances.
- Determination of the origins and concentrations of background ozone entering the city.
- Design of supplemental monitoring programs.

CONTENTS

ACKNOWLEDGMENTS

The research reported herein was conducted under NCHRP Project 20-14A by SRI International. Frank L. Ludwig, Staff Scientist, was the principal investigator and author of the report. Other authors of the report are H. S. Javitz, Director, Data Design and Analysis Department; C. Maxwell, Statistical Analyst; A. Valdes, Statistical Analyst; R. Patterson, Senior Environmental Scientist; J. A. Lee, Scientific Programmer; E. Shelar, Research Meteorologist; and R. L. Mancuso, Senior Research Meteorologist. Valuable data were provided by Mssrs. Dale Coventry, Norman Possiel, William Lonneman, and John Summerhays of the U.S. Environmental Protection Agency; Mr. Gary Tannahill of Radian Corp.; Mr. Don Perrine of the California Air Resources Board; and Mr. Rod Moe of the Texas State Department of Highways and Public Transportation. The technical comments of the members of the NCHRP review panel were very helpful.

ESTIMATING EXCEEDANCES AND DESIGN VALUES FROM URBAN OZONE MONITORING NETWORK DATA

SUMMARY

Federal and state regulations mandate air quality studies to develop strategies for achieving compliance with ozone standards. These air quality studies must frequently include monitoring programs to determine ozone concentrations and the degree of compliance with the air quality standards. However, use of historical data, wherever possible, is a more efficient and practical means to quantify ozone problems and minimize requirements for special monitoring. The major objective of the study described here has been to develop the methods needed to analyze existing data and obtain as much information as possible from those data. Corollary objectives have been to provide information about additional data needs and ways in which those needs can be met with the least additional monitoring and to develop methods that provide information necessary for developing effective control strategies. No procedures were developed for designing control strategies, which is probably the major topic where further research will be fruitful.

All the objectives of the study were met. The rather specialized definitions of "design value" and "expected number of exceedances" that were developed by the U.S. Environmental Protection Agency (EPA) for the case of a single monitor have been broadened for application to regionwide conditions. The approach was to use available data to estimate numbers of exceedances and design values for points throughout the region of interest. Initially, points are very widely spaced in order to provide an overall picture of the distribution of these two parameters in the area. The points for which estimates are obtained are then more densely spaced in those areas where the highest design values and the greatest numbers of exceedances have been estimated. In this way it has been possible to estimate the maximum numbers of exceedances and the highest design values occurring in the area and the region in which they are found. These values satisfy the definitions of expected number of exceedances and design value that were derived for a network. Computer programs have been written for processing data to obtain the estimates previously discussed. These computer programs and the directions for their use, which are included in the appendixes to the report, are among the major products of this study.

The deterministic approach to the estimation of regional design values and expected numbers of exceedances served as the basis for a probabilistic approach which used the day-to-day estimates for each grid point of values generated by the deterministic method as a basis for developing conditional probability distributions of ozone concentration. Monte Carlo simulations were used to generate daily estimates of peak-hour ozone concentrations at key locations (those areas where higher design values and greater numbers of exceedances were expected and which had no nearby monitors), based on observed data. This probabilistic method provides a measure of the uncertainty and variability in the deterministic approach. The computer program and directions for its use to obtain the probabilis-

tic estimates of design value and expected numbers of exceedances are included in the appendixes to this report.

The methods developed here not only provide estimates of design value and expected numbers of exceedances for the region, but also identify those days when the highest concentrations occurred, which, in turn, allows the analyst to determine the meteorological conditions associated with high ozone concentrations in the region. The air quality data and meteorological information for the high-ozone days can be examined and used to estimate the transported background-ozone concentrations entering the region. The estimation methods are fully described in this report. The determination of the origins of the precursors to the transported ozone through air trajectory analysis is also discussed.

The foregoing methods were applied to data from four urban regions: Houston, St. Louis, Philadelphia, and Los Angeles. Each of these areas had relatively dense ozone monitoring networks that had been operated for at least a few months. With the availability of data from these unusually dense monitoring networks, the method could be applied to determine the sensitivity of the results to the number of stations in the monitoring network. A network of about ten sites was found to be adequate if the sites are properly located. There is a tendency to underestimate the expected number of exceedances when the number of monitoring sites is reduced. However, the design-value estimates are generally within the range of estimates for a single site, as derived from different EPA-recommended methods.

Studies have shown that a complete monitoring network need not be operated throughout the year. There is a close relationship between peak-ozone value and maximum temperature; if data are collected for all days when the maximum temperature in the region exceeds about 20C (68F) the estimates of design value and expected numbers of exceedances will be accurate. It appears that the most efficient way to collect adequate ozone-monitoring data in an urban region is to operate about five fixed stations: one in the central part of the city and four in different directions a few tens of kilometers outside the highly urbanized region. This fixed network should be supplemented by mobile monitors operated during warm weather to fill-in the area between the central monitor and the peripheral monitors in the downwind direction.

CHAPTER ONE

INTRODUCTION AND RESEARCH APPROACH

Federal and state regulations currently require that air quality studies be conducted to develop strategies to obtain compliance with ozone standards. These air quality studies must frequently include a monitoring program to measure urban ozone concentration and determine the degree of compliance with the air quality standards. Efficiency and other practical considerations dictate that as much historical data as possible be used to quantify the ozone problem and minimize requirements for special monitoring. Methods are needed to analyze existing data and obtain as much information as possible from them and to specify requirements for additional data. Furthermore, because of the requirements for compliance, the conditions that are associated with the exceedance of standards must also be defined.

The general objective of the research reported here was to develop methods to quantify the ozone problem. Specifically, methods were needed to evaluate existing data; to produce estimates of important parameters, such as transported ozone, exceedances, and design values for ozone; and to indicate the amount and location of additional monitoring that may be required.

Some of the parameters, such as number of exceedances and design values, have rather specialized definitions that were developed by the U.S. Environmental Protection Agency (EPA); those specialized definitions needed to be translated into specific data analysis and interpretation requirements. In particular, it was necessary to develop definitions of design value and expected number of exceedances that are applicable to ozone concentrations through a large urban area and its surroundings and, yet, are consistent with the original definitions that apply to a single monitoring site.

During the course of the work, one other important constraint was observed: All the methods, computer programs, and other project results were developed so that they would be usable by others. It was assumed that the intended users would have access to reasonably large computing facilities and that they had technical training, but not particularly strong backgrounds in statistics, meteorology, air quality analysis, or the other disciplines that contributed to development of the methods described here.

The approach taken to fulfill the research objectives was through the accomplishment of the following six major research tasks:

1. Methods for assessing data adequacy were developed.
2. Automated procedures for estimating exceedances and design values of ozone were devised.
3. The estimation methodologies were applied to existing data bases to test their performance.
4. Guidelines for supplemental monitoring were established, based on the applications of the methods to selected data bases.
5. A procedure was developed for estimating ozone levels that are transported into a city, especially during periods of high ozone concentration.
6. Potential applications of the methodologies to transportation planning were noted.

This report describes the foregoing efforts.

Chapter Two contains extended discussions of the theory underlying many of the practical results, including the original definitions of design value and expected number of exceedances that have been developed for single monitoring sites by the U.S. Environmental Protection Agency (EPA). Because the fundamental definitions are so important, Chapter Two contains comprehensive discussion of the original EPA definitions, the alternatives for expanding those definitions to a whole network of monitors, the rationale for the definitions that were chosen, and objective methods by which those definitions can be applied to monitoring data collected from stations in a monitoring network.

Chapter Two also outlines some of the criteria by which the adequacy of a data base can be judged. In that regard, it anticipates the material in Chapter Three, which provides an overview of how the findings are applied and then proceeds to interpret the results of the application of the methods to four specific urban data bases that had unusually dense ozone monitoring networks. The interpretation of the results includes discussions of the effect of station density, which bears on the adequacy of the data base. Chapter Three also deals with the problem of temporal completeness and defines conditions under which one can be assured that the ozone standard will not have been exceeded and, hence, those conditions under which monitoring will not be necessary.

Throughout the course of the research, the ultimate goal has been to define impacts of transportation projects and to work toward strategies (especially as they involve the transportation system) that will correct air quality problems. The results provide the methods required to define current ozone problems, but they do not address the design of appropriate control measures. The final chapter of this report suggests that additional research will be necessary to extend the results in that direction. Chapter Four also includes other suggestions for further research and reviews the conclusions that can be drawn from the work.

The appendixes to the report serve as a complete guide to the practical application of the major project results. Appendixes A through D list and describe the necessary computer programs along with directions for their use. Suggestions are also given for those modifications that may be needed to implement the computer codes on different types of machines, or to apply them to somewhat different data types.

In some cases, subjective methods seemed to be more useful than objective computerized techniques. For example, determining transported and background ozone (App. E) or designing supplemental monitoring programs (App. F) requires judgmental decisions based on whatever relevant information may be at hand. The amount and kind of information vary from case to case. The design of supplemental monitoring programs must sometimes be performed within budgetary, political, or geographical constraints that are impossible to computerize. Thus, some of the procedures described in the appendixes can give only step-by-step guidance and definitions of decision criteria. Although it would be simpler for the user if objective computer schemes were available, some decisions are best made on the basis of informed good judgment.

The appendixes are intended to be usable with little or no reference to the theory and the results of applications that are discussed in the body of the report; however, reading the whole report before using the methods should improve the user's interpretation of the results that are obtained from the methods. To expedite publication the appendixes included herein are reproduced as submitted by the research agency.

CHAPTER TWO

FINDINGS

DEFINITION OF EPA METHODS

Determining Expected Number of Exceedances

In the "Guideline for the Interpretation of Ozone Air Quality Standards" (*1*), the EPA presents a relatively simple procedure for determining whether a monitoring site is in compliance with the National Ambient Air Quality Standard (NAAQS) requirement that the expected number of yearly exceedances be one or less. This section explains this procedure, summarizes the EPA comments concerning its application, and provides some examples. It begins by discussing the EPA method for determining the expected number of exceedances for a single year of data. This is followed by a discussion of the EPA method for determining the expected number of exceedances when more than one year of data is available.

The determination of the expected exceedances for a single year uses the following formula:

$$e = V + (V/n)(N - n - z) \qquad (1)$$

in which

N = the number of required monitoring days in the year;
n = the number of valid daily maxima;
V = the number of measured daily values above the level of the standard;
z = the number of days assumed to be below the standard level; and
e = the expected number of exceedances for the year.

A few additional comments are warranted concerning the definition of the terms in the formula. N is always equal to the number of days in the year less the number of days for which monitoring requirements have been waived by the appropriate Regional Administrator (*1*, p. 10).

The number of valid daily maxima (n) is the number of days in which at least 75 percent of the hourly values from 9:01 am to 9:00 pm local standard time (LST) were measured or at least one hourly value exceeded the level of the standard (*1*, pp. 3-4).

A day may be assumed to be below the standard if: the day does not meet the criteria for having a valid daily maxima, both the preceding and following days have valid daily maxima, and the daily maxima on both the preceding day and the following day do not exceed 75 percent of the level of the standard (*1*, p. 9).

Finally, the data set used in the formula will be deemed "adequate" only if it is 75 percent complete for the peak pollution potential seasons (*1*, p. 11).

It is instructive to perform a sample calculation. Consider a hypothetical ozone monitor (e.g., in Chicago) where 293 valid daily maxima were observed during 1978. Of these, 2 days had daily maxima above the standard level. An additional 30 days were exempt from monitoring because of a monitoring waiver granted by the EPA Regional Administrator. Finally, suppose that 16 of the remaining 42 days without valid daily maxima may be assumed to be below the standard level. The expected number of exceedances during 1978 would then be

$$\begin{aligned} e &= 2 + (2/293)(335 - 293 - 16) \\ &= 2.18 \end{aligned}$$

Thus, for this particular year the expected number of exceedances is greater than one. The expected number need not be an integer.

The determination of the expected exceedances in multiple-year data must be addressed because the EPA (*1*, p. 12) recommends that a data set spanning 3 years be used to determine compliance with NAAQS. For example, to determine the compliance of a particular station with the NAAQS expected-exceedance requirement on 1 January 1981, the EPA recommends that a data set spanning 1 January 1978 through 31 December 1980 be compiled. For each of the three calendar years preceding 1 January 1981, the expected number of yearly exceedances is computed applying the formula given above. If the average of these three expected numbers of yearly exceedances is 1.0 or less, the site is said to be in compliance.

The following simplified example of this type of calculation was derived from material in Ref. (*1*, pp. 14-15). Suppose that a hypothetical site in New York has observed the following for 1978–1980:

> 1978: 365 valid daily maxima; 3 days above the standard level.
> 1979: 285 valid daily maxima; 2 days above the standard level; 21 missing days assumed to be below the standard level.
> 1980: 287 valid daily maxima; 1 day above the standard level; monitoring requirements waived for a 60-day period by the Regional Administrator; seven missing days assumed to be below the standard. Note that 1980 is a leap year with 366 days.

The expected number of exceedances for each of these three years may be derived as follows:

for 1978: $e = 3 + (3/365)(365 - 365 - 0) = 3.0$

for 1979: $e = 2 + (2/285)(365 - 285 - 21) = 2.4$

for 1980: $e = 1 + (1/287)(366 - 287 - 7) = 1.0$

Averaging these three numbers (3, 2.4, and 1.0) gives 2.1 as the estimated expected number of exceedances and completes the required calculation. Because the estimated expected number of exceedances is greater than 1.0, the site fails to comply with the NAAQS expected exceedance required on 1 January 1981.

Reference (*1*) does not definitely answer the question of how complete the data set must be to assert compliance with the NAAQS nor does it provide guidance on how to use the data in a year that is not yet complete. The following suggestions are in general concordance with the views expressed by Curran (*1*):

1. A data set must contain at least three "adequate" years of data before it can be used to demonstrate compliance with the NAAQS. However, noncompliance can be demonstrated with fewer than three "adequate" years of data.
2. There appears to be no reason why noncalendar years may not be used in assessing compliance. For example, to assess the compliance of a site on 15 March 1981 with the NAAQS expected exceedance requirement, a data set spanning 15 March 1978 through 14 March 1981 may be compiled. This data set comprises 3 years of data, where each year spans from 15 March through 14 March. All previously described formulas may be applied to these noncalendar years.

Determining Design Value

Conceptually, the design value for a particular site is the value that should be reduced to the standard level to ensure that the site will meet the standard. If $P(X > c)$ denotes the probability that a randomly selected observation X is greater than a concentration c, the design value d satisfies the relationship $P(X > d) = 1/365 = 0.00274$ (*1*, pp. 17-18).

The EPA appears to have a more flexible attitude toward estimation of the design value than toward determination of the expected number of exceedances. One simple procedure is defined for determining the expected number of exceedances and, thus, for determining compliance with the NAAQS. On the other hand, the design value serves as a means to the end of meeting the applicable air quality standards. According to Ref. (*1*, p. 17), "As long as this final goal is kept in mind, any of the . . . approaches are satisfactory." The analyst is much freer to exercise judgment in the determination of the design value, being able to choose among four methods. In fact, in two of the methods the analyst may choose which statistical distributions to use and how they will be fit. The data requirements are also less stringent for the determination of the design value. More or less than 3 years of data may be used and the data need encompass only 50 percent of the daily maxima for the peak season.

The four basic methods recognized by the EPA for determining the design value are: (1) Use a graph to interpolate the upper 0.274 percentile of the plotted data. (2) Use a table to approximate the upper 0.274 percentile of the data. (3) Fit a distribution to the data set and calculate the upper 0.274 percentile of this distribution. (4) Fit a separate distribution to each year's data, combine these distributions via conditional probabilities, and calculate the upper 0.274 percentile of the combined distribution.

In general, these four methods should give comparable design values. None of the methods require a computer, although the third and fourth methods entail enough calculation to warrant the use of a computer or a programmable calculator. The following paragraphs describe and illustrate the four methods recognized by the EPA.

The *Graphical Method* consists of graphing the observed concentrations to interpolate the upper 0.274 percentile. Usually it is only necessary to plot a half-dozen concentrations to accomplish this interpolation. To illustrate the plotting technique, consider the previous hypothetical of a New York site that monitored 937 daily maxima during 1978, 1979, and 1980, with an additional 88 days assumed to be less than the standard or granted a waiver by the Regional Administrator. The sum of these two counts is 1025. Suppose that the six largest daily maxima were 184, 164, 153, 146, 136, and 129 ppb. These values are plotted in Figure 1 against the upper percentile of the distribution of daily maxima which these values estimate. The highest observed concentration is an estimate of the upper 1/1026 × 100 percentile, the second-highest observed concentration is an estimate of the upper 2/1026 × 100 percentile, and so on. A smooth curve approximating the plotted values has been drawn. The intersection of this curve and the line representing the upper 1/365 × 100 percentile is the estimated design value; in this example the estimated design value is 155.5 ppb.

This EPA method slightly underestimates the design value. Instead of plotting the ith largest concentration against $i/1026$, it is plotted against

$$\frac{i}{937 + 88 \times \dfrac{937}{937 + 70} + 1} = \frac{i}{1019.9}$$

The number 70 in the formula is the number of days with missing data that have neither been waived by the Regional Administrator nor otherwise assumed to be less than the standard. In most cases the difference between the design values obtained using the EPA method and the adjusted method should be small.

The rationale behind this adjustment is best illustrated by the use of an example. In this example assume, for the sake of simplicity, that the design value is defined as the upper 0.4 percent of the yearly distribution of daily maxima rather than the upper 0.274 percent. Suppose that there were 1,000 days in the monitoring period at a monitoring station and that 250 of these 1,000 days have daily observed maxima, 500 were either waived by the Regional Administrator or otherwise assumed to be less than the standard, and 250 were missing daily maxima but could not be assumed to be less than the standard. If the largest observed daily maxima is 129 ppb, the second largest is 118 ppb, the third largest is 106 ppb, and the fourth largest is 97 ppb, then according to the EPA method, these observations should be plotted approximately against 1/751, 2/751, 3/751, and 4/751, respectively. A curve is drawn through the plotted points to assist in interpolation (and smooth out any irregularities in the plotted points). Because these points lie along a smooth curve, and $3/751 \cong 4/1000$, the EPA method estimates the design value to be the third largest observation of 106 ppb.

The adjusted method plots the observations against 1/501, 2/501, 3/501, and 4/501, respectively. Once again these

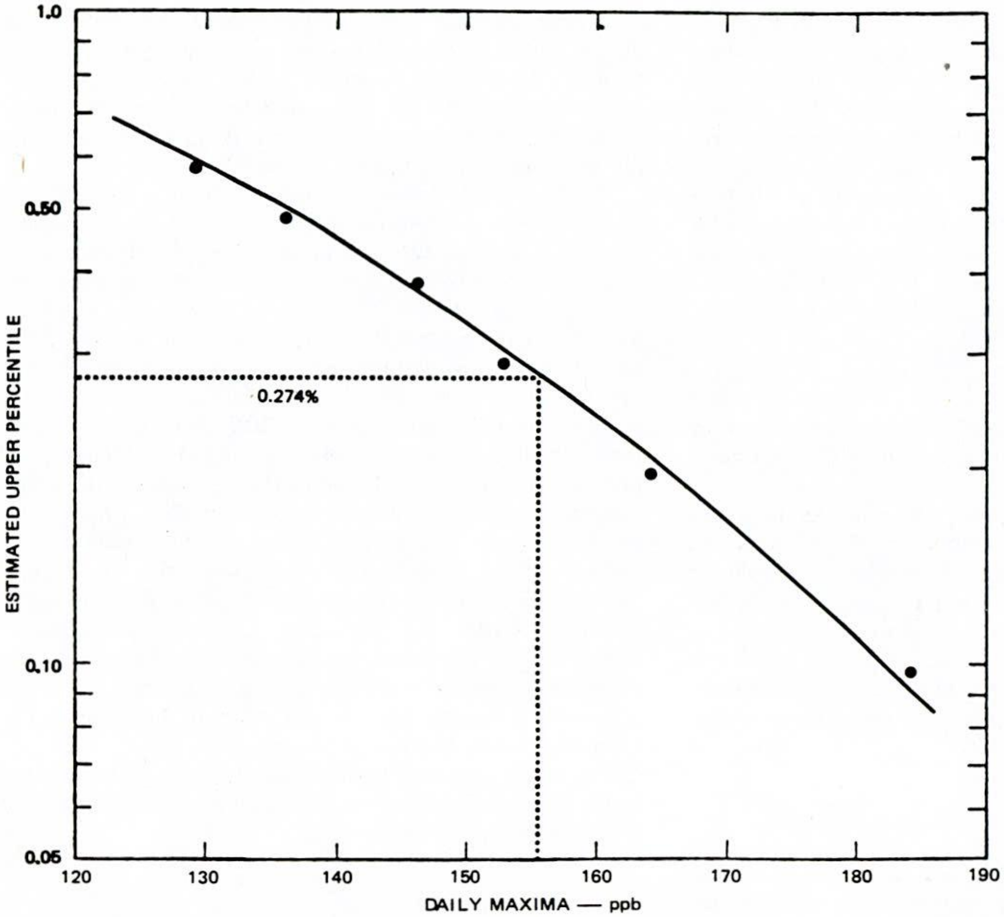


Figure 1. Graphic method applied to hypothetical New York City data.

points lie along a smooth curve, and because 2/501 ≅ 4/1000, the adjusted method estimates the design value to be the second largest observation of 118 ppb.

The EPA method should be corrected to adjust for the 250 days with missing data that cannot be assumed to be less than the standard. The EPA method for calculating the expected number of exceedances does include such an adjustment, under the assumption that the missing days were similar to the observed days. Indeed, if in the current example all 500 of the unwaived days, rather than 250, had been observed, one would have probably seen an additional observation approximately equal to 129 ppb, an additional observation approximately equal to 118 ppb, an additional observation approximately equal to 106 ppb, and so on. In this case, the upper 0.4 percentile would correspond to the fourth largest observation because one would have essentially complete monitoring (all days except those assumed less than the standard) and 4/1001 ≅ 0.4 percent. Because the best estimate of the fourth largest observation out of 500 observations is 118 ppb, this is the estimated design value. Note that this is the value obtained by the adjusted method, but even in this fairly extreme example the difference between the design values obtained by the EPA method and the adjusted method is not large.

The *Table Look-Up Method* recommended by the EPA for determining the design value consists of a "table look-up" (*1*, pp. 24-26), but does not in fact require a table. The table look-up method is conceptually based on the previous graphical method. In the New York example used to illustrate the graphical method, the estimated design value of 155.5 ppb (the upper 0.274 percentile point) lay between the second largest daily maximum of 164 ppb (the 2/1026 × 100 or 0.195 percentile) and the third largest daily maximum of 153 ppb (the 3/1026 × 100 or 0.293 percentile). In the table look-up method, the design value would be estimated using the second largest daily maxima of 164 ppb because it is the larger endpoint of that interpolation interval. Of course, the estimate of 164 for the design value is too large, but it is rapidly obtained.

The general table look-up method is as follows: If the ith largest observed daily maxima would be plotted against an upper percentile less than or equal to 0.274 percent, and the $(i + 1)$th largest observed daily maxima would be plotted against on upper percentile greater than 0.274 percent, the design value is (over) estimated to be the ith largest observed daily maxima.

Stated in this fashion, the table look-up method can be applied using either the EPA graphical method percentile estimates or the adjusted graphical method percentile estimates—although the latter are recommended. If the denominator used in the percentile estimates is between 365 and 729 inclusively, the design value is (over) estimated to be the largest concentration; if the denominator is between 730 and 1094 inclusive, the design value is (over) estimated to be the second largest concentration; if the denominator is between 1095 and 1459 inclusively, the design value is (over) estimated to be the third largest concentration, etc. If the denominator is 364 or less, by convention the design value is (under) estimated to be the largest observed daily maxima. The other methods are more accurate and not much more time-consuming; therefore, use of the table look-up method is not recommended.

The *Distribution Fit Method* recommended by the EPA consists of fitting a distribution to the data set and calculating the upper 0.274 percentile of this distribution (*1*, pp. 21-25). There is ambiguity concerning the amount of the data that must be fit. The illustrations suggest that the entire data base should be used, but the commentary indicates that because of the intended use of the distribution, only the degree of approximation in the top few percent of the data is important. Specifically, the EPA recommends that the fit be carefully examined for the two (or fewer) daily maxima corresponding to upper percentiles less than 0.274 percent, and for the two largest daily maxima corresponding to estimated percentile larger than 0.274 percent. The authors of this report believe that because the emphasis is on the upper tail of the distribution, it is sufficient to fit the distribution through the half-dozen or so largest observations. In fact, it may be misleading to do otherwise, because the tail and body of the distribution of daily maxima may be dissimilar.

With the convention that one is fitting the tail of the distribution of daily maxima, the distribution fit method is a variant of the graphical method. An approximation curve was drawn in Figure 1 using the graphical method; the distribution fit method differs from the graphical only in using curves with specific shapes. Assuming that the data follow a particular distribution in the upper tail constrains the shape of the approximating curve. For example, if the tail distribution is assumed to be exponential, the curve on a semilog coordinate graph must be a straight line. In Figure 2 an approximating line was drawn to the same data used in Figure 1 and the estimated design value is found to be 154 ppb. The other two popular distributions—the lognormal and the Weibull—do not correspond to easily specified curves when the upper percentile is graphed in logarithmic units and the daily maxima is graphed in linear units. However, they do correspond to straight lines when graphed on appropriate axes and therefore can also be fit by hand. For the lognormal distribution, the daily maxima should be plotted in logarithmic units and the estimated percentiles should be plotted as Z scores obtainable from a table of the standard normal distribution. (For example, the upper 1.0 percentile corre-

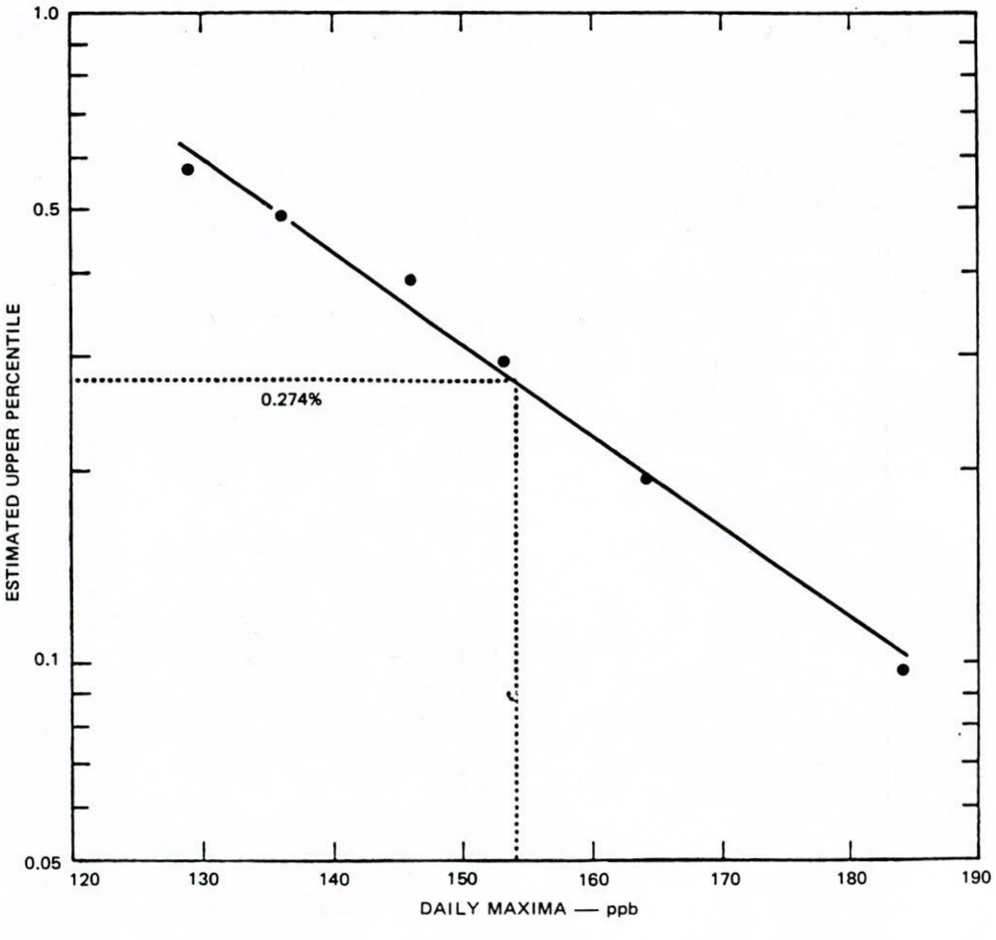


Figure 2. Exponential fit to hypothetical New York City data.

sponds to a Z score of 2.33 and the upper 0.274 percentile corresponds to a Z score of 2.775). Figure 3 shows a plot of the data on the axes appropriate to the lognormal distribution. The approximating line was drawn, yielding the estimated design value to be 152.9 ppb. For the Weibull distribution, the daily maxima should be plotted in logarithmic units against the logarithm of the logarithm of the reciprocal of the estimated upper percentile. For example, $\log[\log(1/0.00274)] = 1.775$. Figure 4 shows the data from the New York example plotted on the axes that are appropriate to the Weibull distribution. The approximating line was fit by eye to give an estimated design value of 155.2 ppb.

The graphical method seems likely to result in as accurate an estimate of the design value as the distribution fit method. If using a nonlinear curve is objectionable because of subjectivity, it is suggested that the exponential distribution be assumed and a straight line be fitted by least squares. Breiman (*2*) has shown that for most air quality data, the exponential distribution fits the upper tail as well as the lognormal or Weibull distributions.

The *Conditional Probability Method* can be used when the individual yearly distributions are dissimilar. The other methods were designed to estimate the design value under the assumption that the data from the separate years could be pooled. Pooling is legitimate if, for example, upper tails of the distributions of daily maxima are comparable over the years in the data base. Comparability can be visually assessed by applying the curve-fitting technique of the graphical method to the (approximately) sixth largest observed daily maxima from each year's data. As long as the curves are similar, the yearly distributions are comparable in the upper tail and can be pooled.

When the yearly distributions of daily maxima are dissimilar, it *may* be necessary to use the EPA method that is based on conditional probabilities to estimate the design value. The word "may" is emphasized because two conditions must be satisfied before the first three methods fail to work well: the yearly distributions of daily maxima must be dissimilar, and the number of days in each year with missing data that cannot be assumed to be below the standard must be substantially unequal. Reference (*1*) does not fully recognize the interrelationships of these two conditions, recommending (*1*, p. 29) that the conditional probability method be used whenever "the number of measurements during the oxidant season differs by more than 20 percent from one year to another." It may not prove worthwhile to use this more complicated method unless the upper 0.274 percentiles of the yearly distributions also differ by more than 10 percent. The EPA (*1*) example of the conditional probability method does not satisfy the second condition—the 3 years in the example data

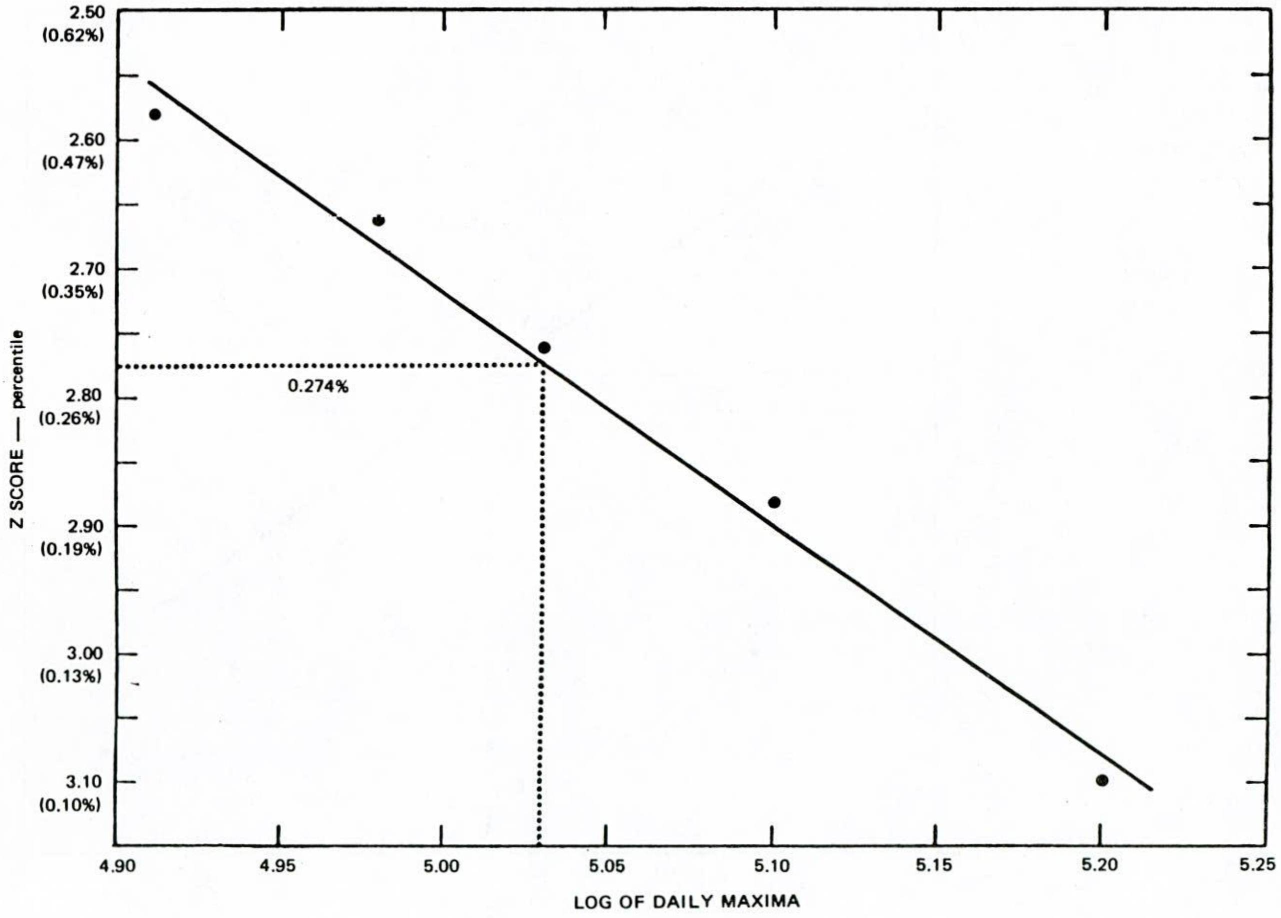


Figure 3. Lognormal fit to hypothetical New York City data.

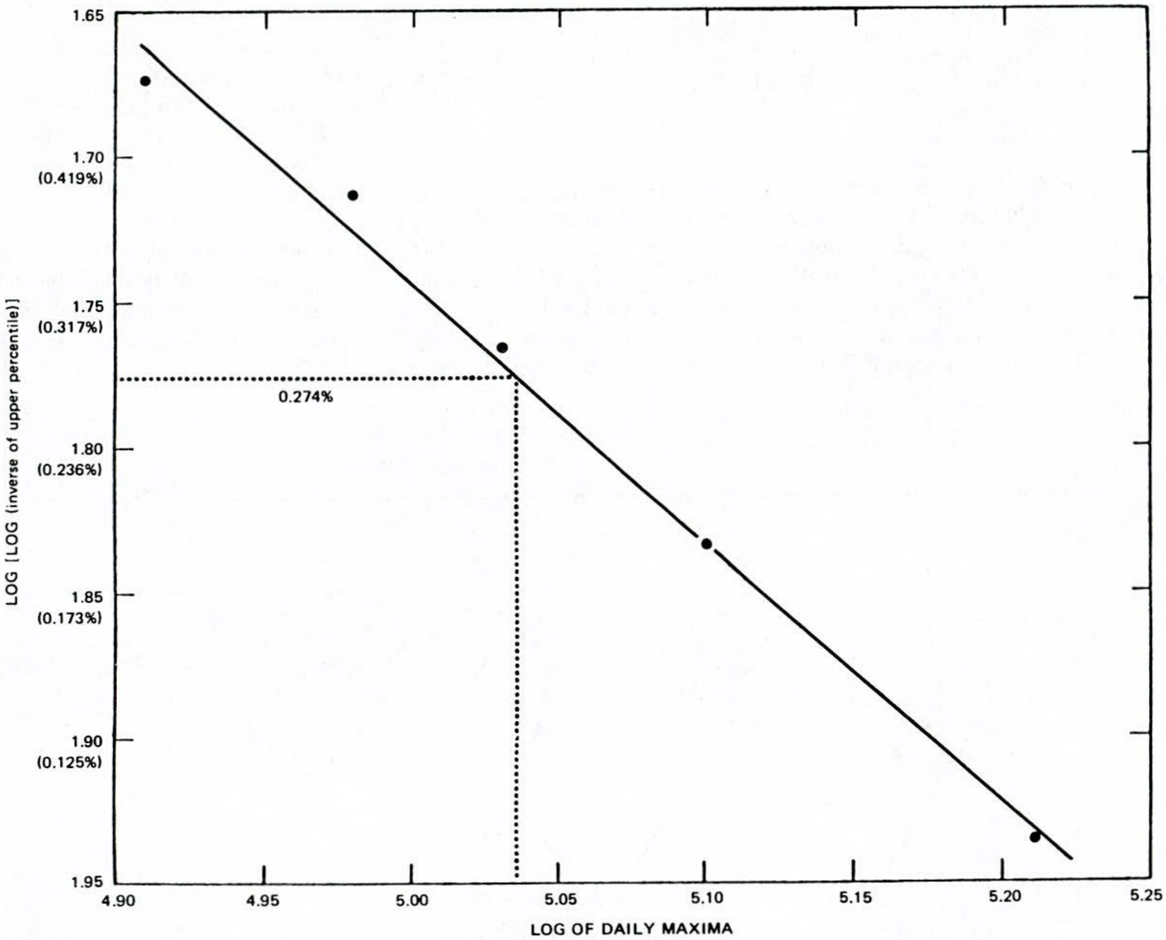


Figure 4. Weibull fit to hypothetical New York City data.

base have 365, 303, and 349 ppb daily values respectively. Thus, even though the distributions are dissimilar, the design value obtained using the first method (147 ppb) is nearly identical to the design value obtained using the conditional probability method (150 ppb).

The conditional probability method is based on the following logical argument. Suppose that a day were arbitrarily selected out of a population of days following the same underlying distribution as the days in the data base. The design value d is defined by the condition that the probability that the concentration X on that day would exceed d is 1/365. Symbolically,

$$P\{X > d\} = 1/365 \qquad (2)$$

Let M be the number of years in the data base. Because X is being drawn from a population of days following the same underlying distribution as the data base, there is a $1/M$th probability that X follows the same distribution as the data from the first year in the data base, a $1/M$th probability that X follows the same distribution as the data in the second year in the data base, and so on. Thus, d satisfies

$$P\{X > d\} = \sum_{i=1}^{M} \frac{1}{M} P\{X_i > d\} = \frac{1}{365} \qquad (3)$$

where X_i is a daily maxima from the ith yearly distribution.

The design value d may be calculated in the following fashion. Define

$P_i(c) = P\{$a daily maximum from the ith yearly distribution exceeds a given concentration value $c\}$

The value of $P_i(c)$ may be determined easily by application of the first or third EPA method for the data for one year. For example, consider a hypothetical monitoring site in Detroit with the following data:

1977: 250 valid daily maxima, 80 days assumed less than standard, 35 days with missing values that cannot be assumed less than the standard, the largest observed daily maxima were 195, 175, 164, 154, and 151 ppb.

1978: 170 valid daily maxima, 30 days assumed less than the standard, 165 days with missing values that cannot be assumed less than the standard, the largest observed daily maxima were 177, 170, 167, 165, and 162 ppb.

1979: 300 valid daily maxima, 50 days assumed less than the standard, 15 days with missing values that cannot be assumed less than the standard, maximum observed daily maxima were 162, 155, 153, 150, and 147 ppb.

These data are plotted in Figure 5 to approximate the yearly distributions. These lines determine the values of $p_i(c)$, as described in the following.

First guess at d is 180 ppb. By using Figure 5 P_{1977} (180 ppb) = 0.00515, P_{1978} (180 ppb) = 0.00337, and P_{1979} (180 ppb) = 0.000385. The average probability is $P\{X > 180 \text{ ppb}\}$ = 0.00297. Second guess at d is 185 ppb. Now P_{1977} (185 ppb) = 0.00430, P_{1978} (185 ppb) = 0.00190, and P_{1979} (185 ppb) = 0.000225. The average probability is $P\{X > 185 \text{ ppb}\}$ = 0.00214. Because this probability is smaller than 0.00274, the design value d must be between 180 ppb and 185 ppb. The third guess at d is 182 ppb; P_{1977} (182 ppb) = 0.00480, P_{1978} (182 ppb) = 0.00270 and P_{1979} (182 ppb) = 0.00310. The average probability is $P\{X > 182 \text{ ppb}\}$ = 0.00260. Thus the design value is between 180 ppb and 182 ppb, and compromising on an approximate design value d = 181 ppb. In contrast to the conditional probability method, the design value could have been computed for the pooled data, as shown in Figure 6. The estimated design value using the first method is 176.5 ppb.

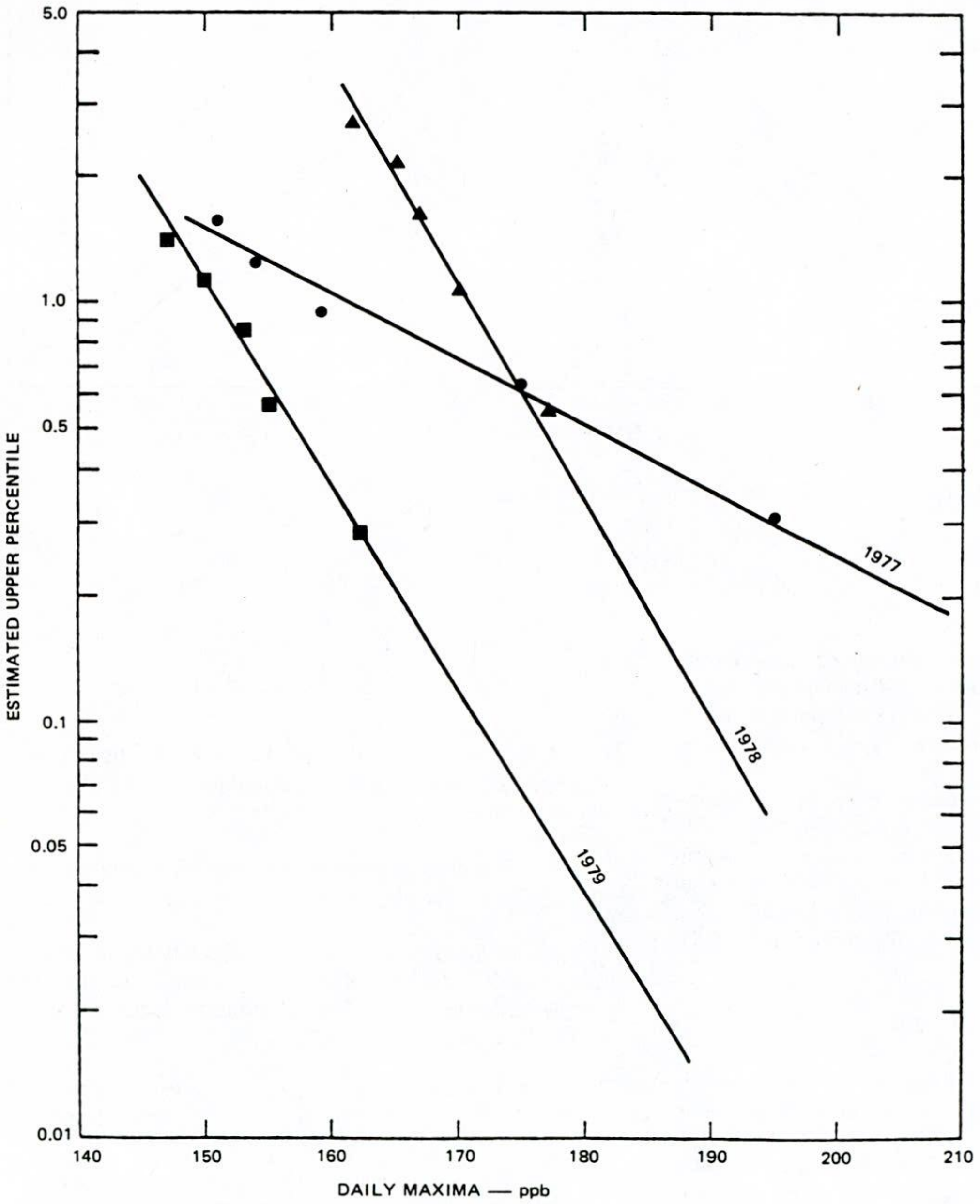


Figure 5. Conditional probability method applied to hypothetical Detroit data.

Extension of the Concept of Design Value to a Spatial Network

To determine whether an entire area is in compliance with the ozone standards, it is necessary to extend spatially the EPA definition of exceedance and design value. A number of alternative definitions are available, each of which simplifies to the EPA definition when there is only a single monitoring site. Table 1 gives the type of matrix of maximum daily ozone values that would be obtained from a spatial network. The numbers on the left-hand vertical axis denote the grid cells of the spatial network, with a total of *N* cells. For example, if the geographic region being modeled were rectangular, say 10 cells by 15 cells in area, *N* would be equal to 150. If each of those 150 cells was subdivided into quarters (i.e., reducing their dimensions by one-half), *N* would be equal to 600. Consequently, the magnitude of *N* depends not only on the total area being modeled but also on the size of the cells, and *N* can become arbitrarily large as the size of each cell shrinks. A few of the grid cells will have a monitoring station in them. It is assumed that a methodology has been developed for estimating the daily maximum ozone concentration for every cell and every day. The numbers on the upper horizontal axis denote the days in the data base, and there are a total of *M* days. For example, if the data base contained 3 years of data, *M* would equal 1095. All of the squares in the matrix are assumed to contain either measured or estimated daily maximum ozone concentrations. The marginal results are displayed on the far right-hand side and the bottom of the table. The design values for each grid cell, obtained using an EPA method on the data in the matrix row representing that grid cell, are displayed on the right-hand side of the table. The highest ozone values over all grid cells on a particular day are displayed at the bottom of the table.

Three basic definitions have been identified for the design value for the spatial network:

1. The maximum of the design values for the individual grid cells (e.g., the maximum value in the right-hand column).
2. The design value (upper 99.726 percentile) of the distribution of daily maxima in the bottom row (i.e., the EPA methodology is applied to the distribution of the *M* daily maximum ozone values, where the maximum extends over both the hours in the day and the grid cells in the spatial network).
3. The design value (upper 99.726 percentile) of the distribution of all daily maxima in the matrix (i.e., the EPA methodology is applied to the distribution of the $N \times M$ daily ozone values in the matrix).

Table 1. Design value matrix for a network.

Grid Cell	Observed Daily Maximums for Designated Day (ppb)							Design Value for Grid Cell (ppb)
	1	2	3	4	...	M-1	M	
1	135	171	210	150	...	82	98	185
2	114	129	220	83	...	57	138	192
3	15	38	72	59	...	101	63	97
N-1	44	62	159	222	...	93	81	201
N	38	115	92	81	...	77	183	156
Maximum over all locations	135	189	237	250	...	101	195	

It can be shown that the design value obtained from the third definition is no larger than the value obtained from the first definition, which in turn is no larger than the design value obtained from the second definition.

Definition 1 has been adopted for a network design value. Definition 1 is the most natural extension of the EPA work because it defines a network design value in terms of the design values of the individual sites and grid cells in the network. In the other two definitions, the network design value is derived without computing the design values for individual sites or grid cells.

Definition 1 is also more closely related to real ozone exposure conditions. The ultimate purpose of air quality standards is to protect living organisms from the adverse effects of pollutant exposure. If the first definition is used and the design value of a network is defined to be the maximum design value over all locations within the network, then one is quantifying the exposure to the organisms that reside at the worst location. The use of a metropolitan design value that was lower than what occurred at some location in the area could mean that the standard had not been attained at all sites, even though the design value had met the standard. If the appropriate regulatory agencies can reduce that network

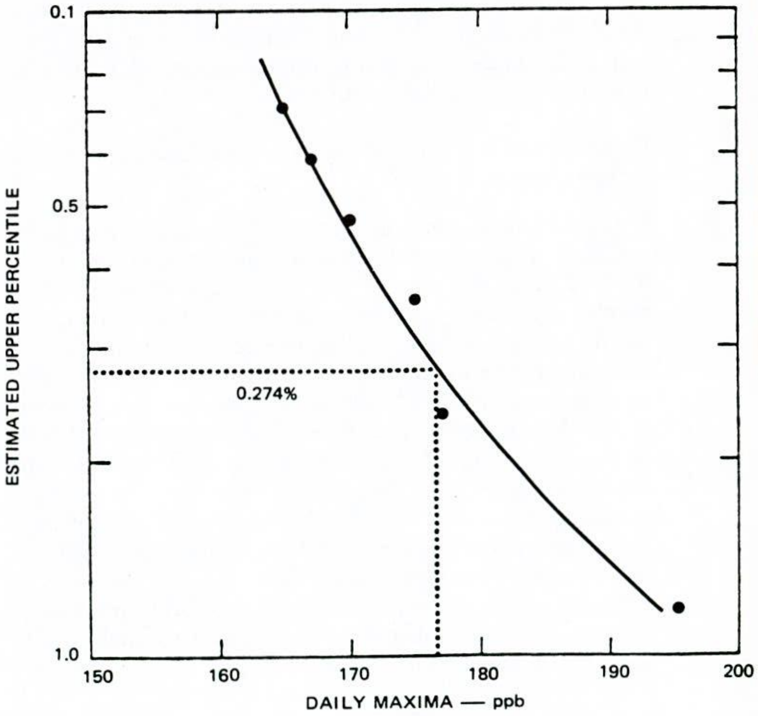


Figure 6. Graphical method applied to pooled hypothetical Detroit data.

design value (using definition 1) to an acceptable standard level, all organisms residing at the worst location, or any other location, are being provided a minimum level of protection.

Generally, organisms that migrate throughout an area are also being provided with the minimum level of protection. For example, human beings tend to cover a wide area in their daily commute to and from work and other activities, and may be exposed to ozone in two, three, or more locations. Their individual exposures tend to be less, however, than if they had stayed in the worst location. That is, humans tend to spend only a portion of their time in highly polluted areas, spending the balance of their time in less polluted areas. The exception may be air pollution monitoring personnel, who actively seek out many areas with the highest concentration levels. These personnel tend to receive a higher exposure than do the organisms who reside in the worst single location. Definition 2 for network design value is appropriate to those personnel because, by construction, it assumes that the organism finds, every day, the location with the highest ozone concentration.

With respect to definition 3, the pooling of the data from all of the grid cells implies an averaging effect. The design value from that definition is appropriate to an organism that has an equal probability of being at any location on any day. Clearly this definition would underestimate the exposure of organisms that spend most of their time in areas with high concentrations.

There are other possible definitions for network design value. For example, the average could be used instead of the maximum in the first and second definitions. The network design value would then be defined as either the average of the design values of the grid cells or the 99.726 percentile of the distribution of the average (over grid cells) of the daily ozone maxima. It is believed that both of these definitions lead to an underestimation of the ozone exposure to living organisms, so they have not been considered.

Extension of the Concept of Expected Exceedance to a Spatial Network

The foregoing results concerning the design value may be modified to accommodate expected number of exceedances. Table 2 gives the type of information regarding expected number of exceedances that would be obtained from a spatial network. The entries in the matrix cells are either a "1," denoting that an exceedance occurred; a "0," denoting that an exceedance did not occur, or was assumed not to occur; or a probability, between zero and one, denoting an estimate of the probability that an exceedance occurred. The row totals, multiplied by $M/365$, are the expected number of exceedances per year at the individual monitoring sites and grid cells. The numbers on the bottom are the maximum values in the columns representing single days.

The three basic definitions for the expected number of exceedances for the spatial network, which are analogous to those presented for the design value, are:

1. The maximum of the expected number of exceedances for the individual grid cells (i.e., the maximum value in the right-hand column).

Table 2. Expected exceedance matrix for a network.

Grid Cell	Probability of an Exceedance for Designated Day							Number of Expected Exceedances for Grid Cell
	1	2	3	4	...	M-1	M	
1	0	0	0.5	0	...	1	0	1.5
2	0	0	0	0.1	...	0.1	0	0.6
3	0	0	0	0	...	0	0	0
N-1	1	0	0	0	...	0.2	0	2.4
N	0	0	0.7	0	...	0.1	0	1.8
Maximum over all locations	1	0	0.7	0.1	...	1	0	

2. The number of days per year in which an exceedance would be experienced by an organism that migrates to the location with the largest probability of an occurrence (i.e., the sum of the numbers in the bottom row multiplied by $M/365$).
3. The average number of expected exceedances per year (i.e., the total of the numbers in the $N \times M$ squares in the matrix multiplied by $M/365N$).

As before, definition 3 yields the smallest value for the expected number of exceedances, and definition 2 yields the largest value. For the reasons given previously, definition 1 has been adopted, which corresponds to the authors' choice for design value definition. The incorporation of these definitions into schemes for estimating design values and exceedances for networks is discussed later.

DETERMINING DATA ADEQUACY

Useful Components for a Data Base

Before applying the definitions, those parameters must be identified which could reasonably be expected to improve one's ability to estimate ozone concentrations at locations from which there are no ozone data. It should be noted that factors influencing ozone measurements at those sites where ozone data are collected are also important, because existing ozone measurements can be used to infer conditions at other locations without monitors. Those parameters which appear, on the basis of past studies, to have the greatest potential for inferring important information about ozone for locations from which there are no data have been identified. The identification of such parameters have been based on known physical or statistical relationships between a parameter and ozone concentration. This knowledge has also been used as the basis for a subjective ranking of the various parameters within each category of parameter (e.g. meteorological data, air quality data, siting information, and so forth). This subjective ranking system forms the basis for evaluating data bases.

Table 3 summarizes those elements that may have some importance in a data base and provides some comments regarding the reasons for the subjective ratings given in the table. The last two categories of data given in the table—emissions and site descriptions—are fundamentally different from the air quality and meteorological data be-

Table 3. Important elements of a data base to be used for determining design values for ozone.

Data Category	Parameter	Subjective Rating of Importance[1]	Remarks
Air quality	Ozone	A+	This is the parameter of concern and some data must be available.
	NO, NO_2, NO_x	B+	Titration of ozone by NO causes large negative correlations.
	NMHC	B	NMHC and NO_x are both precursors to ozone formation.
	CO, SO_2, Pb	C	Of little direct use; may be important as surrogates for NO_x and NMHC.
	TSP	D	Generally heavily influenced by sources that are not relevant to ozone.
Meteorological	Wind (speed/direction)	A-	While not as essential as ozone data, wind direction is needed to determine which observations are influenced by nearby emissions areas.
	Temperature	B	Strong correlations between temperature and ozone are frequently observed and might prove useful (3,4).
	Mixing heights	C+	Affects dilution and also vertical distribution of ozone.
	Stability	C+	Similar in effects to mixing depth; will relate to how well surface O_3 observations describe conditions aloft (5).
	Solar radiation	C	Affects ozone formation but may be redundant if temperature information is also available.
	Air trajectories	B	Probably valuable for determining sources of transported ozone.
Emissions	NO_x, NMHC	B+	Emissions data will not be archived directly with other types of data, but spatial and temporal (average diurnal, weekly, and annual cycles) are likely to be very important to the development of techniques for identifying transported ozone.
Site description	Location	A	Knowledge of where measurements are made is essential.
	Descriptions of sites and surroundings	B+	Very important to data interpretation; could be interferences.

[1] Ratings:
A = absolutely essential
B = very valuable
C = useful
D = of marginal use

cause they are not being used or archived on an hourly or daily basis. They represent supplemental data that are assumed to be reasonably stable with time. Undoubtedly there are temporal variations, especially in emissions, but it must be assumed that for any given season, day of the week, and hour of the day, the emissions are relatively constant. Thus, these data will be tabulated separately as average values (spatially distributed) with appropriate seasonal, weekly, and diurnal correction factors. For permanent monitoring stations, the site location remains unchanged, but some temporary or mobile monitors might have to be treated specially to incorporate information about their location at any given time. Furthermore, data of those types are not likely to be incorporated directly into the process of estimating design values and exceedances, but will be necessary for developing control and monitoring strategies.

Table 3 contains elements that are needed to develop and evaluate a methodology for estimating ozone exceedances and design values; most are not required to apply such methodologies. None of the data items is particularly extraordinary; virtually all are easily available for the large- and medium-sized urban areas that will be of greatest interest. Some elements, such as mixing height and stability, may have to be derived from other directly measured meteorological parameters. However, the techniques for doing this are readily available in both cases. Holzworth (*6*) gives methods for determining mixing height from radiosonde information. Methods for deriving stability category from conventional meteorological observations have been described by Turner (*7*) among others. Air trajectories must be calculated, and the computer program of Heffter and Taylor (*8*) is available for this purpose. Although solar radiation is sometimes measured, a methodology for determining design values that is to be widely applied should not depend on such measurements because they are not widely available. However, there are methods for estimating insolation categories from cloud-cover observations and solar-elevation angle (7). All the items in Table 3 not only are linked physically or statistically to ozone, but can be considered for use in practical methodologies because they are generally either measured or calculated from parameters that are measured.

Most of the data listed in the table are needed only for a few selected days. The methodology described later identifies key days when concentrations were high enough to affect design-value calculations. The meteorological data, air quality, and other ozone measurements are very important for these days, because they are needed to define contributions of transported ozone and the processes that contributed to the formation of high ozone concentrations. These data are also needed to develop control strategies that would be effective in limiting ozone concentrations for the kinds of conditions observed on the key days.

Ensuring Data Quality

General

The major concern regarding quality of the data is with the ozone measurements because those data will be directly involved in the estimation of exceedances and design values. Barring gross errors, the other data will be of less concern because they are not involved directly in the estimation procedure. The two aspects of quality that are of greatest concern are the accuracy and precision of the data and the representativeness of the data. By accuracy, is meant the systematic errors or bias that may be found in the measurements; precision is used to denote random errors. The assessment of data quality focuses on data anomalies, experimental procedures, and representativeness. The importance of these considerations are discussed in the following.

Detection of Anomalies

Data anomalies are most easily identified by careful examination of the complete data set to determine if the observed values are consistent with those at other locations and times and for other pollutants. The following approach can be used for examining the consistency of the data:

1. Hourly ozone concentrations are graphed for days when the ozone concentration exceeded 120 ppb.
2. Other related parameters that might be available are graphed.
3. The plots are scanned visually to identify questionable data (according to criteria discussed below).
4. Any other supplementary information that might be available is consulted to determine whether the data should be accepted, rejected, or flagged as questionable.

Visual scanning is one of the quickest and most effective ways for identifying anomalous data points. Such scanning is frequently enough to identify misplaced decimal points or other malfunctions of the data collection or archiving system. Of course, a specific analysis of outliers, as suggested by the EPA (9), would be very helpful. That document provides the guidance and computer programs necessary to identify extreme values of doubtful validity.

The graphing of other parameters is often useful for assessing the reasonableness of the ozone data. For example, the chemical links among NO, NO_2, and ozone can be invoked to evaluate reasonableness of changes in ozone concentrations relative to changes in concentrations of the other two pollutants. As another example, a sudden change in ozone concentration that accompanied a sudden shift in wind direction might be considered more reasonable if the wind had suddenly shifted so that it came from an area of precursor emissions.

The following are examples of attributes that would cause data to be suspect: sudden increases, drops, or spikes in the ozone data record; very high concentrations occurring immediately before or following a period without data, such as might be indicative of an instrument malfunction; very high concentrations at only one station in a network of stations; and very high ozone concentrations, especially when wind, temperature, or NO data suggest that high ozone concentrations are unlikely. In the last-cited criterion, low temperatures or high NO concentrations would suggest low ozone concentrations; wind directions that were not from an area of known precursors would do likewise.

A single location monitoring high ozone concentrations, while all others in the region are observing low concentrations, would generally be suspect because ozone is a secondary pollutant—not emitted directly, but formed from other pollutants—so it is not subject to small-scale, localized pockets of high concentrations. Also, the precursors from which it has been formed will have been subjected to considerable mixing during the time taken for the ozone to form.

Another consideration in detecting ozone anomalies in the prevailing meteorological situation is that certain meteorological conditions are known to be associated with low ozone concentrations. Thus, for example, an observation of more than 120 ppb ozone when the temperature is below 20 C (68 F) is suspect. Even if such an observation proves valid, it may have been caused by natural processes and, hence, should be subject to exclusion from consideration when control strategies are developed.

Precision of Measurement and Calibration Procedures

The screening procedures previously described are directed at identifying anomalous data points that indicate concentrations greater than what was probably present in the ambient air. Those screening procedures will not be able to identify deviations from true values that arise from the fundamental imprecision of the measurement process. The literature contains some information from which it is possible to estimate lack of precision. The recommended calibration procedure for ozone monitoring has been described in the *Federal Register (10)*. The procedure requires that a stable ozone concentration be generated and measured by ultraviolet (UV) photometry. The photometer used must have a precision of ± 5 ppb, or 3 percent of the concentration, whichever is greater. Such a calibration procedure would result in an uncertainty of between about 5 and 10 ppb, depending on concentration. If it is assumed that similar imprecision is associated with UV instruments that are used for monitoring, one might deduce the probable imprecision of the data derived from such monitors. The instruments used for routine monitoring might well have somewhat larger uncertainties because they do not receive the special care and maintenance that is prescribed by the regulations for the instrument that is used in establishing the standard concentrations. If it is assumed that the uncertainties add as the square root of the sum of their squares, the total uncertainty in this type of measurement is likely to be between 7 and 15 ppb.

The other major type of ozone monitor, the chemiluminescent instrument, has a noise level of around 1 or 2 ppb according to the manufacturer's specifications that have been published by the Lawrence Berkeley Laboratory (*11*). That same document indicates a reproducibility for this type of instrument of about 5 to 10 ppb, with similar deviation from linearity. Thus, the precision of a carefully operated chemiluminescent ozone instrument could be reasonably estimated to be about 10 ppb. In combination with the recommended calibration procedures, one might expect a well-operated chemiluminescent instrument to measure ozone concentrations within about 10 to 20 ppb.

The preceding discussion defines the precision expected for carefully operated ozone monitoring networks. The data analysis procedures will have to recognize the limitations in the data, and the data screening and selection procedures will have to include recommendations for estimating how carefully operated a monitoring system might be.

Before the new calibration procedure was recommended in the *Federal Register* (*10*), other calibration methods were used. Beard (*12*) analyzed the differences between readings obtained using gas-phase titration (GPT) of excess ozone with nitric oxide compared with calibration with the UV photometry method. His results indicated that the gas-phase titration calibration method resulted in readings that were 2 to 7 percent higher than were obtained using UV photometric calibration from the same gas mixtures. On the other hand, Hogeson (*13*) found "excellent agreement between O_3 determination by UV or GPT." It appears that if there is an inaccuracy in data obtained using the gas-phase titration method of calibration, it is likely to be a relatively small one.

The accuracies resulting from the other major class of calibration techniques in use prior to 1979 were not generally as good as those obtained using the gas-phase titration calibration technique. Three iodometric techniques were examined by Hogeson (*13*) and compared with the gas-phase titration and the UV calibration methods. The three techniques were: the EPA technique using 1 percent neutral-buffered potassium iodide reagent; a California Air Resources Board method that uses 2 percent neutral-buffered potassium iodide reagent, and the Los Angeles Air Pollution Control District method using 2 percent unbuffered potassium iodide. The following relationships with UV measurements were derived by linear regression:

Gas-phase titration (GPT):

$$[O_3]_{GPT} = 1.09[O_3]_{UV} - 3 \text{ ppb}$$

California Air Resources Board (CARB):

$$[O_3]_{CARB} = 1.29[O_3]_{UV} - 5 \text{ ppb}$$

EPA:

$$[O_3]_{EPA} = 1.24[O_3]_{UV} - 35 \text{ ppb}$$

Los Angeles Air Pollution Control District (LAAPCD):

$$[O_3]_{LAAPCD} = 0.96[O_3]_{UV} - 32 \text{ ppb}$$

The foregoing expressions are applicable to 50 percent relative humidity. Hogeson concluded from this analysis (*13*) that the neutral-buffered potassium iodide procedures indicated ozone concentrations that were from 18 to 30 percent higher than those determined by UV procedures in the presence of moist air. Judging by this, there was a systematic bias toward higher readings that arose from older calibration procedures. Thus, the calibration procedures that were used should enter into the process of judging data quality and interpreting the observations.

Representativeness of Siting

The remaining factor that needs to be considered in evaluating data quality is representativeness. The authors of this report have drawn heavily on Ludwig's and Shelar's (*14, 15*) studies of representativeness as it affects the selection of ozone monitoring sites and the determination of ozone exceedances. Basically, they develop criteria of two sorts. One set of criteria is used to prevent undue influence by nearby emissions on the measured values. The other criteria are used to identify areas where the highest concentrations are likely to occur, or where reliable measurements of background ozone can be made.

The identification of sites without significant local influence is not overly difficult for ozone because, as noted earlier, ozone tends to be more evenly distributed than are primary pollutants, and there are few significant sources emitting ozone directly into the atmosphere. However, negative influences are a problem. Ozone reacts rapidly with NO, which is emitted by motor vehicles and major point sources, so ozone concentrations in the vicinity of such sources will be lower than is representative for the area. Ludwig and Shelar (*14*) present rules for separating ozone monitoring sites from NO sources that depend mostly on average daily traffic.

The best locations of suitable sites for measuring background and peak value ozone concentrations are also described by Ludwig and Shelar (*14*). Basically, background values are best measured outside the urban area in the upwind direction (upwind referring to wind directions during those conditions most conducive to ozone formation). Locations to one side of the urban area (relative to these wind directions) are also acceptable, but are not as desirable as an upwind site. The peak concentrations are most apt to be in the downwind direction (again, for those conditions associated with photochemical ozone production) and several tens of kilometers from the city. Monitoring at such sites allows sufficient time for the ozone to form and reach peak values. The selection of an appropriate combination of background and "worst case" sites is discussed in greater detail in Appendix F.

Techniques for Evaluating Data Bases

The preceding sections discussed factors that affect data quality and that are apt to contribute to its usefulness for determining exceedances and design values. Obviously, any procedure for evaluating an ozone data base will include a comparison between the characteristics of that data base and those characteristics deemed to be either desirable or undesirable in the data base. Table 3 rated the various elements that might be available in the data base. The measurement methods and the calibration and maintenance procedures also affect the accuracy and precision of the data, as noted earlier, which affect the usefulness and applicability of the data base and the procedures for determining design values and exceedances. This, in turn, will affect the requirements for supplemental data collection and must be considered in applying the methodology.

The final element in the evaluation procedure relates to the locations of stations and the completeness of the data base, in both the spatial and temporal dimensions. The following questions must be answered. Are the stations placed so that ozone concentrations in air being transported into an area can be identified regardless of wind direction and for wind directions most frequently associated with exceedances? Are the stations located near where one might expect to find the greatest ozone concentrations produced from precursors emitted within the area of interest regardless of wind direction and for those wind directions most likely to be associated with high ozone concentrations? Are there a sufficient number of well-placed stations to define the large-scale features of ozone distribution through the area? Were most of the monitoring sites operating on those days most likely to have experienced ozone concentrations in excess of the standard? Do the operational and calibration procedures meet EPA requirements?

In addition to those questions, which are important to the application of the methodologies described later, other questions are important when the results are obtained and used to develop control strategies. For example, are measurements of precursors available from key areas? Are wind measurements that characterize the general air flow in the area available? Are meteorological data available that can be used to determine whether surface observations are truly representative of the transported ozone?

Ludwig and Shelar (*14*) have provided reasonably detailed discussions of site-selection procedures for ozone and other monitoring purposes. The criteria and procedures that they describe for selecting monitoring sites are also applicable to the problem of evaluating the appropriateness of existing monitoring sites and the spatial completeness and quality of existing data bases. With regard to the importance of temporal completeness, the EPA guidelines (*1*) for the interpretation of the ozone air quality standards can be used to infer criteria for temporal completeness of a data base from a single site, but information from a single site is not as important when data are being collected from a network of stations where the methodology allows one to estimate missing values from other simultaneous observations. Nevertheless, the network itself should be operative during those seasons when exceedances are apt to occur. For example, assurance that ozone concentrations never exceed 120 ppb during the winter season reduces the importance of a complete data set during that season. Similar arguments might be applied to certain hours of the day.

In summary, the steps required to evaluate a data base are as follows:

Step 1—Determine the types of data available.

Step 2—Compare the available data items with those listed in Table 3.

Step 3—Identity important items (e.g., items in Table 3 whose subjective importance rating is either A or B) that are missing.

Step 4—Review measurement, calibration, and maintenance techniques.

Step 5—Identify instances where data accuracy or precision might be inadequate.

Step 6—Review information about station siting.

Step 7—Identify any stations whose data might not be representative of conditions outside their immediate area.

Step 8—Review overall distribution of stations in the network.

Step 9—Identify inadequacies in the overall distribution of stations (e.g., too few stations to define transported ozone concentration or no stations located where peak ozone concentrations might be expected, and so forth).

Step 10—Review the completeness of the data records for each station.

Step 11—Identify instances where the available number of data are not sufficient to meet theoretical requirements.

The preceding list provides a framework for evaluating data bases. Obviously, the procedure described by the steps listed yields not only an evaluation of the data base but also the identification of most of its major shortcomings. Once the shortcomings have been identified, the specification of supplemental data requirements will be much simpler. Appendix E presents guidelines for supplemental monitoring. Chapter Three gives the results of some of the applications of the techniques with reduced amounts of data. The results suggest that good estimates of design value and expected exceedances can be obtained from a nine- or ten-station network operated on days when temperatures exceed about 22 C (72 F). The stations must be carefully located to avoid local interferences. The most efficient approach is to have about five permanent stations, one in the city and four outside the immediate urban area, to monitor peak and background concentrations. This fixed network can be supplemented with several mobile units deployed generally downwind of the city on warm days.

PROCEDURES FOR DETERMINING EXCEEDANCES AND DESIGN VALUES

Definitions and Background

Three types of distributions of concentrations can be defined for a spatial network: "observable," "true," and "interpolated." For reference purposes, abbreviated definitions of these distributions are given in Table 4, and a schematic of the interrelationships between these distributions is shown in Figure 7. The following discussion of the three types of distributions is motivated by the need to specify the nature of the distribution with respect to which the design value and expected number of exceedances will be defined.

To examine any of the above distributions the number and location of all of the grid cells in the spatial network and the time period that the data base should span must be specified. For the moment, it is assumed that the grid cells have been defined so that the daily maximum concentration over a cell is essentially uniform, and there is no more than one monitoring site per cell. Later in the discussion this assumption will be relaxed.

Table 4. Types of distributions.

Distribution	Definition
Observable	Actually observed data plus all those data that would have been gathered, had the existing stations operated continuously without malfunction plus extra data that could have been gathered by a network.
Potentially observable	That part of the observable data that was not actually measured.
True	The stochastic process that generates the observable distribution.
Interpolated	Actually observed data plus estimates of the remainder of the observable data (the data gathered by the monitoring stations plus our estimate of the potentially observable data).

The Observable Distribution

The observable distribution is the sum of two components. The data actually gathered from the existing monitoring sites (including any temporal gaps) constitute the "actually observed" component. Potentially, however, one could have had properly functioning monitors in each grid cell each hour of each day in the desired time period. The extra data gathered from the "new" monitors (in the formerly empty grid cells), together with the formerly missing data from the existing monitors, form the "potentially observable" component. The sum of the actually observed and the potentially observable components constitutes the observable distribution (see Table 4 and Fig. 7).

Figure 8 is a schematic of an observable distribution in a spatial network. The data base under consideration spans 15 days. The spatial network consists of 9 grid cells with four monitoring sites. In each grid cell the daily maxima for the *i*th day have been indicated by the number *i*, and these numbers have been arranged in a histogram. A number is surrounded by a square if the corresponding maxima have actually been observed and is surrounded by a circle if the corresponding maxima have not been observed (i.e., are potentially observable).

How is the concept of an expected number of exceedances related to the observable distribution? If the observable distribution was known, the number of exceedances could be computed that would have been obtained with a complete and properly functioning spatial network. For example, grid cells 1 through 9 would have 1, 2, 3, 1, 2, 1, 2, 1, and 0 exceedances, respectively. Consequently, according to the definition given earlier, the number of exceedances in the spatial network is equal to 3—the maximum number of exceedances experienced by any grid cell. However, one cannot know the entire observable distribution; only the actually observed component is known. Thus, it is possible to compute only an expected number of exceedances. By using the EPA method (*1*), the expected number of exceedances can be computed for cells 2, 3, 4, and 9 during the 15-day period. These calculations are given in Table 5. The expected numbers of exceedances are computed to be $1\frac{1}{6}$, $2\frac{1}{2}$, $1\frac{2}{11}$, and 0, respectively, for the 15-day period. These expected numbers of exceedances can be interpreted to be estimates of the numbers of exceedances that one would encounter if the entire observable distribution in grid cells 2, 3, 4, and 9 was known. Although the EPA method is only applicable to the grid cells with monitoring stations, a method presumably can be developed for computing the expected numbers of exceedances in grid cells without monitoring stations, thus allowing the computation of the expected number of exceedances for the spatial network.

How is the concept of a design value related to the observable distribution? The design value is defined as the upper 1/365th percentile of a distribution of daily maxima. Clearly it is facetious to attempt to discuss the upper 1/365 percentile when the example contains only 15 observable concentrations per grid cell. For the sake of this discussion, the design value can be assumed to be defined as the upper 20th percentile. Then, the design value would equal to the third highest observable value. For grid cells 1 through 9, the design values based on the observable distribution are 110, 110, 130, 100, 110, 100, 110, 90, and 100, respectively. In accordance with the definition selected earlier, the design value for the spatial network would be 130 ppb. Of course, the entire observable distribution is not known, only the observed component, in grid cells 2, 3, 4, and 9. By using a method closely allied to the methods suggested by the EPA, estimated design values can be computed for these four grid cells, as given in Table 6. These estimates are 104, 128, 104, and 108, respectively. Once again, a method to estimate the design value of the grid cells without monitoring sites would make it possible to estimate the design value for the spatial network.

The True Distribution

Until now the observable distribution has been discussed as a separate entity. One could also consider the observable distribution as being the observed and potentially observable results of a stochastic process that constitutes the "true" distribution (see Table 4 and Fig. 7). The observed data base

Table 5. Results of exceedance computations using the EPA method.

Variable	Value	Estimate for Designated Grid Cell Number			
		2	3	4	9
N	Number of required monitoring days	15	15	15	15
n	Number of valid daily maxima	12	12	11	11
v	Number of measured daily values above the standard (120 ppb)	1	2	1	0
Z	Number of days assumed to be less than the standard (e.g., preceding and following day are less than 90 ppb)	1	0	2	0
e	$V + \frac{v}{n}$ x $(N - n - Z)$, the expected number of exceedances	$1\frac{1}{6}$	$2\frac{1}{2}$	$1\frac{2}{11}$	0

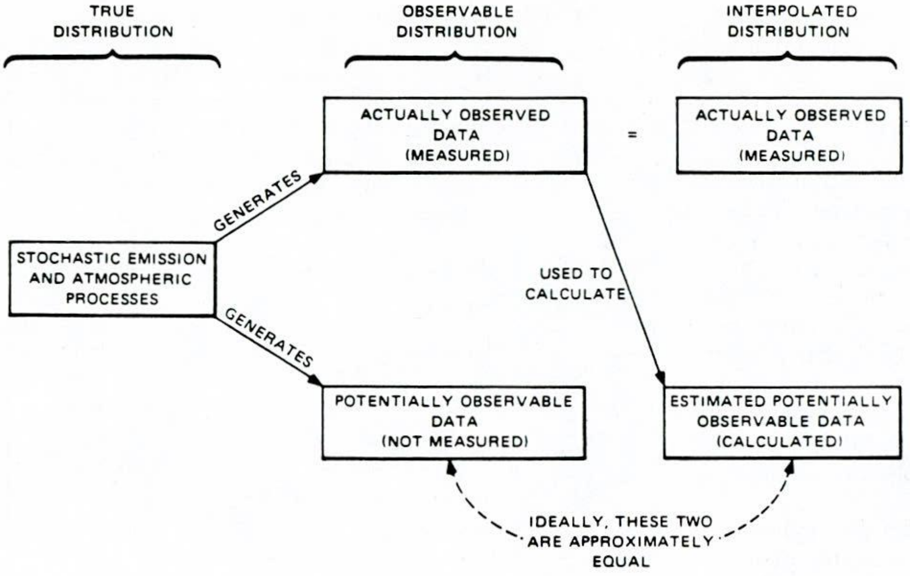


Figure 7. Relationships among the types of distribution.

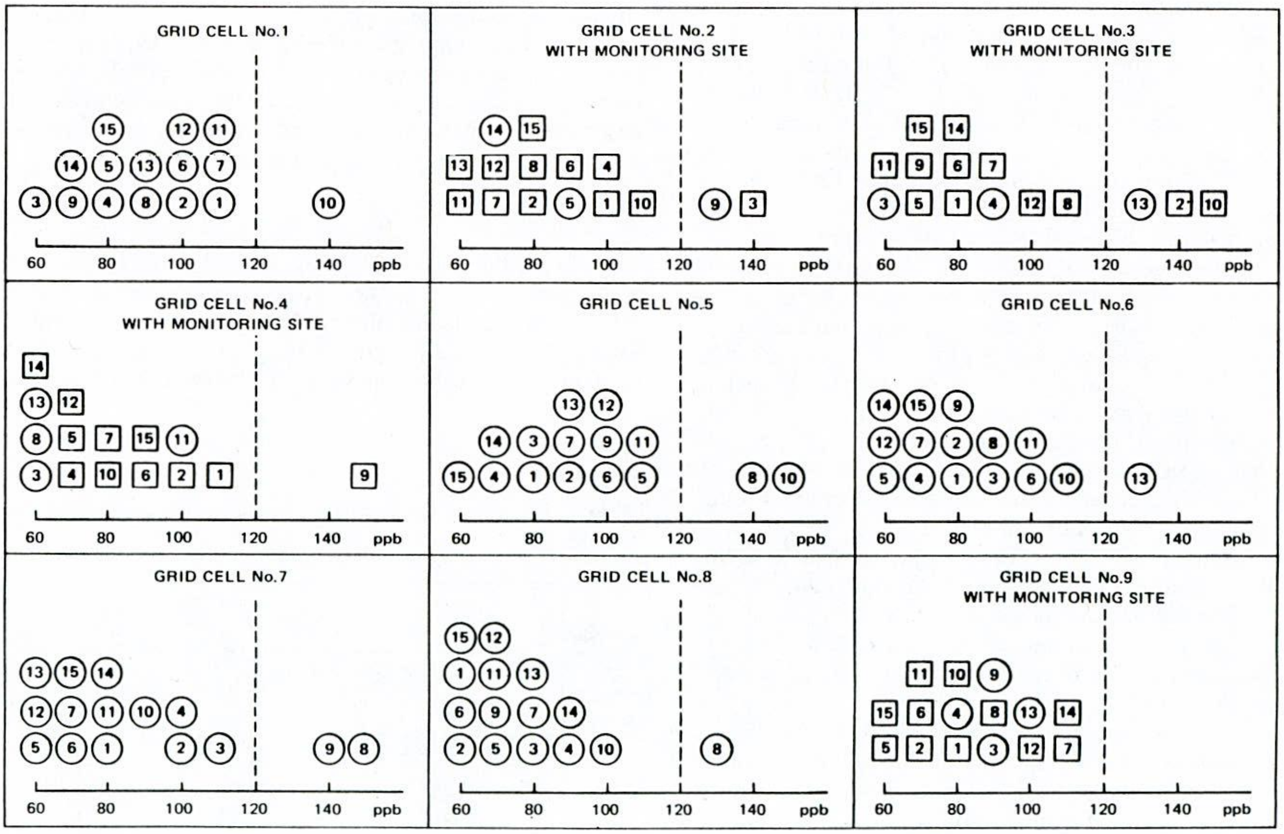


Figure 8. The observable distribution in a spatial network.

Table 6. The estimated design value (calculated by a method closely related to those employed by the EPA).

Variable	Value	Estimate for Designated Grid Cell Number			
		2	3	4	9
n	Number of valid daily maxima	12	12	11	11
Z	Number of days assumed below the standard	1	0	2	0
n + Z	Total used to compute percentiles	13	12	13	11
Ranks of observations straddling the 20th percentile		(2,3)	(2,3)	(2,3)	(2,3)
Corresponding probabilities	(P_1,P_2)	$\frac{2}{13},\frac{3}{13}$	$\frac{2}{12},\frac{3}{12}$	$\frac{2}{13},\frac{3}{13}$	$\frac{2}{11},\frac{3}{11}$
Corresponding observations	(O_1,O_2)	(110,100)	(140,110)	(110,100)	(110,100)
Estimated design value[1]	$O_2 + (O_1 - O_2)\dfrac{P_2 - 0.2}{P_2 - P_1}$	104	128	104	108

[1] In this example, the design value is defined relative to the 20th percentile rather than the 0 274th percentile.

component is then a realization, or random sample, from the true distribution. The potentially observable data base component would also be a realization or random sample from the true distribution, although by definition the monitoring equipment to record this component is lacking. Instead of being interested in the expected number of exceedances and the design value of the observable distribution, one might be interested in the expected number of exceedances and design value of the true distribution.

The EPA does not explicitly state whether they are interested in computing the expected number of exceedances and the design value with respect to the observable distribution or to the true distribution. To a large extent, they have been spared making that decision because, for a single site, the estimates of the expected number of exceedances and the design value are virtually unaffected by the choice of either the observable or true distribution. For example, the calculation given in Table 5 would be unaffected by the choice of the true distribution, and for Table 6 the only changes required would be the customary addition of 1.0 to each number in the third row (denoted *n* + *Z*) and the subsequent modifications in the fifth row (i.e., the addition of 1.0 to the denominator of each fraction) and the final row. With larger numbers of observations, the effect of adding 1.0 to the denominator would be negligible.

Although the estimates of the expected number of exceedances and the design value for a single station are nearly unchanged by the choice of the type of distribution, the confidence that can be placed on the accuracy of those estimates varies considerably. For example, consider a grid cell with a monitoring station that operated every day in a 3-year period. If three exceedances occurred with respect to the observable (and in this case observed) data base, the expected annual number of exceedances is 1.0 with 100 percent confidence. The confidence is 100 percent because there is no uncertainty with respect to the observable distribution—observations are for the number of grid cells in areas of high concentration and the magnitude of the correlation between those grid cells. The reduction is greatest when these are a large number of relatively independent grid cells in areas with high ozone concentrations.

A simplified example can be used to illustrate why the expected number of exceedances is reduced. Consider a grid of five independent cells, for which it is known *a priori* that the number of exceedances at each of these locations in a 3-year period follows a Poisson distribution with a mean of 3.0. With respect to the true distribution, the expected number of exceedances for each cell and for the network is then exactly 3.0. On the other hand there is a 35.3 percent chance that any one station will record four or more exceedances during the 3-year observation period, and an 88.77 percent chance that at least one of the stations will record four or more exceedances during the 3-year observation period. Consequently, with respect to the observable distribution, the expected number of exceedances for the network will probably be four or more.

The authors of this report have elected to define the expected number of exceedances and the design value with respect to the observable distribution (rather than with respect to the true distribution) for four reasons:

1. If emissions change over time, the hypothetical process that generated the data over a given period may never exist again. Inference with respect to that process is therefore unimportant in any practical sense.
2. Individuals in the area covered by the spatial network are exposed to the observable distribution rather than to the true distribution; i.e., the true distribution may never actually be realized, just as the "true" mean of a distribution may never actually occur in the random samples (observable distributions) drawn from that distribution. See Table 4 and Figure 7.
3. A number of assumptions would be necessary to model the true distribution that are not necessary for the observable distribution.
4. The statistical methodology necessary to estimate the expected number of exceedances and the design value for the true distribution would be many times more complicated than that required to model the observable distribution.

The Interpolated Distribution

Having decided to estimate the expected number of exceedances and the design value with respect to the observable distribution, one must now face the reality that only the observed component of that distribution is known. Because the potentially observable component of the observable distribution is not known, one must estimate that component using the observed component. The combination of the estimated potentially observable concentrations and the observed concentrations is denoted as the "interpolated" distribution. The interpolated distribution is an estimate of the observable distribution from which the number of exceedances and the design value are derived to serve as estimates of the same parameters for the observable distribution.

Description of Methods for Estimating the Potentially Observable Distribution

There are two general methods for estimating the potentially observable data. These two approaches will be referred to as "isopleth modeling" and "probabilistic modeling." Isopleth modeling is the conceptually simpler method; the observed data are used to estimate ozone values in the unmonitored grid cells and to estimate missing ozone values in the monitored grid cells. Whatever techniques are used, isopleth modeling must allow for the estimation of ozone concentrations larger than those actually observed: thus, an isopleth modeling technique will have to allow for the extrapolation of spatial gradients of concentration. Certain automated methods for drawing isopleths have this property and would provide an estimated concentration for each potentially observable data point. These estimated concentrations provide the basis for a diagram similar to Figure 8, although the circles would denote estimated concentrations rather than potentially observable concentrations.

The computerized isopleth program that has been developed uses the number of grid cells and the daily maximum ozone value at the centers of the grid cells. If a grid cell contains one or more monitoring station observations on a given day, the maximum from among the observed values within the grid cell and the predicted value at the cell center is taken to be the daily maximum ozone value for the grid cell. This will introduce a bias toward higher values because it represents a conservative approach.

The isopleth program does not have to be run for every day in the data base, because one is concerned primarily with days in which one or more grid cells record ozone values above 120 ppb, and consequently the isopleth program may be restricted to days that would produce high ozone values. On the basis of experience with the isopleth program, the authors found that no predicted ozone values exceeded 120 ppb unless at least one monitoring station recorded an ozone concentration of approximately 100 ppb or more.

How many grid cells should be specified? Ideally, the grid cells should be small enough that the ozone concentration is uniform within the cell, but it is necessary only that the cells be small enough so that further subdivision does not appreciably change the design value or the expected number of exceedances. The following procedures establish the necessary number of cells. Initially, the grid consists of no fewer than nine cells. The expected number of exceedances and the design value for the initial grid are computed. A refined grid, consisting of smaller cells than in the initial grid is defined. These new cells are defined by subdividing those cells with most exceedances and highest design values. The expected number of exceedances and the design value for the refined grid is computed.

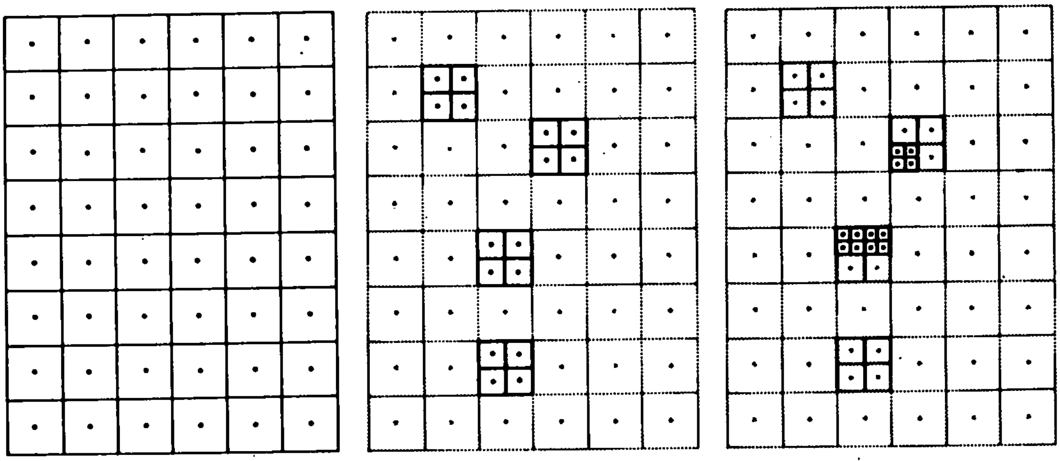
Isopleth Methodology

Basically, the isopleth methodology requires that estimates be generated for ozone concentrations in each cell of a network of cells overlying the area of interest for all those days in the data set when it is likely that exceedances have occurred somewhere in the region. The estimated values for these cells are used to determine the number of exceedances for each cell and the design value for each cell.

As previously suggested, two problems arise with this approach. The cells must be small enough so that ozone concentrations within them can be characterized by a single value. The philosophy has been adopted that this is important only in "critical" cells where the highest design values, or most exceedances, are to be found. Thus, the method begins with relatively large cells that are characterized by the ozone concentration estimates at their center, and proceeds to generate smaller cells in those areas where the design value is high or the number of exceedances great. By using small grid cells only in regions with high design values or large numbers of exceedances, one is forced to identify such areas. This is done iteratively. One starts with a grid of large cells and subdivides those with the most exceedances or the highest estimated ozone values. The process is repeated until further subdivision does not appreciably change the results. This process is described in greater detail later.

Figure 9 shows a schematic representation of the nested grid approach that was used as part of the isopleth methodology. Figure 9(a) shows the initial grid. Concentrations are estimated for the center of each cell (indicated by the points in the figure) for each day when exceedances might have been possible.

After the estimates are obtained for each of the initial cells and each day, the number of exceedances and the design value are estimated for each cell. The definitions of "exceed-



(a) INITIAL GRID (b) FIRST SUBDIVIDED GRIDS (c) SECOND SUBDIVIDED GRIDS

Figure 9. Schematic representation of the nested grid approach used in the isopleth methodology.

ance" and "design value" for a network that were described earlier require that the points with the most exceedances and with the highest ozone concentrations be identified. There can be no assurance that these points coincide with any of those shown in Figure 9(a), for which the estimates are initially made. Therefore, the area is further subdivided to test the possibility that more exceedances or higher ozone concentrations may be found elsewhere in the area. The subdivision is limited to the several cells with the most exceedances and highest estimated ozone concentrations. Figure 9(b) shows that 4 grid cells were selected to be subdivided. This results in 16 new, smaller cells for which concentrations must be estimated for each day in the data set. As before, the estimates are based on the values calculated for the points at the centers of the cells; these points are indicated in Figure 9(b). The number of exceedances and the highest ozone concentrations are determined for each of the new, smaller cells and the results compared with those obtained for the cells in the original grid. If none of the smaller cells has appreciably more exceedances or higher ozone concentrations than were found for cells in the original grid, computations are stopped. It is assumed that further subdivision is not likely to change the results very much. If there are appreciably more exceedances or higher ozone concentrations in the smaller grid cells than were found in the original grid cells, those smaller grid cells with higher concentrations or more exceedances are identified and subdivided further as shown in Figure 9(c).

When an ozone concentration has been observed in a cell, it is compared with the objectively estimated value for that cell. The concentration representing the cell will be the greater of the two. This is only one of several options that could have been taken, but it is the most conservative in that it will not lead to underestimates of numbers of exceedances or of highest ozone concentrations.

The design values are estimated as follows. For each grid cell, the estimate will be based on the ten highest ozone concentrations estimated for that cell, using a log-linear regression. The logarithms of the ten highest ozone concentrations are plotted against the corresponding percentile values, and a straight line is fitted by least-squares linear regression. Based on that line, the design value is defined as the ozone concentration corresponding to the 0.274th (1/365th) percentile. The design value for the network is then the highest design value found for the grid cells or subgrid cells that have been generated. Similarly, the number of exceedances for the network is the most found for any of the grid or subgrid cells.

To this point, little has been said about the methods used to estimate the values for the grid cell centers. An interpolation scheme is used, but it is important to note that it does not use linear interpolation or a simple inverse distance-weighting approach, because such schemes do not yield values that fall outside the range bounded by the smallest and largest values in the input data list. Rather, the objective analysis scheme uses second-degree polynomial fitting to the data so that maxima and minima can be specified that are not coincident with points in the input data set.

In the objective analysis, a grid-point value is determined for the center of each cell by fitting a second-degree polynomial surface by least-squares to K nearby observations. An optimum fit is obtained by minimizing

$$Q = \sum_{i=1}^{K} w_i (q_i - q_i')^2 \qquad (4)$$

where w_i is a weighting factor, q_i is an observed value at the ith location (x_i, y_i) and q' is the polynomial estimate for the same location. The polynomial estimate is of the form $a + bx + cy + dxy + ex^2 + fy^2$. The initial estimate is made for each grid point; and is treated as the first observational value (i = 1). It is given the fixed weight w_1. The program generates the first estimate by computing a simple weighted average using nearby observations. A relatively low value is normally assigned to w_1 so that the initial estimate has a significant influence only in areas where there are few or no nearby data. The initial estimate prevents the scheme from generating extreme values in data-sparse regions. The remaining K-1 observations are those nearest the grid point, except, whenever possible, at least one observation is selected from each of the four angular quadrants around the grid point (i.e., in a direction between north and east, between east and south, and so forth). This approach makes the analysis smoother and more continuous in regions of poorly distributed data.

The weighting factors for a given observation are defined as follows:

$$w_j = \frac{C^2}{C^2 + R_j^2} \tag{5}$$

where C is a constant and R_j is the distance between the jth observation and the point for which the interpolated value is being determined. To use the polynomial interpolation scheme above, different values can be chosen for the following variables: K-1, the number of stations to be fit by the polynomial; w_1, the weight to be assigned to the initial estimate at an interpolation point; and C, the constant used in the weighting function.

The magnitude of the distance-weighting effect will depend on the value assigned to the constant C, but a second-degree polynomial fitting is not very sensitive to the value used. A high value of C will increase the sensitivity of the method to the K value chosen. In the program described in Appendix C, the constant C was assigned a value of 0.3, which is reasonable on the basis of other experience. Sensitivity tests were then made using different values for K and w_1.

The sensitivity runs were made for five cases summarized in Table 7 with the results shown in Figures 10 through 14. For each of these cases, sensitivity runs were made using the values given in Table 8. The objective analyses were compared with subjective hand-drawn analyses. The results for each of the five cases were similar. In general, the effect of increasing K from 7 to 9 (for both $w_1 = 0.05$ and 0.1) is to produce some smoothing of the field and to eliminate detail. Increasing w_1 from 0.05 to 0.1 (for both $K = 7$ and 9) also introduces additional smoothing. Using $K = 7$ rather than $K = 9$ is preferred because fewer data sites are required, but some smoothing beyond that found for $K = 7$ and $w_1 = 0.05$ seems desirable, so the values $K = 7$ and $w_1 = 0.1$ have been used for the calculations.

If fewer than six stations are available, the objective analysis is based on a fitting of a first-degree polynomial, which requires values from only three stations. The results from a first-degree polynomial fitting are also shown in Figures 10 through 14. The linear fitting tends to produce fields that are smoother than the polynomial fits, so its use would not be desirable except in cases of insufficient data. If there are only one or two station values, the analyzed field is uniformly equal to that one value—or each grid value is an inverse distance-weighted average of the two available values.

Probabilistic Methodology

Comparison of the isopleth modeling and the EPA method discussed earlier suggests that a method that can incorporate both spatial and temporal data would be desirable. The probabilistic modeling method, discussed below, incorporates both types of data. It is a synthesis of the isopleth modeling and EPA methods.

The probabilistic modeling method is implemented in a series of steps that are summarized here and explained in more detail in subsequent paragraphs:

1. The interpolated distribution is computed using the isopleth program. The interpolated distribution will be used indirectly to estimate the expected number of exceedances and the design value for the grid network rather than directly as in the isopleth modeling technique.

City	Date	Figure Number
Houston	5 October 1978	10
Houston	7 October 1978	11
St. Louis	1 October 1976	12
St. Louis	13 July 1976	13
St. Louis	8 June 1976	14

Table 7. Summary of sensitivity tests of the isopleth polynomial interpolation methods.

Interpolation Scheme	Constant Value C	K	w_1
Subjective	—	—	—
Objective, linear fit	0.3	a	—
Polynomial fit	0.3	7	0.05
Polynomial fit	0.3	7	0.1
Polynomial fit	0.3	9	0.05
Polynomial fit	0.3	9	0.1

[a] Data from nearest five stations used.

Table 8. Interpolation schemes and constants tested.

2. The interpolated distribution is used to estimate the parameters of a joint distribution for the daily maximum ozone concentrations of the grid cells.
3. The conditional distribution of the ozone concentrations of the grid cells without monitoring stations (given the ozone concentrations of the grid cells with monitoring stations) is derived. There will be a different conditional distribution for each day because the ozone concentrations of the grid cells with monitoring stations change daily.
4. A Monte Carlo simulation is performed using the collection of conditional distributions.
5. The results of the Monte Carlo simulation are used directly to estimate the expected number of exceedances and the design value for the grid network. Figure 15 is a schematic of the probabilistic-modeling process.

The first step is fairly straightforward. The interpolated distributions generated by the isopleth program are used for both types of modeling. To save computer expenses the interpolated ozone values are generated only for days when at least one grid cell is likely to have recorded an ozone value above 120 ppb; the data set used by the isopleth methodology is therefore limited to those days when at least one monitoring site observed an ozone concentration in excess of 80 ppb. This threshold was chosen because it limits the amount of data to be processed, but does not exclude any cases where the isopleth methodology indicates that exceedances of the NAAQS standards are likely.

In addition to treating only a subset of the days, the probabilistic modeling program also uses only data from a subset of all the cells that are generated by the isopleth program. The cells that the probabilistic modeling program uses are divided into two groups. The first group (of size N_1) includes those cells that contain no monitoring station, but are among those with either one of the ten highest design values or one of the ten highest numbers of exceedances, or both. Cells that fall into both categories are considered only once;

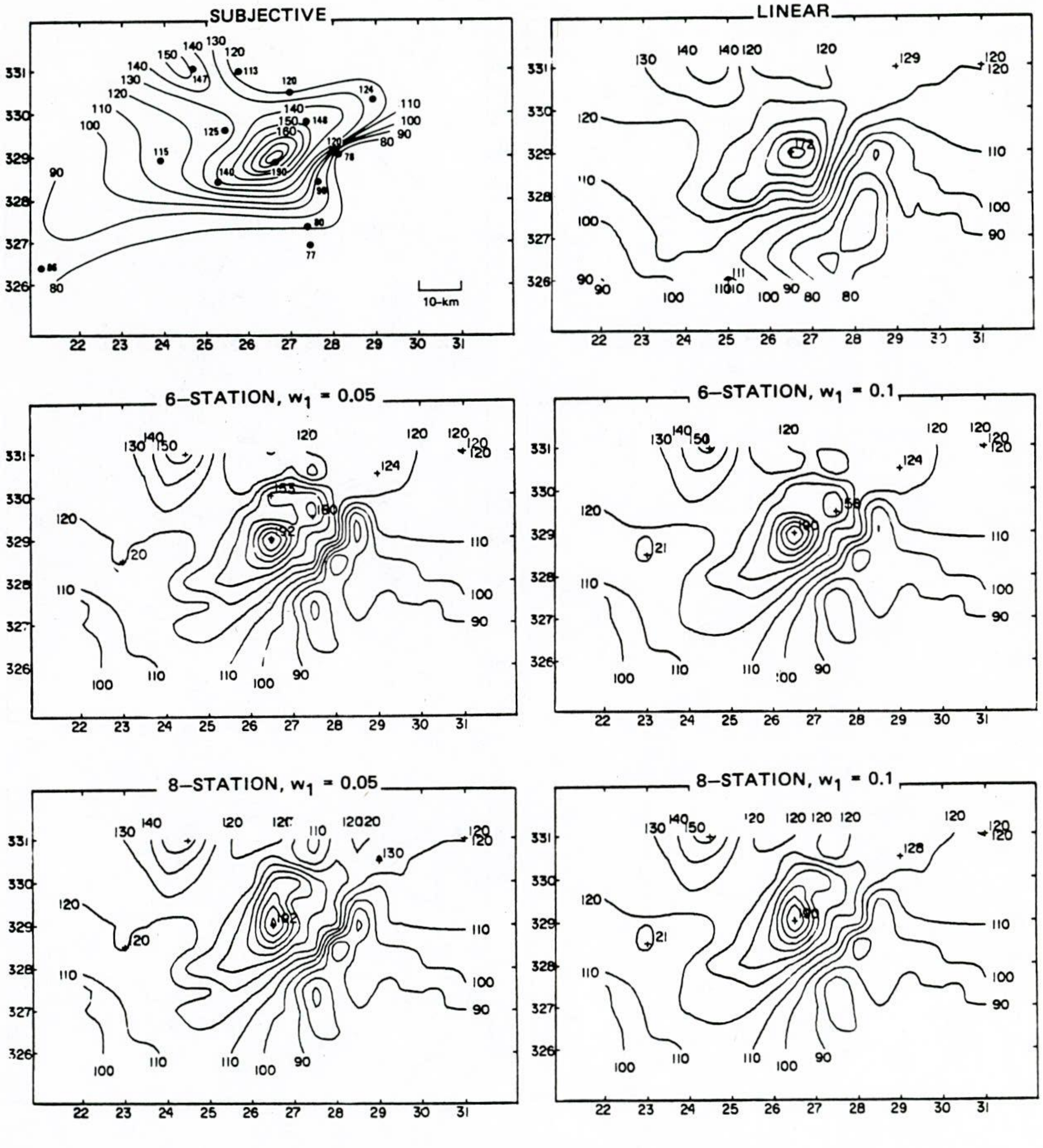


Figure 10. Examples of isopleth analyses of data from Houston for 5 October 1978.

hence, there should be between 10 and 20 of the first type of cells. The second type of cells (of size N_2) are those containing a monitoring station.

In the second step, the interpolated distribution is used to estimate the joint probability distribution of all grid cells. This joint distribution is modeled as a multivariate lognormal distribution for two reasons: the lognormal distribution generally fits the tails of the ozone concentration distribution well, and it is a distribution from which it is reasonably easy to compute conditional distributions. The parameters of the joint distribution include the mean and variance of the (marginal) distribution of ozone concentrations at each grid cell and the correlations between each of the cells. For example, if there are 20 grid cells (that is, if $N_1 + N_2 = 20$), then 20 means, 20 variances, and 190 distinct correlations must be estimated.

The estimation procedure is structured so that the resulting distribution fits the interpolated distribution well in the tails; the fit in the lower part of the distribution is less important because that portion does not affect the estimates of the design value or the expected number of exceedances appreciably. Consequently, the usual estimates of the parameters

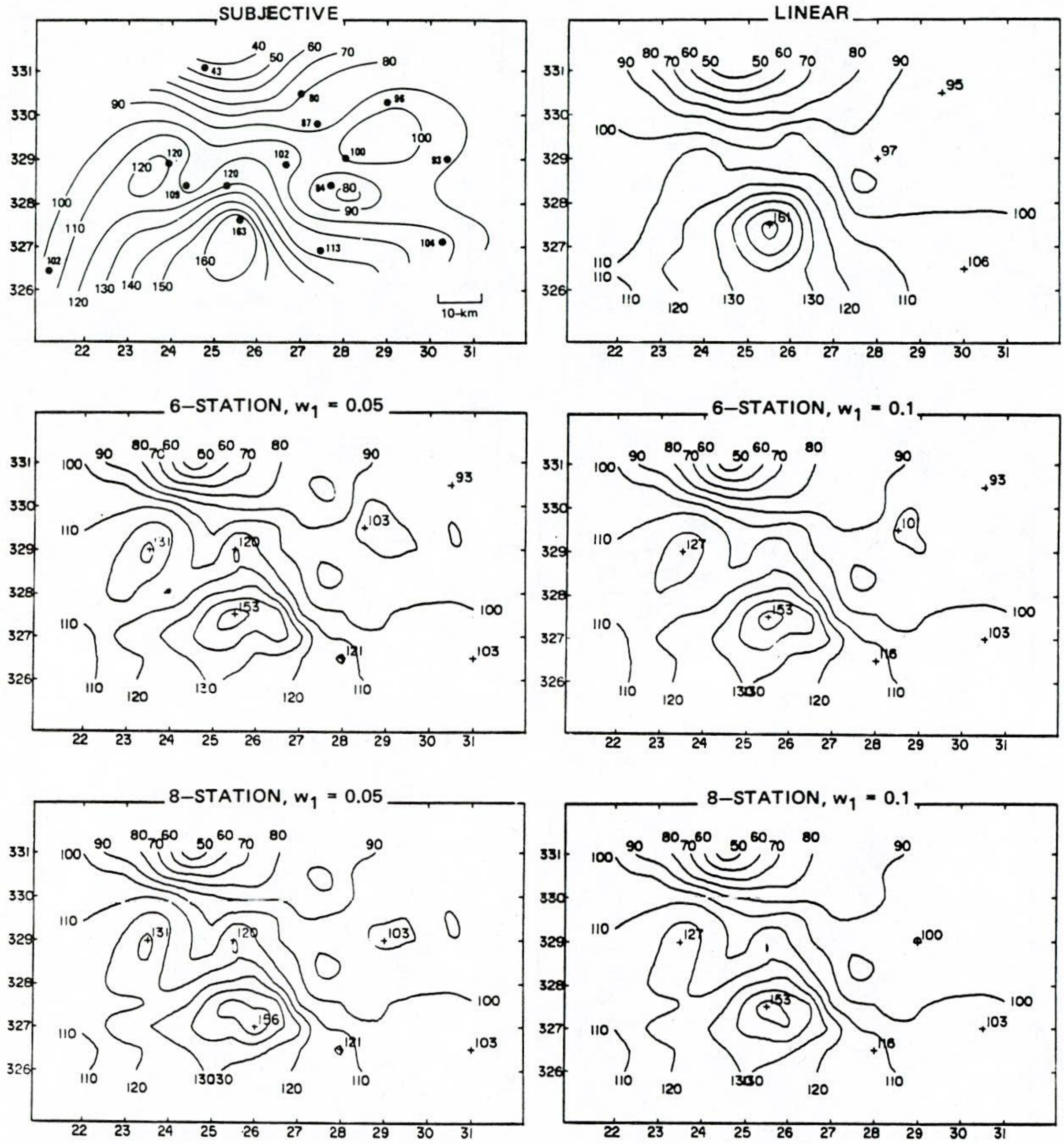
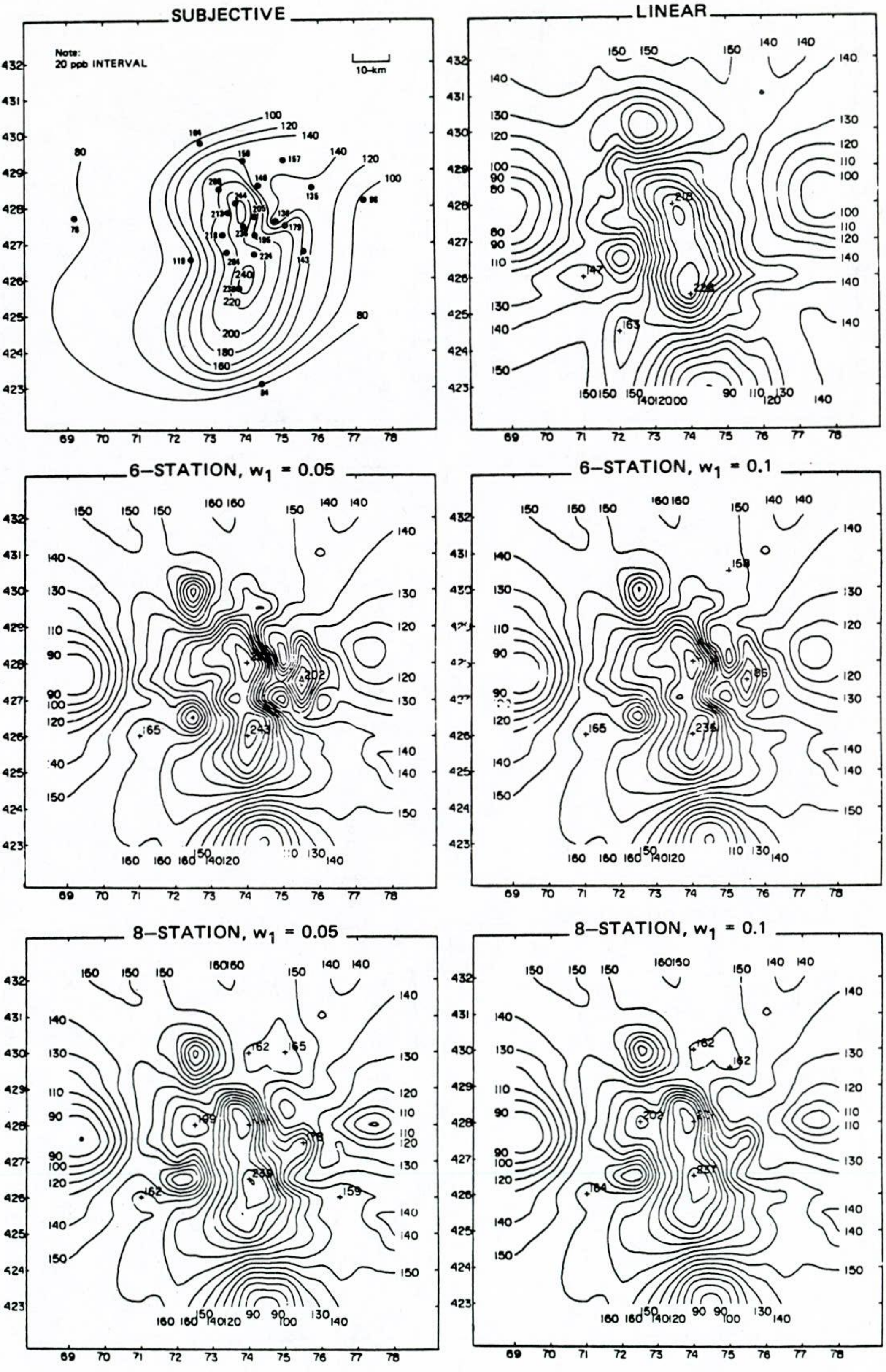


Figure 11. Examples of isopleth analyses of data from Houston 7 October 1978.

of the multivariate lognormal distribution—the sample mean, the sample variance, and Pearson's correlation coefficient computed using the logarithms of the interpolated distribution—are not wholly appropriate, and more appropriate estimators were derived.

Two different techniques were used for estimating the parameters of the joint distribution of the N_1 cells without monitoring stations (but among the cells with the highest number of exceedances or the highest design values) and the N_2 cells with monitoring stations. Let μ denote the $(N_1 + N_2)$ vector of means of the joint distribution and let C denote the $(N_1 + N_2) \times (N_1 + N_2)$ matrix of covariances of the joint distribution. Then both techniques, which are described in the following section, provide estimates of μ and C. The first technique, which is simpler and less rigorous, calculates the means and covariances with only partial accounting for the truncation of the data at a threshold value (e.g., days for which no monitoring site registered above 80 ppb ozone were not subjected to isopleth modeling). The second technique fully accounts for the truncation of the data; unfortunately, in doing so one must further truncate the data and exclude all days for which the average ozone concentration is below 80



SUBJECTIVE

LINEAR

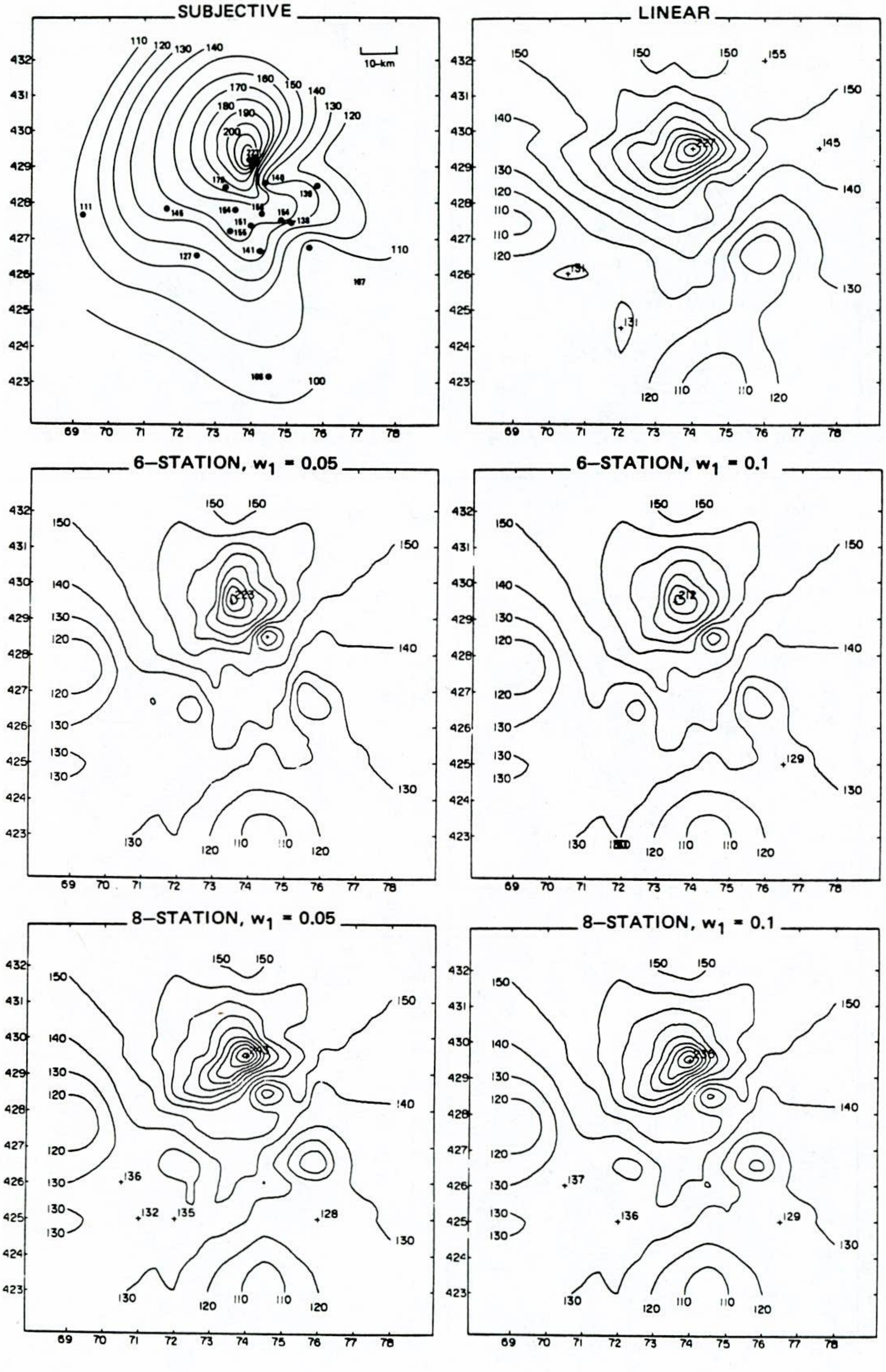
6–STATION, w_1 = 0.05

6–STATION, w_1 = 0.1

8–STATION, w_1 = 0.05

8–STATION, w_1 = 0.1

Figure 12. Examples of isopleth analyses of data from St. Louis for 1 October 1976.



SUBJECTIVE

LINEAR

6–STATION, $w_1 = 0.05$

6–STATION, $w_1 = 0.1$

8–STATION, $w_1 = 0.05$

8–STATION, $w_1 = 0.1$

Figure 13. Examples of isopleth analyses of data from St. Louis for 13 July 1976.

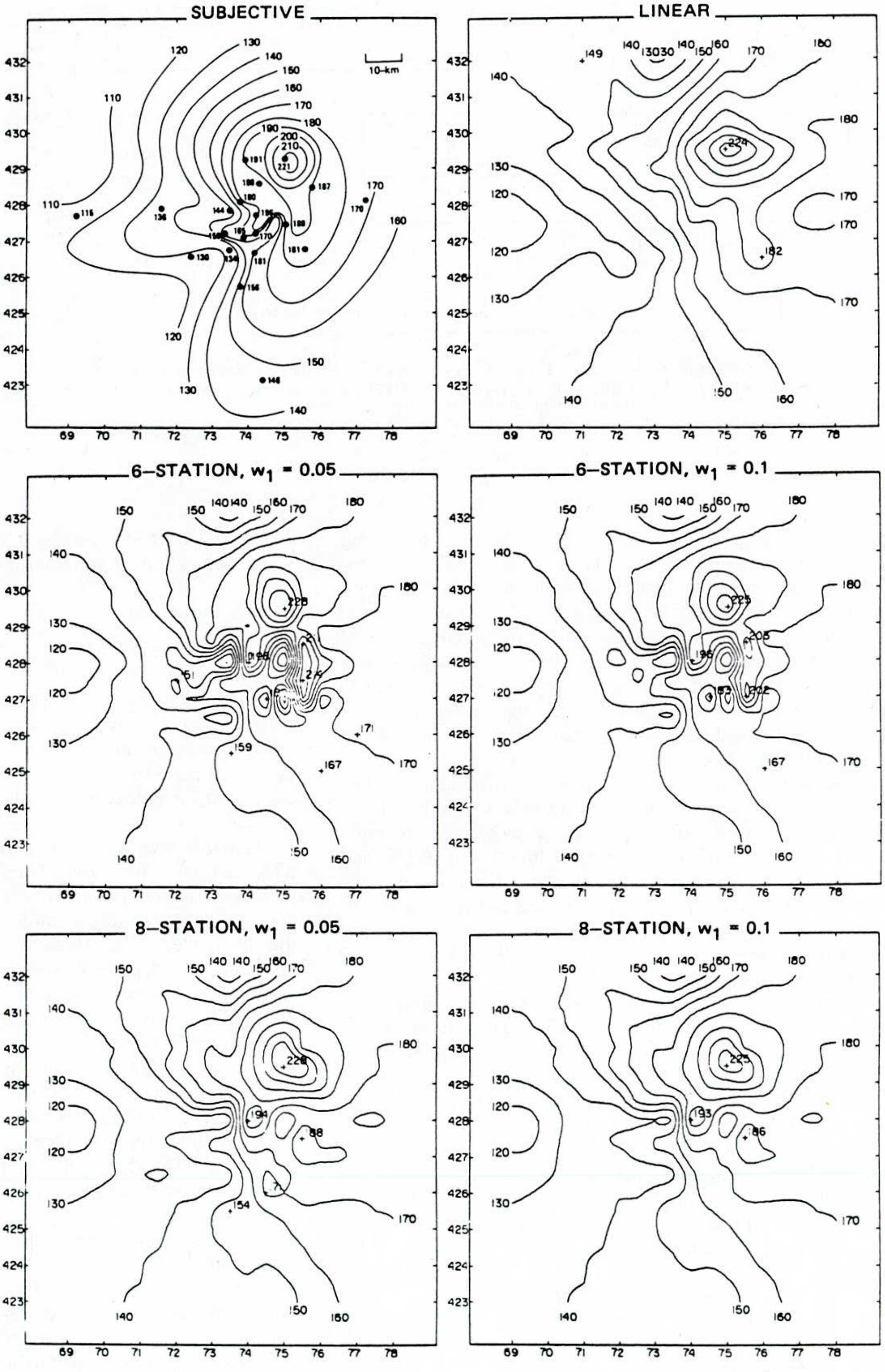


Figure 14. Examples of isopleth analyses of data from St. Louis for 8 June 1976.

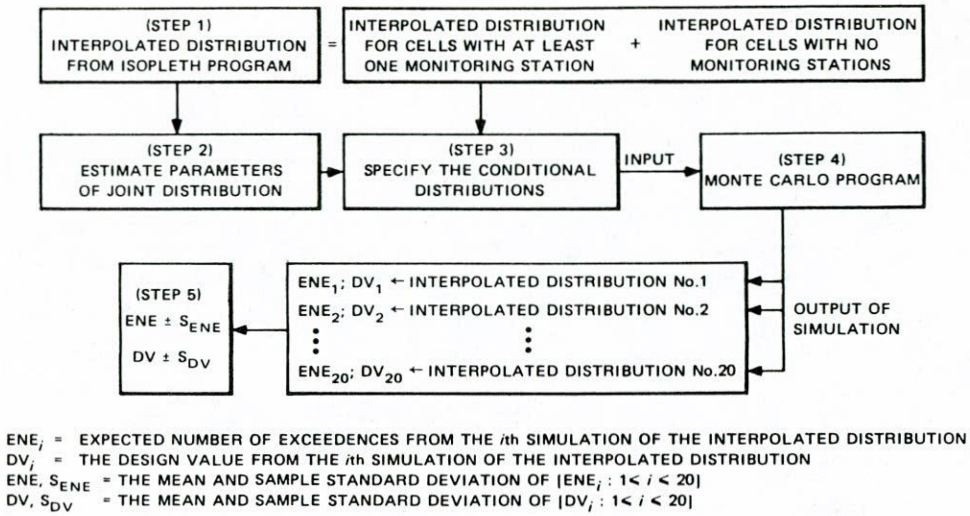


Figure 15. Schematic of probabilistic modeling.

ppb. Although results are presented in Chapter Three using both techniques, it is concluded that the further data truncation of the second technique may result in substantial instability in the parameter estimates.

In the third step, the conditional distribution of the ozone concentrations of the N_1 grid cells without monitoring sites is derived, given the ozone concentrations of the N_2 grid cells with monitoring sites. To describe how the conditional distribution is obtained, one first examines the structure of the covariance matrix C.

The matrix C can be subdivided as shown in Figure 16. The matrix is symmetric along the main diagonal and its parts have the following properties: C_{11} (N_1 rows by N_1 columns) is the covariance matrix of the cells without stations that had the most exceedances or highest design values. C_{22} (N_2 rows by N_2 columns) is the covariance matrix of those cells with stations. C_{12} (N_1 rows by N_2 columns) and C^T_{12} (N_2 rows by N_1 columns) are the covariances of the cells without stations with the cells with stations.

Computation of the conditional distribution requires the following, in addition to the submatrices shown in Figure 16: X_2, a column vector (N_2 rows) of the logs of the observed values for cells with stations for one day; M_1, the column vector (N_1 rows) of the means of the values from cells without monitoring stations (M_1 is the vector of the first N_1 elements of μ); and M_2, the column vector (N_2 rows) of the means of the cells with monitors (M_2 is the vector of the second N_2 elements of μ).

With this information, the parameters for the conditional distribution of the cells without monitors, given the observed values in the cells with monitors, are computed as follows:

$$\text{Mean vector} = M = M_1 + C_{12}(C_{22})^{-1}(X_2 - M_2) \quad (6)$$

$$\text{Covariance matrix} = \Sigma = C_{11} - C_{12}(C_{22})^{-1}(C_{12})^T \quad (7)$$

Following convention, the superscripts $(\ )^{-1}$ and $(\ )^T$ denote the inverse and transpose of a matrix. Because the vector X_2 changes daily, the mean vector M must be computed for each day. However, the Σ matrix need only be computed once, because its components are all derived from the complete data set.

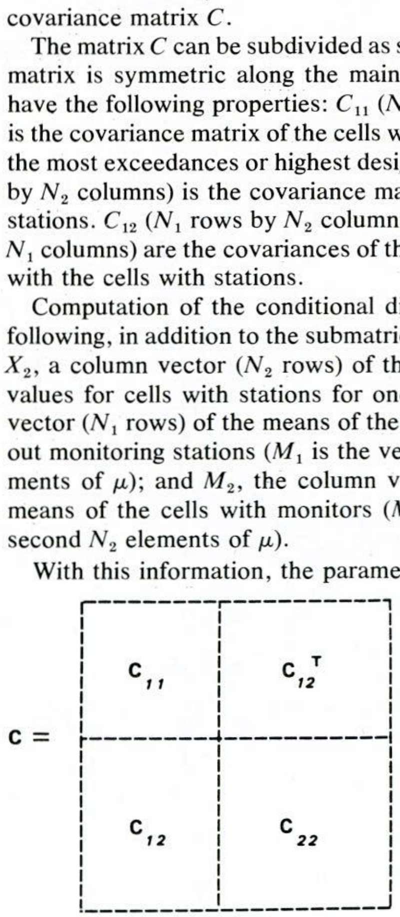


Figure 16. Schematic representation of the covariance matrix C.

The fourth step is the use of the conditional distribution in a Monte Carlo simulation. When the conditional distribution has been computed, a set of simulations is produced for the network by generating a random vector for each day such that the collection of these random vectors has the same distribution as found for the cells without monitors. This is accomplished as follows. A random-number generator produces numbers that are uniformly distributed between 0 and 1. Such numbers are used as inputs to a normal probability inversion routine, which returns numbers that are normally distributed with mean 0 and a variance of 1. Vectors with N_1 elements generated in this way will be distributed with a 0 mean vector, unit variance, and 0 covariances and are denoted Z. The Z vectors are multiplied by the Cholesky decomposition of Σ (denoted L). The vector M is added to LZ.

In summary, if Z is a vector generated randomly as described above, $LZ + M$ has the desired distribution. In general, values simulated from this distribution will tend to be high for those cells and days where the original interpolated values were high and vice versa. This is because monitored values will also have tended to be high, or low, on these days, producing a high mean vector for the conditional distribution.

Parameter Estimation Techniques

The simple method for parameter estimation has two variants. In the first variant the means for the jth station (μ_j) and the covariances between the jth and kth stations (c_{jk}) are defined as follows:

$$\mu_j = \frac{1}{M}\sum_{i=1}^{M} x_{ij} \qquad j = 1 \text{ to } N_1 + N_2 \tag{8}$$

and

$$c_{jk} = \frac{1}{M-1}\sum_{i=1}^{M} (x_{ij} - \mu_j)(x_{ik} - \mu_k) \quad j = 1 \text{ to } N_1 + N_2 \quad k = 1 \text{ to } j \tag{9}$$

where x_{ij} = log of the ith observed value in the jth cell and M = number of days in the truncated sample. The mean vector, μ, and the covariance matrix, C, are defined as:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_N \end{pmatrix} \tag{10}$$

$$c = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{N1} & c_{N2} & \cdots & c_{NN} \end{pmatrix} \tag{11}$$

Note that C is symmetric about the main diagonal (i.e., $C_{ij} = C_{ji}$). This symmetry was used to simplify the computation and use of the matrix.

The second variant uses the values of μ and C obtained from the first variant as a starting point, and then further refines them. In the second variant, the Pearson correlation coefficient is computed from the matrix C as follows:

$$P_{jk} = \frac{C_{jk}}{\sqrt{C_{jj}C_{kk}}} \tag{12}$$

This yields a matrix

$$R = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ P_{21} & P_{22} & \cdots & P_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ P_{N1} & P_{N2} & \cdots & P_{NN} \end{pmatrix} \tag{13}$$

Then extrapolated estimates are computed of the means and standard deviations for each cell by sorting the logs of the daily maxima and fitting the least-squares line through the points

$$\left[x_{ij} \ , \ \Phi^{-1}\left(\frac{M-i}{M}\right)\right] \tag{14}$$

where x_{ij} = the log of the ith daily maximum for cell j after sorting with respect to i in descending order.

The intercept of the least-squares line gives an estimate of the mean μj of the cell, and the 95th percentile (denoted P_{95}) can also be obtained from the least-squares line as described earlier in this chapter. For the normal distribution (univariate),

$$\frac{P_{95} - \mu}{1.645} = \text{standard deviation}$$

Let $S = (\sigma_1, \sigma_2, \ldots \sigma_N)$ be the standard deviations so estimated. Then a new C is estimated from R as

$$C = S^T R S \tag{15}$$

The more complex technique for computing μ and C is referred to here as the E-M (Expectation-Maximization) method. The E-M method is more appropriate for use with truncated data. It uses an iterative procedure, which converges to the maximum likelihood estimators of μ and C. The technique requires the estimation of the statistics, given the condition that the data lie outside some truncation region. If there were no truncation, one could compute the sufficient statistics X and S, defined as follows:

$$\overline{X} = \frac{1}{M}\sum_{i=1}^{M} X_i \tag{16}$$

and

$$S = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T \tag{17}$$

where X_i is the vector of the observed values for the ith day and X_i^T is the transpose of X_i. Note that S is simply $1/M$ times the cross-product matrix. If the statistics X and S are given, the maximum likelihood estimates of the distribution parameters μ and C can be determined as follows:

$$\mu = \overline{X} \tag{18}$$

$$C = \frac{1}{M} S - (\overline{X})(\overline{X}^T) \tag{19}$$

In the case of a truncated data set, the expected values of $\overline{X}$ and S are of interest, given that some of the x_{ij} values have been excluded from the data set T used for the calculations. Having computed starting values for $\overline{X}$ and S as previously described, estimates of μ and C are computed and used in turn to compute new estimates for $\overline{X}$ and S; the procedure is described in the following. This process continues until convergence is achieved.

When x_{ij} is outside the data set T, the expected values of X are computed as follows:

$$E\left[\overline{X} \mid \mu^{(j)}, C^{(j)}, \left\{X_i \notin T\right\}\right] = \frac{1}{M+M_o}\sum_{i=1}^{M} X_i + \frac{M_o}{M+M_o} E\left[\overline{X} \mid \mu^{(j)}, C^{(j)}, \left\{X \in T\right\}\right] \tag{20}$$

$$E\left[S \mid \mu^{(j)}, C^{(j)}, \left\{X_i \notin T\right\}\right] = \frac{1}{M + M_o} \sum_{i=1}^{M} X_i X_i^T + \frac{M_o}{M + M_o} E\left[X X^T \mid \mu^{(j)}, C^{(j)} \left\{X \notin T\right\}\right] \quad (21)$$

where $E\,(\cdot)$ denotes the expectation of the expression within the parentheses, M_0 is the number of observations in the reduced data set T, $M + M_0$ is the total number of observations, and $\mu^{(j)}$ and $C^{(j)}$ are estimates of μ and C at the jth iteration. This method further truncates the data set by considering only those days for which the average observed concentration (not the maximum, as in the isopleth method and the simpler probabilistic method) is above a threshold. Tests indicate that 80 ppb is suitable when the data have already been screened to ensure at least one observation over 80 ppb, so that value has been used. This approach yields a linear truncation region so that the expectations can be expressed as functions of the multivariate normal distribution with 0 mean vector and unit variance written as $N\,(0, I_n)$, where I_n represents an $n \times n$ identity matrix. The truncation can exclude some observations that would have been included in the calculations for the simpler case.

Some terms must be defined before the iteration procedure is discussed. Let:

t = truncation (threshold) level, e.g., 80 ppb;
$t^* = \dfrac{t_n - \mu^{(j)} \cdot 1_n}{\|L\|}$;
n = number of cells;
1_n = vector of length n whose elements are 1;
$\mu^{(j)}$ = estimate of the mean vector at the jth iteration;
L = the lower triangular portion of the Cholesky decomposition of C at the jth iteration, i.e., L is lower triangular and $LL^T = C^{(j)}$;
ϕ = the univariate normal (0, 1) density;
Φ = the univariate normal (0, 1) cumulative distribution;
$E_1(t^*) = -\phi(t^*)/\Phi(t^*)$;
$E_2(t^*) = 1 - t^*\phi(t^*)/\Phi(t^*)$; and
$\gamma = L \times 1_n / \|L \times 1_n\|$, where $\|\,.\,\|$ denotes norm.

Then

$$\mu^{(j+1)} = \left[\Sigma X_i + M_0\,(\mu + L\gamma E_1)\right]/(M + M_0) \quad (22)$$

$$C^{(j+1)} = \left[\Sigma X_i X_i^T + M_0 \left[\mu^{(j)} \mu^{(j)^T} + L\,(I_n + E_2(t^*) - 1)\right] \gamma\gamma^T L^t + E_1(t^*) \left[\mu\gamma^T L^T + L\gamma\mu^T\right]\right\}/(M + M_0) - \mu^{(j+1)} \cdot \mu^{(j+1)^T} \quad (23)$$

In practice, the E-M algorithm generally produces estimates close to their convergence limit on the first iteration, with subsequent steps producing much slower convergence. Therefore, a few iterations (the code given in Appendix D uses ten as a default, but the user may change this) yield results that are comparable to those obtained after many more iterations. As with most iterative procedures, the performance of the E-M algorithm depends on the starting values. Two approaches to specifying the starting values have been tried. One of these uses the results obtained from the simpler method that was described earlier as the starting point for the E-M method. The second approach uses the truncated data set to obtain estimates as follows:

$$\mu^{(o)} = \sum X_i / M \quad (24)$$

$$C^{(o)} = \sum X_i X_i^T / M - \mu\mu^T \quad (25)$$

These starting values tend to overestimate the mean and underestimate the variance because they are obtained from the truncated data. However, the large step taken by the algorithm at the first iteration tends to eliminate these biases; therefore, the estimates change only slightly on subsequent iterations. The program described in Appendix D can use either approach as an option.

In the foregoing discussion, L and its transpose can be replaced by $C^{1/2}$, which is defined as:

$$C^{1/2} = ED^{1/2}E^T \quad (26)$$

where E is the matrix of eigenvectors of C, and $D^{1/2}$ is a matrix with the square roots of the eigenvalues of C on the diagonal and zeros elsewhere. Because the matrices L and $C^{1/2}$ differ only in multiplication by an orthogonal matrix, either can be used to transform one set of vectors to another with the desired covariance. A version of the E-M algorithm was tested using $C^{1/2}$ calculated by a singular valued decomposition that was verified by comparison with results obtained from a standard statistical software package (*16*), but the Cholesky decomposition described here yields an equivalent matrix with a fraction of the computation time.

As will be shown later, results obtained with the isopleth and the probabilistic methodologies tend to be similar. Thus, it will usually be sufficient to apply the isopleth methodology, which simplifies the calculation requirements considerably.

CHAPTER THREE

INTERPRETATION, APPRAISAL, AND APPLICATIONS

HOW METHODOLOGIES ARE APPLIED

The findings derived from this project were described in Chapter Two, which provided the theory underlying the methodologies that have been developed for determining numbers of exceedances and design values for ozone monitoring networks. Although those methodologies are important of themselves, their effective application depends on their incorporation into a more complete evaluation procedure. The appendixes to this report provide a user's guide for the complete methodology. The remainder of this chapter describes some results obtained by applying the methodology to selected data bases. The purpose of those applications has been to define monitoring requirements and to provide information by which transported and background ozone concentrations can be estimated and supplemental monitoring programs can be designed.

Basically, the procedure for estimating exceedances and design values so that control strategies can be developed to meet the National Ambient Air Quality Standards consists of six steps:

1. Acquisition of historical air quality and meteorological data.
2. Preprocessing and screening the data.
3. Design and execution of a supplemental monitoring program if the available data are not adequate.
4. Application of the isopleth and probabilistic methodologies to determine number of exceedances, design value, and critical ozone days. In most instances, application of the isopleth methodology will be sufficient.
5. Examination of the prevailing conditions on ozone days to determine: background ozone concentrations, origins of transported ozone, meteorological conditions conducive to high ozone concentrations, other information which might be required for modeling conditions on critical days.
6. Design and testing of control strategies.

The focus of this report is on the first five items in the list; the last item was outside the scope of this project.

EXAMPLES OF THE APPLICATION OF THE METHODOLOGIES

Description of the Data Bases Used

In order to test the methodologies that were developed, it was necessary to apply them to real data bases. These applications were also used to evaluate the sensitivity of the methodologies to the number and location of stations in a network. In order to do this, the monitoring networks with an especially large number of well-located stations had to be used. (As noted earlier, additional needed information includes data concerning meteorology, oxides of nitrogen, and nonmethane hydrocarbons.) A list of possible data sets was compiled. These included: Regional Air Pollution Study (RAPS), St. Louis—1976; Houston Area Oxidant Study (HAOS), Houston—1977; Houston Oxidant Modeling Study (HOMS), Houston—1978; Los Angeles—1976 through 1978; Philadelphia—1979; Tulsa, Oklahoma—1977; Northeast Oxidant Study—1975; San Francisco Bay Area—1978.

The monitoring program developed for the Tulsa study consisted of eight monitoring sites distributed along a south-to-north line through the Tulsa area. The monitoring period was from 1 July through 30 September 1977. Although the selected configuration is useful for examining the transport of ozone and its precursors when the wind is properly aligned, the density and spatial distribution of the monitors are not sufficient to evaluate the procedures developed in this study; the Tulsa data base was therefore rejected. Because the St. Louis (RAPS) and Houston (HAOS and HOMS) data bases were selected to evaluate the methodology, inclusion of data from the Tulsa area would also tend to overrepresent the south-central United States. The St. Louis and Houston areas are among the most intensively studied areas in the United States. Those data bases were included because they are far more complete than any available elsewhere. Furthermore, St. Louis represents a relatively typical Midwest location, and Houston, with its unusual industrial sources, represents an atypical area.

Both San Francisco and Los angeles have complicated land/sea and complex terrain influences on the air flow. However, Los Angeles historically has worse ventilation. Because of the greater emissions, reduced ventilation, and warmer average temperatures in the Los Angeles area, ozone concentrations tend to be higher in Los Angeles than in San Francisco. For this reason, the Los Angeles data were selected to represent West Coast locations.

The Philadelphia monitoring network was chosen to represent East Coast areas because it is more representative of a single urban area than was the Northeast Oxidant Study network.

A description of each of the data bases is given next. The two Houston data bases, HAOS and HOMS, have been treated as one data base for the Houston area.

St. Louis Regional Air Pollution Study (RAPS)

During the period of 1974 to 1977, the Regional Air Pollution Study (RAPS) was conducted in the St. Louis, Missouri, area (*17*). Twenty-five monitoring locations were selected, and the monitoring sites make up the Regional Air Monitoring system (RAMS) network. The locations of the monitoring stations are shown in Figure 17. The 1976 RAPS data were used because data for that year were most complete and reliable.

Houston Oxidant Modeling Study

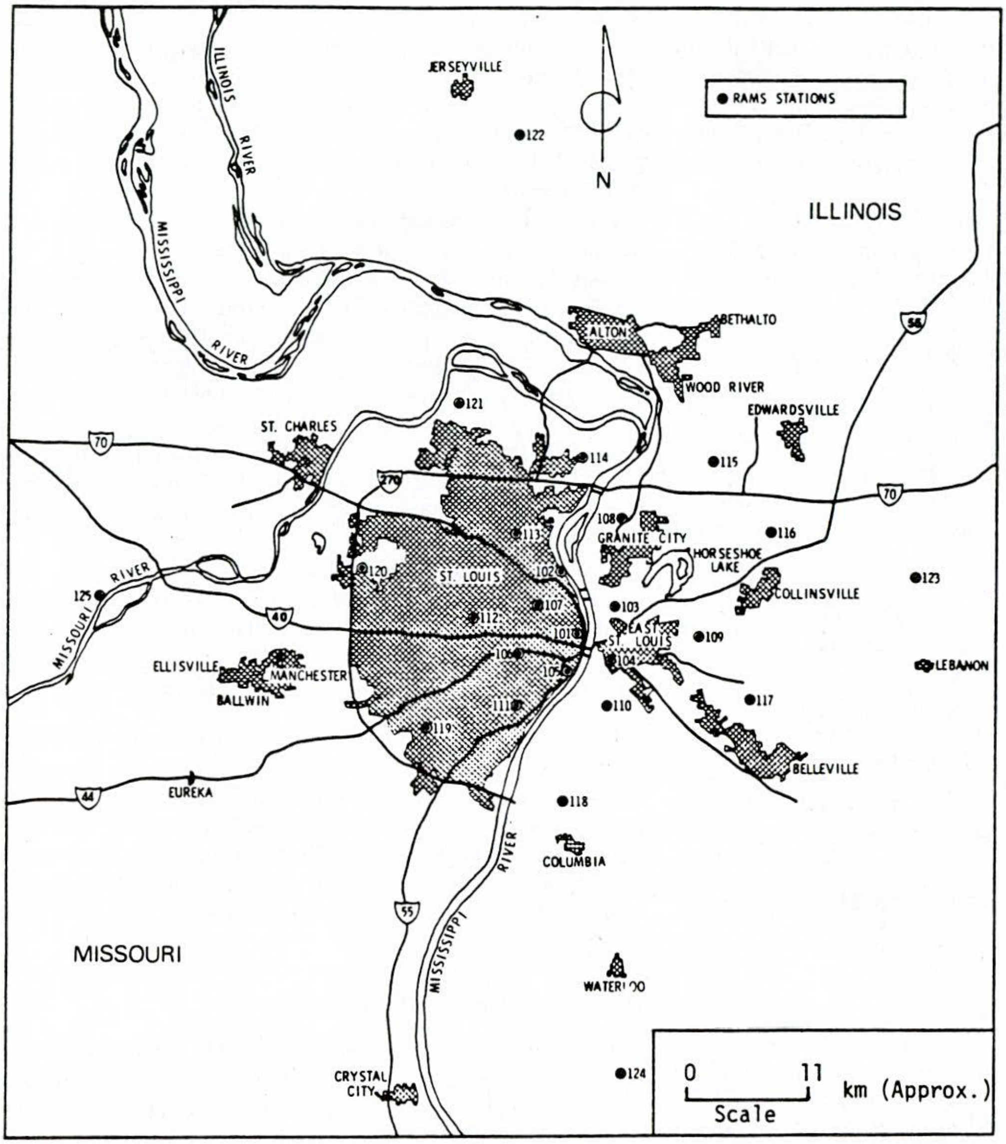
The Houston Oxidant Modeling Study (HOMS) was an intensive sampling program undertaken by the EPA and others in the Houston, Texas, area during the period 15 September to 12 October 1978 (*18*). The surface sampling program operated 17 permanent sites with 24-hour monitoring and three mobile laboratories with 12-hour monitoring (0600-1800 CST).

Houston Area Oxidant Study

The Houston Area Oxidant Study (HAOS) comprises data measured during the period from 1973 through 1977 (*19*). The number of stations operational during the study increased as the study progressed. A total of 93 different sites was used at one time or another during the study, but most of these measured only total suspended particulates (TSP). A maximum of 20 stations was monitored for ozone during the 1977 summer intensive program. The HOMS and HAOS stations that monitored ozone are shown in Figure 18. Because the density of stations monitoring ozone was far greater during 1977 than at other times, it was convenient to focus on those data when the network size was reduced for purposes of testing the sensitivity of the method to number of stations.

Los Angeles: South Coast Air Basin

Los Angeles is historically known for the photochemical



SOURCE: Ref. (17)

Figure 17. Location of the St. Louis Regional Air Monitoring Systems (RAMS) station.

pollutant problems that persist in the region. Los Angeles data are monitored at permanently operated monitoring stations. The most recent data for the year 1978 were used. Los Angeles is located in the South Coast Air Basin as defined by the California Air Resources Board (*20*). Figure 19 shows the spatial distribution of the monitoring sites in the South Coast Air Basin.

Philadelphia

During the summer of 1979 an intensive field monitoring program was conducted in the Philadelphia area (*21*). Seventeen sites were monitored for air pollutants and 12 meteorological parameters (Fig. 20) were measured. As noted earlier, the Philadelphia data provided an East Coast urban data set of comparable quality to those from the other areas.

Effects of Numbers of Stations in a Network

In order to test the effects of network size on the numbers of exceedances and design values, the data bases described in the preceding section were used and data collected at certain observing sites in each of the networks were artificially excluded. By so doing, the researches were able to obtain information about how the number and distribution of monitoring sites affect the estimation of design value and number of exceedances for the networks. Subsets of stations were drawn from the complete networks shown in Figures 17 through 20. In St. Louis and Philadelphia, the original network consisted of 25 monitoring sites; smaller networks of 16 and 9 monitoring sites were selected from among those 25. The selected subsets of monitoring sites in St. Louis are given in Table 9. These subsets were chosen to provide

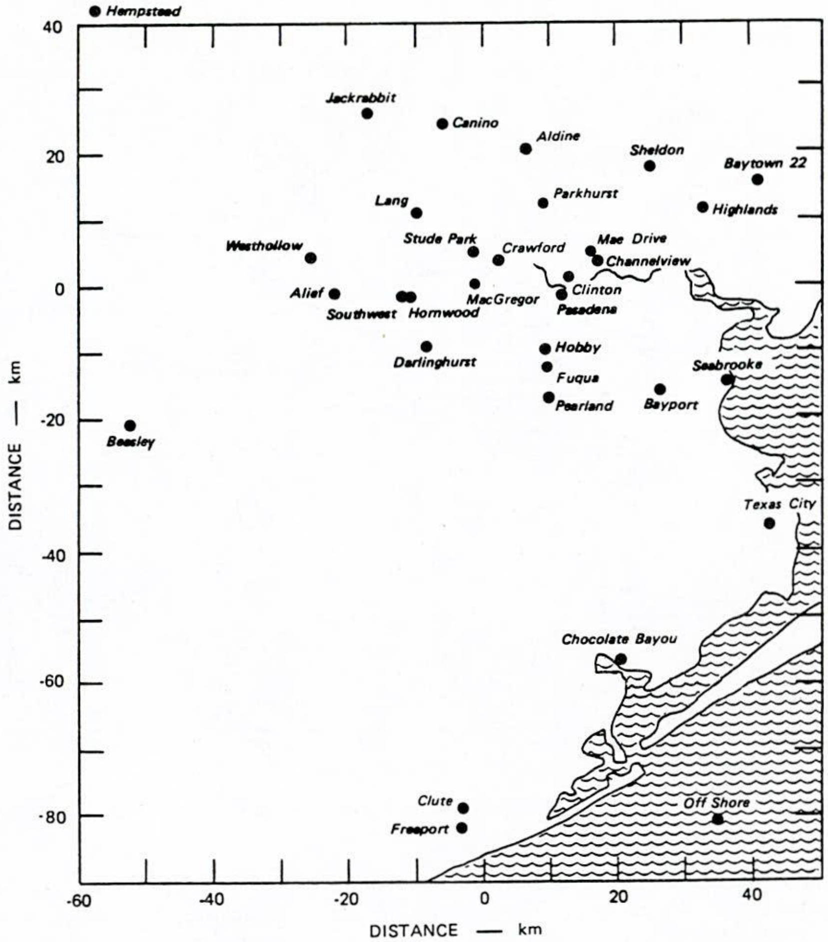
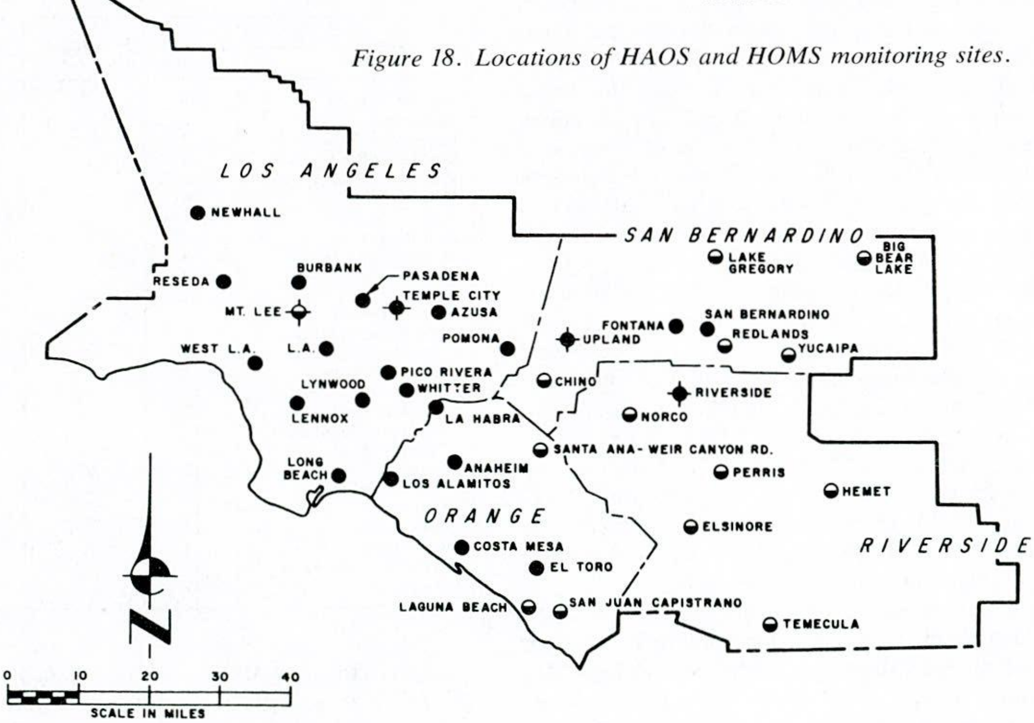


Figure 18. Locations of HAOS and HOMS monitoring sites.



SOURCE: CALIFORNIA AIR RESOURCES BOARD

Figure 19. South Coast air basin operating stations, October–September 1976.

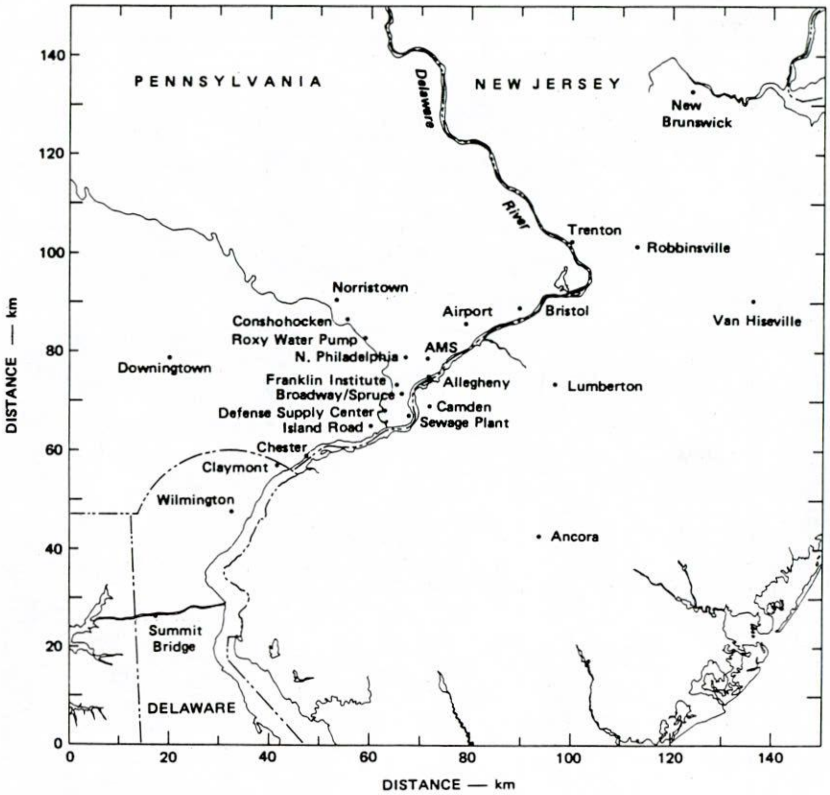


Figure 20. Air quality and meteorological monitoring stations in the Philadelphia area.

Table 9. Summary of St. Louis sites used in reduced network tests.

Station Number	9-Station Subset	16-Station Subset
101	X	X
102		
103		
104		
105		
106		
107		
108		X
109		X
110		
111		
112		X
113		
114		X
115	X	X
116		X
117	X	X
118		X
119	X	X
120		X
121	X	X
122	X	X.
123	X	X
124	X	X
125	X	X

Table 10. Summary of Philadelphia sites used in reduced network tests.

Site Name	Site Number	9-Station Subset	16-Station Subset
Summit Bridge	1		X
Downington	2	X	X
Lumberton	3		X
Robbinsville	4	X	X
Van Hiseville	5		X
Chester	6		
Bristol	7	X	X
NE Airport	8	X	X
Allegheny	9		
Camden	10		X
Ancora	11	X	X
Philadelphia:			
S. Broad and Spruce	12		
Franklin Institute	13	X	X
AMS Laboratory	14		X
Roxy Water Pump	20		X
Island Road	21		X
Defense Support	22		
Sewage Plant	23		
Claymont	15	X	X
New Brunswick	16	X	X
Trenton	17		
Conshohocken	18		
Norristown	19	X	X
North Philadelphia	24		
Wilmington	25		

coverage of the central area and to have monitoring sites well outside the central area. The monitoring sites outside the central area are necessary because ozone concentration maxima are known to occur away from cities in the downwind direction. Furthermore, it is desirable to have monitors located so that their data are suitable for estimating background or transported concentrations. Similar considerations were involved in the selection of the subset of monitoring sites for Philadelphia (Table 10), but the available data sets were not as complete as in St. Louis, so it was necessary to consider the completeness of data sets from the different locations when the subset was selected.

Larger networks of stations were available for Houston and Los Angeles. In Los Angeles, two smaller networks (having 20 to 10 stations as given in Table 11) were selected with the same considerations in mind as previously discussed in connection with the St. Louis and Philadelphia data. Inasmuch as data from the HAOS and HOMS programs had been combined, some of the 32 stations in the Houston area were operated during only one of the two programs. Therefore, in selecting the 22- and 11-station subsets of monitoring sites (given in Table 12), it was intended to emphasize those sites that were common to the two programs.

Table 13 summarizes the effects of changing the size of the network on the calculated design values and the numbers of exceedances determined by the isopleth methodology. For most of the cities, the calculated design values do not change much when the number of stations in the network is decreased. The numbers of exceedances do decrease appreciably in most cases, but this appears to be an artifact of the method that will be discussed later.

It is apparent from the results given in Table 13 that the calculated design values change little when the numbers of stations in the network are reduced, except for St. Louis. In the other three cities the difference between the design values calculated for the full network and for the network with the fewest stations is 10 percent or less of the full network value. The St. Louis results change very little when the number of stations are reduced from 16 to 9; the large changes in design value are between the 25- and 16-station networks.

Table 11. Summary of Los Angeles sites used in reduced network tests.

Site		10-Station Subset	20-Station Subset
Name	Number		
Los Angeles, Downtown	1	X	X
Azuza	2		X
Burbank	3		X
North Long Beach	4	X	X
Receda	5	X	X
Pomona	6		X
Lennox	7		X
Whittier	8		X
Newhall	9	X	X
Pasadena-Wal	10		
Lynwood	11		
Pico Rivera	12		
West Los Angeles-Robertson	13	X	X
Mt. Lee	14		
Anaheim	15		X
La Habra	16		
Costa Mesa-Harbor	17		X
El Toro	18	X	X
Los Alamitos-Orangewood	19		
Santa Ana-Weir Canyon	20		X
Norco-Prado Park	21		
Riverside-Rubidoux	22		
Riverside-Magnolia	23		X
Perris	24	X	X
San Bernardino	25		X
Redlands	26		
Upland-ARB	27	X	X
Fontana-Foothill	28		
Lake Gregory	29	X	X
Yucaipa	30	X	X

Table 12. Summary of Houston sites used in reduced network tests.

Site		11-Station Subset	22-Station Subset
Name	Number		
Aldine	1	X	X
Crawford	2	X	X
Fuqua	3	X	X
Westhollow	4	X	X
Mae Drive	5	X	X
Clinton	6		X
Parkhurst	7		X
MacGregory	8		
Lang	9		X
Jackrabbit	10	X	X
Texas City	11	X	X
Clute	12	X	X
Hempstead	13		X
Offshore	14	X	X
Southwest	15		X
Bayport	16		
Freeport	17		
Baytown	18	X	X
Beasley	19	X	X
Chocolate Bayou	20		X
Darlinghurst	21		X
Homewood	22		X
Pasadena	23		
Channelview	24		
Seabrooke	25		X
Highlands	26		
Camino	27		
Pearland	28		
Sheldon	29		X
Hobby	30		
Alief	31		X
Stude Park	32		

Table 13. Summary of design value and exceedance estimates for various network sizes calculated using the isopleth methodology.

Area	Number of Stations Used	Design Value (ppb)	Number of Exceedances
Houston	32	298	61
	22	299	60
	11	288	46
Los Angeles	30	456	148
	20	453	141
	10	408	131
Philadelphia	25	182	18
	16	182	19
	9	182	10
St. Louis	25	294	28
	16	216	28
	9	210	22

When the St. Louis data were reexamined, it was found that on those days with the highest concentrations, only one station reported very high values. When that single station was removed, the isopleth technique no longer estimated high values for the cell in which that station was located. It appears that the data collected at the one station were either anomalous or in error, although errors were not detected in the original screening. The isopleth methodology determines the dates on which the highest concentrations were observed and allows the determination of background ozone and prevailing meteorological conditions. This also permits identification of days that should be subjected to post-calculation screening. The individual data points for these days can be examined more carefully than is usually possible during the initial screening of the full data set.

With the exception of the change in the design value between the 25- and the 16-station St. Louis networks, the changes in estimated design values shown earlier are within about 10 percent of the value determined for the network with the largest number of stations. In order to put this into perspective, the ten highest values for ten different grid cells in the St. Louis network were examined. The design values derived for each cell by three different EPA methods were compared. In one method, the ten points were plotted versus the corresponding percentiles and smooth curves were drawn according to the graphical method described in Chapter Two. The values determined by this method were compared with those from the table look-up method and the distribution fit method. This provided three different estimates of design value for each of the ten cells. The percentage difference between the largest and smallest of the estimates ranged from about 2.5 to more than 20 percent. In three cases, the differences were less than 5 percent; the differences were between 5 and 10 percent for five of the cases, so the differences arising from reducing the number of stations to about ten are comparable to differences in values obtained by different EPA estimation methods. The results suggest that a network of nine or ten stations should be adequate for determining design values.

The reduction in the number of exceedances calculated as the number of monitoring stations decreases is more striking than are the changes in estimated design value. However, these changes appear to be the result of a very conservative approach to the estimation of concentrations for each cell. It should be recalled that the estimated concentration for a cell on a given day is taken to be the highest value observed within that cell, or the interpolated value, if that is higher than the highest observed value. Thus, for cells that contain several stations, the overall number of exceedances is likely to be more (and cannot be less) than found for any individual station within that cell. If a station is removed from the network and that station had observed one or two exceedances on days when neither the interpolated value nor any of the other stations observed exceedances, the number of exceedances calculated for the cell will be correspondingly reduced. This will be true for each station in the cell. Thus, the number of exceedances for the cell will appear to decrease even if the single station that observed the most exceedances is not removed. The conservative approach that has been taken in the isopleth methodology results in a pronounced tendency to overestimate numbers of exceedances whenever the network includes cells that have more than one monitoring site within them, so the reduction in the number of sites probably leads to better estimates of the actual number of exceedances from a network. Nevertheless, the conservative approach seems prudent and more consistent with EPA philosophy.

Accuracy of the Estimation Methods

The ability of the various methods to model an observed design value for a grid cell containing a monitoring station is a measure of accuracy. Several steps were taken to compare real and modeled design values. The design value for a cell containing one or more monitoring stations was computed from the observed values in that cell using the EPA curve-fitting and table look-up methodologies. (These will be referred to as "the EPA design values" in the following discussions.) The measured values in that cell were ignored, and the isopleth method was used to estimate a design value for the cell. (This estimate will be referred to as "the isopleth design value.") And the results of the isopleth model were input to the probabilistic modeling program, which computed a simulated design value via both the simple and the expectation-maximization (E-M) techniques.

The foregoing procedure was applied to four grid cells from St. Louis. The cells were selected so that each had one or more monitoring stations that had reported values a high percentage of the time. Runs of the isopleth and probabilistic programs were made for each of the cells. The results of these runs are given in Table 14. The figures in this table were obtained by using the full network, and ignoring only the observed values in the cell being tested in that run. The network may be reduced by deleting the data from other stations. The reduced network values for grid cell 20 are given in Table 15. The striking feature of this table is the improvement in the agreement between E-M technique design values and those obtained by other methods for cell 20 when the number of stations is reduced. Generally, it was observed that the design values for these and other cells as estimated by both the simple estimation and the E-M techniques tended to follow the design value as obtained by the isopleth technique as the number of stations in the network are reduced.

Effects of Excluding Selected Days from the Calculations

Only those days when at least one station within the network observed ozone concentrations of 80 ppb or greater were selected for calculating numbers of exceedances and design values. It was assumed that no day would be excluded on which interpolated values might have exceeded 120 ppb. This has proven to be so. The calculations were rerun using the complete network for each of the four cities, excluding all days when no station observed concentrations of 110 ppb or more. The estimated design values were identical with those determined from the data set that excluded only cases with observed concentrations below 80 ppb. One fewer exceedance was found for St. Louis; in other respects the results were the same with regard to numbers of exceedances as well.

The results of excluding selected days from the calculations have practical implications. They suggest that monitoring need only be pursued when concentrations are over some threshold (e.g., 100 ppb). Design of a supplemental monitoring program can make use of this fact to limit the number of

Table 15. Estimated design values (ppb) at Station 20 when network size is reduced.

		Probabilistic			
Network Size	Isopleth	Simple	σ^a	E-M	σ
Full (excluding Station 20)	181.6	180.7	0.03	180.88	0.04
16-station	185.1	182.1	0.08	180.71	0.03
9-station	179.7	178.2	0.04	178.13	0.02

[a] σ = standard deviation

Table 14. Summary of St. Louis design value estimates (ppb) for different methods.

Cell Identifier	EPA Methods			Probabilistic			
	Table Lookup	Curve Fit	Isopleth	Simple	σ^a	E-M	σ
3	168	165	154.5	169.3	0.27	172.71	0.14
14	187	189	177.6	182.7	0.33	179.97	0.17
20	198	201	181.6	180.7	0.03	180.88	0.04
16	160	163	169.5	180.3	0.15	180.42	0.15

[a] σ = standard deviation

days during which monitoring must be undertaken. The following section provides some guidance in determining the days on which monitoring should take place.

SUPPLEMENTAL MONITORING REQUIREMENTS

Spatial Coverage

The results given in Table 13 suggest that good estimates of design value and numbers of exceedances can be obtained if measurements are available from about ten well-placed monitors operated on days with high ozone concentrations. The estimates obtained for design value with such a network will generally be better than the estimates of the number of exceedances. Inasmuch as the main purpose of estimating exceedances and design values is to determine if the region is in compliance with the standards (and if not, to develop a strategy to achieve compliance), the design value is more important because the design value forms the basis for control strategy. Thus, supplemental monitoring should be undertaken when existing networks have fewer than about ten stations. However, in recommending that supplemental monitoring be undertaken when the network has fewer than ten sites, one is only referring to large urban areas like those previously discussed. Furthermore, it should be possible to tailor the supplemental monitoring to the specific locale and the specific meteorological conditions if mobile monitors are available.

Temporal Coverage

As discussed earlier, the results change very little when days are excluded if no station observed concentrations over 110 ppb. The exclusion of days on the basis of highest observed ozone can be useful for reducing the calculations, but it does little to provide guidance concerning whether or not monitoring should have been conducted. That is, if one were to attempt to plan a supplemental monitoring program, it would be very important that that program be in operation during those days when ozone concentrations above about 100 ppb were likely to be observed. Ludwig and Shelar (*14*) noted that high ozone concentrations were most likely to be associated with high temperatures. A reliable relationship of this type would be very useful for defining periods when a full complement of monitoring stations would be required.

Figures 21 through 24 are scattergrams with maximum ozone concentrations on a given day as a function of the maximum temperature observed at the same time on the same day for Houston (the HAOS data base for 1977), Philadelphia, St. Louis (the RAPS data base for 1976), and Tulsa, respectively. Each asterisk represents a single case; if more than one case has the same combination of maximum ozone and maximum temperature, a numeral is plotted showing the number of cases. Examination of the figures shows that excluding days with maximum temperatures under 22 C (72 F) would not eliminate any days with observed ozone concentrations exceeding 120 ppb. If cases with maximum temperatures less than 23 C (74 F) are excluded, virtually all of those days when maximum ozone concentration exceeded 120 ppb are still retained.

A monitoring criterion based on maximum temperature has considerable potential usefulness. First, it can be used in combination with routinely available climatological information to define the seasons when ozone must be measured. Second, such a criterion could be used operationally to select days when equipment should be deployed or operated. Routine daily forecasts of maximum temperature would then serve to guide monitoring operations.

The isopleth methodology was applied to the 25-station Philadelphia data base to determine the effect of excluding days from consideration when the maximum temperature at a key location within the area was 22C (72F) or less. The results are given in Table 16. It appears from the table that calculated design value is virtually unchanged if the days with the lower temperatures are not included in the calculations, but exceedance estimates are lowered. Of course, in an operational application, it would be more appropriate to use a lower temperature as a criterion for monitoring in order to compensate for errors in the forecasting of maximum temperature. A logical choice for such a criterion would be about 20C (68F).

DEFINITION OF SUPPLEMENTAL MONITORING PROGRAM

The results discussed in the preceding section provide a basis for defining a supplemental monitoring program. If the appropriate climatological statistics are available, it is therefore possible to define months when monitoring for ozone is required in any given locale. A monitoring program should be operational during any month when the climatological records indicate that temperatures are likely to exceed about 22C (about 72F) on more than one day during the month.

Unfortunately, easily available statistics are not generally compiled in this form. Instead, the climatological records more frequently give the average daily maximum temperature for the month, the highest value ever observed, and the average number of days when the temperature exceeds 32C (90F). Two alternatives are available. First, the National Climatic Center in Asheville, North Carolina, can be requested to prepare a special summary listing the frequency of occurrence of temperature maxima greater than 22C (72F) for each month in the locality of interest. A second, more convenient but not quite as accurate, approach is to use published cooling degree day data as a guide. Cooling degree days are calculated using the average of the maximum and minimum temperatures for a day; the number of cooling degree days for any given day is equal to the number of degrees by which the average for that day exceeds 18C (65F). For example, a day with maximum and minimum temperatures of 23C (74F) and 15C (60F), respectively,would have an average temperature of 19C (67F) and two cooling degree days. Cooling degree days are not wholly reliable as criteria for determining what seasons should be subject to monitoring because they only reflect the observed maximum temperatures indirectly; but, as a rule of thumb, it appears that months with an average of ten or fewer cooling degree days are quite unlikely to have ozone concentrations exceeding the standard.

One must be careful that the site from which the meteorological data were obtained is truly representative of the whole area. For example, cooling degree day data for San Francisco suggest that the only months when exceedances are likely to occur are September and October. This is be-

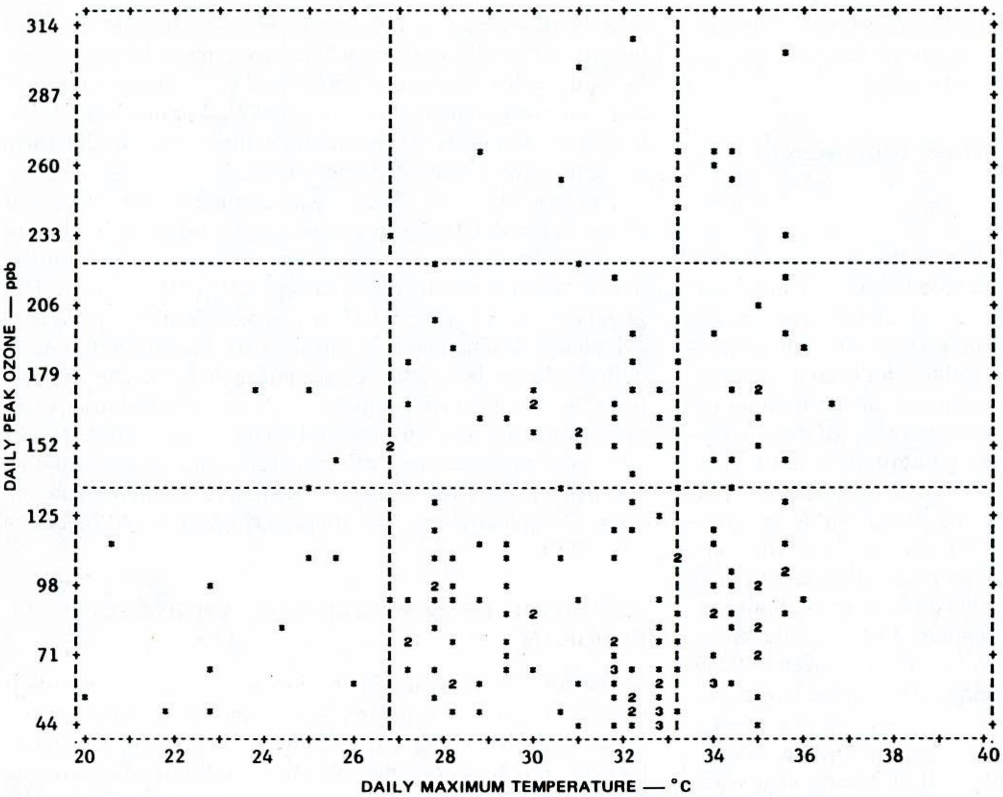


Figure 21. Scattergram of ozone as a function of daily maximum temperature for Houston (HAOS 1977) data.

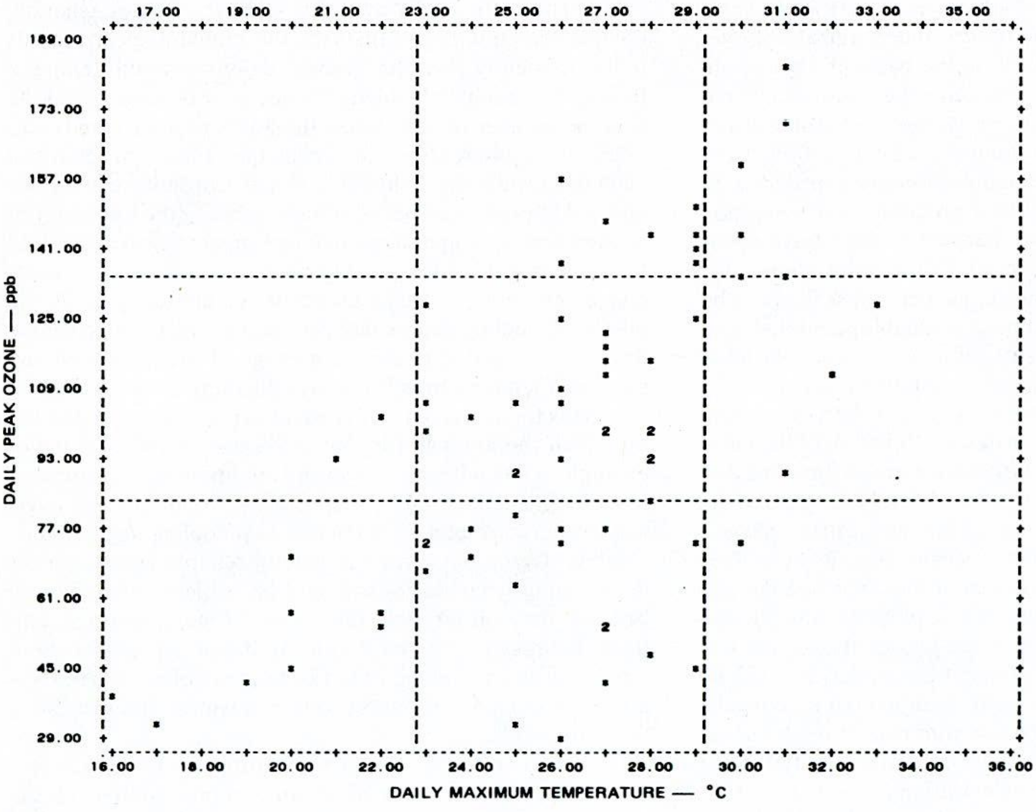


Figure 22. Scattergram of ozone as a function of daily maximum temperature for Philadelphia.

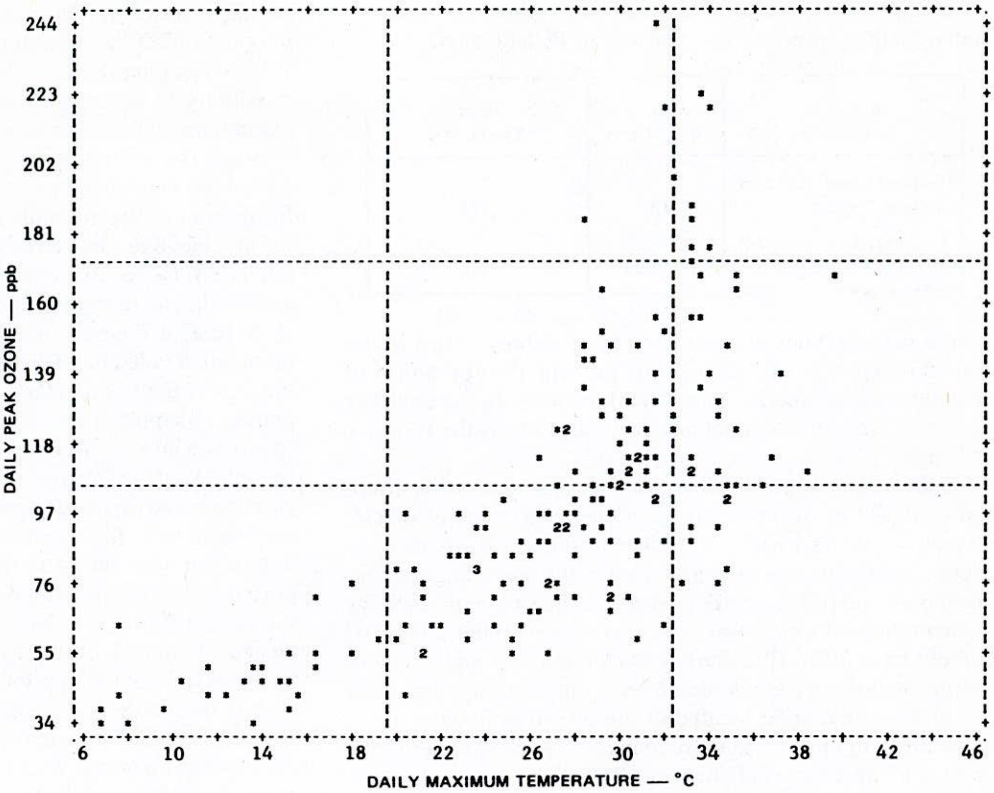


Figure 23. Scattergram of ozone as a function of daily maximum temperature for St. Louis (RAPS) data.

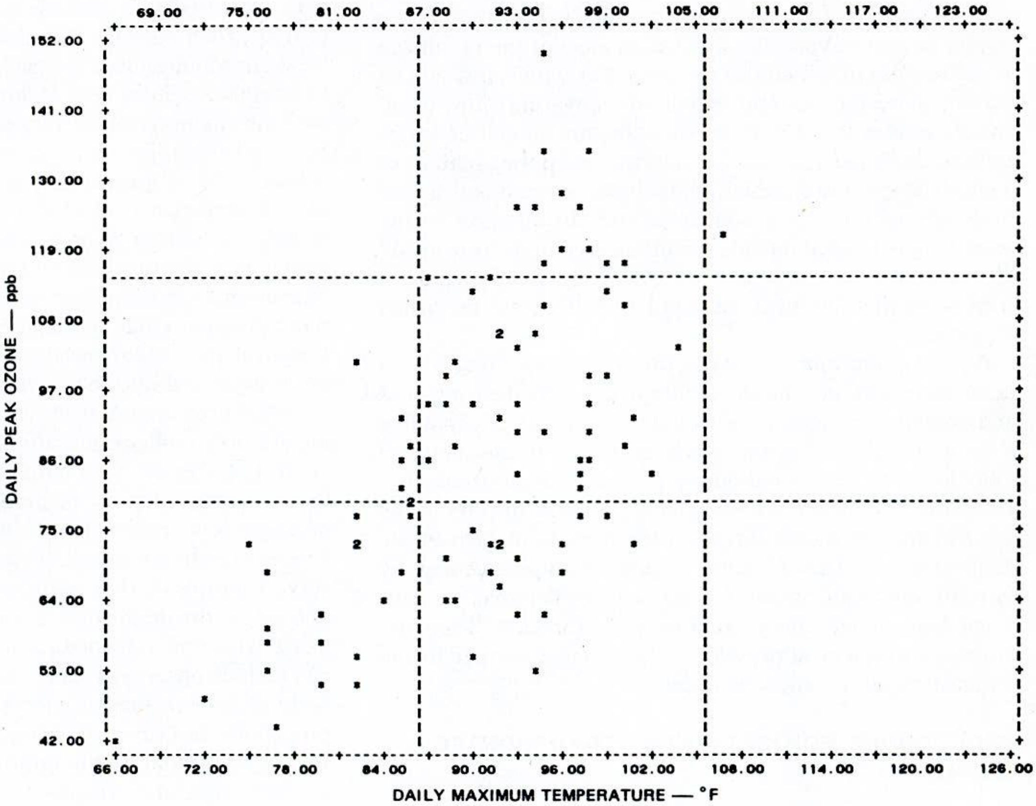


Figure 24. Scattergram of ozone as a function of daily maximum temperature for Tulsa.

Table 16. The effect of excluding days when the maximum temperature was 22 C or less in Philadelphia.

Parameter	Using All Data	With Days >22°C Excluded
Calculated design value (ppb)	182	178
Calculated number of exceedances	18	10

cause cool summer breezes keep temperatures down in the San Francisco area, but do not prevent the formation of ozone at other nearby Bay Area locations that are farther inland. One source of cooling degree day data is the *Weather Almanac*.

Once it has been decided what months are critical and when it will be necessary to be prepared to monitor ozone, it may not be necessary to operate all stations every day. As suggested by the analysis discussed in the preceding section, it should be sufficient to deploy the instruments on days when temperatures are expected to exceed about 20 to 21C (about 68 or 70F). If resources are limited, but mobile monitoring stations are available, at least one station, permanent or mobile, should be located in the direction forecast to be upwind from the city, and the remainder located either in the city or in the downwind direction. The best locations for such sites are described in more detail in Appendix F; still more detail is given by Ludwig and Shelar (*14*). Basically, there should be at least one station about 5 hours travel time downwind from the upwind edge of the urban area. If the urban area is so large that this distance falls within it, the site should be just beyond the downwind edge of the urbanized area. Because of difficulties in forecasting winds, it would be better to have three or four mobile sites covering a downwind arc of perhaps 90°. For fixed sites, the arc should be larger and should be located where it is downwind during periods of highest temperature (which should also correspond to periods of highest ozone concentration). In addition to the several sites located outside the urban area in the downwind direction, and the one or two sites in the upwind direction, at least some measurements should be made within the urban area itself.

A good compromise between the cost of operating a large network of stations and the ability to estimate design values and exceedances accurately would be a network of about five fixed stations—one in the urban area, one in the direction most likely to be upwind during periods of high ozone concentration, and three covering an arc outside the city in the most likely downwind direction for periods of high ozone concentration. This network would be supplemented by three or four mobile monitors that could be deployed on days when high ozone concentrations were forecast. Their deployment would be at preselected locations, governed by the forecast meteorological conditions.

PROCEDURES FOR ESTIMATING TRANSPORTED OZONE

Appendix E provides a complete guide to the estimation of the concentrations and origins of background ozone transported into an area. The remainder of this section provides an overview of the approach that has been adopted.

In developing the approach described in Appendix E, the possibility of developing an objective methodology for estimating transported ozone was considered, but discarded because the meteorological and photochemical conditions varied too much from day to day to permit a reliable scheme. Furthermore, the amounts and kind of information required for an objective method of estimating transported ozone are not likely to be consistently available. A more subjective approach was therefore chosen.

At first, a method of estimating transported ozone that involved detailed hand analysis did not appear to be feasible, but it was then recognized that the main purpose for determining transported and background ozone concentrations is to provide information that can be used to develop strategies for achieving compliance with ozone standards. Such strategies are based on the design value concept, so it is not necessary to address the question of transported ozone for every day in the data set. It is necessary to determine the transported ozone component only for those days used to derive the design value; the design value plays the key role in the design of control strategies to achieve compliance with the standards. The presumption in designing a control strategy is that if the control strategy can reduce the concentrations used to derive the design value to the point where the concentrations no longer exceed the standard, there will be generally compliance with the standard. Thus, one has only to examine the meteorological conditions that prevailed for the ten highest ozone-concentration days from which the design value is estimated.

In order to develop an effective control strategy, the transported ozone entering the area must be known so that the local contribution to the observed ozone concentrations can be evaluated. Some sort of modeling is then applied to evaluate the effectiveness of various proposed control strategies. Several basic approaches to modeling are possible, including rollback, the empirical kinetic modeling approach—EKMA (*23*), Lagrangian (trajectory) photochemical modeling, and Eulerian (airshed) photochemical modeling. Each type of model has slightly different requirements for transported ozone inputs, but the first three are quite similar. The Eulerian approach tends to require more information about the temporal and spatial variability in ozone concentrations that are transported into the area than do the other three.

Differences arising from varying requirements for modeling are only one consideration in developing and evaluating control strategies. Simulating historical conditions requires that one knows the ozone amounts transported into the city on those few critical days that determine the design value. This is a relatively straightforward, but not necessarily easily solved, problem, that requires some procedures by which one may estimate the ozone concentrations entering the area using whatever information is available. That information can include observed ozone concentrations and a substantial body of meteorological information in many instances. The procedure described in Appendix E uses corollary data of this type to generate the information required to devise and evaluate control strategies.

It is important to remember that many regions will impose control strategies in the future, so background ozone concentrations transported into any area will be affected by control measures that have been imposed in upwind regions. Thus, the evaluation of transported ozone must include provisions for estimating how the incoming ozone concentrations might be affected by future control measures imposed upwind—or in the area itself, when the air is relatively stagnant. Thus, the techniques for treating transported ozone provide for calculating historical air trajectories to reveal influences that might be reduced by future control measures in other areas.

Another factor has also been considered. In certain situations, ozone concentrations reach relatively high values from natural causes: The best known of these are associated with intrusions of stratospheric air to relatively low altitudes. Some guidelines have been provided to help users recognize meteorological situations that might accompany high ozone concentrations from natural causes. The guidelines for the interpretation of the ozone standard (*1*) provide that some days may be excluded from consideration by an EPA Regional Administrator. Presumably, days when ozone concentrations exceeded the standard from natural causes would be among such excluded days and would be removed from the data set before calculating exceedances and design values.

One of the first guides given for estimating transported ozone was included with the discussions of the EKMA (*23*), which pointed out that surface ozone readings are most indicative of ozone concentrations through the mixing layer when they are made upwind (or outside the urban plume) after the nocturnal radiative inversion has broken and before photochemical ozone has formed from residual precursors in the transported air. EPA suggests that the best indications of transported ozone are obtained from concentrations averaged over a 3-hour period between about 1100 and 1300 LDT (local daylight time). Some ways that commonly available meteorological data can be used to define the time of breakup of the nocturnal inversion are described in Appendix E.

A somewhat different approach to the estimation of background ozone for use with EKMA was taken by Maxwell and Martinez (*24*), who suggest selecting background ozone concentrations on the basis of the vector-averaged 0600-1400 wind direction; the background ozone concentration is assumed to be the peak ozone concentration recorded at the most distant upwind station. Essentially, this method includes another component in the background or transported ozone concentration. It is assumed that the peak ozone concentration at an upwind site would have been observed whether or not the city existed. This is tantamount to a redefinition of background ozone to include not only that ozone which is transported as ozone, but also that ozone which can be generated photochemically from transported precursors. The authors of this report believe that this is a more valid definition for purposes of designing control strategies, but the choice can be made at the time the procedure is applied, according to the user's requirements. The use of peak ozone concentrations or, to a lesser extent, use of 1100 to 1300 LDT concentrations, to define the transported ozone presupposes that the material being transported later in the day is similar to that arriving early in the morning. The material arriving early in the day is directly involved with the local emissions that produce the highest ozone concentrations. If the source of the transported ozone (and precursors) is a nearby city, it will be incorrect to assume relative uniformity throughout the day. If the origins of the transported material are sufficiently distant, the mixing that takes place during transport should be sufficient to support an assumption of relative uniformity between the morning and afternoon hours.

The northeast corridor of the United States is an area where the cities are close together and frequently aligned with the wind direction when ozone concentrations are high. Therefore, it would be difficult to justify the assumption that conditions observed later in the day are similar to those during the critical morning rush hour. Under such conditions, it would be wise to use the earliest time when a valid estimate can be obtained for the ozone concentrations through the depth of the mixed layer. Ludwig (*25*) has analyzed more than 250 vertical profiles taken in the vicinity of urban areas and has developed some guidelines for judging when surface observations of O_3 can be used to characterize conditions through the mixed layer. Those guidelines form the basis for the recommendations made elsewhere in this report.

Figure 25 is a schematic outline of the procedure for identifying transported and background ozone components. As shown in the figure, it is first necessary to identify "critical" days (i.e., the days from which estimates of the design value have been obtained). The next step is to acquire all the important information available for those days, because the best estimate of background ozone will be obtained when all the meteorological and air quality data are carefully considered. Figure 25 lists some of the important data that should be assembled and examined to determine if the observed ozone concentrations could have arisen from natural causes, such as intrusions of ozone-rich stratospheric air. The procedure described in Appendix E provides descriptions of conditions commonly associated with the stratospheric intrusions. If there is reason to believe that natural causes are indicated, the next step would be to contact the EPA Regional Administrator and request that the days in question be examined and exempted from consideration for purposes determining the design value. The monitoring data on these days should also be reexamined to ensure their validity.

If no natural causes are evident and all data are valid, the next step is to determine the airflow in the region. The way in which this is done will depend on the data that are available, but the purpose is to determine whether air movement (and the direction of movement) is generally well defined or whether conditions are more stagnant.

Figure 26 shows types of stations that might be used to define transported ozone concentrations when the wind direction is well-defined (Figure 26(a)) and when it is stagnant (Figure 26(b)). In the latter case, the ozone concentration that is sought is not a transported ozone, but rather a background value. In either event, the next step is to determine when the inversion was broken so that measurements of ozone at the surface are representative of concentrations through the greater depths of the atmosphere.

Among the "signatures" used to determine whether or not vigorous vertical mixing has taken place are (*25*): sudden increases in wind speed, which signal transport of momentum downward with the onset of mixing; sudden changes in temperature, which indicate downward transport of heat; relatively rapid rises in ozone concentrations at the surface,

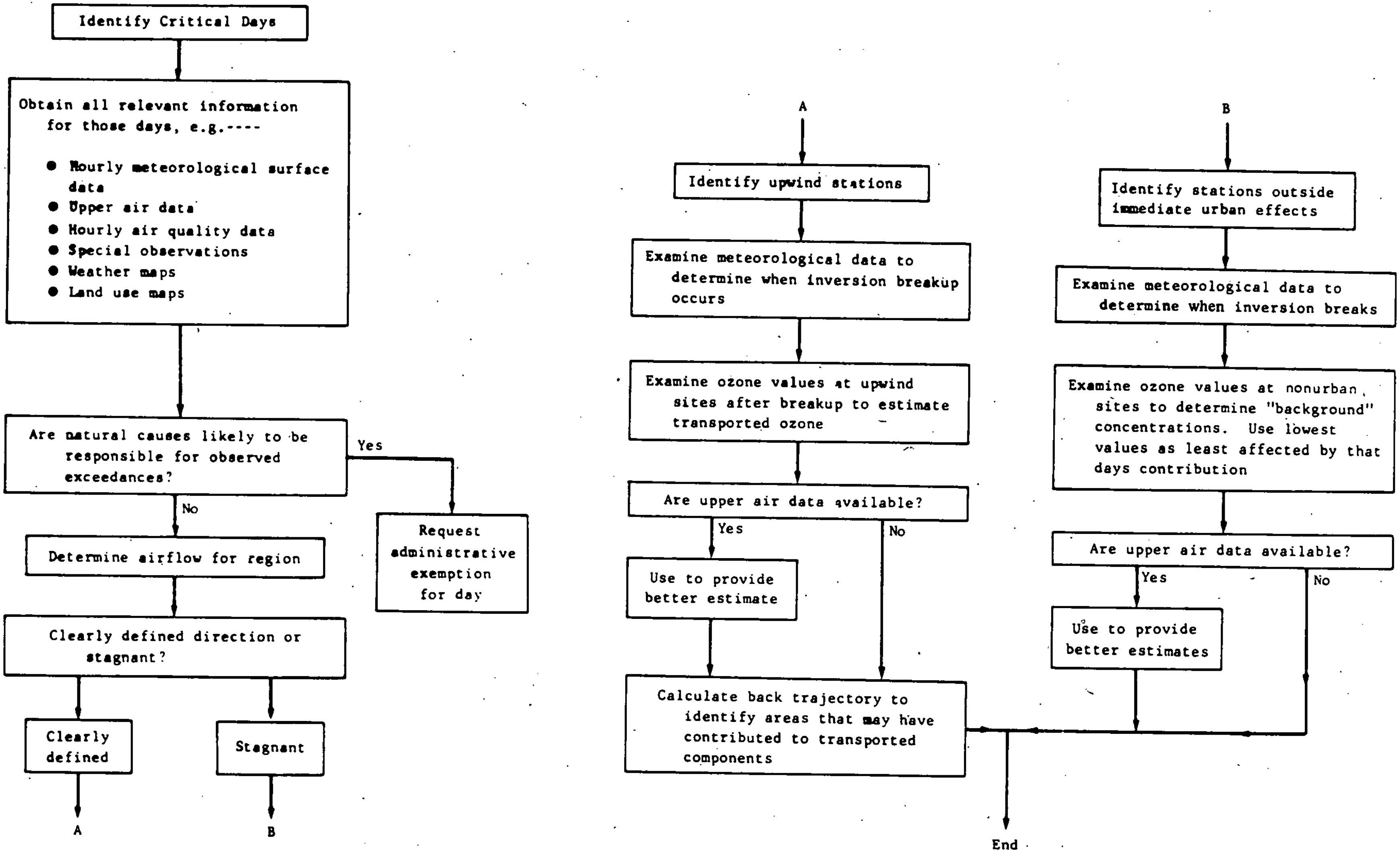
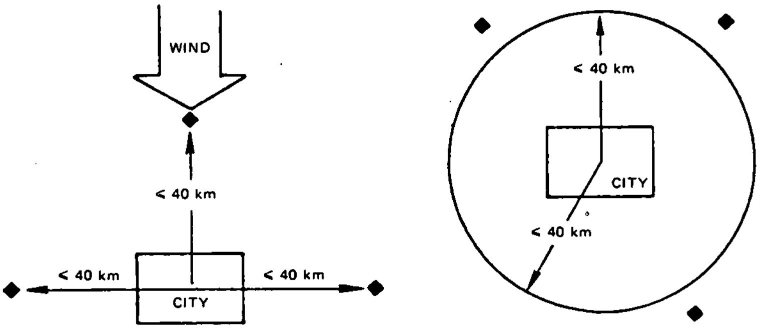


Figure 25. Schematic diagram of a procedure for estimating transported and background ozone.



(a) PREFERRED LOCATIONS WITH PREVAILING WINDS

(b) PREFERRED LOCATIONS IN STAGNANT CONDITIONS

SOURCE: Ref. (23)

Figure 26. Example of acceptable monitoring locations for estimating transported ozone.

which indicate downward transport of ozone from aloft; and surface temperatures that reach levels at which the morning balloon sounding (rawinsonde) measurements indicate that a neutral lapse rate should have been established so that vertical motions are no longer damped by atmospheric stability.

Keys like those previously listed are used to determine when surface observations at the selected sites are representative of concentrations through a relatively great depth in the atmosphere. If special observations have been made from aircraft or balloons, so that the average concentration of ozone through the depth of the mixing layer can be determined directly, they should be used.

One step remains in the procedure if air entering the city has been transported from elsewhere. That step is to identify the possible source regions for that air by calculating historical trajectories to identify any cities that may have contributed to the background concentrations on the critical days. This identification provides a basis for estimating how the transported background ozone might change when future control measures are implemented. Appendix E describes some relatively simple graphical techniques that provide adequate estimates for this purpose, but some computer programs are also available in the public domain that can be used to calculate trajectories (*8*).

POTENTIAL APPLICATIONS TO TRANSPORTATION PLANNING

State Implementation Plans (SIPs) to achieve ambient air quality standards were first required by the Clean Air Act of 1970. The Act also required Transportation Control Plans (TCPs) for the most severely polluted areas, but the early TCPs were largely ineffective because of a lack of coordination with the ongoing transportation planning process. The need for coordination and involvement of local agencies was recognized in the Clean Air Act Amendments of 1977. The Amendments required state and local governments to revise the SIPs, as necessary, and to submit the revised SIPs to EPA by 1 January 1979. The revisions must provide for attainment of the National Ambient Air Quality Standards by December 1982, with incremental emissions reductions mandated over the 4-year time period. Areas that face problems meeting the ozone or carbon monoxide standards may receive an extension to December 1987.

On the transportation side, the Department of Transportation (DOT), under its joint planning regulations (23 C.F.R. 450), has established the urban transportation planning process. The regulations require that highway projects be based on a "Continuing, Comprehensive transportation planning process carried on Cooperatively," referred to as the "3C" process. It is "continuing" in that plans, programs, and projects are reviewed and revised as needed; it is "comprehensive" in scope; and it is "coordinated" among federal, state, and local agencies. Participation by public and local elected officials is to be accommodated in all phases of activity. The 3C process is administered by a Metropolitan Planning Organization (MPO), defined as being a forum for cooperative decision-making by principal elected officials of general purpose local government. The MPO provides a policy level function through review and approval of the outputs of the 3C process, as well as technical management of the 3C process through coordination and information transfer functions.

In many major urban areas of the country, the revised SIPs require the implementation of transportation controls. The 1977 Clean Air Act Amendments recognized that local involvement is essential for ensuring that the revised plans are implemented: In particular, Section 174 of the Amendments encourages that the transportation portions of the SIPs be prepared by the MPO responsible for transportation planning, by the organization responsible for air quality planning, or by the organization with both responsibilities. Through Section 175, the amendments provide funding for the transportation elements of air quality planning. Through a "Presidential Allowance for Integration of Transportation and Air Quality Planning," this funding was denied until the Environmental Protection Agency and the Department of Transportation had achieved integration of transportation and air quality planning. The resulting "Transportation–Air Quality Planning Guidelines" required that SIP revisions "provide for expeditious implementation of reasonably available control measures." The transportation control measures considered by the EPA to be reasonably available include (at a minimum):

1. Inspection and maintenance programs.
2. Vapor recovery.
3. Improved public transit.
4. Exclusive bus and carpool lanes.

5. Areawide carpool programs.
6. Private automobile restrictions.
7. Long-range transit improvements.
8. On-street parking controls.
9. Park-and-ride and fringe parking lots.
10. Pedestrian malls.
11. Employer programs to encourage formation of car pools and van pools, use of mass transit and bicycles, and walking.
12. Bicycle lanes and storage facilities.
13. Staggered work hours ("flextime").
14. Road pricing (tolls) to discourage single-occupancy automobile trips.
15. Controls on extended vehicle idling.
16. Traffic flow improvements.
17. Alternative fuels or engines and other fleet vehicle controls.
18. Other than light-duty vehicle retrofit.
19. Extreme cold-start emission reduction programs.

With the exception of inspection and maintenance, vapor recovery, retrofits, and fleet controls, these measures are similar to those of the Transportation System Measure (TSM) strategies of the Federal Highway Administration (FHWA) and Urban Mass Transportation Administration (UMTA) TSM regulations. Table 17 gives the characteristics of commonly implemented TSM strategies and tactics that traffic engineers and urban transportation planners may consider in preparing SIP revisions and air quality alternatives analyses (*26*). Of particular interest are the columns designating the air quality impacts as either local or regional.

Ozone presents a regional pollution problem, and hence measures impacting on regional emissions of ozone precursors and concentrations of ozone are of interest. Project level analyses have therefore been excluded from further consideration. On the systemwide level, a recent report (*27,28*) has shown that TSM strategies are only minimally effective in reducing hydrocarbon emissions when compared with the effects of the Federal Motor Vehicle Emission Control Program (FMVECP), inspection and maintenance, and control of stationary source emissions. The potential reduction is about 2 or 3 percent. Another recent study (*28*) has reported similar effects for the reduction of oxides of nitrogen from motor vehicles. Hence,the results of the current research are not particularly important to the investigation of TSM strategies in the transportation planning process when the intent is to reduce ozone concentrations. (This is not true with regard to carbon monoxide, where the emission reduction impact is potentially much greater.) These results are quite useful, however, in the context of analyzing the effects of the FMVECP, inspection and maintenance programs, and other control technology measures on transportation-related air-quality impacts, and of analyzing the relative impacts of alternative systemwide transportation plans.

The research of the current study basically yields three results. It provides methods for estimating the numbers of exceedances and the design values for a region. It identifies those conditions associated with high ozone concentrations. It provides procedures for estimating the levels of transported ozone.

As an input to the transportation planning process, the research thus serves two main functions. It fills-in "data gaps" spatially and temporally. And it establishes culpability in terms of locally caused and transported ozone for meteorological conditions identified as being conducive to exceedances.

The key area of application of the first function is air quality modeling and monitoring studies that are used to examine the impacts of technological controls or transportation system changes. The establishment of baseline conditions is enhanced by the ability to estimate missing data so that concentrations may be mapped accurately under a range of conditions. A similar application is in maximizing the information obtained from short-term monitoring studies, in terms of indicating optimal siting and in analyzing sampling data. A third application is in validating regional photochemical modeling results and in expanding the results of a limited number of model runs, both through supplying "missing" data.

The second function is straightforward in that the research results can be used to establish the extent of locally caused ozone concentrations and thus the portion of the total levels that are amenable to control.

Some very important questions are not addressed by the research reported here. After the procedures described here have identified the degree to which local contributions cause the observed exceedances and the conditions under which those exceedances have occurred, there still remains the problem of identifying effective control measures. The concluding chapter of this report presents recommendations for future research, which discuss this problem in more detail.

Table 17. Characteristics of commonly implemented TSM tactics.

STRATEGY	TACTIC	APPLICABLE CITY SIZE, POPULATION				AREA OF APPLICATION						AIR QUALITY IMPACTS		IMPLEMENTATION COST THOUSAND $					
		<50,000	<250,000	<1,000,000	>1,000,000	Spot	Rdwy. Fac.	Corridor	Local Area	Major Area	Region	Local	Regional	<10	<50	<100	<500	<1,000	>1,000
Traffic Operations	Intersection & Roadway Widening	x	x	x	x	x	x					x			x	x	x	x	x
	One-Way Streets	x	x	x	x				x			x			x	x			
	Turn Lane Installation	x	x	x	x	x						x		x	x				
	Turning Movement & Lane Use Restrictions	x	x	x	x	x						x		x					
	New Freeway Lane Using Shoulders		x	x	x		x					x	x			x	x		
Traffic Signalization	Local Intersection Signal Improvement	x	x	x	x	x						x		x	x				
	Arterial Signal System	x	x	x	x		x					x				x	x		
	Area Signal System	x	x	x	x				x	x		x	x				x	x	x
	Freeway Diversion & Advisory Signing			x	x			x				x					x		
	Freeway Surveillance & Control			x	x		x	x				x	x				x	x	x
Pedestrian and Bicycle	Widen Sidewalks	x	x	x	x	x	x					x			x				
	Pedestrian Grace Separation	x	x	x	x	x						x					x	x	
	Bikeways	x	x	x	x				x			x			x	x			
	Bike Storage		x	x	x	x			x	x		x		x	x				
	Pedestrian Control Barriers			x	x	x						x		x					
Roadway Assignment	Exclusive Bus Lane-Arterial											x							
	• Take-a-Lane		x	x	x		x					x		x	x				
	• Add-a-Lane		x	x	x		x					x					x	x	x
	Bus-Only Street			x	x		x					x			x				
	Contra-Flow Bus Lane		x	x	x		x					x			x	x			
	Reversible Lane Systems		x	x	x		x					x			x	x			
	Freeway HOV Bypass			x	x		x					x			x				
	Exclusive HOV Lane-Freeway																		
	• Take-a-Lane			x	x		x					x	x		x	x			
	• Add-a-Lane			x	x		x					x	x				x	x	x
Route Diversion	Area Licensing			x	x					x		x	x		x	x	x		
	Auto Restricted Zones				x					x		x	x		x	x	x		
	Pedestrian Malls		x	x	x				x			x				x	x		
	Residential Traffic Control	x	x	x	x				x			x		x					
Parking Management	Curb Parking Restrictions	x	x	x	x	x	x					x	x	x					
	Residential Parking Control		x	x	x				x			x		x	x				
	Off Street Parking Restrictions			x	x				x	x		x	x	x					
	HOV Preferential Parking			x	x			x	x				x	x					
	Parking Rate Changes		x	x	x				x	x	x	x	x	x					
Transit Operations	Bus Route & Schedule Modifications	x	x	x	x		x	x	x	x		x	x		x	x			
	Express Bus Service		x	x	x		x	x				x	x		x	x			
	Bus Traffic Signal Preemption			x	x		x					x	x		x	x	x		
	Bus Terminals	x	x	x	x				x			x			x	x	x	x	x
	Simplified Fare Collection	x	x	x	x						x		x		x	x			
Transit Management	Marketing Program	x	x	x	x						x		x		x	x	x		
	Maintenance Improvements	x	x	x	x						x		x			x	x		
	Vehicle Fleet Improvement	x	x	x	x						x		x			x	x	x	
	Operations Monitoring Program		x	x	x						x		x		x	x	x	x	x
Intermodal Coordination	Park-Ride Facilities		x	x	x			x				x	x	x	x	x	x		
	Transfer Improvements	x	x	x	x			x			x		x	x	x				
Commercial Vehicles	On-Street Loading Zones		x	x	x	x			x			x		x					
	Off Street Loading Areas		x	x	x				x	x		x			x	x	x		
	Peak Hour On-Street Loading Prohibition		x	x	x	x			x			x		x					
	Truck Route System	x	x	x	x			x				x	x	x	x				
Work Schedule	Staggered Work Hours & Flex Time			x	x					x	x	x	x	x	x				
	4-Day Week			x	x					x	x	x	x	x	x				
Pricing	Peak-Hour Tolls				x		x		x			x	x	x	x	x			
	Low-Occupancy Vehicle Tolls				x		x		x			x	x		x	x			
	Gasoline Tax	x	x	x	x						x		x		x	x			
	Peak/Off-Peak Transit Fares		x	x	x						x		x		x	x	x		
	Elderly and Handicapped Fares	x	x	x	x						x		x		x	x			
	Reduce Transit Fares	x	x	x	x						x		x	x	x				
Paratransit	Carpool Matching Programs		x	x	x				x	x	x		x	x	x	x			
	Vanpool Programs	x	x	x	x				x				x		x	x			
	Taxi/Group Riding Programs			x	x					x	x		x	x	x				
	Dial-a-Ride	x	x	x	x				x	x			x			x	x		
	Jitney Service		x	x	x		x	x					x		x				
	Elderly and Handicapped Service	x	x	x	x					x	x		x	x	x	x	x		

Source: Ref. (26)

CHAPTER FOUR

CONCLUSIONS AND SUGGESTED RESEARCH

The major focus of the research described in this report and its appendixes has been the development of procedures and definitions. The concepts of number of exceedances and design value for a whole region have been defined. Procedures have been developed for estimating those parameters, for estimating background and transported ozone concentrations, and for developing supplemental monitoring programs.

The most important results of this research effort have been procedures, rather than "conclusions" in the usual sense. Testing those procedures has yielded other results that are closer to the usual definition of "conclusions." The following section enumerates those. Although the research has been quite successful in fulfilling the objectives of this study, certain questions related to the determination of expected numbers of exceedances and design values for monitoring networks remain unanswered, and the procedures described here have some missing components with regard to their ultimate practical application. This chapter concludes with some suggestions for research that would provide the missing components and answer the remaining questions.

CONCLUSIONS

The following conclusions can be drawn from the results of applying the procedures that have been developed:

1. It is possible to extend the EPA guidelines for calculating expected exceedances and design values to a whole region and maintain consistency with the definitions developed for a single station.
2. A monitoring network of about 10 stations (using the siting techniques developed here) estimates the design value reasonably well, but the stations must be well located. They need not be operated at all times.
3. Changes in the estimated network design value caused by reducing the number of available monitors in the network are usually only somewhat greater than differences between the values obtained for a single station by different methods recommended in the EPA guidelines for the estimation of ozone design values (*1*).
4. There appears to be little loss in the ability to estimate number of exceedances and design values for a network when days with temperatures less than about 22C (72F) are ignored; monitoring schedules can be designed accordingly to reduce the amount of data that must be collected.
5. Although the methods used to estimate peak-hour ozone concentrations where no stations are located can (using second-degree interpolation techniques) detect regions of concentration that are higher than any observed at stations in the network, there is a tendency to underestimate both parameters when the number of monitoring sites is reduced. The underestimation appears to be more pronounced for number of exceedances than for design value, but this discrepancy is probably an artifact of the method.
6. The probabilistic methodology produced results substantially similar to the isopleth, especially with simple estimates. With temporally and spatially adequate data, both the E-M and the simple estimates closely followed the isopleth estimates. Both methods encountered difficulty when the number of days sampled was small (less than half the days during the period when high ozone values are likely). Although close to the isopleth values, simulated cell design values nonetheless had small standard deviations, and in cases where this could be compared to a design value determined by EPA guideline methods, the latter was generally not within two standard deviations of the former. A possible approach to resolving this variation is discussed in the next section.

RECOMMENDATIONS FOR FURTHER RESEARCH

The research described in this report has produced a reliable practical method for estimating numbers of exceedances and design values from data collected at a network of monitoring stations. However, the results deal only with ozone, and the problem of corrective measures remains unaddressed. Furthermore, the methods have been successfully applied by their designers, but they have not been tested by the intended users. These shortcomings are the basis for the research suggestions offered below. Further research seems most needed in:

1. Refinement of the methods to correct any deficiencies uncovered during practical applications by the intended users.
2. Adaptation of the methods to treat other pollutants, individually and in combination.
3. Development of procedures that can be used to design control strategies that will improve air quality such that the expected number of exceedances will not be in violation of the standard.

Refinement and Improvement of Methods

When the methods described in this report are applied by others, it is virtually inevitable that some difficulties will be encountered, or that there will be cases for which the methods are not appropriate. The usual sequence of events with a methodology like this is that various users develop ad hoc solutions to their particular difficulties and no concerted attempt is made to identify and correct the problems that are most frequently encountered. A valuable service would be performed for the user community if an effort were made to identify and interview persons and organizations who had used these methods. After a list of problems encountered (and solutions developed) is compiled, the most common problems could be identified and the necessary revisions made in the methodology. Solutions that had been developed would serve as the starting point for further refinement.

Further investigation of possible "oversmoothing" by the isopleth program is needed. Oversmoothing may cause artificially low variances within the cells and artificially high correlations between cells in the probabilistic simulation procedure, with the result that the values from simulation to simulation change little and standard deviations of cell design values are often less than 1 ppb. An appropriate procedure should be devised to model the true covariance of the observed concentrations and the relation of these to the covariance of the network after isopleth smoothing has been applied. Such a model would probably have to be developed by an iterative process; its final form cannot be defined at this time, but it is reasonable to assume that it would include a comparison of the correlations between the observed and modeled values for two cells as a function of the distances between cells.

Required modifications would probably be rather simple, because the methodology has been applied extensively during the course of this project. The research required should not be so large as to warrant a full-scale research effort, but it could be included as part of other more substantive research efforts, such as those described in the following.

Extension of the Methods to Other Pollutants

As noted earlier, ozone is particularly amenable to the isopleth technique developed here for estimating concentrations throughout a network of cells, because ozone has fewer small-scale features in its concentration pattern than do primary pollutants that are emitted directly into the atmosphere. The only other pollutant in the same category for which air quality criteria have been established is nitrogen dioxide. The current criteria for nitrogen dioxide are only for annual averages, so a methodology dealing with short-term peak concentrations is not now required. However, consideration is being given to the establishment of a short-term nitrogen dioxide standard.

Another pollutant type for which the methods described here might be useful includes sulfuric acid and other sulfate aerosols. However, even if short-term standards for these pollutants were established, they would not be of particularly great interest to the transportation community because the various modes of transportation are not major emitters of sulfate particulate precursors.

The preceding paragraph discussed only the adaptation of the current methods for use with other pollutants. Such adaptation would be straightforward, but if the methods were extended to treat several pollutants simultaneously, a major restructuring of the methodology would be required. It is uncertain how the isopleth methodology could be adapted to treat more than one pollutant at a time. Any future research would have to determine that. Perhaps the isopleth method would simply treat the individual pollutants, for example ozone and nitrogen dioxide, separately, delivering two sets of grid point values for each day to the probabilistic methodology. Obviously, when more than one pollutant is being examined by the isopleth methodology it would be necessary to develop new criteria for subdividing the grids at each iteration.

Any change in the probabilistic methodology to include correlations among different pollutants and the statistical distribution, conditioned on observed values of more than one parameter, could conceivably be extended to make use of nonpollutant parameters. For example, wind speed and temperature might possibly be incorporated into such a system. To a large extent, the methods that have been developed for preprocessing the data in the current system have anticipated possible extension into multiparameter systems.

Procedures for Developing Effective Control Strategies

The methods developed here provide the kinds of information that will be necessary for the development of effective control strategies for meeting ozone standards. However, it was beyond the scope of this effort to define the procedures for such development, which seems a logical and important next step. Regardless of how the problem is approached, development of procedures for control strategies is likely to be quite complex. It can be simplified somewhat if only one pollutant (e.g. ozone) is considered at a time, in which case the following questions (at the very least) will have to be answered: How much must precursor emissions be reduced in order to meet the standard? (Here, the design value concept would be invoked to define the critical concentration that must be reduced to the standard level.) Which precursors should be reduced (e.g., oxides of nitrogen or reactive hydrocarbons)? Would reductions of precursor emissions in some parts of the urban region be more effective than in others? What are the major sources of precursor emissions? Which source categories of precursor emissions are most amenable to control? What techniques are available for reducing precursor emission?

If one chooses to broaden the problem by developing a comprehensive control strategy that addresses all the critical pollutants (which is probably a wise course of action), the foregoing questions must be answered for each pollutant and some new questions added to the list: Is it possible that some set of controls that would lead to desirable trends in one pollutant would cause there to be countertrends in another? To what extent are dual purpose strategies possible? Can we be sure that we are not just shifting problems from pollutant to pollutant or place to place?

It should be recognized that although this report has dealt with ozone because current air quality standards are written in terms of ozone, that compound serves as a surrogate for all of those oxidants that are harmful to human health. The amounts and types of the other oxidants that are present along with a given amount of ozone are likely to vary considerably according to the origin of the ozone. Ozone from the stratosphere, ozone formed from precursors transported over a long distance, and locally generated ozone are all likely to be accompanied by a different admix of other oxidants. Given the underlying health motivation for the air quality standards, future research may be required to address these differences in the overall mixture of oxidants.

Any procedure for developing effective control strategies will have to address questions like those previously given. It must provide step-by-step methods that will require only simple tools to screen out the least effective and most costly strategies. As the simple tools reduce the possibilities to more manageable numbers, then steps involving more sophisticated screening procedures will have to be used. At this time, full-scale modeling appears to be potentially useful, but because of its high cost and difficulties of application, it should be reserved for a limited number of possibilities.

Any future research into the route from design value estimates to control strategies must emphasize simple practical techniques. The best course would completely avoid any need for highly sophisticated and expensive modeling efforts. This may never be possible, especially in some larger urban areas, but a concerted research effort could at least minimize such requirements.

REFERENCES

1. CURRAN, T. C., "Guideline for the Interpretation of Ozone Air Quality Standards." *EPA Report 450/4-79-003* (1979).
2. BREIMAN, L., "Statistical Analysis and Interpretation of Peak Air Pollution Measurements." Technology Service Corporation, Santa Monica, Calif. (1978).
3. MEYER, E. L., FREAS, W. P., SUMMERHAYS, J. E., and YOUNGBLOOD, P. L., "The Use of Trajectory Analysis for Determining Empirical Relationships Among Ambient Ozone Levels and Meteorological and Emissions Variables." *International Conference on Photochemical Oxidant Pollution and Its Control*, Raleigh, N. C. (September 12–17, 1976).
4. LUDWIG, F. L., REITER, E., SHELAR, E., and JOHNSON, W. B., "The Relation of Oxidant Levels to Precursor Emissions and Meteorological Features, Part 1: Analysis of Findings." *EPA Report 450/3-77/0229* (1977).
5. LUDWIG, F. L., JAVITZ, H. S., RUFF, R. E., and MARTINEZ, J. R., "Statistical Analyses of Ozone Data for Transportation/Air Quality." *Amplified Research Plan Project 8922*, SRI International, Menlo Park, Calif. (1979).
6. HOLZWORTH, G. C., "Mixing Heights, Wind Speeds, and Potential for Urban Air Pollution Throughout the Contiguous United States." *EPA Report AP-102* (1972).
7. TURNER, B., "Diffusion Model for an Urban Area." *J. Appl. Meteorol.*, Vol. 3 (1964) pp. 83–91.
8. HEFFTER, J. L., and TAYLOR, A. D., "A Regional-Continental Scale Transport Diffusion and Deposition Model, Part I: Trajectory Model." *National Oceanic and Atmospheric Administration Tech. Memo, ERL ARL-50* (1975) pp. 1–16.
9. CURRAN, T. C., "Screening Procedures for Ambient Air Quality Data." *EPA Report 450/2-78-037* (1978).
10. "National Primary and Secondary Ambient Air Quality Standards." *Federal Register*, Vol. 44, No. 28 (Feb. 8, 1979) pp. 8202–8237.
11. "Instrumentation for Environmental Monitoring." Lawrence Berkeley Laboratory, Univ. of Calif., Berkeley (1973).
12. BEARD, M. E., "Evaluation of the Gas-Phase Titration of Excess Ozone With Nitric Oxide Calibration Procedure for Ozone Analyzers." Draft Report, Environmental Monitoring and Support Laboratory, EPA, Research Triangle Park, N. C. (1978).
13. HOGESON, J. A., "A Survey of Calibration Techniques for Atmospheric Ozone Monitors." *National Bureau of Standards Report NBSIR 76-1191* (1976).
14. LUDWIG, F. L., and SHELAR, E., "Site Selection for the Monitoring of Photochemical Air Pollutants." *EPA Report 450/3-78-013* (1978).
15. LUDWIG, F. L., and SHELAR, E., "Assessing the Representativeness of Ozone Monitoring Data." EPA Contract 68-02-2548 (Task Order II), SRI International, Menlo Park, Calif. (1979).
16. BARR, A. J., ET AL., "SAS User's Guide, 1979 Edition." SAS Institute, Cary, N. C. (1979).
17. STROTHMANN, J. A., and SCHIERMEIER, F. A., "Documentation of the Regional Air Pollution Study (RAPS) and Related Investigations in the St. Louis Air Quality Control Region." *EPA Report 600/4-79-076*, U.S. Environmental Protection Agency (1979).
18. NITZ, K. C., and MARTINEZ, J. R., "Data Base Guide for 1978 Houston Oxidant Modeling Study." Draft Final Report, EPA Contract 68-02-2984, SRI International, Menlo Park, Calif. (1980).
19. LUDWIG, F. L., MARTINEZ, J. R., and NITZ, K. C., "Data Summaries for the Houston Area Oxidant Study." Interim Report, SRI International for Houston Area Oxidant Study, Contract DA-1 (1978).
20. "California Air Quality Data (October, November, December, 1976)." California Air Resources Board, Vol. VIII, No. 4 (1976).
21. ALLARD, D., CHAN, M., and MARLIA, C., "Philadelphia Oxidant Data Enhancement Study." *Report AV-FR-80/588*, Draft Final Report, EPA Contract 68-02-3332, AeroVironment Inc., Pasadena, Calif. (1980).
22. RUFFNER, J. A., and BAIR, F. E., *The Weather Almanac*, Avon Books, N.Y., N.Y. (1979) 728 pp.
23. "Uses, Limitations and Technical Bases of Procedures for Quantifying Relationships Between Photochemical Oxidants and Precursors." U.S. Environmental Agency, *EPA Report 450/2-77-0021a* (1977).
24. MAXWELL, C., and MARTINEZ, J. R., "Characterization of HOMS Data Quality and EKMA Evaluation." *Monthly Progress Report 3*, EPA Contract 68-02-2984, SRI International, Menlo Park, Calif. (1980).
25. LUDWIG, F. L., "Assessment of Vertical Distributions of Photochemical Pollutants and Meteorological Variables in the Vicinity of Urban Areas." *EPA Report No. 450/4-79-017* (1979).

26. DiRenzo, J. F., "Travel and Emissions Impacts of Transportation Control Measures." *Transportation Research Record 714* (1979) 17–24.
27. Federal Highway Administration, "The Effectiveness of Transportation Control Measures to Reduce One of the Principal Auto-Related Pollutants—Hydrocarbons." *Bulletin* (Aug. 8, 1980).
28. Ludwig, F. L., Patterson, R. M., and Goen, R. L., "A Study of the Relationship Between Motor Vehicle Emissions and Attainment of National Ambient Air Quality Standards, Part 1: Nitrogen Dioxide." Final Report, *Project 1975*, SRI International for the National Commission on Air Quality, Contract 20-AQ-1945 (Feb. 1981).
29. U.S. Environmental Protection Agency, "User's Manual: SAROAD (Storage and Retrieval of Aerometric Data." *EPA Report APTD-0663*, Office of Air Programs, Research Triangle Park, N. C. (1971).
30. U.S. Environmental Protection Agency, "Procedures for Quantifying Relationships Between Photochemical Oxidants and Precursors." *EPA Report 450/2/77-021b* (1978).
31. Pooler, F., "RAPS Site Descriptions." (1979) (unpublished).
32. Ludwig, F. L., Cavanagh, L. A., and Ruff, R. E., "Evaluation of Ozone Monitoring Network Operations for the GM Trend Study, Phase 1—Michigan, 1978." Prepared for General Motors, Troy, Michigan, by SRI International Project 1206, Menlo Park, Calif. (1980) 172 pp.
33. Giroux, H. D., Maas, S. J., and Gouze, S. C., "Air Monitoring Resources Identification and Evaluation." Technical Report, American Petroleum Institute Project 903-76, Meteorology Research Inc., Altadena, Calif. (1977).
34. Radian Corporation, "Air Quality Monitoring Resources Identification and Evaluation—Northeast Corridor and Texas Gulf Coast Regions." Prepared for the American Petroleum Institute, Radian Corporation, Austin, Texas (1978).
35. Martinez, J. R., and Nitz, K. C., "Analysis of High NO_2 Concentrations in California, 1975–1977. Final Report, Project 6780-12, SRI International, Menlo Park, Calif. (1979).
36. Butson, K. D., and Hutch, W. L., "Selective Guide to Climatic Data Sources." U.S. Department of Commerce, National Climatic Center, Asheville, N. C. (1979).
37. Singh, H. B., Ludwig, F. L., and Johnson, W. B., "Ozone in Clean Remote Atmospheres: Concentrations and Variabilities." Final Report, Project 5561, SRI International for Coordinating Research Council, Menlo Park, Calif. (1977) 157 pp.
38. Went, F. W., "Organic Matter in the Atmosphere and its Possible Relation to Petroleum Formations." *Proc. Nat. Acad. Sci.*, Vol. 46 (1960) p. 212.
39. Ripperton, L. A., Jeffries, H., and Worth, J. J. B., "Natural Synthesis of Ozone in the Troposphere." *Environ. Sci. and Tech.*, Vol. 5 (1971) pp. 246–248.
40. Crutzen, P. J., "Ozone Production Rates in an Oxygen-Hydrogen-Nitrogen Oxide Atmosphere." *J. Geophys. Res.*, Vol. 76 (1971) pp. 7311–7327.
41. Chameides, W. L., and Stedman, D. H., "Ozone Formation from NO_x in 'Clear Air.'" *Env. Sci. and Tech.*, Vol. 10 (1976) pp. 150–154.
42. Fishman, J., and Crutzen, D. J., "A Numerical Investigation of Tropospheric Photochemistry Using a One-Dimensional Model." *Proc. Non-Urban Tropospheric Composition Symposium*, Hollywood, Fla. (Nov. 10–12, 1976) pp. 23.1–23.15.
43. Weinstock, B., and Chang, T. Y., "Methane and Non-Urban Ozone." Paper presented at National Meeting of Air Poll. Cent. Assoc., Portland, Ore. (June 29, 1976).
44. Viezee, W., and Singh, H. B., "The Distribution of Berryllium-7 In the Troposphere: Implications on Stratospheric/Tropospheric Air Exchange." *Geophys. Res.*, Vol. 7, No. 10 (1980) pp. 805–808.
45. Reiter, E. R., "Lower-Tropospheric Ozone of Stratospheric Origin." *Arch. Met. Geoph. Biokl.*, Ser. A, Vol. 26 (1977) pp. 179–186.
46. Singh, H. B., Ludwig, F. L., and Johnson, W. B., "Tropospheric Ozone: Concentrations and Variabilities." *Atmos. Environ.*, Vol. 12 (1978) pp. 2185–2196.
47. Danielsen, E. F., and Mohnen, F. A., "Project Dustorm Report: Ozone Transport, In-Situ Measurements, and Meteorological Analyses of Tropopause Folding." *J. Geophys. Res.*, Vol. 82 (1977) pp. 5867–5877.
48. Danielsen, E. F., "Project Springfield." *DASA Report 1517*, Defense Atomic Support Agency, Washington, D.C. 20301 (1964).
49. Danielsen, E. F., "Stratospheric-Tropospheric Exchange Based on Radioactivity, Ozone, and Potential Vorticity." *J. Atmo. Sci.*, Vol. 25 (1968) pp. 502–518.
50. Johnson, W. B., Viezee, W., Cavanagh, L. A., Ludwig, F. L., Singh, H. B., and Danielsen, E. F., "Measurements of Stratospheric Ozone Penetrations into the Lower Troposphere." *Proc. Fourth Symp. Turbulence, Diffusion, and Air Pollution*, Reno, Nev. (Jan. 15–18, 1980).
51. Lamb, R. G., "A Case Study of Stratospheric Ozone Affecting Ground-Level Oxidant Concentrations." *J. Appl. Meteor.*, Vol. 16 (1977) pp. 780–794.

APPENDIX A

OVERVIEW OF THE PROCEDURES NECESSARY TO DETERMINE DESIGN VALUES AND EXCEEDANCES OF THE NATIONAL AMBIENT AIR QUALITY STANDARD FOR OZONE

The appendices to this report describe a number of procedures and computer programs that can be used to estimate numbers of exceedances of the National Ambient Air Quality Standard for ozone and to determine design values. This appendix provides an overview of the entire procedure. Subsequent appendices provide guidance and, where necessary, computer programs for each step. Figure A-1 is a schematic diagram showing how the individual procedures are interconnected.

Each box in the flowchart of Figure A-1 represents one of the steps in the overall methodology. The appendix in which the instructions are found for executing each step is also indicated within the box. The first two steps are to acquire the necessary data and to process and screen them for errors, after which some judgement must be made regarding whether or not the available data are sufficient to determine numbers of exceedances and design value for the network. If data are sufficient, then the techniques and computer programs described in the body of this report, and in Appendices C and D, are applied.

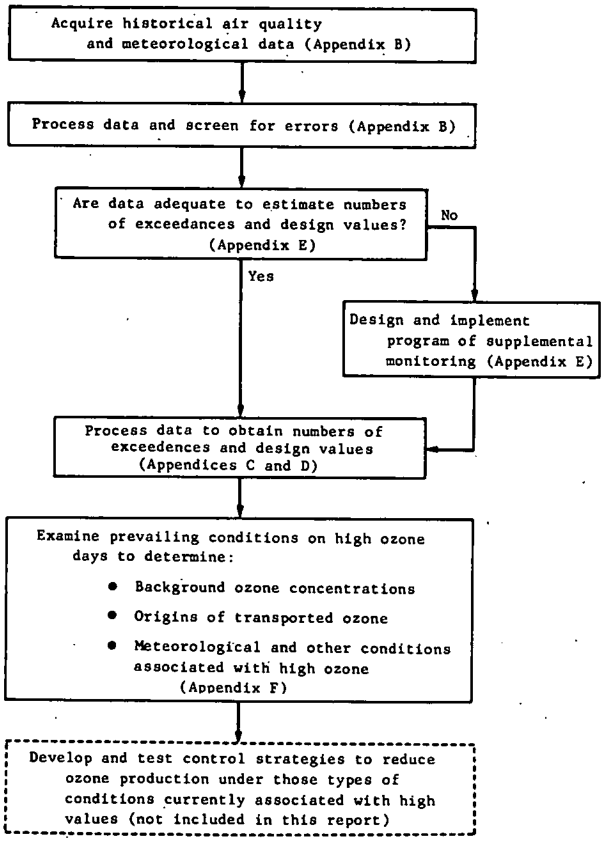


FIGURE A-1 OVERVIEW OF PROCEDURES DESCRIBED IN THE APPENDICES

If the existing data are not adequate for determining exceedances and design values, then the supplemental monitoring program must be designed and implemented. Supplemental monitoring will provide enough data that the techniques can be applied to determine exceedances and design values. The techniques described in Appendices C and D identify the days when high ozone concentrations were observed, so that the conditions on those days can be examined to determine background ozone concentration and the origins of transported ozone. This information—along with the summary of the meteorological conditions and patterns of emissions associated with the high ozone concentrations—will be used to develop and test strategies for controlling ozone in the area. The guidelines given in the appendices stop at that point. It has not been an objective of this study to develop methodologies for developing ozone control strategies and testing them, only to provide a procedure that gives the necessary information.

APPENDIX B

GUIDE TO DATA ACQUISITION AND SCREENING, AND TO THE DATA PROCESSING PROGRAM

COMPUTER PROCESSING PROGRAM

Introduction

The RDSARD ("Read SAROAD") program is a data preprocessing routine that produces an input data file for the EXCEED program (described in Appendix C). RDSARD is written so that only minor modifications are needed for tape format differences, unit number differences, and machine-specific commands. A discussion of each of these possible alterations is presented later. The program listing is given at the end of this appendix.

RDSARD is designed to read standard SAROAD (Storage and Retrieval of Aerometric Data) formatted tapes. The section below gives a brief overview of SAROAD data. The data processing is done in blocks of 12 station-hours of data (either the first 12 or the second 12 hours of a day) at a time for each pollutant. Non-SAROAD formatted data tapes can be processed, but it would be necessary for the user to alter the subroutine SAROAD.

The storage requirements of RDSARD are large, so the program listed at the end of this appendix processes two month's data for each execution. It will be necesssary for the user to execute RDSARD several times to preprocess all the available data and to combine the outputs before executing the EXCEED program, unless virtual memory or very large core storage is available, in which case more data could be handled at one time.

SAROAD Format

The U.S. Environmental Protection Agency archives air quality and meteorology data from the National Air Monitoring Stations (NAMS) operations throughout the United States. These data can be obtained by contacting the National Air Data Branch of the EPA.*

The data are available on magnetic tape in a standard format known as SAROAD (Storage and Retrieval of Aerometric Data). The normal SAROAD format has 80 character (card image) records. The number of records per block or physical record can vary, but is usually 50. The number of files varies depending on the data requested. Figure B-1 presents the SAROAD record format. Each SAROAD record identifies the monitoring station, the air pollutant or meteorological data code of the data being input, the date, the starting hour of the input data (either 00 or 12), 12 hours of data, and additional parameters that identify the units of measure and the location of the decimal point. More detailed discussions of SAROAD data and formats are given by the U.S. EPA (29).

---

*U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, National Air Data Branch (MD-14), Research Triangle Park, North Carolina 27711, (919) 541-5395.

```
1TTSSSSSSSAJJ1YYMMDDHHPPPPPQQUUEXXXXYYYYXXXXYYYYXXXXYYYY . ETC

      1 = 1
     TT = STATE CODE
SSSSSSS = SAROAD SITE CODE
      A = AGENCY CODE
     JJ = PROJECT CLASSIFICATION CODE
 YYMMDD = DATA (YEAR, MONTH, DAY)
     HH = BEGINNING HOUR FOR THIS RECORD (EITHER 00 OR 12)
  PPPPP = PARAMETER CODE
     QQ = METHOD OF MEASURE
     UU = UNIT OF MEASURE
      E = EXPONENT
   XXXX
   YYYY = DATA VALUES (TWELVE 4-COLUMN FIELDS)
```

FIGURE B-1 SAROAD LOGICAL RECORD STRUCTURE

The RDSARD program processes data having the SAROAD codes:

- 44201 (ozone)
- 42602 (nitrogen dioxide)
- 42603 (oxides of nitrogen)
- 43102 [nonmethane hydrocarbons (NMHC)]
- 61101 (wind speed)
- 61102 (wind direction)
- 62101 (temperature).

In addition, the user can specify the type of humidity and solar radiation parameters to be processed depending on the data available. The available humidity data may include dew point temperature (62103) or relative humidity data (62201). The available solar radiation data may include ultraviolet radiation data (63302) or total solar radiation data (63301).

The EPA screening programs given by Curran (9) can be applied at this point to identify data points that need careful checking.

Input Data

The RDSARD program requires two sets of input data:

- Hourly air quality and meteorological data
- Run information data.

The hourly air quality and meteorological data are input from tape, read from Logical Unit 1. The program, as presented, reads standard SAROAD formatted data in the form of 50 logical records per block and one file. If the user has SAROAD data with a different number of records per block or more files, changes must be made in subroutine SAROAD, as discussed below (Program Modifications).

The required run characteristic inputs are read from Logical Unit 5. They control several operations within RDSARD. The required run information and formats are presented in Table B-1. Although the inputs are discussed in terms of card images, other input media can be used.

Output Data

Two sets of output are produced by RDSARD. The two sets contain basically the same information, but one is disposed to Logical Unit 7 (which can be assigned to a line printer) and the other to Logical Unit 6 with a binary write statement. Typically, Logical Unit 6 is assigned to a binary disk file that will be used later as input to the EXCEED program. The user will probably need to run the RDSARD program several times unless adequate storage is available and the program is modified to take advantage of it. Multiple runs create multiple disk files that must be merged into a single file for input to the EXCEED program.

Table B-1

RUN INFORMATION VARIABLES AND THEIR FORMATS

Card	Format	Columns	Variable Name	Typical Value	Remarks
1	2A4	1-8	HDR	PHILOX	A short title (up to 8 characters) used to identify the daily data summary outputs.
2	2I5[a]	1-10	INDAY	182	First Julian date for which a daily summary will be created during execution of RDSARD.
			IOUTDY	212	Last Julian date for which a daily summary will be created during execution of RDSARD.
3	I2[a]	1-2	NSTNS	25	Number of monitoring stations for which data are to be entered.
4	2F10.0	1-20	AMET1	660007	SAROAD site code of the station from which 24-hourly meteorological data (wind, temperature, humidity, and solar radiation information) are to be obtained.
			AMET2	180070	SAROAD site code of the station from which 24-hourly wind data are to be obtained.
5	F10.0	1-10	HUMID	62103	SAROAD parameter of the available humidity data--either 6210300 for dew point temperature or 622010 if relative humidity data is available. Enter zero (0.) if neither is available.

Table B-1 (Concluded)

6	F10.0	1-10	SOLAR	63301	SAROAD parameters of the available solar radiation data--either 633010 for total solar radiation or 633020 if ultraviolet radiation data available. Enter zero (0.) if neither is available.
7	2I5[a]	1-10	IBSVTM	119	Julian date of the first day of Daylight Savings Time (should be the last Sunday of April).
			IESVTM	301	Julian date of the last day of Daylight Savings Time (should be the last Sunday of October).
8-12[b]	8F10.0	1-10	SITES	180070,...	SAROAD site identification codes
13	2I4[a]	1-8	IPCON	1	Allows the user to specify whether the daily summaries created by RDSARD are to be printed on the line printer (0 = no printout; summaries are printed for any other value).
			IPLIM	100	User can specify that only days with observed peak ozone concentrations (ppb) equalling or exceeding the value assigned for IPLIM will be printed to the lineprinter.

[a] Right justify.

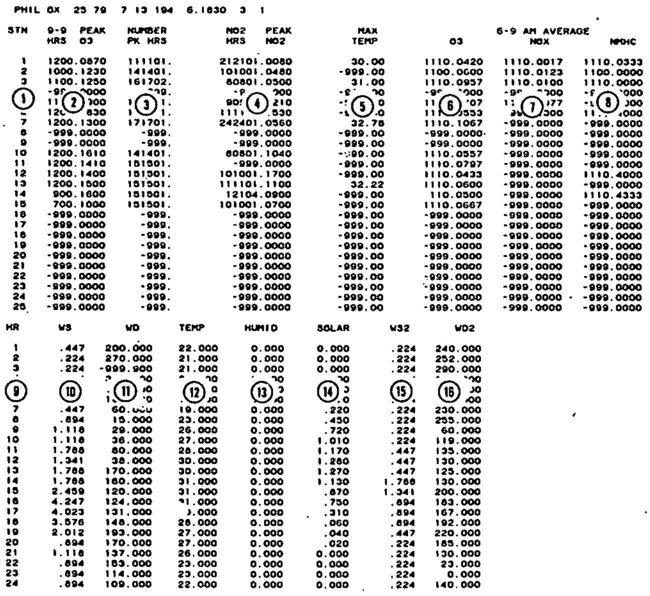
[b] If all four cards are not needed, add blank cards to make a total of four.

The preprocessing of the data and the resultant daily summary format conforms to EPA data requirements for the determination of expected number of exceedences and design values of ozone (1), and serves directly as input for the EXCEED program described in Appendix C. EPA requires that 75 percent of the hours from 0900 to 2100 have observed concentrations, or that the standard shall have been exceeded, for a day to be considered a valid monitoring day. Note that even if 75 percent of the hours are not available, but an exceedence of the standard is observed during one of the existing hours of data, then the day is accepted as a valid monitoring day. Tests for these criteria are made while preprocessing the data, and this information is provided in the daily summary data format shown in Table B-2 (with explanations of the meanings of the data entries).

The data records contain information that is not used by the progam EXCEED, but that is of considerable potential use when interpreting the results. For example, 0600-0900 average concentrations of hydrocarbons and oxides of nitrogen are useful in modeling ozone production potential for the day (23,30). Maximum temperature is also related to ozone formation potential, meteorological data are be useful for assessing background ozone concentrations and transport. The intent is to save the analyst some of the effort of locating this information later. The files written to Logical Unit 6 contain the same data shown in Table B-2 and in the same order (but without the labeling information). They are written with a binary write statement. The data are written row by row as shown in Table B-2.

Table B-2

EXAMPLE OF DATA RECORD FROM PROGRAM RDSARD[1]



```
PHIL OX   25 79   7 13 194   6.1830   3   1

STN    9-9   PEAK    NUMBER          NO2   PEAK        MAX                        6-9 AM AVERAGE
       HRS   O3      PK HRS          HRS   NO2         TEMP          O3              NOX            NMHC

  1    1200.0870     111101.         212101.0080       30.00       1110.0420       1110.0017       1110.0333
  2    1000.1230     141401.         101001.0480     -999.00       1100.0600       1110.0123       1100.0000
  3    1100.1250     161702.          80801.0500       31.00       1110.0957       1110.0100       1110.0000
       -9[illegible]0000   [illegible]99.   -[illegible]000   -[illegible]0   -9[illegible]000   -9[illegible]000   -[illegible]000
 ①     11 ② 000    1 ③ 1.    90[illegible] ④ 210    -[illegible] ⑤ 0    11 ⑥ 07    1[illegible] ⑦ 77    -[illegible] ⑧ 000
       12[illegible]830    1[illegible]1.    111[illegible].530    -[illegible].0    11[illegible]553    99[illegible]300    11[illegible]000
  7    1200.1300     171701.         242401.0560       32.78       1110.1067       -999.0000       -999.0000
  8    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
  9    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 10    1200.1610     141401.          80801.1040     -999.00       1110.0557       -999.0000       -999.0000
 11    1200.1410     151501.          -999.0000      -999.00       1110.0797       -999.0000       -999.0000
 12    1200.1400     151501.         101001.1700     -999.00       1110.0433       -999.0000       1110.4000
 13    1200.1500     151501.         111101.1100       32.22       1110.0600       -999.0000       -999.0000
 14     900.1600     151501.          12104.0900     -999.00        110.0500       -999.0000       1110.4333
 15     700.1000     151501.         101001.0700     -999.00       1110.0667       -999.0000       -999.0000
 16    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 17    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 18    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 19    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 20    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 21    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 22    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 23    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 24    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000
 25    -999.0000       -999.          -999.0000      -999.00       -999.0000       -999.0000       -999.0000

HR       WS        WD        TEMP       HUMID      SOLAR       WS2        WD2

 1      .447    200.000     22.000      0.000      0.000      .224    240.000
 2      .224    270.000     21.000      0.000      0.000      .224    252.000
 3      .224   -999.900     21.000      0.000      0.000      .224    290.000
                [illegible]00    [illegible]00    [illegible]00    [illegible]00    [illegible]00
 ⑨     ⑩      ⑪ [illegible]0   ⑫ 0    ⑬ [illegible]    ⑭ [illegible]    ⑮    ⑯ [illegible]0
               1[illegible]0    [illegible]0    [illegible]0    [illegible].0    [illegible]    [illegible]00
 7      .447     60.000     19.000      0.000       .220      .224    230.000
 8      .894     15.000     23.000      0.000       .450      .224    255.000
 9     1.118     29.000     26.000      0.000       .720      .224     60.000
10     1.118     36.000     27.000      0.000      1.010      .224    119.000
11     1.788     80.000     28.000      0.000      1.170      .447    135.000
12     1.341     38.000     30.000      0.000      1.280      .447    130.000
13     1.788    170.000     30.000      0.000      1.270      .447    125.000
14     1.788    160.000     31.000      0.000      1.130     1.788    130.000
15     2.459    120.000     31.000      0.000       .870     1.341    200.000
16     4.247    124.000     31.000      0.000       .750      .894    183.000
17     4.023    131.000     [illegible].000      0.000       .310      .894    167.000
18     3.576    148.000     28.000      0.000       .060      .894    192.000
19     2.012    193.000     27.000      0.000       .040      .447    220.000
20      .894    170.000     27.000      0.000       .020      .224    185.000
21     1.118    137.000     26.000      0.000      0.000      .224    130.000
22      .894    183.000     23.000      0.000      0.000      .224     23.000
23      .894    114.000     23.000      0.000      0.000      .224      0.000
24      .894    109.000     22.000      0.000      0.000      .224    140.000
```

[1] For explanation of numbered symbols, see attached notes.

NOTES —EXPLANATION OF THE DATA IN TABLE B-2

General identifying information is presented in the header and Col. 1:

Header	PHIL OX (title); 25 (25 stations); 79 7 13 (date); 194 (Julian day); 6.1830 [of the 25 stations and on this day, station 6 6.1830) had the highest observed ozone concentration of 0.183 ppm (6.1830)]; 3 1 [stations 3 and 1 are the stations from which meteorological data are extracted (see the notes for Columns 9 through 16 below). These stations should be chosen to be meteorologically representative of the area as a whole as is possible.
Col. 1	Station number column; the following examples apply to the first row, i.e. Station 1.

Air quality data are presented in Columns 2 through 8:

Col. 2	1200.0870—During the 12 hours from 9 am to 9 pm, 12 hours (1200.0870) had ozone values reported. The maximum ozone value recorded was 0.087 ppm (1200.0870). If the value is negative, less than 9 hours of data between 9 am and 9 pm were recorded, but an exceedance of the standard was observed.

Col. 3	111101.—The maximum recorded ozone value listed in Column 2 was first recorded during the hour 1000-1100 (111101.). The last hour when it was observed was also during that hour (111101.0). The maximum value was recorded once (111101) during the day.
Col. 4	212101.0080—The maximum NO_2 concentration recorded was 0.008 ppm (212101.0080); it was observed during one hour (212101.0080). The first hour of the observed maximum was 2000-2100 (212101.0080), as was the last (212101.0080).
Col. 5	30.00—The maximum temperature recorded at the site on this day was 30.00°C. (Extra digits beyond the decimal may be shown when there has been conversion from °F to °C.)
Col. 6	1110.0420—The 6 am to 9 am (LST) average ozone concentration at Station 1 was 0.0420 ppm (1110.0420). The "1's" in the thousands, hundreds, and tens columns (1110.0167) indicate that values were recorded respectively for 6-7 am, 7-8 am, and 8-9 am hours. A zero in the hundreds column (e.g., 1010.0420), would indicate that the 6-9 am ozone value was calculated using data only for the 6-7 am and 8-9 am hours and that the 7-8 am data point was missing. Similarly, zeroes in the other columns indicate data was missing for the corresponding hours.

Col. 7	1110.0017—The same as Col. 6, but refers to the 6-9 am average NO_x concentration.
Col. 8	1110.0333—The same as Col. 6, but for 6-9 am average NMHC. The NMHC concentrations are in parts-per-million carbon (ppmc) instead of parts-per-million (ppm).

Meteorological data then follow:

Col. 9	Hour of the day.
Col. 10-14	Wind speed (m/s), wind direction, temperature (°C), humidity [either relative humidity (percent) or dewpoint temperature (°C) as defined when file created], and solar radiation (either total or ultraviolet) for the first station listed in the header—in this case, Station 3.
Col. 15-16	Hourly wind speed (m/s) and wind direction for the second station listed in the header—in this case, Station 1.

Program Modifications

The RDSARD code might require modification if:

- The SAROAD data tape contains more than one file.
- The SAROAD data tape has a block size different than 400 words.
- Logical unit numbers different from those currently specified in RDSARD are desired.
- The data are not in SAROAD format.
- There are computer specific statements that will not work on the user's machine.

The necessary modifications are discussed below.

SAROAD Data Tape Contains More Than One File. The only changes required are in the SAROAD subroutine. The fifth and sixth lines from the end of the subroutine are two comments that will allow the program to handle multiple file tapes if they are included in the FORTRAN code. The statement

```
IF(IEOF.GT.2) GO TO 96
```

should be changed by replacing the "2" with a value equal to one fewer than the number of files the user needs to read from the data tape. For example, if there are four data files on the tape, the statement should read

```
IF(IEOF.GT.3) GO TO 96.
```

SAROAD Data Tape With Block Size Different From 400 Words. Three changes will be necessary in the SAROAD subroutine. First, the dimension of the variable BLOCK at line 7 must be set to the block size of the users data tape. Second, the second assignment of BLOCK in the BUFFER IN statement--currently BLOCK (400)--must be set to the block size (or larger) of the user's tape. Finally, the upper limit of the loop

```
DO 90 I=1, 400, 8
```

(i.e., 400) should be changed to the block size of the user's tape.

Different Logical Unit Numbers. The logical unit numbers currently used in the RDSARD code are:

- Logical Unit 1 (ITP) is used to read the input data tape.
- Logical Unit 5 (IRD) is used to read the 13 card-image run specification inputs.
- Logical Unit 6 (IWRB) is where the binary data summary is written.
- Logical Unit 7 (IWRI) is used for output of the form shown in Table B-2; this will usually be assigned to a line printer.

All the unit numbers are defined at the beginning of the READIN subroutine. Any needed changes can be made by changing the value of the appropriate variable, as listed above.

Data in Non-SAROAD Format. RDSARD was developed to make conversions as convenient for the user as possible. The input data are read by the SAROAD subroutine. All preliminary handling of the data from tape--such as checking for missing values (blanks, 9998, and 9999 are SAROAD codes for trace and missing data), determining the location of decimal points, determining the index number of the station to be read and determining the index number for the parameter to be read--is performed in SAROAD. However, data statements containing the values needed for converting units of measure and for determining the index number of the air quality and meteorological parameters, are stored in the main RDSARD routine. If the air pollutant concentrations can be converted to ppm and if the pollutant and the meteorological data are available in groups of twelve hourly average values (beginning at 0000 or 1200), then no changes will need to be made to any other subroutines. The value of IH should be set to either 00 or 12, so that the subroutines will recognize the data as either the first or second twelve hour data group.

Computer Specific Statements. RDSARD was developed on a CDC-6400 computer; at least four statement types in RDSARD may not be correctly interpreted by other computers. These statements are:

- PROGRAM RDSARD (INPUT, OUTPUT, TAPE1, TAPE5, TAPE6, TAPE7=OUTPUT).
- BUFFER IN (ITP,0) (BLOCK(1), BLOCK(400).

- DECODE (80, 100, BLOCK(I)) STN AG TPC, IYR, IMO, IDY, IH, PARA, 1METH, IN, EE, (DATA(KKY), KKY=1,12).
- IF (UNIT (ITP) 10, 95, 5.

The PROGRAM statement, the first line in RDSARD, may need to be altered depending on the machine being used. On the CDC, the parameters listed on the program card, initialize the logical unit numbers (specified as TAPE #) to be used in the program.

The BUFFER IN statement, in subroutine SAROAD, commands the computer to read one block or physical record of data from unit ITP with parity = 0 and store the data in the array BLOCK.

The

```
DECODE (80, 100, BLOCK(I)) STN; AG, IPC, IYR, IMO, IDY, IH,
        PARA, METH, IU, EE (DATA(KKY), KKY=1,12)
```

statement is found in the SAROAD subroutine. This statement extracts logical records (in the case of SAROAD data, 80-character card images) from the array--BLOCK--which serves as storage for the logical records read from the data tape using the BUFFER IN statement described above. Many machines use the DECODE statement; however, on other machines the normal procedure is to dump the data from tape to disk and then use READ statements to read the data from disk.

The

```
IF(UNIT (ITP)) 10, 95, 5
```

statement found in subroutine SAROAD is a status check for the BUFFER IN command. The variable UNIT can have the following values:

- -1 = Logical Unit ITP ready
- 0 = End-of-file on Logical Unit ITP
- +1 = Parity error on read from Logical Unit ITP.

Some changes in the CDC-machine-specific statements may not be relatively straightforward; however, with the exception of the statements discussed above, RDSARD is written in ANSI standard FORTRAN and should cause no additional problems.

Description of the Subroutines and Functions

The RDSARD program has eight subroutines and one function. RDSARD (the main routine) has no executable statements. RDSARD is used to specify data and to route the program. Figure B-2 shows which routines call other routines. As the figure shows, there are three main subroutines in RDSARD. The first, READIN, reads the run information data and prints the information to verify correct reading. The second, SAROAD, reads the SAROAD data from tape, determines indices, and calls the other subroutines listed below it in Figure B-2 to process the data. The third, GROUP, combines the data and writes the output to the two logical units as discussed above.

A brief description of the routines in RDSARD follows.

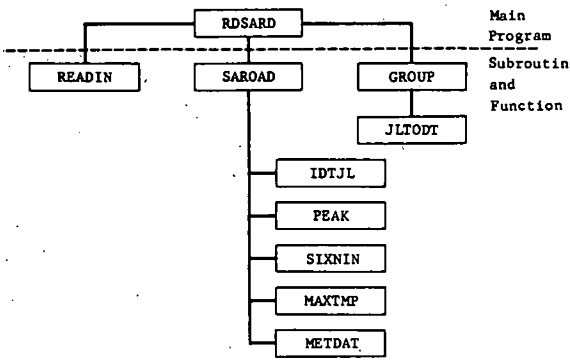


FIGURE B-2 SCHEMATIC DIAGRAM OF PROGRAM RDSARD ORGANIZATION

RDSARD--This is the main routine; RDSARD identifies certain data values and directs operations of the program. RDSARD calls subroutines READIN, SAROAD, and GROUP.

READIN--The run information data are read and printed by this routine.

SAROAD--The air quality and meteorological data are read from Logical Unit 1 by this routine. The station parameters, and missing data are identified in SAROAD. SAROAD then calls processing routines that depend on the data parameter being processed. SAROAD calls function IDTOJL and subroutines PEAK, SIXNIN, MAXTMP, and METDAT.

PEAK--This subroutine determines the daily peak ozone and nitrogen dioxide concentration for each monitoring station measuring these pollutants. In addition, the first hour, last hour, and number of hours during which the peak was observed are determined.

SIXNIN--The 0600-0900 (local time) average concentrations of ozone, oxides of nitrogen, and nonmethane hydrocarbons for each site are calculated in SIXNIN. If data are reported for only one hour, then the 0600-0900 average concentration is considered missing.

MAXTMP--Finds the daily maximum temperature observed at each site.

METDAT--This routine accumulates and stores the 24-hourly average records of wind data for two selected sites. Temperature, humidity information, and solar radiation data are stored for one selected site.

GROUP--The GROUP routine combines data to form the daily summaries. The daily data summaries are output to the logical units described earlier.

JLTODT--Determines the month and day, given the Julian date and the year.

IDTJL--This function determines the Julian date given the year, month, and day.

INSPECTION AND SCREENING OF DATA

Introduction

The major concern regarding quality of the data is with the ozone measurements, because those data will be used directly in the estimation of exceedances and design values. Barring gross errors, the other data are of less concern, because they are not directly incorporated into the estimation procedure. The most important aspects of quality are the data's accuracy and precision and its representativeness. By accuracy, we mean the systematic errors or bias that may be found in the measurements; precision is used to denote random errors. Procedures for assessing data quality focus on:

- Data anomalies
- Measurement accuracy

- Measurement methods.

Methods that can be used to identify the most serious anomalies are discussed later.

Measurement Methods and Accuracy

The accuracy of the various measurement methods was discussed in Chapter 2. As noted there, the recommended calibration procedure for ozone monitoring, as described in the Federal Register (10), results in an uncertainty of about 7 to 15 ppb with UV photometer instruments. The other major type of ozone monitor, the chemiluminescent instrument, when used with the recommended calibration procedures, should measure ozone concentrations within about 10 to 20 ppb. The instrumentation and calibration methods should be considered carefully when evaluating the data base.

Representativeness of the Monitoring Sites

The remaining factor that must be considered in evaluating data quality is representativeness. Ludwig and Shelar (14,15) studied representativeness as it affects the selection of ozone monitoring sites and the determination of ozone exceedances. Their work describes the criteria that should be met if an ozone monitoring site is to be representative of its surroundings. These findings are summarized briefly below, but the reports should be consulted when evaluating monitoring sites.

Very briefly, Ludwig and Shelar define three types of ozone monitoring site (14):

- The general urban monitoring site, representative of conditions in urban neighborhoods (with dimensions of a few kilometers).
- The urban monitoring site, located where concentrations are expected to be highest, again representative of conditions on a scale of a few kilometers.
- The regional background site, representing larger scale conditions, relatively unaffected by precursor emissions within the city.

The identification of appropriate background ozone monitoring locations is discussed in greater detail in Appendix E, dealing with the estimation of transported and background ozone concentrations. The other two types of sites are required to provide a good comprehensive monitoring network. Selection of all types of sites is discussed in Appendix F, which deals with the related topic of supplemental monitoring. Briefly, the general urban ozone monitor should be located in an area of reasonably homogeneous land use and sufficiently distant from individual sources of nitric oxide (NO serves as a short-term sink for ozone) so that the prevailing ozone concentrations are not appreciably affected.

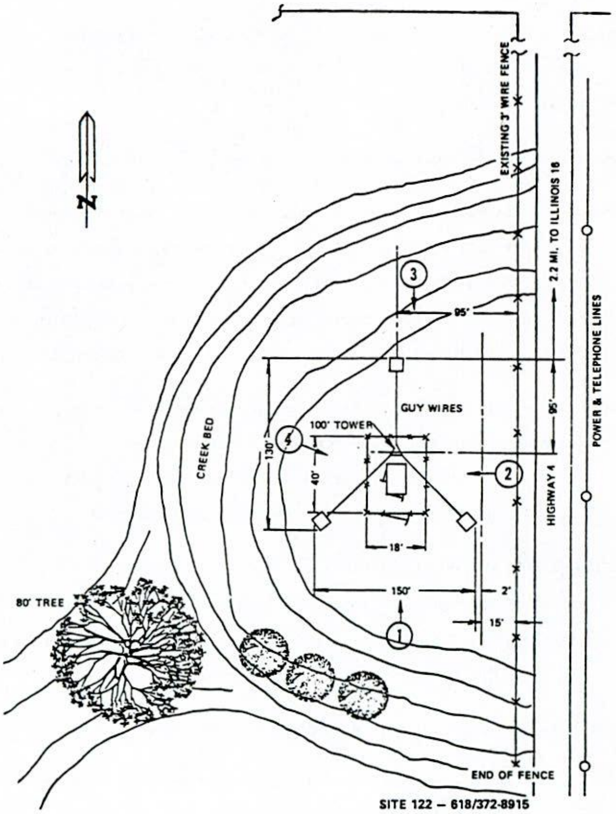
Two types of NO source are particularly relevant: traffic sources and large, elevated point sources. Although the effects of the point sources can be quite pronounced, those effects are usually of limited duration and spatial extent. When particular sets of meteorological conditions associated with the design value are examined for purposes of developing control strategies, point source effects should be considered. Otherwise, traffic NO sources are more likely to be a problem. Table B-3 shows Ludwig's and Shelar's recommendations for minimum separations between ozone monitors and roadways with different traffic volumes (14). The locations of stations in a network should be compared with the definitions given above and the specifications in Table B-3. Substantial deviations from the recommendations should suggest that the data be considered very carefully before use.

Ideally, good site descriptions are available for each of the sites from which data are used. Figure B-3 shows the type of description available for each of the St. Louis RAPS sites (31). Figure B-4 from Ludwig, Cavanagh, and Ruff (32) shows a less elegant, but adequate, map of the area around an ozone monitoring site in Michigan; similar descriptions (with photos and other corollary information) were prepared for most of the monitors in Michigan. Descriptions have also been compiled for Los Angeles and San Francisco Bay area stations by Meteorology Research Incorporated (33), and for the northeastern United States and the Houston area by Radian Corporation (34). Unfortunately, the locations of many ozone monitors are not as well documented as those shown in the figures and repots cited here, which provide examples of the

Table B-3

RECOMMENDED SEPARATION BETWEEN OZONE MONITORING SITES AND ROADWAYS [after Ref. (14)]

Average Daily Roadway Traffic (vehicles/day)	Site/Roadway Separation (m)
<1,000	20
1,000-10,000	20-250
>10,000	>250

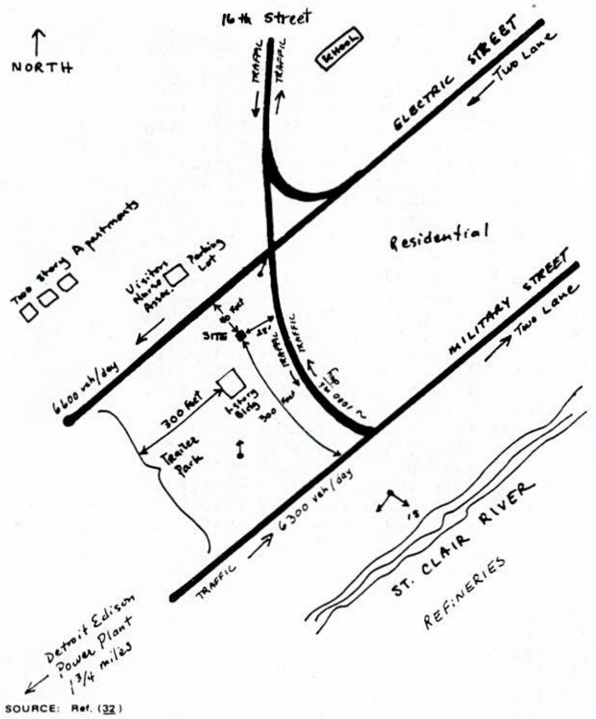


SITE 122 — 618/372-8915
Norton Farm
W. side of Hwy. 4, 2.2 mi. S. of Ill. 16
Jerseyville, Ill.

SOURCE: Ref. (31)

FIGURE B-3 EXAMPLES OF RAPS SITE DESCRIPTION (SITE 122)

PORT HURON



SOURCE: Ref. (32)

FIGURE B-4 EXAMPLES OF MICHIGAN SITE DESCRIPTION (PORT HURON)

kinds of information that should be assembled in order to evaluate a site.

Recognizing Anomalous Data Points

In general, the approach outlined below was adapted from the work of Martinez and Nitz (35), who developed subjective techniques for screening data for anomalies. As noted earlier, the EPA methods (9) should also be considered. Inasmuch as a major goal in this procedure is to determine exceedances of the ozone standard, the emphasis in the screening procedures discussed below is on the higher ozone concentrations, especially those in excess of 120 ppb.

The procedure consists of three steps:

- Hourly ozone concentrations are graphically plotted or tabulated for days when the ozone concentration exceeded 120 ppb.
- The tabulations or plots are scanned visually to identify questionable data (according to criteria discussed below).
- Any available supplementary information is consulted to determine whether the data should be accepted or rejected.

Scanning of the data is usually effective for identifying anomalous data points. Frequently such scanning is enough to identify misplaced decimal points or other malfunctions of the data collection or archiving system. The following attributes cause data to be suspect:

- Sudden increases, drops, or spikes in the ozone data record.
- Very high concentrations occurring immediately before or following a period without data, such as might be indicative of an instrument malfunction.
- Very high ozone concentrations occurring only at a single location in a network, especially if they persist through morning and evening hours.
- Very high ozone concentrations, especially when wind, temperature, or NO data suggest that high ozone concentrations are unlikely.

In the last-cited criterion, low temperatures or high NO concentrations would argue against high ozone concentrations; wind directions that were not from an area of known precursors would do likewise. Obviously there is some subjectivity involved in the methodology, but the above criteria provide guidance for data screening. These screening procedures are directed at identifying anomalous data points that indicating concentrations greater than those actually present in the ambient air. They will not be able to identify deviations from true values that arise from the imprecision of the measurement process that was discussed earlier. They also ignore readings that are in error on the low side. Such errors will not affect design value estimates, and the isopleth methodology (Appendix C) will substitute a better estimate automatically.

The data review process can begin most easily after the execution of the RDSARD program. As output from RDSARD, the user will have daily data summaries of the input data base. The daily summary format lists on one page the daily peak ozone concentrations at all monitoring sites, the time they occur, and hourly meteorological data at two sites. Therefore, by using the daily summaries, most of the errors described above can be identified quickly. The following discussion presents an example, using the daily summary example of Table B-2. to identify an error.

Table B-4 depicts the daily summary for 13 February 1976 (Julian date 44) for St. Louis. (See notes for Table B-2.) The highest observed ozone concentration reported for this date was 233 ppb at Site 20 at 1700. Examining the column listing the daily peak ozone concentrations observed elsewhere in the monitoring network, it is found that the second-highest ozone concentration reported was only 45 ppb. Neither the season nor the maximum temperatures observed on this day were conducive to the formation of high ozone concentrations. That fact plus the ozone observed elsewhere in the network and the incomplete record at this site on this day all make the 233 ppb ozone concentration suspect and suggest that it should be discarded.

The daily summaries can also indicate the occurrence of data spikes. If an ozone concentration is suspect, the user should first check the number of hours between 0900 and 2100 that had ozone concentrations reported. (This information is in the daily summary.) If hours

Table B-4 CONDENSED DATA FORMAT OF THE 13 FEBRUARY 1976 RAPS DATA

```
STLU RPS 25 76  2 13  44 20.2330 24 25

   HRS  MAX O    NO HRS      MAX NO2       MAX T     AVG NO    AVG NO2   AVG NMHC     6-9 OZ    6-9 NOX   6-9 NMHC
 1  1200.0250 151501.0   90901.0480    13.1886      .0167      .0280      .2953 1110.0030 1010.2225 1110.2322
 2  1200.0220 212101.0   60601.0440    12.0728      .0028      .0200      .0685 1110.0030 1110.2568 1110.1284
 3  1200.0400 151501.0   90901.0790    12.3845      .0032      .0230      .4626 1110.0030 1110.2109 1110.1944
 4  1200.0250 151501.0   90901.0650    12.5844      .0159      .0260 -999.9000 1110.0030 1110.3509 -999.9000
 5   500.0110 222201.0   90901.0690    12.6025      .0078      .0310      .0858 1110.0030 1110.4219 1110.3083
 6  1200.0200  10101.0  232301.0480    13.0903      .0027 -999.9000      .2349 1110.0030 -999.9000 1110.2122
 7  1200.0160 151501.0   80801.0510    12.7311      .0282      .0290      .3640 1110.0030 1110.2865 1110.2251
 8  1200.0450 141401.0   50501.0320    11.9351      .0042      .0190      .4796 1110.0030 1110.1241 1110.0417
 9  1200.0330 161601.0  101001.0630    12.8479      .0026      .0080      .0814 1110.0030 1110.1149 1110.0288
10   500.0030  22413.0  101001.0460    12.6516      .0201      .0305      .1905 1110.0030 1110.0802 -999.9000
11  1200.0210 131502.0   90901.0520    13.2004      .0301      .0380      .3733 1110.0030 1110.3587  110.1901
12  1100.0200 161601.0   90901.0810    12.7752      .0103      .0310      .1674 1110.0030 1110.2112 1110.1262
13  1200.0170 161601.0   80801.0360    12.9385      .0147      .0190      .6123 1110.0033 1110.1305 1110.0875
14  1200.0310 171701.0    -999.9000    11.8904      .0025 -999.9000 -999.9000 1110.0077 -999.9000 1110.0095
15  1200.0300 151501.0   30301.0320    11.8819      .0025      .0050 -999.9000 1110.0040 1110.0380 1110.1436
16  1200.0330 151501.0    -999.9000    13.0308 -999.9000 -999.9000      .0183 1110.0050 -999.9000 1110.0252
17  1200.0320 161601.0  232301.0280    12.9108      .0035      .0090     -.0561 1110.0037 1110.0270 1110.0024
18  1200.0320 121402.0  141401.0470    12.4478      .0091      .0415      .6501 1110.0083 1110.0268 1110.0138
19   100.0030  71004.0   10505.0030    13.7344      .1057 -999.9000 -999.9000 1110.0030 -999.9000 -999.9000
20  -800.2330 171701.0  161601.0450    11.9550      .0071      .0150 -999.9000 1110.0030 1110.1460 -999.9000
21  1000.0310 181801.0  141401.0240    12.4790      .0025      .0100 -999.9000 1110.0170 1110.0119 -999.9000
22  1200.0390 161601.0  232402.0070    10.5453      .0025 -999.9000 -999.9000 1110.0077   10.0025 -999.9000
23  1200.0310 161601.0  232301.0320    12.1472      .0027      .0110 -999.9000 1110.0050 1110.0472 -999.9000
24  1200.0390 131301.0  121201.0240    14.0956      .0129      .0220      .4153 1110.0127 1110.0194 1110.0071
25  1200.0360 171701.0   21707.0030    13.9861      .0025      .0030 -999.9000 1110.0193 1110.0106 -999.9000

          WS          WD        TEMP       DEWPT          UV        WS2        WD2
 1     5.839     215.806      11.867       8.461       0.000      7.019    233.845
 2     5.079     218.250      11.512       8.278       0.000      7.173    233.653
 3     2.918     210.869      10.848       7.967       0.000      5.705    232.602
 4     5.039     207.245       9.996       7.309       0.000      6.178    231.780
 5     5.382     211.905      10.316       8.054       0.000      5.308    233.720
 6     3.901     211.045       9.896       7.394       0.000      4.291    256.907
 7     4.050     216.804       9.841       6.766       0.000      2.367     19.358
 8     2.944     208.480       9.591       7.558       0.000      5.242      6.831
 9     2.116       4.113      10.630       8.759       0.000      3.196     19.808
10     1.208       5.429      12.813       7.435       0.000      5.603     12.817
11     2.973      15.688      14.096       8.541       0.000      4.380     27.556
12     3.576      10.282      13.638       7.682       0.000      5.268     22.291
13     4.790       6.916      12.997       7.061       0.000      5.543     30.465
14     4.269       4.118      13.013       6.415       0.000      5.660     20.789
15     4.794     357.651      12.846       5.119       0.000      5.886     16.907
16     6.491       4.773      11.974       3.708       0.000      5.848     18.045
17     6.137       9.922      11.197       3.303       0.000      5.230     21.842
18     5.325      11.833       9.804       3.211       0.000      4.598     25.705
19     5.541      25.116       8.026       2.738       0.000      4.663     38.374
20     4.906      32.123       7.008       2.103       0.000      5.744     40.855
21     5.580      41.371       6.946       1.326       0.000      5.911     45.279
22     5.250      40.483       5.451        .066       0.000      6.180     47.801
23     5.056      43.838       4.151       -.985       0.000      5.569     65.436
24     3.715      43.700       2.221      -2.191       0.000      4.441     71.796
```

are shown as less than 12, a period of equipment failure or downtime is often at fault. Data spikes often occur immediately preceding or following a period of equipment failure. Therefore, the original data record should be examined for these cases.

Appendix E, which discusses the determination of transported ozone, presents a discussion of the meteorological signatures that identify a well-mixed layer, e.g., increased wind speed, increased temperature, and so forth. Those same signatures can be used to determine when vertical mixing is taking place at night. If ozone is stored in a layer aloft, it can be mixed to ground and cause an apparently anomalous nighttime increase in ozone concentration. The corollary meteorological records must be examined before discarding high ozone concentrations occurring at unusual hours of the day.

One final comment is in order regarding the screening of the data. After the program EXCEED (Appendix C) has been run, those days when the highest ozone was observed will be identified. The ozone data for these days should be reexamined with special care, because these observations determine the design value estimates.

LISTING OF PROGRAM RDSARD

PROGRAM RDSARD

```
       PROGRAM RDSARD(INPUT,OUTPUT,TAPE1,TAPE5,TAPE6,TAPE7=OUTPUT)
       COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
      1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
      2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
      3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
      4,IPCON,IPLIM
C
       DATA PAR/44201.,42601.,42603.,43102.,42602./
       DATA AMETDT/61101.,61102.,62101./
       DATA UNITS/-99.,-99.,-99.,-99.,-99.,-99.,1.0,.001,-99.,-99.,
      A1.0,.44704,.51479,1.0,1.0,-99.,1.0,1.0,-99.,-99.,-99.,-99.,
      B1.0,-99.,1.0,-99.,-99.,-99.,-99.,-99.,-99.,-99.,-99.,-99.,
      C-99.,1.0,-99.,-99.,-99.,100./
       CALL READIN
       CALL SAROAD
       CALL GROUP
       STOP
       END
```

SUBROUTINE READIN

```
       SUBROUTINE READIN
C
       COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
      1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
      2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
      3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
      4,IPCON,IPLIM
C UNIT NUMBERS ARE :  ITP=TAPE UNIT
C                     IRD=INPUT INFORMATION FOR INDIVIDUAL RUNS (MAY BE
C                          CARD READER).
C                     IWRI=LINE PRINTER
C                     IWRB=UNIT NUMBER TO WRITE BINARY FILE TOO.
C
       ITP=1
       IRD=5
       IWRI=7
       IWRB=6
C
C
C
C READ INPUT DATA
C
C
      READ(IRD,499)(HDR(IJ),IJ=1,2)
  499 FORMAT(2A4)
      READ(IRD,500)INDAY,IOUTDY
      NDAYS=IOUTDY-INDAY+1
      READ(IRD,501)NSTNS
      READ(IRD,503)AMET1,AMET2
      READ(IRD,504)HUMID
      READ(IRD,505)SOLAR
      READ(IRD,506)IBSVTM,IESVTM
      READ(IRD,502)(SITES(I),I=1,32)
      READ(IRD,507)IPCON,IPLIM
```

```
  500 FORMAT(2I5)
  501 FORMAT(I2)
  502 FORMAT(8F10.0)
  503 FORMAT(2F10.0)
  504 FORMAT(F10.0)
  505 FORMAT(F10.0)
  506 FORMAT(2I5)
  507 FORMAT(2I4)
C
C DETERMINE THE INDEX NUMBER OF THE SELECTED MET STATIONS.
C
      DO 10 J=1,NSTNS
      IF(AMET1.EQ.SITES(J))IAMT1=J
      IF(AMET2.EQ.SITES(J))IAMT2=J
   10 CONTINUE
C
C PRINT INPUT DATA.
C
      WRITE(IWRI,800)(HDR(NN),NN=1,2)
  800 FORMAT(1H1,15HDATA HEADING : ,2A4)
      WRITE(IWRI,805)INDAY
  805 FORMAT(1H 15HSTARTING DAY : ,I4)
      WRITE(IWRI,810)IOUTDY
  810 FORMAT(1H ,13HENDING DAY : ,I4)
      WRITE(IWRI,815)NDAYS
  815 FORMAT(1H ,17HNUMBER OF DAYS : ,I3)
      WRITE(IWRI,820)NSTNS
  820 FORMAT(1H ,21HNUMBER OF STATIONS : ,I3)
      WRITE(IWRI,825)AMET1,AMET2
  825 FORMAT(1H ,32HSELECTED METEOROLOGICAL SITES : ,2F10.0)
      WRITE(IWRI,830)HUMID
  830 FORMAT(1H ,35HSELECTED HUMIDITY OR DEW PT CODE : ,F10.0)
      WRITE(IWRI,835)SOLAR
  835 FORMAT(1H ,34HSELECTED UV OR TOTAL SOLAR CODE : ,F10.0)
      WRITE(IWRI,840)IBSVTM,IESVTM
  840 FORMAT(1H ,37HSTART AND ENDING SAVINGS TIME DAYS : ,2I5)
      WRITE(IWRI,845)
  845 FORMAT(//1H0,14HSTATION NUMBER,7X,11HSAROAD CODE/)
      DO 20 JA=1,NSTNS
      WRITE(IWRI,850)JA,SITES(JA)
  850 FORMAT(1H ,7X,I2,15X,F8.0)
   20 CONTINUE
      RETURN
      END
```

SUBROUTINE SAROAD

```
       SUBROUTINE SAROAD
C
       COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
      1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
      2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
      3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
      4,IPCON,IPLIM
       DIMENSION BLOCK(400)
C
C
C THIS PROGRAM READS STANDARD SAROAD TAPES,DETERMINES THE STATION,
C TYPE OF DATA,AND CALLS THE APPROPRIATE SUBROUTINES BASED ON THE
C TYPE OF DATA. IN ADDITION, MISSING DATA ARE IDENTIFIED AND SET
C EQUAL TO -999.90.
```

```
C
C
      DO 3 IJA=1,62
      DO 2 IJB=1,NSTNS
      DO 1 IJC=1,13
      STORE(IJA,IJB,IJC)=-0.
    1 CONTINUE
    2 CONTINUE
    3 CONTINUE
    5 BUFFER IN(ITP,0)(BLOCK(1),BLOCK(400))
C
C THE "IF(UNIT(ITP))" IS A CDC STATUS CHECK OF THE BUFFER IN OPERATION.
C          -1 = UNIT READY
C           0 = END-OF-FILE
C          +1 = PARITY ERROR
C
      IF(UNIT(ITP))10,95,5
   10 CONTINUE
      DO 90 I=1,400,8
      DO 13 IXX=1,12
      DATA(IXX)=9999.
   13 CONTINUE
      DECODE(80,100,BLOCK(I))STN,AG,IPC,IYR,IMO,IDY,IH,PARA,
     1METH,IU,EE,(DATA(KKY),KKY=1,12)
  100 FORMAT(3X,F7.0,A1,I2,1X,4I2,F5.0,2I2,F1.0,12F4.0)
      IF(IPC.EQ.9)GO TO 90
C
C CALCULATE JULIAN DAY.
C
      IDAY=IDTJL(IYR,IMO,IDY)
      IF(IDAY.LT.INDAY)GO TO 90
      IF(IDAY.GT.IOUTDY)GO TO 90
      IDDAY=IDAY
      IDAY=IDAY-INDAY+1
C
C  CHECK FOR STATION
C
      DO 25 ISTN=1,NSTNS
      IF(STN.EQ.SITES(ISTN))GO TO 26
   25 CONTINUE
      GO TO 90
C
C
C   CHECK PARAMETER CODES
C
C
   26 DO 28 K=1,5
      IF(PARA.EQ.PAR(K))GO TO 30
   28 CONTINUE
      DO 29 KA=1,3
      IF(PARA.EQ.AMETDT(KA))GO TO 30
   29 CONTINUE
      IF(PARA.EQ.HUMID.OR.PARA.EQ.SOLAR)GO TO 30
      GO TO 90
   30 CONTINUE
C
C   CHECK FOR MISSING DATA
C
      EXP=UNITS(IU)/(10.**EE)
C
C CHECK FOR TEMPERATURE IN DEGREES FARANHEIT.
C
      IF(IU.EQ.15)GO TO 36
      DO 35 L=1,12
      IF(DATA(L).EQ.9999. .OR. DATA(L).EQ.9998.)GO TO 32
      X=SIGN(5.,DATA(L))
      IF(X.LT.0.)GO TO 32
      DATA(L)=DATA(L)*EXP
      GO TO 35
```

```
   32 DATA(L)=-999.90
   35 CONTINUE
      GO TO 40
C
C CONVERT DEGREES FARENHEIT TO CENTIGRADE.
C
   36 DO 38 LL=1,12
      IF(DATA(LL).EQ.9999.  .OR.  DATA(LL).EQ.9998.)GO TO 37
      Y=SIGN(5.,DATA(LL))
      IF(Y.LT.0.)GO TO 37
      DATA(LL)=((DATA(LL)*EXP)-32.)*.5556
      GO TO 38
   37 DATA(LL)=-99.90
   38 CONTINUE
C
C   BRANCH TO APPROPRIATE SUBROUTINE
C
   40 IF(PARA.EQ.44201.  .OR.  PARA.EQ.42602.)CALL PEAK
      IF(PARA.EQ.44201.  .OR.  PARA.EQ.42603.  .OR.  PARA.EQ.43102.)
     1CALL SIXNIN
      IF(PARA.EQ.62101.)CALL MAXTMP
      IF(PARA.EQ.61101.  .OR.  PARA.EQ.61102.  .OR.  PARA.EQ.62101.)
     1CALL METDAT
      IF(PARA.EQ.SOLAR.OR.PARA.EQ.HUMID)CALL METDAT
   90 CONTINUE
      GO TO 5
   95 IEOF=IEOF+1
C
C IF MORE THAN ONE FILE IS TO BE READ FROM TAPE, CHECK HERE FOR
C THE NUMBER OF END-OF-FILE (IEOF) MARKS TO BE READ. THE FOLLOWING LINES
C ARE EXAMPLES OF CODE TO USE IF THERE ARE THREE TAPE FILES TO READ.
C     IF(IEOF.GT.2)GO TO 96
C     GO TO 5
C
   96 CONTINUE
  205 RETURN
      END
```

SUBROUTINE PEAK

```
      SUBROUTINE PEAK
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
C
C
C
C THIS PROGRAM FINDS THE DAILY PEAK OZONE OR NO2 CONCENTRATION AT
C EACH MONITORING SITE. IN ADDITION, IT RECORDS THE HOUR(S) OF THE PEAK
C AND THE NUMBER OF MEASURED VALUES BETWEEN 9 AM AND 9 PM.
C
C
C WHICH POLLUTANT?
C
      KE=0
      IF(PARA.EQ.42602.)KE=5
      KA=KE+1
      KB=KE+2
      KC=KE+3
      KD=KE+4
```

```
      HRCNT=3.0
      IF(IH.EQ.12)HRCNT=9.0
      PK=0.0
      PKHR1=0.0
      PKHR2=0.0
      PKSUM=0.0
C
C FIND THE PEAK VALUE AND RECORD THE HOUR(S) THAT IT OCCURS.
C
      DO 50 I=1,12
      IF(DATA(I) .LT. 0.)GO TO 40
      IF(DATA(I).GT.PK)GO TO 35
      IF(DATA(I).EQ.PK)GO TO 30
      GO TO 50
C
C
C THIS VALUE EQUALS THE PEAK
C
   30 PKSUM = PKSUM + 1.
      PKHR2=IH+I
      GO TO 50
C
C THIS VALUE IS THE NEW PEAK VALUE.
C
   35 PKHR1=IH+I
      PKHR2=PKHR1
      PKSUM=1.0
      PK=DATA(I)
      GO TO 50.
C
C  THIS VALUE IS LESS THAN THE CURRENT PEAK VALUE.
C
   40 IF(IH.EQ.12)GO TO 45
      IF(I.LT.10)GO TO 50
      HRCNT=HRCNT-1.0
      GO TO 50
   45 IF(I.GT.9)GO TO 50
      HRCNT=HRCNT-1.0
   50 CONTINUE
C
C COMPARE PEAK VALUE TO PEAK FOUND EARLIER FOR THIS DATE IF ANY
C AND STORE THE NEW PEAK AND THE HOUR(S) OF ITS OCCURRENCE.
C
      IF(PK.LT.STORE(IDAY,ISTN,KA))GO TO 100
      IF(PK.EQ.STORE(IDAY,ISTN,KA))GO TO 90
      STORE(IDAY,ISTN,KA)=PK
      STORE(IDAY,ISTN,KB)=PKHR1
      STORE(IDAY,ISTN,KC)=PKHR2
      STORE(IDAY,ISTN,KD)=PKSUM
      GO TO 100
C
C CURRENT PEAK IS LESS THAN PREVIOUS PEAK FOR THIS STATION ON
C THIS DATE.
C
   90 IF(PKHR1.LT.STORE(IDAY,ISTN,KB))STORE(IDAY,ISTN,KB)=PKHR1
      IF(PKHR2.GT.STORE(IDAY,ISTN,KC))STORE(IDAY,ISTN,KC)=PKHR2
      STORE(IDAY,ISTN,KD)=STORE(IDAY,ISTN,KD)+PKSUM
  100 IF(PARA.EQ.42602.)GO TO 101
      STORE(IDAY,ISTN,5)=STORE(IDAY,ISTN,5)+HRCNT
  101 RETURN
      END
```

SUBROUTINE GROUP

```
      SUBROUTINE GROUP
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
      DIMENSION MET(24,7),STOR2(32,7)
      EQUIVALENCE (STORE(1,32,7),STOR2(32,7))
      EQUIVALENCE (STRMET(1,24,7),MET(24,7))
      REAL MET
C
C
C THIS PROGRAM COMBINES DATA IN ARRAY STORE TO FORM THE DAILY
C SUMMARY FORMAT TO PRINTED. IN ADDITION,THE DAILY PEAK OZONE
C VALUE FOR THE MONITORING NETWORK IS FOUND AND THE DATA ARE PRINTED
C TO A HARD COPY UNIT AND TO AN AUXILLARY UNIT IN BINARY FORMAT.
C
C
      DO 200 I=1,NDAYS
C
C FIND NETWORK DAILY PEAK OZONE.
C
      PKSTN=0.0
      PKOZ=0.0
      DO 90 IA=1,NSTNS
      IF(STORE(I,IA,5).LT.9.)GO TO 75
      IF(STORE(I,IA,1).GT.PKOZ)GO TO 70
      GO TO 75
   70 PKOZ=STORE(I,IA,1)
      PKSTN=IA
C
C COMBINE DATA IN ARRAY STORE INTO OUTPUT ARRAY STOR2.
C
   75 XY=1.
      IF(STORE(I,IA,5).LT.9.  .AND.  STORE(I,IA,1).GT..120)XY=-1.
      STOR2(IA,1)=XY*(STORE(I,IA,1)+(STORE(I,IA,5)*100.))
      STOR2(IA,2)=(STORE(I,IA,2)*10000.)+(STORE(I,IA,3)*100.)
     1+STORE(I,IA,4)
      STOR2(IA,3)=(STORE(I,IA,7)*10000.)+(STORE(I,IA,8)*100.)
     2+STORE(I,IA,9)+STORE(I,IA,6)
      STOR2(IA,4)=STORE(I,IA,10)
      DO 85 IB=5,7
      IC=IB+6
      STOR2(IA,IB)=STORE(I,IA,IC)
   85 CONTINUE
   90 CONTINUE
      DO 105 JK=1,NSTNS
      DO 104 JKL=1,7
      IF(STOR2(JK,JKL).LT..001)STOR2(JK,JKL)=-999.00
  104 CONTINUE
  105 CONTINUE
C
C COMBINE DATA IN ARRAY STRMET.
C
      DO 120 J=1,24
      DO 110 JA=1,7
      MET(J,JA)=STRMET(I,J,JA)
  110 CONTINUE
  120 CONTINUE
      PKSTN=PKSTN+PKOZ
```

```
C
C OUTPUT STATEMENTS.
C
      IDAY=I+INDAY-1
      CALL JLTODT
C
C BINARY WRITE TO TAPE OR DISK.
C
      IF(PKOZ.LT..08)GO TO 190
      WRITE(IWRB)HDR,NSTNS,IYR,IMO,IDY,IDAY,PKSTN,IAMT1,IAMT2,
     1((STOR2(IX,JX),JX=1,7),IX=1,NSTNS),MET
  190 IF(IPCON.EQ.0)GO TO 200
      PLIM=FLOAT(IPLIM)/1000.
      IF(PKOZ.LT.PLIM)GO TO 200
      WRITE(IWRI,600)(HDR(N),N=1,2),NSTNS,IYR,IMO,IDY,IDAY,PKSTN,
     1IAMT1,IAMT2
  600 FORMAT(1H1,1X,2A4,4I3,I4,F8.4,2I3)
      WRITE(IWRI,605)
  605 FORMAT(1H0,3HSTN,3X,3H9-9,2X,4HPEAK,3X,6HNUMBER,9X,3HNO2,3X,
     14HPEAK,10X,3HMAX,18X,14H6-9 AM AVERAGE)
      WRITE(IWRI,610)
  610 FORMAT(1H ,6X,3HHRS,2X,2HO3,5X,6HPK HRS,9X,3HHRS,3X,3HNO2,10X,
     14HTEMP,11X,2HO3,10X,3HNOX,9X,4HNMHC/)
      DO 650 M=1,NSTNS
      WRITE(IWRI,601)M,(STOR2(M,MA),MA=1,7)
  601 FORMAT(1H ,I3,2X,F10.4,3X,F7.0,7X,F11.4,6X,F8.2,3X,3(3X,F9.4))
  650 CONTINUE
      WRITE(IWRI,606)
  606 FORMAT(1H0,2HHR,6X,2HWS,8X,2HWD,6X,4HTEMP,6X,5HHUMID,6X,5HSOLAR,
     16X,3HWS2,7X,3HWD2/)
      DO 660 MM=1,24
      WRITE(IWRI,603)MM,(MET(MM,MN),MN=1,7)
  603 FORMAT(1H ,I2,7F10.3)
  660 CONTINUE
  200 CONTINUE
      RETURN
      END
```

SUBROUTINE SIXNIN

```
      SUBROUTINE SIXNIN
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
C
C
C
C THIS PROGRAM CALCULATES THE DAILY 6 AM TO 9 AM (LT) AVERAGE POLLUTANT
C CONCENTRATIONS. THIS PROGRAM IS USED FOR OZONE, NOX, AND NMHC
C CONCENTRATIONS.
C
C
C CORRECT FOR DAYLIGHT SAVINGS TIME.
C
      IC=0
      IF(IDDAY.GE.IBSVTM.AND.IDAY.LT.IESVTM)IC=1
      IA=7-IC
      IB=9-IC
      CNT=3.0
      SUM=0.0
      AK=4.
      SET=1110.
```

```
C
C CALCULATE 6-9 AM AVERAGE.
C
      DO 50 I=IA,IB
      AK=AK-1.
      IF(DATA(I).LT.0.)GO TO 45
      SUM=SUM+DATA(I)
      GO TO 50
   45 CNT=CNT-1.0
      SET=SET-(10.**AK)
   50 CONTINUE
C
C IF LESS THAN 2 OF THE 3 HOURS OF DATA AVAILABLE THEN AVERAGE IS
C CONSIDERED MISSING.
C
      IF(CNT.LT.2.)GO TO 55
      SUM=(SUM/CNT)+SET
      GO TO 60
   55 SUM=-999.90
C
C DETERMINE WHICH POLLUTANT AVERAGE IS BEING CALCULATED.
C
   60 IF(PARA.EQ.44201.)STORE(IDAY,ISTN,11)=SUM
      IF(PARA.EQ.42603.)STORE(IDAY,ISTN,12)=SUM
      IF(PARA.EQ.43102.)STORE(IDAY,ISTN,13)=SUM
      RETURN
      END
```

SUBROUTINE MAXTMP

```
      SUBROUTINE MAXTMP
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
C
C
C
C FINDS THE MAXIMUM TEMPERATURE.
C
C
      PKTEMP=-999.90
      DO 50 I=1,12
      IF(DATA(I).LE.PKTEMP)GO TO 50
      PKTEMP=DATA(I)
   50 CONTINUE
      IF(PKTEMP.GT.STORE(IDAY,ISTN,10))STORE(IDAY,ISTN,10)=PKTEMP
      RETURN
      END
```

SUBROUTINE METDAT

```
      SUBROUTINE METDAT
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
C
C
C
C THIS PROGRAM STORES 24 HOURLY METEOROLOGICAL DATA FOR TWO SPECIFIED
C SITES IN AN ARRAY NAMED STRMET(62,24,7). THE USER MUST SPECIFY THE
C TYPE OF HUMIDITY AND SOLAR RADIATION DATA WANTED.
C
C
C DETERMINE IF SITE IS A SELECTED METEOROLOGICAL SITE.
C
      IF(STN.EQ.AMET1.OR.STN.EQ.AMET2)GO TO 5
      GO TO 90
C
C DETERMINE WHICH METEOROLOGICAL PARAMETER IS BEING PROCESED.
C
    5 DO 10 I=1,3
      IF(PARA.EQ.AMETDT(I))GO TO 20
   10 CONTINUE
      IF(PARA.EQ.HUMID.AND.STN.EQ.AMET1)GO TO 20
      IF(PARA.EQ.SOLAR.AND.STN.EQ.AMET1)GO TO 20
      GO TO 90
C
C SPECIFY ARRAY INDEX NUMBER DEPENDING ON THE METEOROLOGICAL PARAMETER
C CODE.
C
   20 IF(PARA.EQ.62101.  .AND. STN.EQ.AMET2)GO TO 90
      IF(PARA.EQ.HUMID)I=4
      IF(PARA.EQ.SOLAR)I=5
      IF(I.LT.3.AND.STN.EQ.AMET2)I=I+5
C
C STORE METEOROLOGICAL DATA IN STRMET.
C
      DO 50 J=1,12
      IHR=IH+J
      STRMET(IDAY,IHR,I)=DATA(J)
   50 CONTINUE
   90 RETURN
      END
```

SUBROUTINE JLTODT

```
      SUBROUTINE JLTODT
C
C
C THIS PROGRAM CONVERTS JULIAN DAYS TO MONTH AND DAY.
C
C
      COMMON /ALL/STORE(62,32,13),STRMET(62,24,7),DATA(12),NDAYS,IWRB,
     1UNITS(40),PAR(5),SITES(32),INDAY,IOUTDY,NSTNS,AMET1,AMET2,IWRI,
```

```
     2HUMID,SOLAR,IBSVTM,IESVTM,STN,IYR,IMO,IDY,IH,PARA,IU,ISTN,IDDAY,
     3IDAY,HDR(2),IAMT1,IAMT2,AMETDT(3),ITP
     4,IPCON,IPLIM
      DIMENSION NDYS(13)
      DATA NDYS /0,31,59,90,120,151,181,212,243,273,304,334,365/
      KY=0
      ICHK=MOD(IYR,4)
      IF(ICHK.EQ.0)KY=1
      IF(KY.EQ.1.AND.IDAY.GT.60)IDAY=IDAY-1
      DO 50 I=2,13
      IF(IDAY.LE.NDYS(I))GO TO 51
   50 CONTINUE
   51 IF(KY.EQ.1.AND.IDAY.EQ.60)GO TO 70
      IMO=I-1
      IDY=IDAY-NDYS(IMO)
      GO TO 75
   70 IMO=2
      IDY=29
      GO TO 76
   75 IF(KY.EQ.1.AND.IDAY.GT.59)IDAY=IDAY+1
   76 RETURN
      END
```

FUNCTION IDTJL

```
      FUNCTION IDTJL(IYR,IMO,IDY)
C
C
C THIS PROGRAM USES THE MONTH, DAY, AND YEAR DATA TO CALCULATE THE
C JULIAN DATE.
C
C
      DIMENSION IDAYS(12)
      DATA IDAYS/0,31,59,90,120,151,181,212,243,273,304,334/
      KEY=0
      ICHK=MOD(IYR,4)
      IF(ICHK.EQ.0)KEY=1
      IDTJL=IDAYS(IMO)+IDY
      IF(KEY.EQ.1.AND.IMO.GE.3)IDTJL=IDTJL+1
      RETURN
      END
```

APPENDIX C

USER'S GUIDE FOR THE ISOPLETH METHODOLOGY COMPUTER PROGRAM

OVERVIEW

The "EXCEED" program uses the isopleth method to determine design values and numbers of exceedances. This method uses data from a network of monitoring sites to estimate values throughout the area of interest. The following techniques are used to determine a grid-point value:

- If there are less than six stations (i.e., observations), the objective analysis is based on a fitting of a first-degree polynomial that requires values from only three stations. This method is used where there is insufficient data.
- If there are six or more observations, the objective analysis determines a grid-point value for the center of each cell from a least-squares fit of a second-degree polynomial.

The interpolated values are then used to determine exceedances and calculate design values for grid cells. If a value observed within a cell is higher than the interpolated value, then the observed value is used; otherwise the interpolated value stands. The theory and approach has been described in detail in the body of this report.

The program EXCEED provides for the specification of an ozone level that must be exceeded at one or more monitoring sites before the data are used as a basis for interpolation and subsequent calculations. A value must also be specified for the standard; generally the current NAAQS (120 ppb) will be used. Finally, there is an option to ignore data from certain sites. This option might be invoked, for example, when it is known that the site is unrepresentative or that the instrument was not operating properly.

A listing of the program is included at the end of this appendix.

PROGRAM INPUT

The program input specifies the criteria by which the calculations are made. Table C-1 summarizes the inputs used to define the program operation. Although Table C-1 refers to card format, any input medium is suitable.

The program also uses the observed peak-hour ozone concentration for each selected date at each monitoring site. These data are read from a separate logical unit in the form supplied by the program RDSARD that is described in Appendix B. Table C-2 summarizes the daily ozone information read from Logical Unit 1 by the program EXCEED.

Table C-2

SUMMARY OF PEAK HOUR OZONE AND OTHER INFORMATION READ FROM LOGICAL UNIT 1 BY PROGRAM EXCEED

Word Number	Variable Name	Remarks
1	HDR	Alphanumeric identifier for data
2	LSTA	Number of stations for which data are included
3	IYR	Year of data
4	MO	Month of data
5	MDAY	Day of data
6	JDAY	Julian date of data
7	OZMX	Peak ozone value for day (ppm)
8	JS1	Station number of site for general meteorological data
9	JS2	Station number of site for wind data
10+	OZM(J),(DUMS(I),I=1,7)	Array of ozone information (see discussion of program RDSARD for form) for each station and dummy variables read to skip data that are not used.

Table C-1

SUMMARY OF INPUTS USED TO SPECIFY EXCEED RUN CHARACTERISTICS
(Read from Logical Unit 5)

Card Type	Format	Columns	Variable Name	Units	Typical Value	Remarks
1	20A4	1-80	HEADER	--	--	An alphanumeric header of 80 characters or less.
2	I5	1-5	NSTA	--	10	Total number of stations from which data are input. Must be less than 36.
	F5.0	6-10,11-15	XORIG,YORIG	km	14.0,192.	Coordinates of the origin (southwest corner) of the initial grid using the same system used to define station coordinates.
	F5.1	16-20	GSIZE	km	20.0	Grid size, length of an edge.
	F10.3	21-30	CUTOFF	ppb	80.0	Ozone level which must be exceeded by at least one monitoring site on a given day for the data to be used.
	F5.1	31-35	VIOLAT	ppb	120.0	Ozone level which must be exceeded for an exceedance to be recorded.
	I5	36-40,41-45	NX,NY	--	5	Initial number of grids in the x and y directions.
	F5.2	46-50	FACTOR	--	1.05	Amount by which number of exceedances must increase from previous division, for subdividing to continue.
	I5	51-55	NDAY	--	170	Number of days of data to be read.
	I5	56-60	NONO	--	0	Number of sites not to be used in calculations.
	F6.1	61-66	BAD	--	-999.9	Identifier for bad data.
	I5	67-71	NSUBD	--	4	Maximum number of cells to be divided per cycle (i.e., for each pass through the data).
	I5	72-76	NVALID	--	365	Number of valid days in total sample, used for calculating the design values.
3a,3b,...	2F8.2	1-80	XUSE,YUSE	km	--	X,y coordinates for the monitoring sites in kilometers; one pair per card.
4a,4b,...	16I5	1-80	ILK	--	--	Array of monitoring sites to be ignored. Read only if NONO is greater than zero.

PROGRAM OUTPUT

The first page of output lists the variables read by the program. The array ILK is listed only if NONO is greater than zero. The next page of output lists the number of exceedances for the network, followed by a list of all the cells that were generated during the running of the program and pertinent information about those grid cells. The coordinates of the center of each grid cell are given, along with the number of exceedances found in that grid cell, the ten highest ozone concentrations for the cell, and the design value calculated from the ten highest ozone concentrations. Finally, the highest design value and the grid point at which it occurred are listed. The last page of output identifies the ten cells with the greatest number of exceedances, the highest concentration for each cell, and the design value for each cell. That information was previously given on the second page of output, but the tabulation here also includes the dates on which the data used to calculate the design value were collected.

Information is written on Logical Unit 10 for the probabilistic method program. The information written includes:

- Number of days from which the data set was drawn.
- Initial number of grid cells before subdividing.

```
ANALYSES OF O3 VIOLATIONS FOR A PERIOD IN AN AREA
NSTA = NO. OF STATIONS :   25
COORDINATES OF ORIGIN :   XORIGN=  680.00   YORIGN= 4230.00
GRID SPACING :   20.00
MINIMUM PK HR VALUE FOR INCLUSION IN CALCULATIONS :  80.000
LEVEL ABOVE WHICH EXCEEDANCES OCCUR : 120.0
INITIAL NO. OF X & Y GRIDS :      5     5
FACTOR OF INCREASE TO CONTINUE SUBDIVISION :  1.05
NO. OF DAYS OF DATA TO BE READ :  170
NO. OF SITES TO BE IGNORED :     9
SITE NOS. TO BE IGNORED :     2     3     4     5     6     7
                             10    11    13
BAD DATA IDENTIFIER : -999.9
MAXIMUM NO. OF UNITS TO BE SUBDIVIDED PER STEP :     4
NO. OF VALID DAYS IN TOTAL SAMPLE :    365
STATION COORDINATES         STA.     X              Y .
                              1    744.30        4279.90
                              2    742.50        4286.00
                              3    747.60        4282.50
                              4    747.30        4277.30
                              5    743.80        4276.50
                              6    738.70        4277.60
                              7    740.20        4283.60
                              8    748.40        4291.00
                              9    755.80        4279.90
                             10    747.20        4271.90
                             11    739.90        4272.60
                             12    735.00        4280.90
                             13    737.70        4289.80
                             14    744.30        4297.50
                             15    755.10        4297.80
                             16    762.80        4290.20
                             17    760.70        4272.80
                             18    743.10        4262.40
                             19    729.70        4270.70
                             20    721.10        4284.20
                             21    732.40        4302.40
                             22    741.60        4329.20
                             23    777.30        4286.40
                             24    749.30        4236.60
                             25    697.50        4282.30
```

EXAMPLE OF OUTPUT FROM PROGRAM EXCEED

```
                    ANALYSES OF O3 VIOLATIONS FOR A PERIOD IN AN AREA

NUMBER OF DAYS= 149

PT. X,Y COORDINATES   EXC                                                                                         DESIGN VALUE
                                                      TEN HIGHEST
                               1       2       3       4       5       6       7       8       9      10
  1  690.00 4240.00   10     139.    135.    133.    132.    129.    126.    126.    125.    124.    124.         137.0
  2  690.00 4260.00   11     136.    136.    136.    135.    134.    129.    128.    126.    124.    124.         138.4
  3  690.00 4280.00   13     168.    148.    145.    143.    143.    142.    134.    133.    127.    124.         158.3
  4  690.00 4300.00   12     147.    146.    146.    140.    137.    135.    130.    130.    128.    128.         148.0
  5  690.00 4320.00   12     158.    157.    146.    142.    136.    135.    132.    128.    128.    126.         155.7
  6  710.00 4240.00   10     142.    140.    137.    132.    130.    126.    125.    124.    122.    121.         141.0
  7  710.00 4260.00   11     140.    135.    134.    133.    130.    130.    129.    127.    127.    126.         137.3
  8  710.00 4280.00   12     154.    145.    142.    140.    139.    137.    137.    131.    128.    123.         150.4
  9  710.00 4300.00   11     152.    151.    150.    144.    139.    136.    135.    131.    128.    127.         153.8
 10  710.00 4320.00   13     173.    164.    153.    141.    140.    138.    136.    128.    126.    126.         165.3
 11  730.00 4240.00   13     157.    148.    144.    139.    133.    130.    129.    126.    124.    121.         152.1
 12  730.00 4260.00   12     161.    148.    140.    137.    137.    133.    132.    127.    124.    124.         152.4
 13  730.00 4280.00   14     176.    167.    160.    156.    155.    154.    148.    145.    145.    141.         170.5
 14  730.00 4300.00   14     187.    173.    173.    154.    150.    149.    139.    133.    130.    128.         182.4
 15  730.00 4320.00   16     182.    175.    157.    155.    137.    136.    135.    134.    134.    133.         172.5
 16  750.00 4240.00   10     161.    153.    151.    148.    138.    138.    135.    129.    125.    122.         159.6
 17  750.00 4260.00   20     239.    160.    155.    154.    149.    144.    143.    141.    134.    133.         186.4
 18  750.00 4280.00   19     236.    225.    189.    159.    151.    149.    145.    142.    141.    134.         215.7
 19  750.00 4300.00   28     223.    221.    192.    186.    179.    177.    172.    170.    169.    159.         215.1
 20  750.00 4320.00   23     198.    190.    163.    157.    155.    148.    148.    142.    142.    139.         185.8
 21  770.00 4240.00   13     168.    161.    142.    133.    131.    129.    129.    128.    128.    127.         155.5
 22  770.00 4260.00   15     177.    173.    149.    149.    133.    132.    129.    126.    125.    125.         168.9
 23  770.00 4280.00   18     188.    181.    166.    143.    143.    141.    137.    137.    137.    133.         177.0
 24  770.00 4300.00   15     192.    180.    167.    155.    149.    146.    146.    142.    136.    135.         181.8
 25  770.00 4320.00   16     182.    165.    160.    159.    156.    153.    147.    141.    140.    134.         175.8
 26  745.00 4295.00   25     223.    191.    180.    177.    172.    170.    169.    166.    159.    155.         201.5
 27  745.00 4305.00   16     193.    192.    181.    163.    163.    156.    154.    150.    149.    147.         190.0
 28  755.00 4295.00   21     221.    192.    186.    179.    166.    164.    157.    157.    146.    143.         205.9
 29  755.00 4305.00   16     205.    181.    174.    168.    165.    156.    155.    152.    151.    141.         190.6
 30  745.00 4315.00   17     182.    173.    169.    165.    153.    145.    141.    140.    138.    136.         179.1
 31  745.00 4325.00   21     198.    190.    152.    149.    148.    145.    142.    139.    138.    138.         180.3
 32  755.00 4315.00   17     185.    176.    163.    162.    162.    151.    147.    145.    140.    137.         180.0
 33  755.00 4325.00   17     181.    177.    161.    158.    153.    149.    143.    137.    134.    132.         178.2
 34  745.00 4255.00   17     204.    156.    151.    143.    141.    139.    139.    136.    134.    128.         171.2
 35  745.00 4265.00   20     239.    158.    154.    152.    149.    144.    143.    141.    134.    134.         185.0
 36  755.00 4255.00   14     177.    162.    161.    143.    134.    133.    128.    128.    127.    125.         167.6
 37  755.00 4265.00   16     185.    173.    170.    149.    134.    133.    130.    127.    125.    125.         177.0
 38  745.00 4275.00   13     236.    174.    160.    151.    145.    140.    126.    126.    125.    124.         193.4
 39  745.00 4285.00   12     187.    179.    163.    151.    147.    144.    139.    138.    133.    128.         179.3
 40  755.00 4275.00   20     225.    189.    179.    159.    152.    141.    141.    138.    137.    134.         199.2
 41  755.00 4285.00   17     198.    191.    170.    157.    142.    141.    141.    136.    135.    134.         187.1
```

EXAMPLE OF OUTPUT FROM PROGRAM EXCEED (Continued)

```
RANK  POINT ID        XG       YG    EXCEEDS     LRGST O3   DVALUE              DVALUE DATES
  1      19       750.00 4300.00       28         223.00    215.07    760713 760608 760825 761002 760826
                                                                        760827 760607 760730 760821 761001
  2      26       745.00 4295.00       25         223.00    201.55    760713 760608 761002 760827 760607
                                                                        760730 760821 760825 761001 760813
  3      20       750.00 4320.00       23         198.00    185.75    760607 760825 760608 761002 760713
                                                                        760730 760719 761001 760804 760826
  4      28       755.00 4295.00       21         221.00    205.87    760608 760825 761002 760826 760813
                                                                        760730 760713 761001 760607 760824
  5      31       745.00 4325.00       21         198.00    180.26    760607 760825 761002 760730 760719
                                                                        760713 760804 761001 760626 760608
  6      17       750.00 4260.00       20         239.00    186.36    761001 760608 760813 760917 760904
                                                                        760731 760812 760708 760901 760607
  7      35       745.00 4265.00       20         239.00    184.95    761001 760608 760917 760813 760904
                                                                        760731 760812 760708 760607 760901
  8      40       755.00 4275.00       20         225.00    199.21    760813 760608 761001 760730 760812
                                                                        760521 760607 760713 760825 760824
  9      18       750.00 4280.00       19         236.00    215.74    761001 760813 760608 760730 760713
                                                                        760812 760825 760607 760521 760824
 10      23       770.00 4280.00       18         188.18    177.01    760813 760608 760812 760730 761001
                                                                        760723 760607 760731 760726 760825
```

EXAMPLE OF OUTPUT FROM PROGRAM EXCEED (Concluded)

- Number of days actually processed.
- Number of grid cells after subdividing.
- Half the final number of cells.
- Total number of cells with stations.
- Total number of the cells with stations plus the number of cells without stations that are included among those ten with the most exceedances and/or the ten with the highest design values.
- X and y coordinates of final grid cells.

Also written out on Logical Unit 10 are the interpolated values for each of the final number of grid cells for each day when exceedances occurred.

LOGICAL UNITS

Seven logical units are used in program EXCEED. The input shown in Table C-1 is read from Unit 5. Printed output is disposed to Unit 6. Unit 1 is the ozone data input file. Units 9, 11, and 12 are scratch files used during the program for writing and reading the data through several stages of processing. Unit 10 is used to output specific data for the probabilistic modeling program.

MAIN PROGRAM AND SUBROUTINE DESCRIPTIONS

The information below is included so that, if necessary, the user can more easily modify the program. A flowchart, which is given in Figure C-1, provides the overall organization of the computer program. The sections below describe the program and each subroutine, explain the function of each, list the parameters that are passed, and the parameters that are returned.

The following list shows schematically how the subroutines and functions are interconnected:

Main Program EXCEED
- BIGGYS
- BIG10
- CHOOSE
- COORXY
- IDPTS
- NEXTD
- USES
  - BMOD
- DSNVAL
- GRID
  - BOUNDS
  - MESH
  - MES2

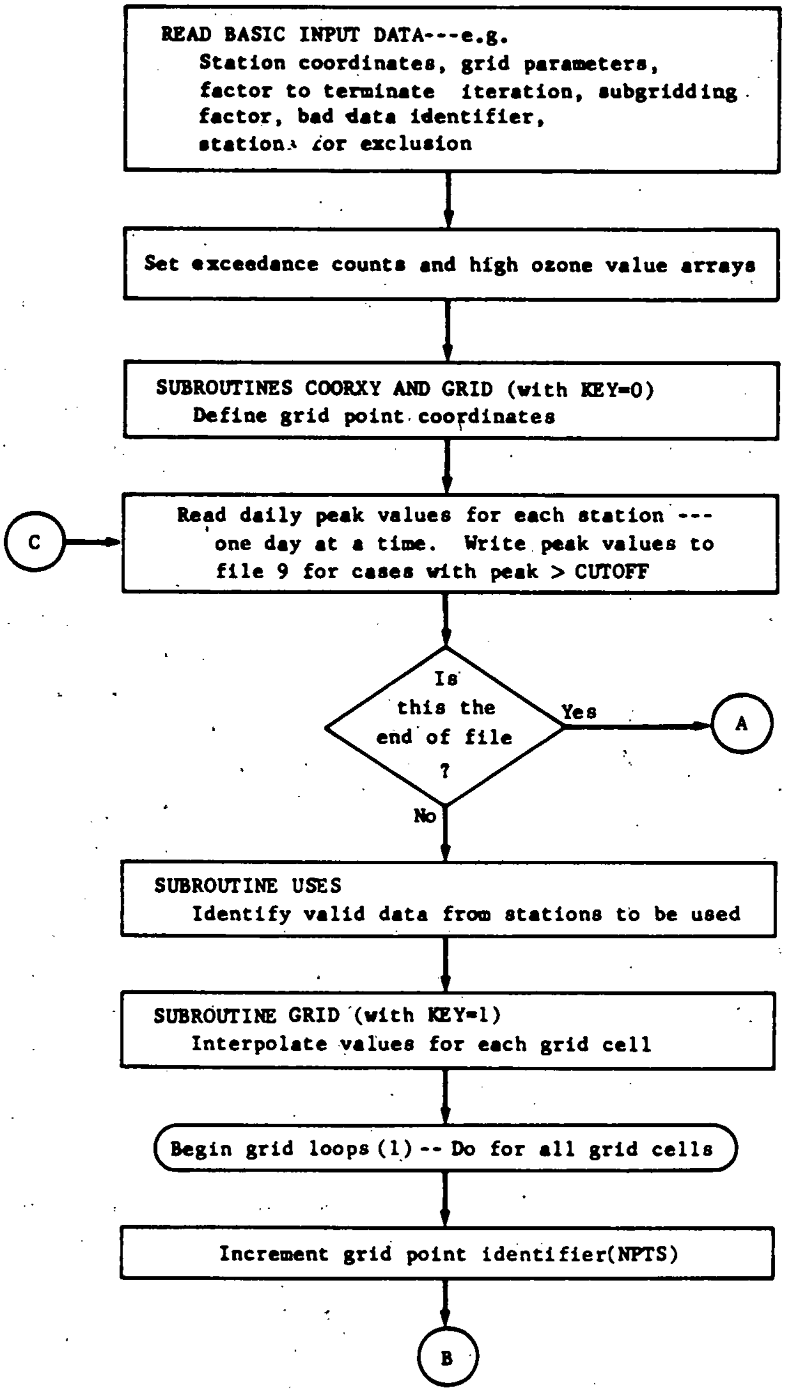


FIGURE C-1 FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED

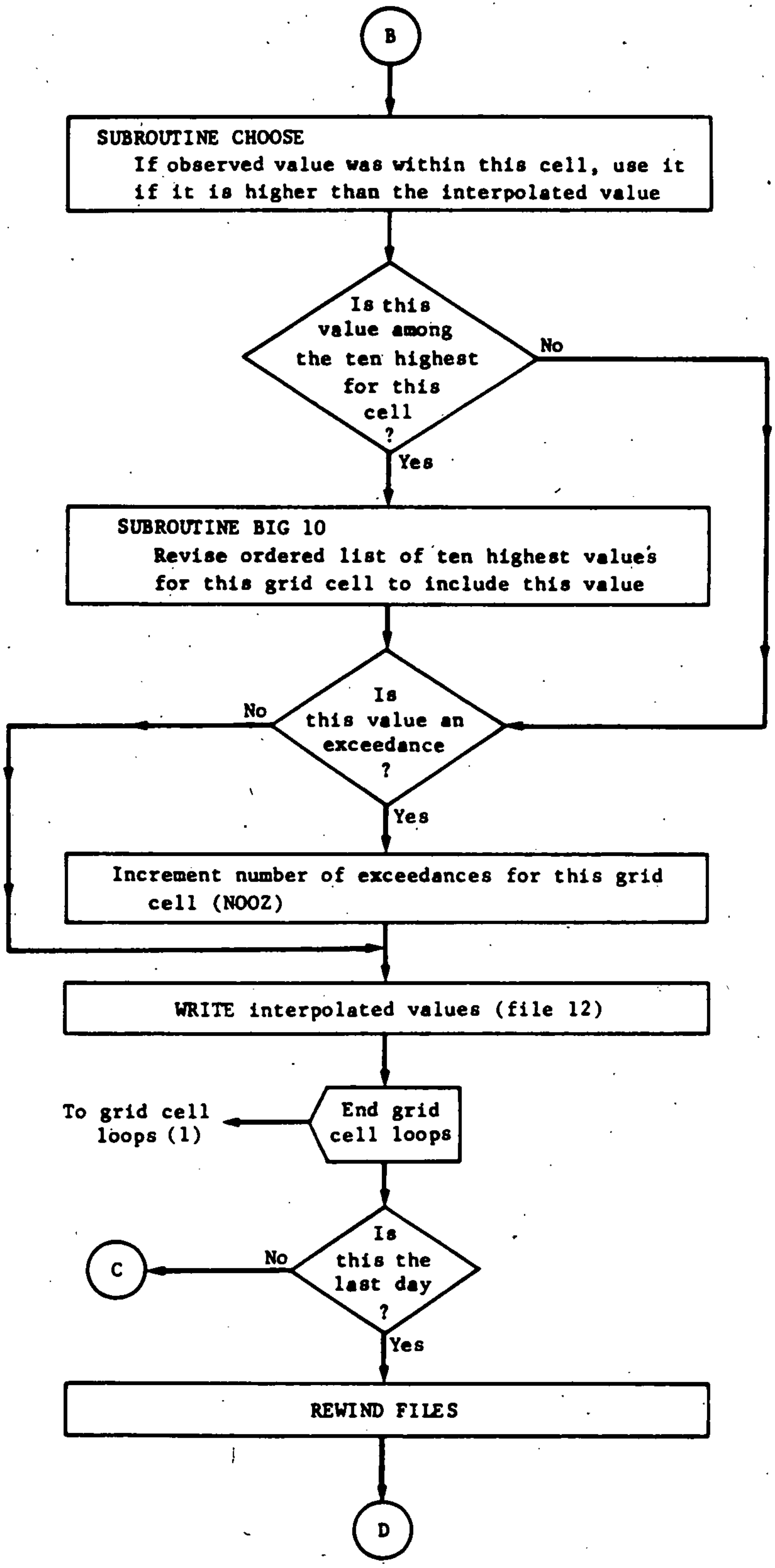


FIGURE C-1 FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED (Continued)

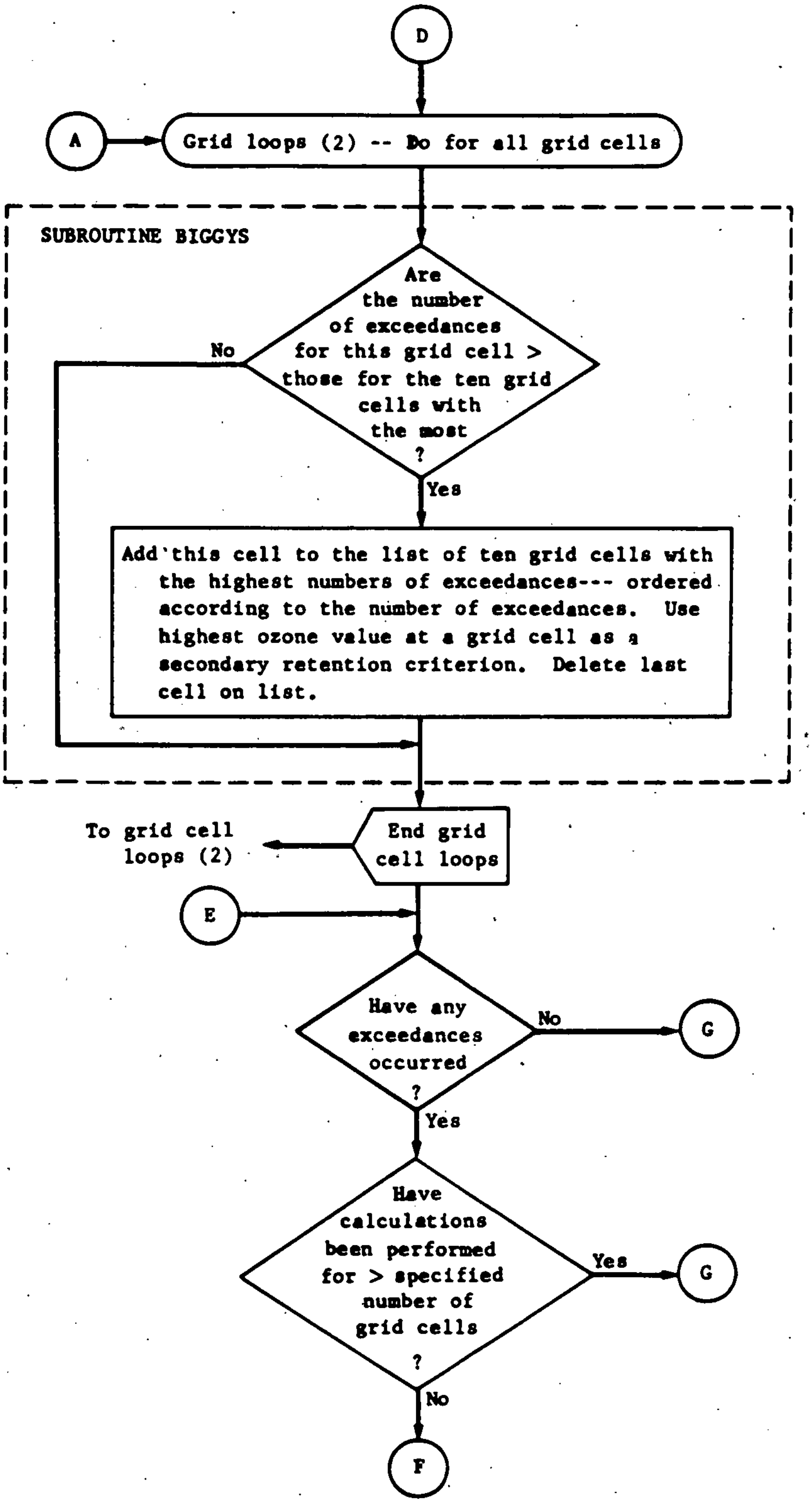


FIGURE C-1 FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED (Continued)

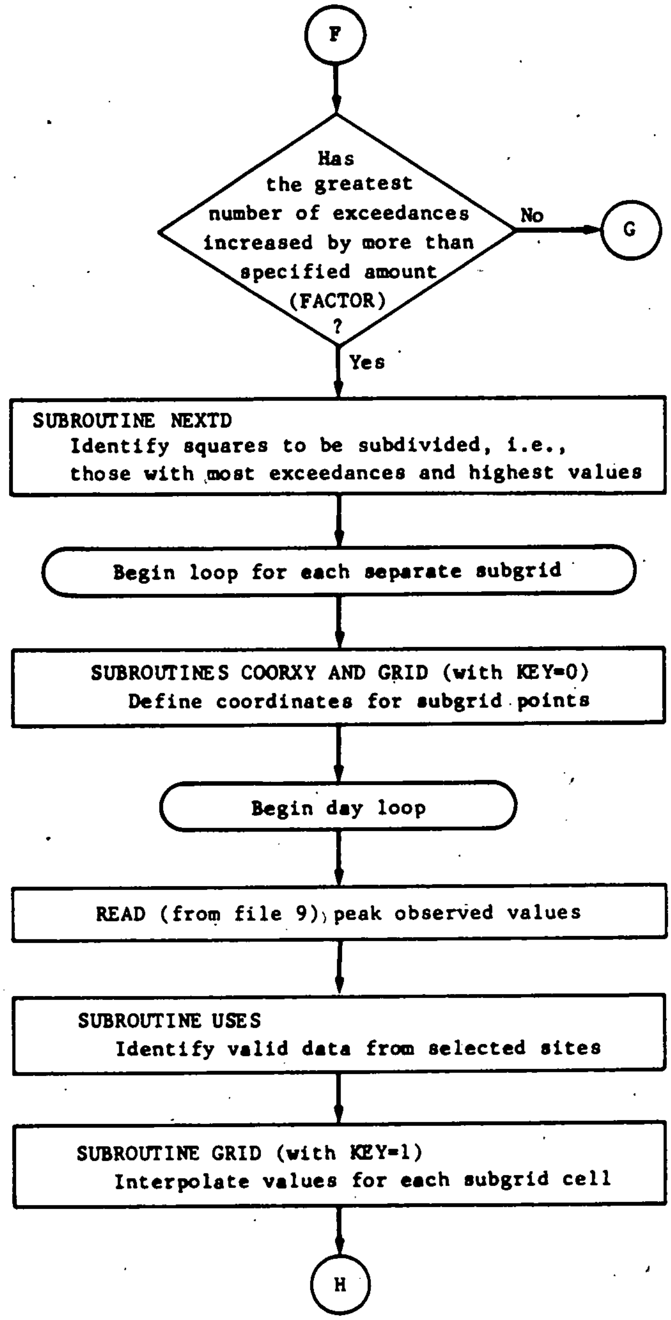


FIGURE C-1 FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED (Continued)

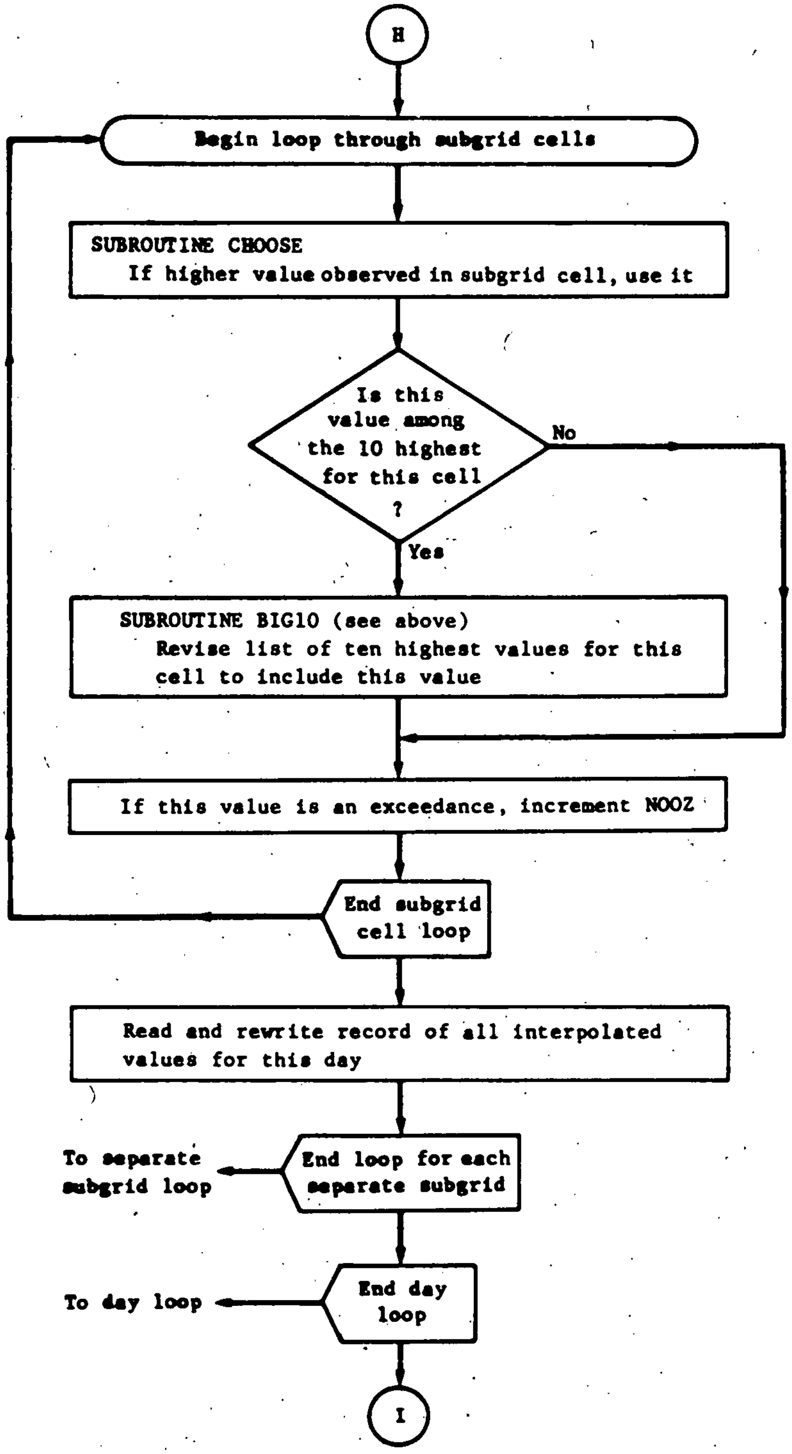


FIGURE C-1 FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED (Continued)

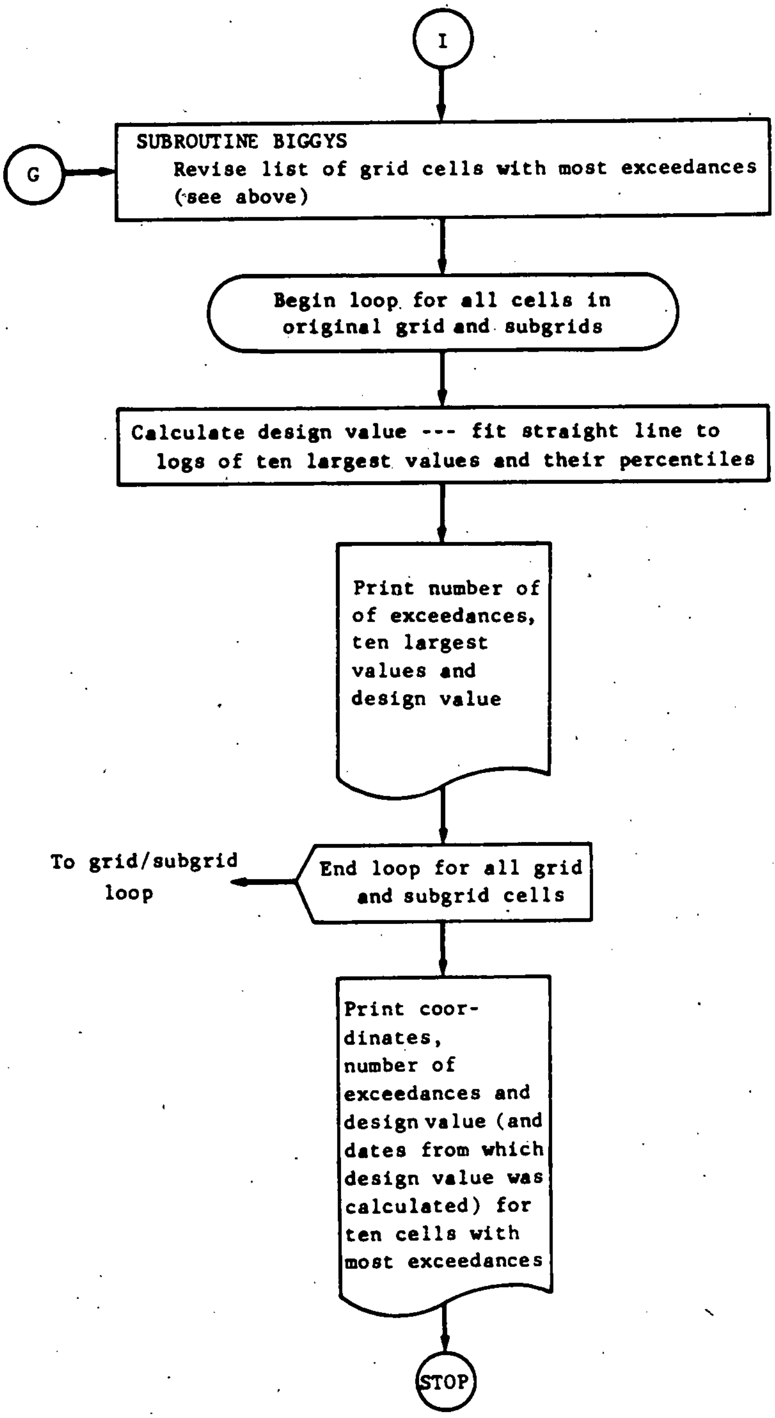


FIGURE C-1. FLOWCHART FOR ISOPLETH METHODOLOGY — COMPUTER PROGRAM EXCEED (Concluded)

The indentations show subroutines or functions called by the routine listed above them.

Program EXCEED--The main program controls the reading of input data, sets initial values for variables. It controls the iterative subdividing of the cells and the writing of output results. Directly or indirectly, all the subroutines and functions below are called by the main program.

BIGGYS (NOOZ,NPTS,TOPOZ,NBIGPT,NTOP,BIGOZ)--This subroutine keeps track of the number of exceedances for each grid cell, identifies the ten grid cells with the most exceedances and highest ozone observed in those ten cells. The parameters for BIGGYS are:

NOOZ--array containing number of exceedances for each grid cell. Dimension (399).

NPTS--index of grid cell for which exceedances are being updated.

TOPOZ--array containing highest ozone value for each of the ten grid cells with the most exceedances. Dimension (10).

NBIGPT--ordered array of the indices of the ten grid cells in which the most exceedances have occurred; index=1 denotes cell number with the most exceedances. Dimension (10).

NTOP--array giving the number of exceedances for the ten grid cells (NBIGPT) with the most exceedances. Dimension (10).

BIGOZ--array containing ten highest ozone values for each grid cell. Dimension (10,399).

BIGGYS is called from the main program and calls no other subroutines.

BIG10 (BIGOZ,NPTS,BOZ,IDATE,KDATE)--This subroutine keeps track of the ten largest observations for each grid cell and the dates when they occurred. The parameters for BIG10 are:

BIGOZ--array of ten highest ozone values for each grid cell. Dimension (10,399).

NPTS--index of grid cell for which list of ten largest values is being updated.

BOZ--ozone value for the grid cell with index NPTS.

IDATE--date for current ozone value, BOZ.

KDATE--array of dates corresponding to ten highest ozone values in each cell. Dimension (10,399).

BIG10 is called from the main program and calls no other subroutines.

CHOOSE (LL,BOZ,OZUSE,XUSE,,YUSE,X,Y,GSIZE)--This subroutine determines if an observing site is within the square for which an interpolated value was calculated and returns the larger (observed or interpolated) value. The parameters for CHOOSE are:

LL--number of stations.

BOZ--calculated ozone value for the grid point. If a higher value was observed in the cell, BOZ is changed to that value.

OZUSE--array of valid ozone data for monitoring stations. Dimension (35).

XUSE,YUSE--arrays of x,y coordinates for stations. Dimension (35).

X,Y--x,y coordinates of grid point.

GSIZE--size of each grid cell.

CHOOSE is called from the main program and calls no other subroutines.

COORXY (NX,NY,GSIZE,XORIG,YORIG,XG,YG,NPTS)--This subroutine calculates the coordinates of the mid points (XG,YG) of each cell in an NX by NY array of square cells with dimensions equal to GSIZE and the origin at XORIG,YORIG. NPTS is the index of each point in a list of such points. The parameters for COORXY are:

NY,NX--number of rows and columns.

GSIZE--size of each grid cell.

XORIG,YORIG--southwest corner or origin of grid.

XG,YG--calculated midpoints of the grid cells. Dimension (399).

NPTS--identifying number of the first cell in the grid.

This subroutine is called from the main program and calls no other subroutines. The values of the parameters are changed during the course of the calculations as grid cells are subdivided.

IDPTS (MANYPT,NWHICH,NBIGPT,DVAL,XUSE,YUSE,XG,YG,NPTS,NSTA,MSTA)-- This subroutine returns a list of cells to be used for probabilistic method. The list includes cells with monitors, but only the smallest cell containing a monitor is listed, plus those cells not listed as containing monitors that are among the ten with the most exceedances, the ten with the highest design values or both. The parameters for IDPTS are:

MANYPT--total number of cells that have either a monitor, one of the ten highest number of exceedances, or one of the ten highest design values.

NWHICH--list of cell numbers with monitors.

NBIGPT--array of the identification numbers for the ten grid cells in which the most exceedances occurred; index=1 denotes cell with most exceedances.

DVAL--array of design values corresponding to cells with the ten largest number of exceedances.

XUSE,YUSE--arrays of x,y coordinates for stations.

XG,YG--arrays of the x,y coordinates of the midpoints of the grid cells.

NPTS--number of grid cells.

NSTA--number of monitors.

MSTA--number of cells containing monitors that are within the boundaries of the grid. If a cell containing monitors has been subdivided, only the smaller cells are counted.

IDPTS is called from the main program and calls no other subroutines.

NEXTD (NOOZ,BIGOZ,NDIV,NSUBD,MANY,NPTS,NTOP,NBIGPT,ISTART)--This subroutine identifies cells to be subdivided (NDIV). The cells with the most exceedances and the cells with the highest observed values are selected, but duplications are eliminated. The parameters for NEXTD are:

NOOZ--array of number of exceedances for each grid cell. Dimension (399).

BIGOZ--array containing ten highest ozone values for each grid cell.

NDIV--array identifying cell numbers to be subdivided.

NSUBD--maximum number of cells to be subdivided per cycle.

MANY--number of cells to be subdivided.

NPTS--number of grid cells before subdividing.

NBIGPT--array of the identification numbers for the ten grid cells in which the most exceedances ocurred; index=1 denotes cell with most exceedances.

NTOP--ordered array giving the number of exceedances for the ten grid cells with the most.

ISTART--the current number of cells that have been generated.

NEXTD is called from the main program and calls no other subroutines.

USES (NSTA,OZM,LL,BAD,ILK,OZUSE)--This subroutine selects data to be used. It causes bad data and sites defined by array ILK to be ignored by setting OZUSE = 999. The parameters for USES are:

NSTA--number of stations.

OZM--ozone data initially read from unit 1.

LL--number of ozone values in OZUSE.

BAD--bad data identifier.

ILK--array of station numbers to be ignored.

OZUSE--array of good ozone data to be returned and used in later calculations.

This subroutine is called from the main program and calls no other subroutines.

BMOD (OZ)--This function takes a floating point number as the argument and returns the first three digits to the right of the decimal as a whole number. This function serves much the same purpose as the FORTRAN function AMOD. The parameter for BMOD is:

OZ--floating point number, of which the right-hand part will be returned as a whole number.

BMOD is called from the subroutine USES.

DSNVAL (BIGOZ,I,NDAYS)--This function calculates the design value using a linear fit to the logarithm of the ten largest values as a function of their corresponding percentiles:

BIGOZ--array of ten largest ozone values for each grid point.

I--current grid cell.

NDAYS--number of valid days from which sample was drawn (e.g., 365).

DSNVAL is called from the main program.

GRID (HS,XS,YS,JT,GSIZE,XMIN,YMAX,NT,MT,H,U,V,KEY)--This subroutine calls three other subroutines that establish the grid and calculate the grid-point values for ozone. The parameters for GRID are:

HS--an array of ozone data for the stations.

XS,YS--arrays of x,y coordinates of stations.

JT--number of stations.

GSIZE--grid-cell size.

XMIN,YMAX--x,y coordinates of upper-left corner of the grid for which values are being interpolated.

MT,NT--number of rows and columns.

H--calculated array of ozone concentrations at grid points.

U,V--wind components associated with the calculated ozone data (not used in EXCEED).

KEY--entry flag for initializing variables on the first call to grid.

This subroutine is called from the main program. Subroutines called by GRID are BOUNDS, MESH, and MES2.

BOUNDS (MT,NT,YMAX,XMIN,YD,XD,Y,X)--This subroutine must be called before any grid calculations are attempted in order to establish the analysis region, the grid network and its boundaries. The parameters for BOUNDS are:

MT,NT--number of rows and columns.

YMAX,YMIN--left-top (or northwest) corner grid-point coordinates.

YD,XD--y and x grid-spacing intervals (equal to GSIZE).

Y,X--arrays containing the y and x locations of columns and rows. Dimension Y(2*MT), X(2*NT).

BOUNDS is called from the subroutine GRID and calls no other subroutines.

MESH (KS,KSS5,HNIL,JJ,M9,N9,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)--This subroutine calculates values at a network of regularly spaced points, based on values given at a set of irregularly spaced points. This routine has an entry point (KEY=0) that must be used initially to establish various features and to determine the closest stations to each grid point. When KEY≠0, values KQ and IDS are substituted for KS and KSS5, respectively. The parameters for MESH are:

KQ--if KQ=1, uses area data weighted averaging analysis (first guess analysis); if KQ=2, uses least-squares fitting of first degree polynomial.

KS--number of stations closest to grid point to be stored.

KSS5--maximum number of stations to be used in the analysis of a grid-point value.

IDS--this parameter is not used in this application; it allows wind direction to be considered in the interpolation.

HNIL--maximum allowable absolute value for an observation. If greater than set value (e.g., 999), it is not used in the analysis.

JJ--number of stations.

M9,N9--number of rows and columns in the grid.

YS,XS--arrays of the x,y coordinates of stations.

VS,US--wind-component values (not used in this application).

HS--array of ozone values for stations.

Y,X--arrays containing the y,x coordinates of the columns and rows.

V,U--arrays of grid-point values for wind components (not used in this application).

H--array of calculated grid-point values for ozone.

IS--indices for the KS stations closest to each grid point. Dimension (KS*MT*NT).

KEY--determines entry point into subroutine. If KEY=0, the routine only calculates the necessary constants for the KS stations closest to each grid point. If KEY≠0, it interpolates grid-point values.

MESH is called twice from the subroutine GRID. Once to initialize the routine and once for the interpolation. MESH does not call other subroutines.

MES2 (KS,KSS5,HNIL,JJ,M9,N9,YS,XS,VS,US,HS,YL,XL,V,U,H,IS,KEY)--This subroutine calculates values at a discrete set of regularly spaced network of grid points based on values given for a set of irregularly spaced set of points. It is similar to subroutine MESH, but calculates grid-point values by a least-square fitting of a second-degree polynomial rather than a first degree. It requires a first-guess analysis for the arrays H, V, and U. However, V and U are wind components and are not used in this application. This routine has an entry point (KEY=0) which must be used initially to determine basic constants. When KEY≠0, values KQ and IDS are substituted for KS and KSS5, respectively. The parameters for MES2 are as follows:

KQ--if KQ=1, uses area data weighted averaging analysis (first-guess analysis); if KQ=2, uses least-squares fitting of second-degree polynomial.

KS--number of stations closest to grid point to be stored.

KSS5--maximum number of stations to be used in the analysis of a grid-point value.

IDS--this parameter is not used in this application; it allows wind direction to be considered in the interpolation.

HNIL--maximum absolute value for an observation. If greater than this value (e.g., 999), it is not used in the analysis.

JJ--number of stations.

M9,N9--number of rows and columns in the grid.

YS,XS--arrays of the x,y coordinates of stations.

VS,US--wind-component values (not used in this application).

HS--array of ozone values for stations.

Y,X--arrays containing the y,x coordinates of the columns and rows.

V,U--arrays of grid-point values for wind components (not used in this application).

H--array of interpolated grid-point values for ozone.

IS--indices for the KS stations closest to each grid point. Dimension (KS*MT*NT).

KEY--determines entry point into subroutine. If KEY=0, the routine only calculates the necessary constants for the KS stations closest to each grid point. If KEY≠0, it interpolates grid-point values.

MES2 is called by GRID twice: once to initialize the routine and once for the inperpolation. This subroutine calls no other subroutine.

LISTING OF PROGRAM EXCEED

A listing of the program EXCEED with all the subroutines follows. The program is written in a general FORTRAN that should be operable with most compilers with only minimal modification. Where we believe there may be machine dependent features, we have tried to include alternative code as comments. The version given runs on a Control Data Corporation 6400. A version of the program has been run on a Digital Equipment Corporation PDP 11/40: Some problems were encountered with overflow, but they have been corrected by scaling some of the intermediate values

generated in the subroutine MESH, so that the numbers are small enough to fit in a 32-bit word.

PROGRAM EXCEED

```
      PROGRAM EXCEED (INPUT,OUTPUT,TAPE5,TAPE6=OUTPUT,TAPE1,
     1TAPE2,TAPE9,TAPE10,TAPE11,TAPE12)
C
      DIMENSION HEADER(8),NWHICH(55),HDR(2)
     1, XG(399),YG(399),NOOZ(399),OZM(35),OZ2D(225),
     1NBIGPT(10),NTOP(10),TOPOZ(10),BIGOZ(10,399),NDIV(10),ILK(30),DVAL
     2(399),KDATE(10,399),XUSE(35),YUSE(35),OZUSE(35),GROZ(399)
     3,U(25,25),V(25,25),DUMS(35,6)
 1    FORMAT(8A10)
 2    FORMAT (I5,2F5.0,2F5.1,2I5,F5.2,2I5,F6.1,I5,F10.3,I5)
 3    FORMAT (2F8.2)
 4    FORMAT (16I5)
 6    FORMAT (1H ,I4,2F8.2,I5,10F8.0,F12.1)
 7    FORMAT (/,1H1,4HRANK,2X,9HPOINT ID ,8X,2HXG,6X,2HYG,3X,7HEXCEEDS,
     1 4X,8HLRGST O3,3X,6HDVALUE ,14X,12HDVALUE DATES )
 8    FORMAT (1H ,I3,I9,7X,2F8.2,I6,2X,2F10.2,4X,5I7,/,70X,5I7)
    9 FORMAT(24H0LARGEST DESIGN VALUE =  ,F8.1,17H AND IS AT POINT ,
     XI3)
 12   FORMAT(17H0 NUMBER OF DAYS= ,I4,/,
     128H0 PT. X,Y COORDINATES   EXC ,81X,12HDESIGN VALUE ,
     2/,1H ,56X,12HTEN HIGHEST ,/,25X,10I8)
   29 FORMAT(1X,I3,(10F10.4))
   30 FORMAT((1X,10F10.5))
   41 FORMAT(1H1,8A10)
   42 FORMAT(1H ,25HNSTA = NO. OF STATIONS :  ,I4,/,
     11H ,24HCOORDINATES OF ORIGIN : ,2X,7HXORIGN=,F8.2,3X,7HYORIGN=,
     2F8.2)
   43 FORMAT(1H ,14HGRID SPACING :,F8.2,/,1H ,
     151HMINIMUM PK HR VALUE FOR INCLUSION IN CALCULATIONS :,F11.3,/,
     21H ,37HLEVEL ABOVE WHICH EXCEEDANCES OCCUR :,F6.1)
   44 FORMAT(1H ,28HINITIAL NO. OF X & Y GRIDS :,2I6,/,1H ,
     144HFACTOR OF INCREASE TO CONTINUE SUBDIVISION :,F6.2,/,
     21H ,32HNO. OF DAYS OF DATA TO BE READ :,I5)
   45 FORMAT(1H ,28HNO. OF SITES TO BE IGNORED :,I5)
   46 FORMAT(1H ,25HSITE NOS. TO BE IGNORED :,(10I5))
   47 FORMAT(1H ,21HBAD DATA IDENTIFIER :,F7.1,/,1H ,
     148HMAXIMUM NO. OF UNITS TO BE SUBDIVIDED PER STEP :,I5,/,
     21H ,35HNO. OF VALID DAYS IN TOTAL SAMPLE :,I6)
   48 FORMAT(1H ,20HSTATION COORDINATES ,5X,6H STA. ,6H   X  ,7X,6H   Y
     1,/,(27X,I2,2X,F8.2,5X,F8.2))
   49 FORMAT(1H1,25X,8A10)
 1001 FORMAT(2A4,4I3,I4,F8.4,2I3)
 1002 FORMAT(2X,F11.4,F9.1,F12.4,7F10.4)
 6002 FORMAT (3I6,(10F10.6))
 6003 FORMAT(7I10)
 6004 FORMAT(10F8.2)
 6005 FORMAT(16I5)
 7086 FORMAT (1H1,32HDAYS DO NOT MATCH--DATE  DAY IDS ,3I7)
 7186 FORMAT (28H1TOO MANY OF SOMETHING--NX =,I3,4X,4HNY =,I3,4X,17HNO.
     1OF STATIONS = ,I3,/,26H NO. OF STATIONS IGNORED = ,I3,/,
     2 13H NO. OF DAYS= ,I4)
 7286 FORMAT(1H0,I2,14H STATIONS ON  ,3I2)
C
C      INPUT ...
C
C           NSTA = NO. OF STATIONS.
C           XORIG, YORIG = COORDINATES OF ORIGIN.
C           GSIZE = GRID SPACING.
C           NX, NY = GRID NUMBER (X,Y DIRECTIONS)
C           FACTOR = FACTOR OF INCREASE TO CONTINUE SUBDIVISION.
```

```
C           NDAY = NO. OF DAYS OF DATA TO BE READ.
C           NVALID = NO. OF VALID DAYS IN TOTAL SAMPLE (FOR CALCULATING
C           DESIGN VALUE).
C           NONO = NO. OF SITES NOT TO BE USED.
C           BAD  = BAD DATA IDENTIFIER (E.G. -99.0)
C           NSUBD = MAXIMUM NO. OF UNITS TO BE SUBDIVIDED PER STEP.
C           XUSE, YUSE = STATION COORDINATES
C           ILK  = SITE NOS. TO BE IGNORED (ORDERED AS IN LIST OF STA).
C
      DO 52 I=1,30
      ILK(I)=0
   52 CONTINUE
C
      READ (5,1) (HEADER(I),I=1,8)
      READ (5,2) NSTA,XORIG,YORIG,GSIZE,VIOLAT,NX,NY,FACTOR,NDAY
     1 ,NONO,BAD,NSUBD,CUTOFF,NVALID
      DO 53 J=1,NSTA
      READ (5,3) XUSE(J),YUSE(J)
   53 CONTINUE
      IF(NONO .GT. 0) READ (5,4) (ILK(J),J=1,NONO)
C
      WRITE (6,41) (HEADER(I),I=1,8)
      WRITE (6,42) NSTA,XORIG,YORIG
      WRITE (6,43) GSIZE,CUTOFF,VIOLAT
      WRITE (6,44) NX,NY,FACTOR,NDAY
      WRITE (6,45) NONO
      IF (NONO .GT. 0) WRITE (6,46) (ILK(I),I=1,NONO)
      WRITE (6,47) BAD,NSUBD,NVALID
      WRITE (6,48) ((I,XUSE(I),YUSE(I)),I=1,NSTA)
C
      NXNY=NX*NY
      IF (NXNY .GT. 225) GO TO 186
      IF (NONO .GT. 30) GO TO 186
C
C          SET THE NO. OF O3 VIOLATIONS AND 10 LARGEST OBSERVED PEAK
C     HOURLY O3 VALUES EQUAL TO 0 FOR ALL GRID POINTS.
C
      INIGRID=NX*NY
      NWR = 12
      NUMDAY=0
      DO 105 I=1,399
      NOOZ(I)=0
      GROZ(I)=0.0
      DO 100 J=1,10
      BIGOZ(J,I)=0.0
 100  CONTINUE
 105  CONTINUE
      ITER=0
      LASTHI=0
      ISTART=0
      NPTS=ISTART
      YN=NY-1
      YMAX=YORIG+(YN+0.5)*GSIZE
      XPTO= XORIG+0.5 * GSIZE
C
C          CALL BOUNDS (NX,NY,YMAX,XORIG,GSIZE,GSIZE,YY,XX)
C
      CALL COORXY (NX,NY,GSIZE,XORIG,YORIG,XG,YG,NPTS)
      KEY=0
      CALL GRID(OZUSE,XUSE,YUSE,NSTA,GSIZE,XPTO,YMAX,NX,NY,OZ2D,U,V
     1,KEY)
      KEY=1
      DO 140 IDAY=1,NDAY
C  UNFORMATTED READ
      READ (1) HDR,LSTA,IYR,MO,MDAY,JDAY,OZMX,JS1,JS2,
     1(OZM(I),(DUMS(I,J),J=1,6), I=1,NSTA)
      IF(EOF(1)) 115,120
```

```
 115  CONTINUE
      REWIND 1
      REWIND 9
      REWIND 11
      REWIND 12
      GO TO 145
  120 CONTINUE
C     OZMX=BMOD(OZMX)
      OZMX=1000.*AMOD(OZMX,1.)
      IF (OZMX .LT. CUTOFF) GO TO 140
  310 FORMAT(2X,*NUMDAY, IYR,MO,MDAY, JDAY, OZMX *,I4,1X,3I2,I4,F12.6)
C
C  WRITE GOOD DATA TO BE USED LATER ON TAPE 9
C
      CALL USES(NSTA,OZM,LL,BAD,ILK,OZUSE,XUSE,YUSE)
      IF(LL.LT.4.)WRITE(6,7286)LL,IYR,MO,MDAY
      IF(LL.LT.1.)GO TO 140
      NUMDAY=NUMDAY+1
      WRITE (9) HDR,LSTA,IYR,MO,MDAY,JDAY,OZMX,JS1,JS2,
     1 (OZM(I),I=1,NSTA)
      CALL GRID (OZUSE,XUSE,YUSE,NSTA,GSIZE,XPTO,YMAX,NX,NY,OZ2D,U,V
     1,KEY)
      NPTS = ISTART
      IDATE=10000 * IYR + 100*MO + MDAY
      DO 135 IX=1,NX
      DO 130 JY=1,NY
      NPTS=NPTS+1
C
C  COMPUTING CORRECT INDEX FOR ARRAY VALUE RETURNED FROM GRID.
C  GRID COMPUTES COORDINATES FROM UPPER LEFT CORNER.
C  EXCEED COMPUTES COORDINATES FROM LOWER LEFT CORNER.
C
      IY=(NY-JY)+1
      IXIY=(IY-1)*NX+IX
      BOZ=OZ2D(IXIY)
C
C  USE OBSERVED VALUE IF THERE IS ONE IN THE SQUARE AND IT IS LARGER.
C
      CALL CHOOSE(NSTA,BOZ,OZUSE,XUSE,YUSE,XG(NPTS),YG(NPTS),GSIZE)
      GROZ(NPTS)=BOZ
C
C .        CHECK GRID SQUARES TO SET IF OZONE HIGHER THAN THOSE AT OTHER
C     TIMES.
C
C          GROZ IS GRID CELL VALUE FOR THIS DAY.
C
      IF (BOZ.LE.BIGOZ(10,NPTS)) GO TO 125
      CALL BIG10 (BIGOZ,NPTS,BOZ,IDATE,KDATE)
 125  IF (BOZ.LE.VIOLAT) GO TO 130
      NOOZ(NPTS)=NOOZ(NPTS)+1
 130  CONTINUE
 135  CONTINUE
      WRITE (12)  JDAY,IDATE,(GROZ(KPTS),KPTS=1,NPTS)
 140  CONTINUE
      REWIND 12
      REWIND 11
      REWIND 9
      REWIND 1
 145  NPTS=ISTART
      IF (NUMDAY .LE. 0) GO TO 186
      DO 155 IX=1,NX
      DO 150 IY=1,NY
C          CHECK FOR WHICH 10 SQUARES HAVE GREATEST NO. OF EXCEEDANCES.
      NPTS=NPTS+1
      CALL BIGGYS (NOOZ,NPTS,TOPOZ,NBIGPT,NTOP,BIGOZ)
 150  CONTINUE
 155  CONTINUE
```

```
C
C           ITERATIONS CEASE IF ...
C                NO EXCEEDANCES DETECTED OR
C                NO. OF PTS. TESTED GT 300 OR
C                FRACTION INCREASE IN EXCEEDANCES LT FACTOR
C
      IF (NTOP(1) .EQ. 0) GO TO 285
  160 IF (NPTS.GE.300) GO TO 285
      XLAST=LASTHI
      MUST=FACTOR*XLAST
      IF (MUST.GE.NTOP(1)) GO TO 285
      LASTHI=NTOP(1)
C
C          GETTING CELLS TO BE DIVIDED ON THE NEXT ITERATION.
C
      CALL NEXTD (NOOZ,BIGOZ,NDIV,NSUBD,MANY,NPTS,NTOP,NBIGPT,ISTART)
      ISTART=NPTS
      JSTART=NPTS
      NX=2
      NY=2
      GSIZE=GSIZE*0.5
      DO 180 I=1,MANY
      NWR=12
      ITER=ITER+1
      NRD=11+MOD(ITER,2)
      IF (NRD.EQ.12) NWR=11
      J=NDIV(I)
      XORIG=XG(J)-GSIZE
      YORIG=YG(J)-GSIZE
      CALL COORXY (NX,NY,GSIZE,XORIG,YORIG,XG,YG,NPTS)
      YN=NY-1
      YMAX=YORIG + (YN+0.5) * GSIZE
      XPTO= XORIG+0.5 * GSIZE
      KEY=0
      CALL GRID(OZUSE,XUSE,YUSE,NSTA,GSIZE,XPTO,YMAX,NX,NY,OZ2D,U,V
     1,KEY)
      KEY=1
      DO 175 IDAY=1,NUMDAY
      NPTS=JSTART
      READ (9) HDR,LSTA,IYR,MO,MDAY,JDAY,OZMX,JS1,JS2,
     1(OZM(IK), IK=1,NSTA)
      IDATE=10000*IYR+100*MO+MDAY
      CALL USES (NSTA,OZM,LL,BAD,ILK,OZUSE,XUSE,YUSE)
      IF (LL .LT. 4) WRITE(6,7286) LL,IYR,MO,MDAY
      CALL GRID (OZUSE,XUSE,YUSE,NSTA,GSIZE,XPTO,YMAX,NX,NY,OZ2D,U,V
     1,KEY)
      DO 170 IX=1,NX
      DO 165 JY=1,NY
      NPTS=NPTS+1
C
C  COMPUTING CORRECT INDEX FOR ARRAY VALUE RETURNED FROM GRID.
C  GRID COMPUTES COORDINATES FROM UPPER LEFT CORNER.
C  EXCEED COMPUTES COORDINATES FROM LOWER LEFT CORNER.
C
      IY=(NY-JY)+1
      IXIY=(IY-1)*NX+IX
      BOZ=OZ2D(IXIY)
C
C  USE OBSERVED VALUE IF THERE IS ONE IN THE SQUARE AND IT IS LARGER.
C
      CALL CHOOSE(NSTA,BOZ,OZUSE,XUSE,YUSE,XG(NPTS),YG(NPTS),GSIZE)
      GROZ(NPTS)=BOZ
      CALL BIG10 (BIGOZ,NPTS,BOZ,IDATE,KDATE)
      IF (BOZ .LE. VIOLAT) GO TO 165
      NOOZ(NPTS)=NOOZ(NPTS)+1
  165 CONTINUE
```

```
  170 CONTINUE
      READ (NRD) KDAY,IDATE,(GROZ(KPTS),KPTS=1,JSTART)
      IF (KDAY.NE.JDAY) GO TO 86
      WRITE (NWR) KDAY,IDATE,(GROZ(K),K=1,NPTS)
  175 CONTINUE
      JSTART=NPTS
      REWIND 9
      REWIND 11
      REWIND 12
  180 CONTINUE
      I1=ISTART+1
      DO 200 I=I1,NPTS
      CALL BIGGYS (NOOZ,I,TOPOZ,NBIGPT,NTOP,BIGOZ)
  200 CONTINUE
      GO TO 160
  285 CONTINUE
      BIGDV = 0.
      DO 290 I=1,NPTS
      IF(MOD(I,48).NE.1)GO TO 287
      WRITE (6,49) (HEADER(IZ),IZ=1,8)
      WRITE (6,12) NUMDAY, (L,L=1,10)
  287 DVAL(I)=DSNVAL (BIGOZ,I,NVALID)
      WRITE (6,6) I,XG(I),YG(I),NOOZ(I),(BIGOZ(J,I),J=1,10),DVAL(I)
      IF(BIGDV.GT.DVAL(I))GO TO 290
      BIGDV=DVAL(I)
      LOCDV=I
  290 CONTINUE
      WRITE(6,9)BIGDV,LOCDV
      WRITE (6,7)
      ITMP = 0
      DO 295 I=1,10
      ITMP = NBIGPT(I)
      WRITE (6,8) I,ITMP,XG(ITMP),YG(ITMP),NTOP(I),TOPOZ(I),
     1DVAL(ITMP),(KDATE(J,ITMP),J=1,10)
  295 CONTINUE
      REWIND 10
      NHALF=NPTS/2
C
C  THIS ROUTINE FINDS THE CELLS WITH MONITORS, MOST EXCEEDANCES AND
C  THE HIGHEST DESIGN VALUES.
C
C     PRINT 5000,MANYPT,NPTS,NSTA,MSTA,NWHICH(1),NBIGPT(1),
C    1DVAL(1),XUSE(1),YUSE(1),XG(1),YG(1)
 5000 FORMAT(* *,4I5,2I5,5F10.3)
      CALL IDPTS(MANYPT,NWHICH,NBIGPT,DVAL,XUSE,YUSE,XG,YG,NPTS,NSTA
     1,MSTA)
C
C
C  WRITE OUT TO TAPE10 :
C                             1)INITIAL DAYS
C                             2)INITIAL GRID CELLS
C                             3)FINAL # OF DAYS
C                             4)FINAL # OF GRID CELLS
C                             5)HALF OF FINAL # OF CELLS
C                             6)TOTAL # OF CELLS WITH MONITORS (STATIONS)
C                             7)TOTAL # OF CELLS WITH MONITORS & 10 HIGHEST
C                               EXCEEDANCES AND DESIGN VALUES.
C
      WRITE (10,6003) NDAY,INIGRID,NUMDAY,NPTS,NHALF,MSTA,MANYPT
      DO 300 I=1,MANYPT,16
      IP15=I+15
C
C  WRITE OUT TO TAPE10 LIST OF CELLS WITH MONITORS.
C
      WRITE (10,6005) (NWHICH(J),J=I,IP15)
  300 CONTINUE
```

```
C
C  WRITE OUT TO TAPE10 THE X & Y COORDINATES OF EACH CELL (NPTS).
C
      WRITE (10,6004) ((XG(I),YG(I)),I=1,NHALF)
      NHALF=NHALF+1
      WRITE (10,6004) ((XG(I),YG(I)),I=NHALF,NPTS)
C
C  WRITE OUT TO TAPE10 NPTS, DAY, DATE & OZONE VALUES FOR EACH CELL.
C
      DO 400 I=1,NUMDAY
      READ (NWR) IDAY,IDATE,(GROZ(L),L=1,NPTS)
      WRITE (10,6002) NPTS,IDAY,IDATE,(GROZ(L),L=1,NPTS)
  400 CONTINUE
      END FILE 10
      REWIND 10
      STOP300
  86  WRITE (6,7086) IDATE,KDAY,JDAY
  187 STOP
  186 WRITE (6,7186) NX,NY,NSTA,NONO,NUMDAY
      END
```

SUBROUTINE IDPTS

```
      SUBROUTINE IDPTS(MANYPT,NWHICH,NBIGPT,DVAL,XUSE,YUSE,XG,YG,NPTS
     1,NSTA,MSTA)
C
C  IDPTS RETURNS LIST OF CELLS TO BE USED FOR PROBABILISTIC CALCULATIONS
C      MANYPT = TOTAL NO. OF CELLS
C      NSTA   = NO. OF MONITORS
C      NWHICH = LIST OF CELLS WITH MONITORS (1ST NSTA ITEMS IN LIST--)
C      ONLY SMALLEST CELL CONTAINING THE MONITOR IS INCLUDED PLUS CELLS
C      WITHOUT MONITORS THAT ARE AMONG THE 10 WITH THE MOST EXCEEDANCES
C      OR THOSE 10 WITH HIGHEST DESIGN VALUES.
C
      DIMENSION NWHICH(55),DVAL(399),XUSE(35),YUSE(35),XG(399),
     1YG(399),BIGD(10),IBIGD(10),NBIGPT(10)
C
      MSTA=NSTA
      DO 10 I=1,55
      NWHICH(I)=-999
   10 CONTINUE
      INSTEP=0
C
C  LOOP TO 200 IDENTIFIES CELLS WITH MONITORS.  WHEN SEVERAL CELLS
C  CONTAIN MONITOR, THE SMALLEST (LARGEST INDEX) IS USED.
C
      DO 200 I=1,NPTS
      INSTEP=INSTEP+1
C
C  CHECK TO SEE IF GRID SPACING WAS CHANGED -- INDICATING A NEW PART
C  OF LIST WITH SMALLER CELLS.
C
      IF (INSTEP .NE. 1) GO TO 20
      ISP1=I+1
      SPACE=ABS(XG(I) - XG(ISP1))
      IF (SPACE .LE. 0.) SPACE=ABS(YG(I) - YG(ISP1))
   20 IP1=I+1
      IF (IP1 .GT. NPTS) GO TO 25
      IF (ABS(XG(I) -XG(IP1)) .EQ. SPACE) GO TO 25
      IF (ABS(YG(I) - YG(IP1)) .NE. SPACE) INSTEP=0
   25 DO 100 J=1,NSTA
```

```
C
C CHECK WHETHER MONITOR IS IGNORED (XUSE = -999.).
C
      IF(XUSE(J).EQ.-999.)GO TO 100
      HLFSP=0.5*SPACE
      DX=ABS(XUSE(J) - XG(I))
      DY=ABS(YUSE(J) - YG(I))
      IF (DX .GT. HLFSP) GO TO 100
      IF (DY .GT. HLFSP) GO TO 100
      NWHICH(J)=I
  100 CONTINUE
  200 CONTINUE
C
C  REPLACE DUPLICATE CELLS WITH -999
C
      DO 35 J=1,NSTA
      IF (NWHICH(J) .EQ. -999) GO TO 35
      IOUT=NWHICH(J)
      JP1=J+1
      DO 34 K=JP1,NSTA
      IF (NWHICH(K) .EQ. IOUT) NWHICH(K)=-999
   34 CONTINUE
   35 CONTINUE
C
C  MAKE LIST OF CELLS WITH STATIONS
C
      NUMKT=0
      DO 220 IND=1,NSTA
      IF (NWHICH(IND) .EQ. -999) GO TO 220
      NUMKT=NUMKT+1
      NWHICH(NUMKT)=NWHICH(IND)
  220 CONTINUE
      MSTA=NUMKT
      MANYPT=MSTA
C
C  GETTING CELLS WITH MOST EXCEEDANCES THAT DO NOT HAVE MONITORS.
C
      DO 300 I=1,10
      DO 250 J=1,MSTA
      IF (NBIGPT(I) .EQ. NWHICH(J)) GO TO 275
  250 CONTINUE
      MANYPT=MANYPT+1
      NWHICH(MANYPT)=NBIGPT(I)
  275 CONTINUE
  300 CONTINUE
C
C  GETTING CELLS WITH HIGHES DESIGN VALUES (FIRST SET ARRAYS=0)
C
      DO 325 I=1,10
      BIGD(I)=0.
      IBIGD(I)=0
  325 CONTINUE
      DO 400 II=1,NPTS
      IF (DVAL(II) .LE. BIGD(10)) GO TO 400
      DO 340 I=1,10
      INSERT=I
      IF (DVAL(II) .GT. BIGD(I)) GO TO 350
  340 CONTINUE
  350 INVINS=10-INSERT
      IF (INVINS .LE. 0) GO TO 380
      DO 370 I=1,INVINS
      ICOMP=11-I
      ICM1=10-I
      BIGD(ICOMP)=BIGD(ICM1)
      IBIGD(ICOMP)=IBIGD(ICM1)
  370 CONTINUE
  380 BIGD(INSERT)=DVAL(II)
      IBIGD(INSERT)=II
```

```
  400 CONTINUE
C
C  ADD CELLS THAT HAVE HIGHEST DESIGN VALUES AND ARE NOT ALREADY LISTED.
C
      NOWMNY=MANYPT
      DO 500 I=1,10
      DO 450 J=1,NOWMNY
      IF (IBIGD(I) .EQ. NWHICH(J)) GO TO 500
  450 CONTINUE
      MANYPT=MANYPT+1
      NWHICH(MANYPT)=IBIGD(I)
  500 CONTINUE
      RETURN
      END
```

SUBROUTINE COORXY

```
      SUBROUTINE COORXY (NX,NY,GSIZE,XORIG,YORIG,XG,YG,NPTS)
      DIMENSION XG(1),YG(1)
C
C  THIS SUBROUTINE CALCULATES THE COORDINATES OF THE MID POINTS
C  (XG,YG) OF AN NX BY NY GRID OF SQUARE GRID CELLS WITH DIMENSIONS
C  EQUAL GSIZE AND ORIGIN AT XORIG,YORIG.  NPTS IS THE INDEX OF EACH
C  POINT IN A LIST OF SUCH POINTS.
C
      DO 110 IX=1,NX
      X=IX-1
      DO 100 IY=1,NY
      Y=IY-1
      NPTS=NPTS+1
      XG(NPTS)=(X+0.5)*GSIZE+XORIG
      YG(NPTS)=(Y+0.5)*GSIZE+YORIG
  100 CONTINUE
  110 CONTINUE
      RETURN
      END
```

SUBROUTINE BIGGYS

```
      SUBROUTINE BIGGYS (NOOZ,NPTS,TOPOZ,NBIGPT,NTOP,BIGOZ)
      DIMENSION NOOZ(399),TOPOZ(10),NBIGPT(10),NTOP(10),BIGOZ
     1(10,399)
C
C  THIS SUBROUTINE KEEPS TRACK OF THE NUMBER OF EXCEEDANCES FOR
C  EACH GRID CELL, THE 10 GRID CELLS WITH THE MOST EXCEEDANCES
C  AND HIGHEST OZONE OBSERVED IN THOSE 10 CELLS.
C       NOOZ(NPTS)=NUMBER OF EXCEEDANCES IN NPTS GRID CELL
C       TOPOZ(I)= LARGEST OZONE VALUE IN GRID CELL WITH ITH MOST EXCEEDANCES
C       NBIGPT(I)=GRID CELL IN WHICH ITH MOST EXCEEDANCES OCCURRED
C       NTOP(J)=JTH MOST EXCEEDANCES (FOR ALL GRID CELLS)
C       BIGOZ(K,L)=KTH LARGEST OZONE OBSERVATION IN LTH GRID CELL
C
      IF (NOOZ(NPTS).LT.NTOP(10)) GO TO 86
      IF (NOOZ(NPTS).GT.NTOP(10)) GO TO 30
      IF (BIGOZ(1,NPTS).LE.TOPOZ(10)) GO TO 86
```

```
   30 DO 40 I=1,10
      INSERT=I
      IF (NOOZ(NPTS).GT.NTOP(I)) GO TO 50
      IF (NOOZ(NPTS).LT.NTOP(I)) GO TO 40
C
C  IF NO. OF EXCEEDANCES EQUAL, USE HIGHEST VALUE TO DECIDE
C
      IF (BIGOZ(1,NPTS).GT.TOPOZ(I)) GO TO 50
   40 CONTINUE
   50 INVINS=10-INSERT
      IF (INVINS.LE.0) GO TO 80
      DO 70 I=1,INVINS
      ICOMP=11-I
      ICM1=10-I
      NTOP(ICOMP)=NTOP(ICM1)
      NBIGPT(ICOMP)=NBIGPT(ICM1)
      TOPOZ(ICOMP)=TOPOZ(ICM1)
   70 CONTINUE
   80 NTOP(INSERT)=NOOZ(NPTS)
      NBIGPT(INSERT)=NPTS
      TOPOZ(INSERT)=BIGOZ(1,NPTS)
   86 CONTINUE
      RETURN
      END
```

SUBROUTINE BIG10

```
      SUBROUTINE BIG10 (BIGOZ,NPTS,BOZ,IDATE,KDATE)
      DIMENSION BIGOZ(10,399),KDATE(10,399)
C
C  THIS SUBROUTINE KEEPS TRACK OF THE TEN LARGEST OBSERVATIONS
C  FOR EACH GRID CELL AND CORRESPONDING DATES
C
      IF (BOZ.LE.BIGOZ(10,NPTS)) GO TO 86
      DO 40 I=1,10
      INSERT=I
      IF (BOZ.GT.BIGOZ(I,NPTS)) GO TO 50
   40 CONTINUE
   50 INVINS=10-INSERT
      IF (INVINS.LE.0) GO TO 80
      DO 70 I=1,INVINS
      ICOMP=11-I
      ICM1=10-I
      BIGOZ(ICOMP,NPTS)=BIGOZ(ICM1,NPTS)
      KDATE(ICOMP,NPTS)=KDATE(ICM1,NPTS)
   70 CONTINUE
   80 BIGOZ(INSERT,NPTS)=BOZ
      KDATE(INSERT,NPTS)=IDATE
   86 CONTINUE
      RETURN
      END
```

SUBROUTINE BOUNDS

```
      SUBROUTINE BOUNDS (MT,NT,YMAX,XMIN,YD,XD,Y,X)
C     SUBROUTINE BOUNDS--THIS ROUTINE MUST BE CALLED FIRST FOR THE
C     PURPOSE OF ESTABLISHING THE ANALYSIS REGIONS GRID NETWORK AND
C     BOUNDARIES.
C
C     CALL BOUNDS (MT,NT,YMAX,XMIN,YD,XD,Y,X)
C
C     MT, NT:        NUMBER OF ROWS AND COLUMNS
C     YMAX, XMIN:     LEFT TOP (OR NORTHWEST) CORNER GRID POINT
C     YD, XD:        Y AND X GRID SPACING INTERVALS
C     Y, X:          ARRAYS CONTAINING THE Y AND X LOCATIONS OF THE
C              COLUMNS AND ROWS (DIMENSION Y(2*MT), X(2*NT))
C
      DIMENSION Y(1),X(1)
      DO 10 M=1,MT
      Y1=1-M
      YM=YMAX+Y1*YD
      JM=M+MT
      Y(JM)=YM
   10 Y(M)=YM
      DO 20 N=1,NT
      X1=N-1
      XN=XMIN+X1*XD
      JN=N+NT
      X(JN)=XN
   20 X(N)=XN
      RETURN
      END
```

SUBROUTINE GRID

```
      SUBROUTINE GRID (HS,XS,YS,JT,GSIZE,XMIN,YMAX,NT,MT,H,U,V,KEY)
      DIMENSION HS(1),XS(1),YS(1),H(1),U(1),V(1),US(35),VS(35),X(30),
     1Y(30),IS(8000)
      DATA US/35*0.0/, VS/35*0.0/
      IF (KEY .NE. 0) GO TO 10
      XD=GSIZE
      YD=GSIZE
      IDS=0
      WT=1.0
      KSS5=5
      KS=JT
      UNL=900.0
      CALL BOUNDS (MT,NT,YMAX,XMIN,YD,XD,Y,X)
      CALL MESH (KS,KSS5,WT,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)
      KSS5=6
      CALL MES2 (KS,KSS5,WT,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)
      RETURN
   10 KQ=1
      CALL MESH (KQ,IDS,UNL,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)
      KQ=2
      CALL MESH (KQ,IDS,UNL,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)
      CALL MES2 (KQ,IDS,UNL,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,U,H,IS,KEY)
      RETURN
      END
```

SUBROUTINE CHOOSE

```
      SUBROUTINE CHOOSE (LL,BOZ,OZUSE,XUSE,YUSE,X,Y,GSIZE)
C
C  THIS SUBROUTINE DETERMINES IF AN OBSERVING SITE IS WITHIN THE
C  SQUARE FOR WHICH THE INTERPOLATED VALUE BOZ WAS DETERMINED AND
C  USES THE LARGER (OBSERVED OR INTERPOLATED) VALUE.
C
C
      DIMENSION OZUSE(1),XUSE(1),YUSE(1)
C
      GG = GSIZE*0.5
      DO 100 I=1,LL
      IF (OZUSE(I) .GT. 900.) GO TO 100
      IF (ABS(XUSE(I)-X) .GT. GG) GO TO 100
      IF (ABS(YUSE(I)-Y) .GT. GG) GO TO 100
      IF (OZUSE(I) .GT. BOZ) BOZ = OZUSE(I)
  100 CONTINUE
      RETURN
      END
```

SUBROUTINE MES2

```
      SUBROUTINE  MES2(KQ,ID,HN,JJ,M9,N9,YS,XS,VS,US,HS,YL,XL,V,U,H,
     1IS,KEY)
C
C  THIS SUBROUTINE CALCULATES VALUES AT A DISCRETE SET OF REGULARLY
C  SPACED NETWORK OF GRID POINTS BASED ON VALUES GIVEN FOR A SET OF
C  IRREGULARLY SPACED SET OF POINTS.  IT IS SIMILAR TO SUBROUTINE
C  MESH BUT CALCULATES GRID POINT VALUES BY A LEAST SQUARE FITTING
C  OF A SECOND DEGREE POLYNOMIAL RATHER THAN A FIRST DEGREE.  IT
C  REQUIRES THAT BEST GUESS BE INITIALLY PLACED IN GRID POINT ARRAYS
C  H, V AND U (V AN U ARE WIND COMPONENTS IF USED).
C
      REAL  LXS;LYS,LXX,LYY,JI1,JI3,JI4,JI5,JI6,NI3,NJ3,KI3,LI3,NK3;
     2      II3,MI3,LXY
      DIMENSION XL(1),YL(1),V(1),U(1),H(1)
      DIMENSION YS(1),XS(1),VS(1),US(1),HS(1),IS(1)
      DATA SPH1,W1/4.0,0.10/
C
C  IF KEY EQ 0 GO TO 200 TO SET BASIC CONSTANTS.
C  IF KEY NE 0 CALCULATE GRID POINT VALUES.
C
      IF (KEY .EQ. 0) GO TO 200
      IDS=ID
      HNL=HN
C  LOOP THROUGH ALL GRID POINTS.
      M=0
      L=0
      I=0
   75 M=M+1
      IF (M.GT.M9) GO TO 100
      N=0
      YLM=YL(M)
      CM=1.0
   80 N=N+1
      IF (N.GT.N9) GO TO 75
      L=L+1
```

```
C  INITIALIZE SUMMING VARIABLES BY SETTING THEM TO ZERO.
      K=0
      NOD=0
      XLN=XL(N)
      AI1=0.0
      AI2=0.0
      AI3=0.0
      AI4=0.0
      AI5=0.0
      AI6=0.0
      BI1=0.0
      BI2=0.0
      BI3=0.0
      BI4=0.0
      BI5=0.0
      BI6=0.0
      CI1=0.0
      CI2=0.0
      CI3=0.0
      CI4=0.0
      CI5=0.0
      CI6=0.0
      DI1=0.0
      DI2=0.0
      DI3=0.0
      DI4=0.0
      DI5=0.0
      DI6=0.0
      EI1=0.0
      EI2=0.0
      EI3=0.0
      EI4=0.0
      EI5=0.0
      EI6=0.0
      FI1=0.0
      FI2=0.0
      FI3=0.0
      FI4=0.0
      FI5=0.0
      FI6=0.0
      GI1=0.0
      GI2=0.0
      GI3=0.0
      GI4=0.0
      GI5=0.0
      GI6=0.0
      GJ1=0.0
      GJ2=0.0
      GJ3=0.0
      GJ4=0.0
      GJ5=0.0
      GJ6=0.0
      GK1=0.0
      GK2=0.0
      GK3=0.0
      GK4=0.0
      GK5=0.0
      GK6=0.0
C  INITIALLY GO TO 86
      GO TO 86
C  LOOP THROUGH K'S CLOSEST DATA TO GRID POINT.
 84   K=K+1
      IF (K.GT.KS) GO TO 90
      I=I+1
      IF (NOD.GT.KSS5)GO TO 84
      J=IS(I)
      IF (J.LE.0 ) GO TO 84
      HSJ=HS(J)
```

```
C  CHECK FOR MISSING POINTS.
      IF (ABS(HSJ).GT.HNL) GO TO 84
      NOD=NOD+1
      XSJ=XS(J)
      YSJ=YS(J)
      USJ=US(J)
      VSJ=VS(J)
      DYS=YSJ-YLM
      DXS=(XSJ-XLN)*CM
      DXX=DXS*DXS
      DYY=DYS*DYS
      DYS2=DYY+DXX
      IF (IDS.LE.0) GO TO  85
      USK=USJ+USJ
      VSK=VSJ+VSJ
      USK=USJ+U(L)
      VSK=VSJ+V(L)
   88 DXS1=USK*USK+VSK*VSK+0.01
      DXS2=(USK*DYS-VSK*DXS)
      DXS2=DXS2*DXS2/DXS1
      DYS2=DYS2+DXS2*SPH1
C  CALCULATE WEIGHING VALUE.
   85 W=C2/(DYS2+C2)
      GO TO 89
C  CALCULATE WEIGHING VALUE.
C  USE THE INITIAL GRID POINT VALUE H(L) AS THE FIRST DATUM FOR THE FITTING.
C  ALSO USE U(L) AND V(L) IF THEY ARE BEING ANALYZED.
 86   W=W1
      HSJ=H(L)
      VSJ=V(L)
      USJ=U(L)
      DXS=0.0
      DYS=0.0
      NOD=NOD+1
      DXX=0.0
      DYY=0.0
C  MAKE SUMMATION REQUIRED FOR LEAST SQUARES FITTING.
   89 HSJ=HSJ*W
      USJ=USJ*W
      VSJ=VSJ*W
      DXY=DXS*DYS
      LXS=DXS*W
      LYS=DYS*W
      LXX=LXS*DXS
      LYY=LYS*DYS
      LXY=LXS*DYS
      AI1=AI1+W
      FI1=FI1+LXS
      EI1=EI1+LYS
      FI2=FI2+LXX
      EI3=EI3+LYY
      GI1=GI1+HSJ
      GJ1=GJ1+USJ
      GK1=GK1+VSJ
      FI3=FI3+LXY
      GI2=GI2+HSJ*DXS
      GJ2=GJ2+USJ*DXS
      GK2=GK2+VSJ*DXS
      GI3=GI3+HSJ*DYS
      GJ3=GJ3+USJ*DYS
      GK3=GK3+VSJ*DYS
      GI4=GI4+HSJ*DXY
      GJ4=GJ4+USJ*DXY
      GK4=GK4+VSJ*DXY
      GI5=GI5+HSJ*DXX
      GJ5=GJ5+USJ*DXX
      GK5=GK5+VSJ*DXX
      GI6=GI6+HSJ*DYY
```

```
      GJ6=GJ6+USJ*DYY
      GK6=GK6+VSJ*DYY
      DI2=DI2+DXY*LXS
      CI2=CI2+DXX*LXS
      BI2=BI2+DXY*LYS
      BI3=BI3+DYY*LYS
      DI4=DI4+DXY*LXY
      CI4=CI4+LXY*DXX
      BI4=BI4+LXY*DYY
      CI5=CI5+DXX*LXX
      BI6=BI6+DYY*LYY
      GO TO 84
   90 CONTINUE
C  MAKE CALCULATIONS REQUIRED FOR LEAST SQUARES FITTING.
C  (IF NUMBER OF DATA GE 6)
      IF (NOD.LT.6) GO TO 80
      AI2=FI1
      AI3=EI1
      EI2=FI3
      AI4=FI3
      DI1=FI3
      AI6=EI3
      BI1=EI3
      EI5=DI2
      FI4=DI2
      CI3=DI2
      FI5=CI2
      FI6=BI2.
      EI4=BI2
      DI3=BI2
      EI6=BI3
      CI6=DI4
      BI5=DI4
      DI5=CI4
      DI6=BI4
      AI5=FI2
      CI1=FI2
      JI3=CI3*BI2-CI2*BI3
      JI1=CI1*BI2-CI2*BI1
      JI4=CI4*BI2-CI2*BI4
      JI5=CI5*BI2-CI2*BI5
      JI6=CI6*BI2-CI2*BI6
      BJ1=BI1*JI3
      BJ2=BI2*JI3
      BJ4=BI4*JI3
      BJ5=BI5*JI3
      BJ6=BI6*JI3
      NI3=GI3*BI2-GI2*BI3
      KI3=DI3*BI2-DI2*BI3
      LI3=EI3*BI2-EI2*BI3
      NJ3= GJ3*BI2-GJ2*BI3
      NK3=GK3*BI2-GK2*BI3
      II3=AI3*BI2-AI2*BI3
      MI3=FI3*BI2-FI2*BI3
      WI1= GI1*BJ2-GI2*BJ1-NI3*JI1
      WI4=GI4*BJ2-GI2*BJ4-NI3*JI4
      WJ1= GJ1*BJ2-GJ2*BJ1-NJ3*JI1
      WJ4=GJ4*BJ2-GJ2*BJ4-NJ3*JI4
      WK1= GK1*BJ2-GK2*BJ1-NK3*JI1
      WK4=GK4*BJ2-GK2*BJ4-NK3*JI4
      WI5= GI5*BJ2-GI2*BJ5-NI3*JI5
      WI6=GI6*BJ2-GI2*BJ6-NI3*JI6
      WJ5= GJ5*BJ2-GJ2*BJ5-NJ3*JI5
      WJ6=GJ6*BJ2-GJ2*BJ6-NJ3*JI6
      WK5= GK5*BJ2-GK2*BJ5-NK3*JI5
      WK6=GK6*BJ2-GK2*BJ6-NK3*JI6
      TI1= DI1*BJ2-DI2*BJ1-KI3*JI1
      TI4=DI4*BJ2-DI2*BJ4-KI3*JI4
```

```
      TI5= DI5*BJ2-DI2*BJ5-KI3*JI5
      TI6=DI6*BJ2-DI2*BJ6-KI3*JI6
      UI1= EI1*BJ2-EI2*BJ1-LI3*JI1
      UI4=EI4*BJ2-EI2*BJ4-LI3*JI4
      UI5= EI5*BJ2-EI2*BJ5-LI3*JI5
      UI6=EI6*BJ2-EI2*BJ6-LI3*JI6
      SI1= AI1*BJ2-AI2*BJ1-II3*JI1
      SI4=AI4*BJ2-AI2*BJ4-II3*JI4
      SI5= AI5*BJ2-AI2*BJ5-II3*JI5
      SI6=AI6*BJ2-AI2*BJ6-II3*JI6
      VI1= FI1*BJ2-FI2*BJ1-MI3*JI1
      VI4=FI4*BJ2-FI2*BJ4-MI3*JI4
      VI5= FI5*BJ2-FI2*BJ5-MI3*JI5
      VI6=FI6*BJ2-FI2*BJ6-MI3*JI6
      PI5=UI5*TI4-UI4*TI5
      PI1=UI1*TI4-UI4*TI1
      PI6=UI6*TI4-UI4*TI6
      TJ4=TI4*PI5
      TJ1=TI1*PI5
      TJ6=TI6*PI5
      RI5=WI5*TI4-WI4*TI5
      OI5=SI5*TI4-SI4*TI5
      QI5=VI5*TI4-VI4*TI5
      RJ5=WJ5*TI4-WJ4*TI5
      RK5=WK5*TI4-WK4*TI5
      ZI1=WI1*TJ4-WI4*TJ1-RI5*PI1
      ZI6=WI6*TJ4-WI4*TJ6-RI5*PI6
      ZJ1=WJ1*TJ4-WJ4*TJ1-RJ5*PI1
      ZJ6=WJ6*TJ4-WJ4*TJ6-RJ5*PI6
      ZK1=WK1*TJ4-WK4*TJ1-RK5*PI1
      ZK6=WK6*TJ4-WK4*TJ6-RK5*PI6
      XI1=SI1*TJ4-SI4*TJ1-OI5*PI1
      XI6=SI6*TJ4-SI4*TJ6-OI5*PI6
      YI1=VI1*TJ4-VI4*TJ1-QI5*PI1
      YI6=VI6*TJ4-VI4*TJ6-QI5*PI6
C  MULTIPLICATION FACTOR APPLIED TO KEEP FROM WORD OVERFLOW
C
      XI1=1E-36 * XI1
      XI6=1E-36 * XI6
      XIYI=XI1*YI6-XI6*YI1
      IF (XIYI.EQ.0.0) GO TO 80
      XIYI=1.0/XIYI
C  CALCULATE GRID POINT VALUES.
      ZI1=1E-36 * ZI1
      ZI6=1E-36 * ZI6
      HLL=(ZI1*YI6-ZI6*YI1)*XIYI
      ULL=(ZJ1*YI6-ZJ6*YI1)*XIYI
      VLL=(ZK1*YI6-ZK6*YI1)*XIYI
      H(L)=HLL
      U(L)=ULL
      V(L)=VLL
      GO TO 80
  100 CONTINUE
      GO TO 286
  200 CONTINUE
C  END OF CALCULATIONS.
C     ENTRY ME2SET
C  SET BASIC CONSTANTS.
      C2=HN
      KS=KQ
      KSS5=ID
  286 CONTINUE
      RETURN
      END
```

SUBROUTINE MESH

```
      SUBROUTINE MESH (KQ,IDS,HNIL,JJ,M9,N9,YS,XS,VS,US,HS,Y,X,V,U,H,
     1IS,KEY)
C     SUBROUTINE MESH--THIS SUBROUTINE CALCULATES VALUES AT A DISCRETE
C     SET OF REGULATORY SPACED NETWORK OF POINTS BASED ON VALUES
C     GIVEN FOR A SET OF IRREGULARLY SPACED POINTS.
C
C     CALL MESH (KQ,IDS,HNIL,JT,MT,NT,YS,XS,VS,US,HS,Y,X,V,
C         U, H, IS)
C
C        KQ:  (=1) AREA DATA WEIGHTING AVERAGING ANALYSIS (1ST GUESS ANALYSIS)
C             (=2) LEAST-SQUARES FITTING OF FIRST DEGREE POLYNOMIAL
C        IDS:  UP-DOWN STREAM WEIGHTING (IDS>0)
C        HNIL:  MAXIMUM ABSOLUTE VALUE FOR A STATION WIND COMPONENT
C           VALUE; IF GREATER VALUE IS NOT TO BE USED IN ANALYSIS
C        HS, US, VS:  STATION VALUES (IF WIND ANALYSIS, THEN US AND VS
C           ARE THE WIND COMPONENTS)
C        H, U, V:  GRID POINT VALUES (ONE DIMENSIONAL ARRAYS STARTING
C           AT LEFT TOP POINT AND THEN MOVING ACROSS TOP ROW FIRST).
C
C     THIS ROUTINE HAS AN ENTRY POINT MEHSET(KEY=0) THAT MUST INITIALLY BE
C     USED TO ESTABLISH VARIOUS FEATURES AND TO DETERMINE THE KS CLOSEST
C     STATIONS TO EACH GRID POINT (STORED IN ARRAY IS).  FOR EXAMPLE:
C
C     CALL MEHSET (KS, KSS5, UNIL, -----, IS)
C
C        KS:  NUMBER OF CLOSEST STATIONS TO GRID POINT TO BE STORED
C        KSS5:  MAXIMUM NUMBER OF STATIONS TO BE USED IN THE ANALYSIS
C           OF A GRID-POINT VALUE.
C        HNIL:  MAXIMUM ABSOLUTE VALUE FOR A STATION WIND COMPONENT
C           VALUE; IF GREATER VALUE IS NOT TO BE USED IN ANALYSIS
C        DYL   WEIGHTING CONSTANT.  PROGRAM ESTABLISHES VALUE IF DYL SET
C            EQUAL TO OR LESS THAN ZERO
C        IS:  INDICES FOR THE KS CLOSEST STATIONS TO EACH GRID POINT
C           ARE STORED IN THIS ARRAY (DIMENSION = KS.MT.NT)
      DIMENSION  DS(150),JS(150),IQ(150)
      DIMENSION YS(1),XS(1),VS(1),US(1),HS(1)
      DIMENSION Y(1),X(1),V(1),U(1),H(1),IS(1)
      DATA KSW,ALPH,WIN/4,2.5,0.05/
C
C  IF KEY EQ 0 GO TO 200 AND CALCULATE THE K'S CLOSEST DATA TO EACH GRID
C  POINT.  IF KEY NE 0 CALCULATE GRID POINT VALUES.
C
      IF (KEY .EQ. 0) GO TO 200
      KQ5=KQ-1+KSS5
      DCK=250.
C  LOOPS THROUGH ALL GRID POINTS.
      M=0
      L=0
      I=0
   75 M=M+1
      IF (M-M9) 77,77,100
   77 N=0
      YLM=Y(M)
   80 N=N+1
      IF (N-N9) 81,81,75
   81 L=L+1
C  INITIALIZE SUMMATION VARIABLES BY SETTING THEM TO ZERO.
      K=0
      NOD=0
      XLN=X(N)
      IF (KQ-1) 82,82.83
```

```
C  FOR KQ LT 2.
   82 H(L)=0.0
      V(L)=0.0
      U(L)=0.0
      NOD=0
      DNH=0.0
      DHH=0.0
      DUH=0.0
      DVH=0.0
      GO TO 84
C  FOR KQ GT 1.
   83 DXH=0.0
      DYH=0.0
      DXYH=0.0
      DXXH=0.0
      DYYH=0.0
      DXHH=0.0
      DXUH=0.0
      DXVH=0.0
      DYHH=0.0
      DYUH=0.0
      DYVH=0.0
C  FOR KQ GT 1 USE INITIAL GRID POINT VALUES IN H, U AND V FOR FIRST
C  DATA OF FITTING.
      W=WIN
      DNH=W
      NOD=1
      K=0
      DHH=H(L)*W
      DUH=U(L)*W
      DVH=V(L)*W
C  LOOP THROUGH K'S CLOSEST DATA TO GRID POINT..
   84 K=K+1
      IF (K-KS) 85,85,90
  384 I=I+KS-K
      GO TO 90
   85 I=I+1
      IF (NOD-KQ5) 86,384,384
   86 CONTINUE
      J=IS(I)
      IF (J) 84,84,87
   87 CONTINUE
C  CHECK FOR MISSING DATA.
      IF (ABS(HS(J)).GT.HNIL)  GO TO 84
      XSJ=XS(J)
      YSJ=YS(J)
      USJ=US(J)
      VSJ=VS(J)
      HSJ=HS(J)
      DYS=YSJ-YLM
      DXS=XLN-XSJ
      DYS2=DYS*DYS+DXS*DXS
      DXS2=0.5*DYS2
      IF (IDS .EQ. 0) GO TO 385
      USK=USJ
      VSK=VSJ
      IF (KQ .EQ. 1) GO TO 388
      USK=USJ+U(L)
      VSK=VSJ+V(L)
  388 DXS1=USK*USK+VSK*VSK+0.01
      DXS2=(USK*DYS-VSK*DXS)
      DXS2=DXS2*DXS2/DXS1
C  CALCULATE WEIGHTING CONSTANT.
  385 W=DYL/(DYS2+DXS2*ALPH+DYL)
      NOD=NOD+1
      DNH=DNH+W
      HSJ=HSJ*W
      USJ=USJ*W
      VSJ=VSJ*W
```

```
C  MAKE SUMMATIONS REQUIRED FOR DATA WEIGHTED AVERAGING ANALYSIS.
      DHH=DHH+HSJ
      DUH=DUH+USJ
      DVH=DVH+VSJ
C  MAKE SUMMATIONS REQUIRED FOR LEAST SQUARES ANALYSIS.
      IF (KQ-1) 89,84,89
   89 DYH=DYH+DYS*W
      DXH=DXH+DXS*W
      DXYH=DXYH+DXS*DYS*W
      DXXH=DXXH+DXS*DXS*W
      DYYH=DYYH+DYS*DYS*W
      DXHH=DXHH+HSJ*DXS
      DYHH=DYHH+HSJ*DYS
      DXUH=DXUH+USJ*DXS
      DYUH=DYUH+USJ*DYS
      DXVH=DXVH+VSJ*DXS
      DYVH=DYVH+VSJ*DYS
      GO TO 84
   90 CONTINUE
      IF (KQ-1) 94,91,94
   91 IF (NOD-1) 80,92,92
   92 IF (DNH) 80,80,93
C  CALCULATE GRID POINT VALUES FOR DATA WEIGHTED AVERAGING ANALYSIS.
   93 DNH=1.0/DNH
      H(L)=DHH*DNH
      U(L)=DUH*DNH
      V(L)=DVH*DNH
      GO TO 80
   94 IF (NOD-3) 80,80,95
C  MAKE CALCULATIONS REQUIRED FOR LEAST SQUARES ANALYSIS.
   95 D=DYH*DYH-DNH*DYYH
      E=DXYH*DYH-DXH*DYYH
      A=DXH*DYH-DNH*DXYH
      B=DXXH*DYH-DXH*DXYH
      BDAE=B*D-A*E
      IF (BDAE) 97,80,97
   97 BI=1.0/BDAE
      C=DXHH*DYH-DHH*DXYH
      F=DYHH*DYH-DHH*DYYH
      H(L) = (B*F-C*E)*BI
      CU=DXUH*DYH-DUH*DXYH
      FU=DYUH*DYH-DUH*DYYH
      CV=DXVH*DYH-DVH*DXYH
      FV=DYVH*DYH-DVH*DYYH
C  CALCULATE GRID POINT VALUES FOR LEAST SQUARE ANALYSIS.
      U(L)=(B*FU-CU*E)*BI
      V(L)=(B*FV-CV*E)*BI
      GO TO 80
  100 CONTINUE
      GO TO 60
  200 CONTINUE
C  END OF CALCULATIONS.
C
C  THE FOLLOWING PART OF THE PROGRAM IS USED TO DETERMINE THE STATIONS
C  CLOSEST TO EACH GRID POINT, AND FOR SETTING BASIC CONSTANTS.
C
C      ENTRY MEHSET
C  SET SOME BASIC CONSTANTS.
      DCK=250.
      D9=M9*N9
      DYL=DCK*0.25/D9
      IF (HNIL .GT. 0.0) DYL=HNIL
      KS=KQ
      KSS5=IDS
      DLCK=DCK*DCK
      JCT=150
      JS(JCT)=-1
      DS(JCT)=DLCK
```

```
C  LOOP THROUGH ALL GRID POINTS.
      M=0
      L=0
      I=0
   20 M=M+1
      N=0
      YLM=Y(M)
   25 N=N+1
      L=L+1
      J=0
      JT=JJ
      XLN=X(N)
      XLCK=DLCK
C  DETERMINE DISTANCE OF ALL DATA TO GRID POINT DATA.
   29 J=0
      JC=0
   30 J=J+1
      DYS=YS(J)-YLM
      DXS=XLN-XS(J)
      DLS=DYS*DYS+DXS*DXS
      IF (DLS-XLCK)  32,33,33
   32 JC=JC+1
      IF (JC .LT. JCT) GO TO 31
      XLCK=XLCK*0.75
      GO TO 29
   31 DS(JC)=DLS
      JS(JC)=J
      IYS=0
      IXS=1
      IF (DXS .LT. 0.0) IXS=2
      IF (DYS .LT. 0.0) IYS=2
      IQ(JC)=IYS+IXS
   33 IF (J-JT) 30,34,34
C
C  SELECT OUT INDICES OF K'S CLOSEST DATA, HOWEVER, AT LEAST ONE DATA
C  IS SELECTED FROM EACH ANGULAR QUADRANT - INDICES ARE STORED IN
C  ARRAY IS.
C
   34 KP=0
      K=0
      JQ=1
      ISW=1
      JX=JC
      IF (KSW.LT.1)  ISW=2
   35 K=K+1
      SIL=DLCK
      J=0
      I=I+1
      KP=KP+1
      JC=JCT
   40 J=J+1
      IF (J .LE. JX) GO TO (41,42) ISW
      GO TO 45
   41 IF (IQ(J).NE.JQ) GO TO 40
   42 IF (DS(J)-SIL)  44,40,40
   44 SIL=DS(J)
      JC=J
      GO TO 40
   45 DS(JC)=DLCK
      IS(I)=JS(JC)
      JQ=JQ+1
      IF (JQ .GT. 4) JQ=1
      IF (KP .EQ. KSW) ISW=2
      IF (JC .LT. JCT) GO TO 49
      IF (KP .GT. KSW) GO TO 49
      K=K-1
      I=I-1
      GO TO 35
```

```
   49 IF (K-KS) 35,50,50
   50 IF (N-N9) 25,55,55
   55 IF (M-M9) 20,60,60
   60 CONTINUE
C  END OF SEARCH FOR K'S CLOSEST DATA.
      RETURN
      END
```

SUBROUTINE NEXTD

```
      SUBROUTINE NEXTD (NOOZ,BIGOZ,NDIV,NSUBD,MANY,NPTS,NTOP,NBIGPT,
     1ISTART)
      DIMENSION NOOZ(399),BIGOZ(10,399),NDIV(10),NBIG(5),NTOP
     1(10),BIGEST(5),NBIGPT(10)
C
C  THIS SUBROUTINE IDENTIFIES CELLS TO BE SUBDIVIDED (NDIV).  THE CELLS
C  WITH THE MOST EXCEEDANCES AND THE J CELLS WITH THE HIGHEST
C  OBSERVED VALUES ARE SELECTED.  DUPLICATIONS ARE ELIMINATED.
C  NTOP(I) = ITH MOST EXCEEDANCES REGARDLESS OF CELL
C  NBIGPT(I) = ID OF CELL WITH ITH MOST EXCEEDANCES
C  BIGEST(J) = JTH LARGEST MAX O3
C  NBIG(J) = ID OF CELL WITH JTH LARGEST MAX O3
C
      DATA IFLAG /0/
      KSTART=1+ISTART
      XDIV=NSUBD
      HALF=0.5*XDIV
      MBIG=HALF
      MEX=NSUBD-MBIG
      DO 25 I=1,5
      IF (IFLAG.GT.0) GO TO 20
      BIGEST(I)=0.0
 20   NBIG(I)=0
 25   CONTINUE
      IFLAG=1
C
C  IDENTIFY CELLS WITH HIGHEST MAX OZONE AND REVISE LIST WHERE REQUIRED
C
      DO 100 I=KSTART,NPTS
      IF (BIGOZ(1,I).LT.BIGEST(5)) GO TO 100
C
C  IDENTIFY CELLS WITH LARGEST VALUES AND REVISE LIST
C
      DO 60 J=1,5
      INSERT=J
      IF (BIGOZ(1,I).GT.BIGEST(J)) GO TO 65
 60   CONTINUE
 65   INVINS=5-INSERT
      IF (INVINS.LE.0) GO TO 75
      DO 70 J=1,INVINS
      ICOMP=6-J
      ICM1=5-J
      BIGEST(ICOMP)=BIGEST(ICM1)
      NBIG(ICOMP)=NBIG(ICM1)
 70   CONTINUE
 75   BIGEST(INSERT)=BIGOZ(1,I)
      NBIG(INSERT)=I
 100  CONTINUE
      MANY=0
C
C  EXCISE DUPLICATES
```

```
C
      DO 150 I=1,MEX
      DO 150 J=1,MBIG
      IF (NBIGPT(I).EQ.NBIG(J)) NBIG(J)=0
 150  CONTINUE
      DO 175 I=1,MEX
      IF (NTOP(I) .LE. 0) GO TO 175
      MANY=MANY+1
      NDIV(MANY)=NBIGPT(I)
 175  CONTINUE
      DO 180 I=1,MBIG
      IF (NBIG(I).EQ.0) GO TO 180
      MANY=MANY+1
      NDIV(MANY)=NBIG(I)
 180  CONTINUE
      RETURN
      END
```

SUBROUTINE USES

```
      SUBROUTINE USES (NSTA,OZM,LL,BAD,ILK,OZUSE,XUSE,YUSE)
C
C          THIS SUBROUTINE SELECTS DATA TO BE USED.  IT IGNORES BAD DATA
C     (I.E. BAD) AND SELECTED SITES DEFINED BY ILK BY SETTING OZUSE=999..
C
      DIMENSION OZUSE(1),XUSE(1),YUSE(1),ILK(1),OZM(1)
      LL=0
      IGNOR=1
      DO 110 I=1,NSTA
C
C  CHECK FOR STATION TO BE IGNORED
C
      IF (I.EQ.ILK(IGNOR)) GO TO 100
C
C  CHECK TO SEE IF DATA VALID
C
      IF(OZM(I) .EQ. BAD) GO TO 105
C
C  CONVERTING DATA IN PPM+100*(NO. OF HRS) TO PPB
C
      XOZ=ABS(OZM(I))
C     OZUSE(I)=BMOD(XOZ)
      OZUSE(I)=1000.0*AMOD(XOZ,1.)
C
C CHECK FOR FEWER THAN 9 HOURS BETWEEN 0900 AND 2100
C WITHOUT AN OBSERVED EXCEEDANCE.
C
      IF(XOZ .LT. 900. AND. OZUSE(I). LE. 120.)GO TO 105
       LL=LL+1
      GO TO 110
 100  IGNOR=IGNOR+1
      XUSE(I)=-999.
      YUSE(I)=-999.
  105 OZUSE(I)=999.
 110  CONTINUE
   10 FORMAT(1H ,10F12.4)
   11 FORMAT(1H ,10F7.2)
  115 CONTINUE
      RETURN
      END
```

FUNCTION BMOD

```
      FUNCTION BMOD(OZ)
C
C  THIS FUNCTION RETURNS DECIMAL PORTION OF A FLOATING POINT NUMBER.
C
      IOZ=OZ
      IOZ=IOZ*1000
      TESTOZ=IOZ
      OZ=OZ*1000.
      BMOD=OZ-TESTOZ
      RETURN
      END
```

FUNCTION DSNVAL

```
      FUNCTION DSNVAL (BIGOZ,I,NDAYS)
C
C          CALCULATE THE DESIGN VALUES USING LINEAR FIT TO LOGS OF THE
C     LARGEST VALUES VERSUS CORRESPONDING PERCENTILES.
C
      DIMENSION BIGOZ(10,399)
      XDAYS=NDAYS
      YDAYS=0.0
      SUMX=0.0
      SUMY=0.0
      SUMXY=0.0
      SUMX2=0.0
      DO 100 K=1,10
      IF (BIGOZ(K,I).LE.0.0) GO TO 100
      XN=K*100
C          TRANSFORM THE OBSERVATIONS TO LOG.
      Y=ALOG10(BIGOZ(K,I))
      YDAYS=YDAYS+1.0
      X=XN/XDAYS
      SUMX=SUMX+X
      SUMY=SUMY+Y
      SUMXY=SUMXY+X*Y
      SUMX2=SUMX2+X*X
 100  CONTINUE
      IF (YDAYS.LE.0.0) GO TO 86
      XB=SUMX/YDAYS
      YB=SUMY/YDAYS
      B=(SUMXY-(YDAYS*XB*YB))/(SUMX2-(YDAYS*XB*XB))
      DSNVAL=YB+(B*(0.274-XB))
      DSNVAL = 10.**DSNVAL
      GO TO 30
 86   DSNVAL=-99.0
 30   CONTINUE
      RETURN
      END
```

APPENDIX D

USER'S GUIDE FOR THE PROBABILISTIC METHODOLOGY COMPUTER PROGRAM

OVERVIEW

The probabilistic modeling and simulation program was developed as a technique to estimate design values and exceedances incorporating both temporal and spatial data. Its input consists of the data written by the isopleth modeling program to Logical Unit 10. As output, it prints the results of the distribution fit (the lognormal mean vector and covariance matrix) as well as the results of the simulations. It also writes the information to Logical Unit 7 in a form usable by the analysis program, which has as its functions the calculation of the design values for each cell in each simulated network, the design value for each simulated network, mean and standard deviation as well as range for the design values, and generation of a printer plot of the design values for each simulated network.

One-pass computational algorithms are used wherever possible to minimize computing cycles. To further economize on core, the scratch work area is shared by various subroutines and symmetric matrices (e.g., variance/covariance, cross products, Cholesky decompositions) so that redundant entries are omitted.

As presently configured, the program assumes that the work sets, when distilled, are no larger than 50 cells x 200 days. Storage for various intermediate results is recycled, and dynamic subscript computation has been used in the scratch work area. Symmetric matrix storage and manipulation techniques have been used extensively to reduce memory requirements and improve the execution speed. Because the program coding uses structured, modular techniques, most of the subroutines perform a single, clearly defined function. It has been possible to write most of these functional subroutines using about 60 or fewer lines of code. Wherever possible, the matrix results obtained with the routines developed on this project have been verified against those obtained from other sources [e.g., PROC MATRIX in the Statistical Analysis System (16).

All files communicated are written in coded (character) rather than binary mode so that different programs in the series can be run on different computers. Some installations may wish to change this feature in the interest of quicker input/output.

The section below discusses the inputs and outputs for the modeling and simulation program, and describes how the program can be used for generating the variance/covariance matrix and mean vector without generating any simulations. Use of the network analysis program for analyzing the results of the modeling/simulation program or for analyzing "real" networks is also described; a later section describes the program and subroutines. The final section, included for users who

implement the program on other hardware or who wish to modify the program, covers important common blocks, arrays, argument lists, and so on, as well as details on scratch area indexing schemes, symmetric matrix manipulation routines, and algorithms.

A listing of this program is included at the end of this appendix. The programs are written in FORTRAN. In addition to an ANSI FORTRAN 66 compiler, the programs also require the IMSL library,* Edition 7 or 8. All calculations are done in single precision, a feature that will probably require modifications if the programs are to be run on machines with a word-size smaller than that of SRI's CDC 6400 computer.

PROGRAM INPUT

The modeling and simulation program requires data for a spatial network over a period of time and an optional processing parameter card. Although reference is made to card images, they need not, of course, be physical cards. The network data are read from Logical Unit 3, and the optional parameter card from Unit 5 (default input). Report output is written to Unit 6 (default output), and raw output for use by the network analysis or other programs is written to Unit 7.

---

*International Mathematical and Statistical Libraries, Inc., Houston, Texas.

Parameter Card

The parameter card is read from unit 5, the default input unit. The parameter card format is summarized in Table D-1.

Network Input

The network input is the format written to Logical Unit 10 by the isopleth modeling program. Other networks may be analyzed by transforming them to the form shown in Table D-2.

The daily values are arranged in order of cell indexes. For ease of conditional distribution derivation, the data in Table D-2 are arranged as follows:

INDX(J)--becomes the indices in the cells of the analysis group (cells among the 10 with highest exceedances or design values and cells with stations). INDX(J) is ISCRAT[a] arranged so that the cells without stations come first.

X(I,J)--becomes the log daily-value vector in the analysis group for the Ith day. Column J in this matrix corresponds to values for original cell INDX(J), so the cells with stations form the last NG2 columns of X.

JDAY(I) = IDAY--Julian date for Ith day read.

Table D-1

PARAMETER CARD FORMAT

Format	Columns	Variable Name	Units	Typical Value	Remarks
I3[a]	1-3	METH	—	-1	Determines distribution fit method used: < 0, simple estimate only; = 0 (or blank--default), both simple and E-M methods; > 0, E-M methods only
I3[a]	4-6	NSIM	--	10	Determines number of simulations: < 0, none; = 0 (or blank--default) 20 simulations; > 0, NSIM simulations
I3[a]	7-9	METHEM	—	0	Determines initial E-M estimate method: ≠ 0, uses sufficient statistics for truncated data set; = 0 (or blank--default), uses simple estimate to initialize, forces METH = 0
I3[a]	10-12	ITER	--	8	Number of iterations for E-M algorithm: = 0 (or blank--default), 10 iterations; > 0, ITER iterations (ignored when METH < 0)
I3[a]	13-15	METHMU	—	0	Determines mean and standard deviation method: = 0, use sample mean and standard deviations in simple method; = 1, use extrapolated mean and standard deviation method (see Chapter Two) in simple correlation procedure
F110.0[b]	16-25	THRESH	ppb	120.	Threshold value for exceedances: ≤ 0 (or blank--default); 120 ppb; > 0, THRESH
F10.0[b]	26-35	TRUNC	ppb	80.	Data truncation limit. To be used if only data for which some value greater than TRUNC were observed or interpolated. Warning: E-M will not work properly if the data are truncated at a value other than TRUNC. TRUNC ≤ 0, METH = -1 forced. There is no default TRUNC. Typical TRUNC value is 80.0.

[a] Right justify.

[b] Include decimal.

Table D-2

INPUT FORMAT FOR ANALYZING NETWORKS

Record	Format	Variable	Description
1	7I10	NTOTD	Total number of days
		INITC	Initial number of cells
		NDAYS	Number of days remaining after truncation performed
		NTOTC	Final number of cells after smoothing
		IHALF	NTOTC/2 for formatting (truncated if not integer)
		NG2	Number of cells with monitoring stations
		NCELLS	Number of cells with stations plus number of cells without stations but with one of the 10 highest number of exceedances or design values
2	16I5	ISCRAT(I)	Indexes of cells in analysis group (with 10 highest exceedances or design values)
3[a]	10F8.2		X,Y coordinates of cells 1 to IHALF first, then IHALF + 1 to NTOTC
4[a]	(6X,I6,6X,(10F10.6))	IDAY	Julian date and daily values for cells 1 to NTOTC in ascending order of index. There are NDAYS records of this type.

[a] Denotes multiple card images.

PROGRAM OUTPUT

Program output consists of report output with informative headings and output suitable for analysis by the network analysis or other program.

Report Output (written to Logical Unit 6)

The first page of the report output consists of an informative summary of the input data. At the top of the page, the program prints the number of days (after initial truncation but before E-M truncation--see below), the number of days after E-M truncation, the number of cells in the analysis group, the number of analyzed cells in group 1 (no stations), the number in group 2 (with monitoring stations), the number of days before initial truncation, the number of grid cells before smoothing, and the final number of grid cells after smoothing. Then for each cell in the analysis group, the program prints the cell index, the cell coordinates, the number of exceedances, and the number of values over 100 ppb. The order in which the information for each cell is printed reflects the order in which all subsequent program output for the cells is printed.

Simple Parameter Estimation Report Output

Output for this stage of the program, if requested, consists of mean estimates and a variance covariance matrix. The estimates given are for the natural logs of the data values. The output is given in the order of the cells as printed on the informative output page, which will not generally be the original order of the cells. Since the covariance matrix is symmetric, only entries on or below the main diagonal are printed. The Jth mean printed is thus the log mean of cell INDX(J), and the I,J element of the matrix is the covariance of cells INDX(I) and INDX(J).

E-M Parameter Estimation Report Output

For each iteration, the E-M estimation procedure prints the current estimates of the mean vector and variance/covariance matrix with figures for the cells printed in the order in which they appear on the informative output page. As with the simple-estimation method, only the lower half of the variance/covariance matrix is printed. For all iterations but the last, the parameters t*, $P(y \leq t^*)$ where y (0,1), and E_1 and E_2 are printed. Refer to Chapter 2 of the text of this report for an explanation of these parameters.

Simulation Output

Although this output is discussed here for convenience, it is printed immediately after either the simple- or the E-M estimation output, depending on the parameter card. The first item of the simulation output is the original covariance matrix, that is, the matrix as estimated by either method. The final covariance matrix, derived from the original by techniques discussed in Chapter 2 of this report, is the variance/covariance matrix of the conditional distribution. After this, for each simulation, the program prints the simulation number, the cell indexes, and for each day the Julian day and the values. The values in the simulation output for the cells with monitoring stations will be the same as the values on these days derived by the isopleth program. Simulated values are generated for the cells without stations.

Output Designed for Input to Other Programs

This output is written in character mode to Logical Unit 7 without headers or spacing for aesthetics. It consists of the following information:

- General information
- Cell indexes
- Block(s) of networks, each containing several networks
- Trailer record.

The general information and cell index vector contain:

- General information (7I10)--NTOTD, INITC, NDAYS, NTOT, IHALF, NG2, NCELLS (as under program input above.
- Cell indexes (16I5)--Indexes of network cells (ordered).

For each block, there is a block header record stating the number of networks in the block (I3 format). For each network within a block, there is a network header record containing the network number, number of days, and number of cells [in (313) format]. Then follows one record for each day consisting of the Julian day and the cell values, in the format

6X, I6, 6X (10 I 10.6).

A trailer record, which follows the last block, consists of "-1" in columns 2-3; schematically,

```
General (7I10)
Index (16I5)
For each block:
     Block header (I3)
          For each network:
               Network header (3I3)
               Values (6X,I6,6X,(10F10,6))
Trailer (I3)
```

NETWORK ANALYSIS PROGRAM

General

The network analysis program was designed to analyze the results of simulation portion of the modeling/simulation program. As input, it expects a general information record, a cell index record, and a variable number of blocks of networks. For all blocks, the networks must consist of the same cells. Each block has a record indicating the number of networks in the block at its start, which may vary from block to block. The number of days may vary from network to network. (For the format of the input to the analysis program, refer to preceding section.)

The analysis program provides output both for each network and for each block of networks. The output for each network consists of the top 10 values for each cell in the network, the design value for each cell, the number of exceedances in each cell (values over 120 ppb),* descriptive statistics, and a scatterplot produced for the network design values.

MAIN PROGRAM AND SUBROUTINE DESCRIPTION

The program has several common blocks for storing both the daily values and important cell statistics. There is also a common scratch work area that is used and reused by various routines.

When possible, matrices are stored in symmetric storage mode. For example, if A is a symmetric matrix, and B is a representation of A in symmetric storage mode, then:

$$A(I,J) = B(I * (I-1)/2+J), \text{ when } J \leq I \quad ,$$

or

$$= B(J * (J-1)/2+I), \text{ when } J > I \quad .$$

---

*A network design value.

This yields a storage saving of nearly 50 percent on large matrices; symmetric structure can often be used to bring about significant computational savings as well. Finally, this storage mode is compatible with that of IMSL routines for symmetric matrices.

The program estimates a variance/covariance matrix by either the simple or the E-M technique, then passes these results to the simulation routines, which have the capability to compute a variance/covariance matrix from a correlation matrix if this is required.

In the development of the program, some routines were developed and tested and then replaced by a faster or more appropriate algorithm. The routines are left in the code so that the other algorithms may be tried with minor modification, at the option of the user.

Common Blocks

/LIMITS/--Contains the exceedance threshold and the truncation value (actually, the natural logarithms of these).

/UNITS/--Contains logical unit numbers for data input, report output, simulation output, and parameter card input. Changing these in the main program effects a global change.

/DATA/--Contains logs of daily values of cells in the analysis group. Column I contains values for cell INDX(I), where INDX(I) is the vector of cell indexes (see /COUNTS/ below).

/STATS/--Contains the mean and standard deviation vectors (logs).

/ANCOV/--Variance/covariance and cross products in symmetric storage mode.

/SUMS/--Vector of sums of the cell values.

/COUNTS/--Number of exceedances per cell, days over 100 ppb per cell, number of cells in the analysis group, number in the analysis group without stations (N1), number in the analysis group with stations (N2), total number of cells, total number of days, final number of days (after E-M truncation), initial number of cells, cell indexes, and vector INSET (see routines READ1 and COMPRS for use of INSET).

/PCTILE/--Not used.

/RAND/--Used for random number generation.

/WORK/--Scratch work area. Any routine may use it, but work area should always be initialized before use, because it is frequently recycled. If it is necessary to increase problem dimensions, scratch area should be augmented to

$$N1^2 + 2 * N1 * N2 + N2^2 + [MAX\ (N1,N2)]^2 \quad ,$$

where N1 and N2 are as defined in /COUNTS/.

Descriptions of Routines

It has been the design philosophy to construct the network analysis program with a few modules of moderate length (2 to 3 pages including comments) whose function is to coordinate a major block of activity, each of which calls small subroutines (each written to perform one specialized function). The entire code is documented with comments. This approach has produced a program that is short and readable (about 1100 lines with average module size about 40 lines), in spite of the moderate complexity of its storage methods, indexing techniques, and computational algorithms. The modularity should make it possible to modify the program with minimal "ripple effect," that is, minimum generation of errors in other parts of the code. The main program and all the subroutines are briefly described below. The listing is given at the end of this appendix.

MONTE--The main program. Its main function is calling other modules. MONTE first defines the values for input and output units, then calls CRACK to read the parameter card, followed by READ1 to read the data. Based on the parameters, MONTE may then call the simple covariance estimation routine NAIVES, the simulation routine SIMULA, the E-M-estimation routine EMDRIV, and the simulation routine SIMULA again. It also calls utility routines COMPRS, CSEED, INFO, NEWSUM, and WRSYM.

EXTRAP--If METHMU=1, computes mean and standard deviation by extrapolation from the regression line of the percentiles as a function of the logs of the values. If this is selected, the covariance matrix is computed in the standard manner then transformed as follows:

$$\Sigma^* = (\sigma^*)^T \left(\frac{1}{\sigma}\right)^T (\Sigma) \left(\frac{1}{\sigma}\right) \sigma^*$$

where σ^* is the new covariance matrix, σ^* is the vector of new standard deviation estimates, $1/\sigma$ is the vector of sample standard deviation estimates, Σ is the sample variance covariance matrix, and $()^T$ denotes vector transpose.

MUVAR--Used by EXTRAP for computation of the extrapolated means and standard deviations.

CRACK--Reads the parameter card if present. If it is not, CRACK sets defaults according to the parameter function discused above. Note that CRACK has a machine-dependent end-of-file check, isolated for easy modification.

WRSYM (A,NA)--This subroutine writes symmetrically stored matrix A of order NA in lower diagonal form.

READ1--Reads the input data in character (formatted) mode from the input unit. The number of cells and cell indexes are saved in /COUNTS/. The x,y coordinates are saved in SCRAT for printing by INFO. Data are read into SCRAT and reordered before storage in final form.

The original order reflects simply the order in which the cells are numbered. The final order reflects first the cells with stations, then the cells without stations but among those with the ten highest exceedances or design values. Cross products and sums (of the logs) are computed on the fly. Vector INSET is also initialized in READ1 as follows: INSET(K) = I when row I of the original data array will form row $K(\leq I)$ of the final, E-M truncated data. The ghost sum vectors (GCROSS, GSUM) and pointers (IGJ, IGK, IGH5) were used for an ad-hoc initial E-M estimate reflecting data values along the truncation axis. These lines are now converted to comments.

CSEED (DSEED)--This subroutine returns a double-precision, random-number generator seed based on system-clock time. Function SECOND is machine dependent.

SIMULA (NSIM, NDSIM)--NSIM is the number of simultions, NDSIM is the number of days in the simulation. SIMULA is the main control module of the simulation procedure. First, SIMULA computes the pointers to various parts of the work area. If we partition the covariance matrix C (referred to in SIMULA as SIGMA) as described in Chapter 2 of the report into the components C_{11}, C_{22}, C_{12}, and C_{12}^T, we compute starting indices for these matrices in SIMULA as follows:

- I_{11}: Start of C_{11} matrix (later, final conditional covariance matrix)
- I_{12}: Start of C_{12}
- I_{12}^T : Start of C_{12}^T
- I_{22}: Start of C_{22} (later C_{22}^{-1})
- ITERM: Start of $C_{12}\ (C_{22})^{-1}$ matrix
- IWK: More work space.

Simula calls routine SPLIT to partition the matrix as required. IMSL routine LINVIF computes (C_{22}^{-1}). CSIGMA computes the final covariance matrix. For each day of simulated values, CMEAN computes the mean vector, and IMSL routine GGNSM generates a random vector with the same covariance structure. SIMULA adds this to the mean vectors, takes antilogs, and writes the results to units IOUT and ISIM.

COV--This routine will compute a covariance matrix from a correlation matrix given a vector of standard deviations. It is not used in the current version of the program, but is included to provide the user with another option.

CSIGMA--Computes the final covariance matrix (of the conditional distribution) given the components of the covariance matrix as returned by SPLIT. Covariance matrix = $C_{11}-C_{12}(C_{22}^{-1})C_{12T}$. The IMSL routine

VMULFF is used for matrix multiplication of $C_{12}(C_{22}^{-1})$, which is saved in SCRAT (ITERM) for use by CMEAN.

CMEAN--Computes the daily mean vector. It uses SCRAT (ITERM) as computed in CSIGMA. If M is a mean vector for cells without stations and M_2 for cells with, the final mean vector is:

$$M = M_1 + C_{12}\ (C_{22})^{-1}\ (X_2 - M_2)$$

where X_2 is a daily vector of observed values at sites with stations.

SPLIT--Takes a square, symmetrically stored matrix and, given N1 and N2, splits it into four components, N1 x N1, N1 x N2, N2 x N1, and N2 x N2 where the first two components are the first N1 rows. The storage is compatible with ANSI two-dimensional arrays of these dimensions. [If A is dimensioned N x m, and we want entry (i,j), we need location i + m x (j - 1) in A.]

INFO--Provides informative listing of counts for the entire network (for example, number of days, number of cells) as well as information for each cell (such as x,y coordinates, number of exceedances).

EMDRIV--Main driven routine for the EM algorithm. As an option, EMDRIV (based on METHEM) calls on routine EMO to use the sufficient statistics for the truncated data set as starting values. Otherwise, simple estimates for the parameters are used. For each interaction, EMDRIV calls LIDECP for a Cholesky decomposition of the covariance matrix. With this result, subroutine ET computes

vector XNU, conditional expectations E and E_2, and normalized truncation TSTAR. Then EMITER is called to perform an iteration.

EMO--Computes initial estimates for the E-M procedure based on the sufficient statistics for the truncated data set. If SUMX is the vector of sums and CROSSP is the cross-product matrix, these are

$$XMU_x(I) = SUMX(I)\ /\ (N-NO)$$

where N = initial number of observations (days) and NO is the number of observations truncated, and

$$SIGMA(I,J) = CROSSP(I,J)\ /\ (N-NO)\ XMU * XMU^T$$

where the superscript (T) denotes transpose. Actual computation and storage of SIGMA and CROSSP are symmetric storage mode.

EMITER--Performs an interation of the E-M algorithm. E1, E2, and XNU are computed in ET earlier and are passed as arguments. Given these, EMITER calls XXT to compute $XNU * XNU^T$, and then EMITER computes $(E2 - 1.0)(XNU * XNU^T) + I$, where I is an identity matrix of appropriate dimension. This result is stored in WORK2. We then use routine ABATT to compute matrix SIG^T, where SIG is the lower diagonal portion of the Cholesky decomposition of SIGMA, passed as a formal parameter. Updated mean estimates are produced as follows:

$$XMI = (SUMX(I) + NO * (XMU(I9 + SIG*XNU * E1)/N \quad .$$

(This is the new mean estimate for cell I.)

For each I and J, XMI and XMJ are compputed and the new SIGMA becomes

$$SIGMA = [CROSSP + NO*(XMU*XMU^T + SIG*WORK2 + E1*(SIG*XNU\ XNU^T + XMU * XNU^T * SIG))]\ FN - XMU^* * XMU^{*T}$$

where XMU^* is the new mean estimate.

The quantity

$$SIG*WORK2 + E1 * (SIG*XNU*XMU^T + XMU*XNU^T*SIG)$$

is stored temporarily in SIGMA for efficiency. Updated means are then stored in WORK2, because both the current and the previous means are required simultaneously.

ET--Compute conditional expectations E1 and E2, normalized truncation TSTAR, and vector XNU. These correspond to quantities E_1, E_2, t^*, and V as discussed in Chapter 2 of the report. A statement function is used for the univariance normal density. The IMSL routine MDNOR is used to get the required normal-probability-distribution value.

MULTV--Multiplies a lower triangular matrix A stored in symmetric storage mode by a vector X to obtain Y. Algebraically,

$$Y_i = \sum_{j=1}^{n} a_{ij}\, x_j \quad .$$

Since $A_{ij} = 0$ for $j > i$, it suffices to sum only from j=1 to i.

ABATT--Multiplies A (lower triangular) by B (symmetric) by A (transpose). All matrices are in symmetric storage mode. Algebraically,

$$(ABA^T)_{ij} = \sum_{k=1}^{n} (AB)_{ik}\, A^T_{kj}$$

Since $A^T_{kj} = A_{jk}$,

$$(ABA^T)_{ij} = \sum_{k=1}^{n} \left| \sum_{l-1}^{n} A_{il}B_{lk} \right| A_{jk}$$

$$= \sum_{k=1}^{n} A_{jk} \sum_{l=1}^{n} A_{il}\ B_{lk}$$

Since A is lower triangular, we only require the sum over k to go to j, rather than n, and the sum over l to go to i, rather than n.

SUMT--Sums the rows of a lower triangular matrix stored in symmetric mode.

NORM--Computes the norm of a vector.

DOT--Computes dot product of two vectors (inner product).

XXT--Computes the outer product of a vector and itself. The result is a symmetrically stored matrix.

SUMX--Sums the entries of a vector.

SUMM--Sums the rows of a symmetric matrix.

SQRM--Computes the square root of a matrix via singular valued decomposition.

NAIVES--Computes simple parameter estimates from the cross-product matrix and sums vector. The covariance matrix is computed and stored in symmetric mode. A one-pass algorithm is used.

NEWSUM--Computes sums and cross products from the data.

COMPRS--Uses vector INSET to compress the data according to the E-M truncation rule, i.e., a day is truncated if the average of the values (rather than the minimum) is below the truncation limit. Recall that INSET(I) = J when the Ith day of the final set is the Jth day of the original ($J \geq I$). COMPRS uses this to compute both a new data matrix as well as a new vector of Julian dates.

PROGRAM CHARACTERISTICS THAT MIGHT AFFECT INSTALLATIONS ON OTHER MACHINES--STORAGE AND PRECISION CONSIDERATIONS

With the exception of the random number generation of SEED, all floating point quantities are defind to be single precision. This does not pose an accuracy problem on a CDC computer, which has 60-bit words. Converting all these quantities to double precision would allow satisfactory accuracy on smaller word (e.g., 32-bit IBM) machines, and such a conversion is often available as a compiler option. Not all quantities must be double precision, however, and it may be wise to allow at least the X array (10,000 words--that is, the single largest core entity) to be treated as single precision, because no sensitive calculations are required to obtain X; it consists simply of the logs of the daily values, where high accuracy is not required.

It is recommended that double precision be used for the intermediate results used in SIMULA and in the E-M block of routines if the machine word size is 32 bits or less. The simple-parameter-estimation technique probably yields satisfactory results in single precision, as it is not iterative in nature and requires no matrix inversion or decomposition. However, since these locations are reused by the E-M-parameter-estimation modules, it may not prove feasible to define them as single precision.

INCORPORATING ADDITIONAL PARAMETER ESTIMATION TECHNIQUES

It may be desirable to experiment with techniques other than the E-M and simple for parameter estimation. To do so, the programmer would:

(1) Modify CRACK to allow the user to specify new parameters for the new techniques.

(2) Modify the branching logic in the main program.

(3) Modify READ1 to keep track of a new pointer vector analogous to the INSET vector, if the new technique requires truncation.

(4) Call the driver routine for the technique from the main program at the appropriate branch. If truncation is required, call COMPRS with the different pointer vector and then NEWSUM if sums and cross products are needed. If no truncation is required, SUMX and CROSSP, as computed in READ1, may be used. The modules for the new technique should compute a mean vector and a variance/covariance matrix or a correlation matrix. In the latter case, a vector of standard deviations must also be computed. Subroutine COV may be used to compute a covariance matrix from a correlation matrix and a vector of standard deviations.

(5) Call SIMULA from the main program, after returning from the new technique driver, to perform simulations based on the parameter estimates. If SIMULA is to work, the logs of the daily values on which the parameters are based should be in /DATA/ x, with each row corresponding to a day. The covariance matrix should be in /ANCOV/SIGMA, and means should be in /STATS/XMU. If required, standard deviations should be in /STATS/XSD. Storage of these quantities in these locations eliminates the need for much additional storage, especially if the new routines also utilize the common work area.

Models to be used for specialized functions should be from an existing subroutine library or from the existing routines in the program (which have been generally written with this possibility in mind). If it is necessary to write such routines, the programmer should be aware of the nature of various matrices (especially symmetry or lower triangularity) and should optimize the code accordingly.

PROGRAM LISTING

A complete listing for the probabilistic methodology modeling and simulation program is given on the following pages.

PROGRAM ANAL

```
      PROGRAM ANAL(OUTPUT,TAPE7,TAPE6=OUTPUT)
      COMMON/UNITS/IN,IOUT,ISIM
      COMMON/COUNTS/NTOTD,N1,N2,EXCED(50),OVR100(50),NDAYS,NCELLS
      COMMON/POINT/INDX(50)
      INTEGER OVR100,EXCED
      COMMON/DESIGN/DSNET(22),DSN(20,50)
      COMMON/DESCR/XMU(50,3),XSD(50,3)
      COMMON/EST/TECH(2,3)
      INTEGER TECH
      COMMON/DATA/X(200,50)
      COMMON/JULDAY/JDAY(200)
      DIMENSION TOPX(10),SCRAT(20)
C
C     ANALYZES NETWORK VALUES (GENERALLY SIMULATED, BUT POSSIBLY
C     OBSERVED OR INTERPOLATED)
C     STATISTICS GIVEN -
C     FOR EACH SIMULATION, THE CELL MEANS, THE CELL EXCEEDANCES,
C     THE VARIANCE COVARIANCE MATRIX
C
      IOUT=6
      ISIM=7
C     READ IN VALUES
      NUMBLK=0
      CALL READHD
 10   READ(ISIM,106)NETS,NAM1,NAM2
      IF(NETS.LE.0)GO TO 60
      NUMBLK=NUMBLK+1
      DO 20 J=1,NETS
        CALL RDSIM
        DSNET(J)=0
        WRITE(IOUT,104)J
 104    FORMAT(1H1,10X,*N E T W O R K *,I4/2X,*CELL*,4X,
     1     *TOP 10 VALUES*,87X,*DESIGN VALUE*/)
        DO 30 I=1,NCELLS
          CALL TOPN(X(1,I),10,TOPX,NDAYS)
          DSN(J,I)=DSNVAL(TOPX,365)
          DO 32 L=1,10
            SCRAT(L)=EXP(TOPX(L))
 32       CONTINUE
          SCRAT(11)=DSN(J,I)
          WRITE(IOUT,105)INDX(I),(SCRAT(L),L=1,11)
 30     CONTINUE
C      EDIT DESIGN VALUES IF DESIRED
C      CALL EDIT(DSN)
       DO 35 I=1,NCELLS
         IF(DSN(J,I).GT.DSNET(J))DSNET(J)=DSN(J,I)
 35    CONTINUE
C
         CALL INFO(J)
 20    CONTINUE
 40    CONTINUE
C      WRITE THEM OUT
       WRITE(IOUT,100)
       DO 50 I=1,NETS
         WRITE(IOUT,101)NAM1,NAM2,I,DSNET(I)
 50    CONTINUE
 100   FORMAT(1H1,10X,25HD E S I G N   V A L U E S//)
```

```
 101  FORMAT(11X,11HTECHNIQUE -,2A4,9HNETWORK  ,I4,10X,13HDESIGN VALUE
     1  ,F7.3)
      IF(NETS.GT.1)GO TO 15
      IF(NETS.LT.1)GO TO 60
      DMU=DSNET(1)
      DSD=0.0
      DMIN=DSNET(1)
      DMAX=DMIN
      DO 12 I=1,N1
        XMU(I,NUMBLK)=DSN(1,I)
        XSD(I,NUMBLK)=0.0
 12   CONTINUE
      GO TO 16
 15   CALL STATS(DSNET,NETS,DMU,DSD,DMIN,DMAX)
C     COMPUTE DESCRIPTIVE STATS FOR SET OF DESIGN VALUES
      DO 17 I=1,N1
        CALL STATS(DSN(1,I),NETS,XMU(I,NUMBLK),XSD(I,NUMBLK),XMIN,XMAX)
 17   CONTINUE
 16   CONTINUE
      TECH(1,NUMBLK)=NAM1
      TECH(2,NUMBLK)=NAM2
      WRITE(IOUT,102)NAM1,NAM2,NETS,DMU,DSD,DMIN,DMAX
 102  FORMAT(//1X,33HD E S C R I P T I V E   S T A T S//
     1   1X,23HESTIMATION TECHNIQUE   ,2A4/
     1  1X, 23HNO. NETWORKS           ,I3/
     2  1X, 23HMEAN DESIGN VALUE      ,F7.3/
     3  1X, 23HSTD DEV OF DESIGN VALS ,F10.3/
     4   1X,  23HMIN DESIGN VALUE       ,F7.3/
     5    1X,  23HMAX DESIGN VALUE       ,F7.3)
      IF(NETS.GT.1)CALL PLDSN(DSNET,NETS,DMIN,DMAX)
      IF(NDAYS.LE.0)GO TO 60
      GO TO 10
 60   CONTINUE
      IF(NUMBLK.GE.2)CALL RTABL(NUMBLK)
 105  FORMAT(2X,I4,4X,11(F8.4,2X))
 106  FORMAT(I3,2A4)
      STOP
      END
```

SUBROUTINE INFO

```
      SUBROUTINE INFO(NETS)
      COMMON/UNITS/IN,IOUT,ISIM
      COMMON/COUNTS/NTOTD,N1,N2,EXCED(50),OVR100(50),NDAYS,NCELLS
      COMMON/POINT/INDX(50)
      INTEGER OVR100,EXCED
      COMMON/DESIGN/DSNET(22),DSN(20,50)
C
C     PRINTS INFORMATIVE NETWORK INFO
      WRITE(IOUT,100)NETS,NTOTD,NDAYS,NCELLS
 100  FORMAT(1H1,10X,*N E T W O R K   I N F O R M A T I O N*//
     1          11X,*NETWORK NO.          *,I4/
     2          11X,*TOTAL NO. OF DAYS    *,I4/
     3          11X,*FINAL NO. OF DAYS    *,I4/
     4          11X,*NO. OF CELLS         *,I4//)
      WRITE(IOUT,101)
 101  FORMAT(10X,*CELL*,10X,*NO. EXCED*,10X,*OVER 100*,
     1 10X,*DESIGN VALUE*/)
      DO 10 I=1,NCELLS
        WRITE(IOUT,102)INDX(I),EXCED(I),OVR100(I),DSN(NETS,I)
 10   CONTINUE
      RETURN
 102  FORMAT(10X,I4,13X,I4,14X,I4,10X,F7.3)
      END
```

SUBROUTINE EDIT

```
      SUBROUTINE EDIT(DSNET)
      RETURN
      END
```

SUBROUTINE PLDSN

```
      SUBROUTINE PLDSN(DSNET,NETS,DMIN,DMAX)
      COMMON/DATA/IMAG4(10000)
      DIMENSION DSNET(1),X(20),ITITL(144),RAN(4),ICHAR(10)
      DATA ITITL/1HS,1HC,1HA,1HT,1HT,1HE,1HR,1H ,1HP,1HL,
     1 1HO,1HT,1H ,1HO,1HF,1H ,1HD,1HE,1HS,1HI,1HG,1HN,
     2 1H ,1HV,1HA,1HL,1HU,1HE,1HS,43*1H ,
     3 1HN,1HE,1HT,1HW,1HO,1HR,1HK,29*1H ,
     4 1HD,1HE,1HS,1HI,1HG,1HN,1H ,1HV,1HA,1HL;1HU,1HE,
     5 24*1H /,ICHAR/10*1H+/
      RAN(1)=0.0
      RAN(2)=FLOAT(NETS)+1.0
      DEL=DMAX-DMIN
      RAN(3)=DMIN-.1*DEL
      RAN(4)=DMAX+.1*DEL
      IOPT=0
      INC=1
      M=1
      IY=1
      DO 10 I=1,NETS
        X(I)=FLOAT(I)
 10   CONTINUE
      CALL USPLT(X,DSNET,IY,NETS,M,INC,ITITL,RAN,ICHAR,IOPT,IMAG4,IER)
      RETURN
      END
```

SUBROUTINE RDSIM

```
      SUBROUTINE RDSIM
      COMMON/UNITS/IN,IOUT,ISIM
      COMMON/JULDAY/JDAY(200)
      COMMON/DATA/X(200,50)
      COMMON/COUNTS/NTOTD,N1,N2,EXCED(50),OVR100(50),NDAYS,NCELLS
      COMMON/POINT/INDX(50)
      INTEGER OVR100,EXCED
      COMMON/DESIGN/DSNET(22),DSN(20,50)
      READ(ISIM,100)NSIM,NCELLS,NDAYS
      IF(NSIM.LE.0)NDAYS=-1
      IF(NDAYS.LT.0)RETURN
 100  FORMAT(5I3)
 101  FORMAT(6X,I6,6X,(10F10.6))
      DO 10 I=1,NCELLS
        OVR100(I)=0
        EXCED(I)=0
 10   CONTINUE
```

```
C
C      NOW THAT SUMMARY VECTORS AND MATRICES ARE INITIALIZED, READ IN
C      DAILY VALUES AND ACCUMULATE APPROPRIATE TOTALS
       INUM=0
       DO 20 IDAY=1,NDAYS
         READ(ISIM,101)JDAY(IDAY),(X(IDAY,I),I=1,NCELLS)
         IF(JDAY(IDAY).LE.0)GO TO 16
         INUM=INUM+1
         DO 15 I=1,NCELLS
           IF(X(IDAY,I).GT.100.0)OVR100(I)=OVR100(I)+1
           IF(X(IDAY,I).GT.120.0)EXCED(I)=EXCED(I)+1.
           X(IDAY,I)=ALOG(X(IDAY,I))
 15      CONTINUE
 20    CONTINUE
 16    NDAYS=INUM
       IFIRST=0
       RETURN
       END
```

SUBROUTINE READHD

```
      SUBROUTINE READHD
      COMMON/UNITS/IN,IOUT,ISIM
      COMMON/JULDAY/JDAY(200)
      COMMON/DATA/X(200,50)
      COMMON/COUNTS/NTOTD,N1,N2,EXCED(50),OVR100(50),NDAYS,NCELLS
      COMMON/POINT/INDX(50)
      INTEGER OVR100,EXCED
      COMMON/DESIGN/DSNET(22),DSN(20,50)
C
C     READ HEADER INFO
      READ(ISIM,900)NTOTD,INITC,NDAYS,NTOTC,IDUM,N2,NCELLS
      N1=NCELLS-N2
 900  FORMAT(7I10)
      READ(ISIM,902)(INDX(I),I=1,NCELLS)
 902  FORMAT(16I5)
      RETURN
      END
```

SUBROUTINE RTABL

```
      SUBROUTINE RTABL(NUMBLK)
      COMMON/EST/TECH(2,3)
      INTEGER TECH
      COMMON /COUNTS/NTOTD,N1,N2,EXCED(50),OVR100(50),NDAYS,NCELLS
      INTEGER EXCED,OVR100
      COMMON/DESCR/XMU(50,3),XSD(50,3)
      COMMON/POINT/INDX(50)
      COMMON/UNITS/IN,IOUT,ISIM
C
C     .GIVEN THE RESULTS OF THE VARIOUS NETWORK ANALYSES, THIS ROUTINE
C      PRINTS OUT A TABLE OF COMPARISONS. XMU(I,J) CONTAINS THE MEANS
C      FOR CELL I BLOCK J (A BLOCK OF SIMULATIONS CORRESPONDS TO AN ES-
C      TIMATION TECHNIQUE). XSD(I,J) IS ANALOGOUS BUT FOR STANDARD
C      DEVIATIONS. THE TECHNIQUE NAMES ARE IN TECH. PROCESSING VARIES
C      DEPENDING ON NUMBLK (THE NUMBER OF BLOCKS)
```

```
C
      REAL DIFF(2),R(2),COLMAX(7)
 70   FORMAT(9X,1HI,10(1H-),1HI,2(33(1H-),1HI))
 40   FORMAT(9X,1HI,10(1H-),1HI,33(1H-),1HI)
C     GENERALLY, FORMATS ENDING IN 40 ARE FOR NUMBLK=2 (40 COL OUTPUT)
C     WHILE THOSE ENDING IN 70 ARE FOR NUMBLK=3 (0 COL)
C     IT IS NOT ALWAYS NECCESSARY TO HAVE TWO.
 140  FORMAT(4X,6HCELL I,1X,10HVALUE     I,2X,5HMEAN,5X,2HSD,5X,3HABS,
     1  5X,3HPCT,3X,1HI)
 170  FORMAT(4X,6HCELL I,1X,10HVALUE     I,2(2X,5HMEAN,5X,2HSD,5X,
     1  3HABS,5X,3HPCT,3X,1HI))
 100   FORMAT(1H1,9X,45HC O M P A R I S O N   O F   E S T I M A T E S//)
 200  FORMAT(4X,I3,2X,2HI ,F7.2,2X,1HI,2(1X,4(F7.2,1X),1HI))
 270  FORMAT(10X,2A4,16X,2A4,26X,2A4)
 370  FORMAT(4X,3HMAX,2X,2HI ,F7.2,2X,1HI,2(1X,F7.2,9X,F7.2,
     1     1X,F7.2,1X,1HI))
C
C     BEGIN - PAGE EJECT AND HEADER
C
      WRITE(IOUT,100)
      WRITE(IOUT,270)((TECH(I,J),I=1,2),J=1,NUMBLK)
      IF(NUMBLK.EQ.3)GO TO 5
C     SHORT HEADER AND DASHED LINE
      WRITE(IOUT,140)
      WRITE(IOUT,40)
      GO TO 10
C
 5    WRITE(IOUT,170)
      WRITE(IOUT,70)
 10   CONTINUE
C
C     ZERO COLUMN MAXIMA
      DO 20 I=1,7
        COLMAX(I)=0.0
 20   CONTINUE
C
C     NOW PRINT TABLE (FOR VARIABLE PORTION ONLY)
      DO 30 I=1,N1
        IF(XMU(I,1).GT.COLMAX(1))COLMAX(1)=XMU(I,1)
        IF(XMU(I,2).GT.COLMAX(2))COLMAX(2)=XMU(I,2)
        IF(NUMBLK.EQ.2)GO TO 21
        IF(XMU(I,3).GT.COLMAX(5))COLMAX(5)=XMU(I,3)
 21     CONTINUE
C
C       COMPUTE COLUMN DIFFERENCES (ABS AND PCT)
        NM1=NUMBLK-1
        DO 22 J=1,NM1
          JX=J+1
          DIFF(J)=XMU(I,JX)-XMU(I,1)
          R(J)=100.0*DIFF(J)/AMAX1(XMU(I,JX),XMU(I,1))
 22     CONTINUE
C       NOW GET COLUMN MAXIMA OF DIFFERENCES
        IF(ABS(DIFF(1)).GT.ABS(COLMAX(3)))COLMAX(3)=DIFF(1)
        IF(ABS(R(1)).GT.ABS(COLMAX(4)))COLMAX(4)=R(1)
        IF(NUMBLK.EQ.3)GO TO 23
        WRITE(IOUT,200)INDX(I),XMU(I,1),XMU(I,2),XSD(I,2),DIFF(1),R(1)
        GO TO 30
 23     IF(ABS(DIFF(2)).GT.ABS(COLMAX(6)))COLMAX(6)=DIFF(2)
        IF(ABS(R(2)).GT.ABS(COLMAX(7)))COLMAX(7)=R(2)
        WRITE(IOUT,200)INDX(I),XMU(I,1),XMU(I,2),XSD(I,2),DIFF(1),R(1),
     1        XMU(I,3),XSD(I,3),DIFF(2),R(2)
 30   CONTINUE
C
C     NOW COLUMN MAXIMA
      ILIM=1+(NUMBLK-1)*3
      IF(NUMBLK.EQ.3)GO TO 35
```

```
      WRITE(IOUT,40)
      WRITE(IOUT,370)(COLMAX(I),I=1,ILIM)
      WRITE(IOUT,40)
      RETURN
 35   CONTINUE
      WRITE(IOUT,70)
      WRITE(IOUT,370)(COLMAX(I),I=1,ILIM)
      WRITE(IOUT,70)
      RETURN
      END
```

SUBROUNTINE STATS

```
      SUBROUTINE STATS(X,N,XMU,XSD,XMIN,XMAX)
      DIMENSION X(1)
C
C     GIVES UNIVARIATE DESCRIPTIVE STATS FOR ARRAY X. N IS NO. OF OBS
C     IN X. XMU, XSD, XMIN, AND XMAX ARE THE MEAN, STANDARD DEVIATION,
C     MINIMIM, AND MAXIMUM VALUES OF X, RESPECTIVELY.
C     NO MISSING VALUE CHECKS ARE DONE.
C
      XMIN=X(1)
      XMAX=XMIN
      SUMX=0.0
      SUMX2=0.0
      DO 10 I=1,N
        IF(X(I).GT.XMAX)XMAX=X(I)
        IF(X(I).LT.XMIN)XMIN=X(I)
        SUMX=SUMX+X(I)
        SUMX2=SUMX2+X(I)*X(I)
 10   CONTINUE
      FN=FLOAT(N)
      XMU=SUMX/FN
      XSD=(SUMX2-FN*XMU*XMU)/(FN-1.0)
      XSD=SQRT(XSD)
      RETURN
      END
```

SUBROUTINE TOPN

```
      SUBROUTINE TOPN(X,N,TOPX,NX)
      DIMENSION X(1),TOPX(1)
C
C     GIVEN A VECTOR X F LENGTH NX, THIS ROUTINE RETURNS THE TOP N
C     (LE NX) VALUES IN VECTOR TOPX IN DESCENDING ORDER. A SINGLETON
C     SORT ALGORITHM IS USED.
C
      CALL VSRTA(X,NX)
      DO 10 I=1,N
        IX=NX-I+1
        TOPX(I)=X(IX)
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE WRSYM

```
      SUBROUTINE WRSYM(A,NA)
      COMMON/UNITS/IN,IOUT,ISIM
      DIMENSION A(1)
C     PRINTS A MATRIX STORED IN SYMMETRIC STORAGE MODE - LOWER DIAGONAL
C     NA IS THE ORDER OF A
      DO 10 I=1,NA
        ISTART=I*(I-1)/2+1
        IEND=ISTART+I-1
        WRITE(IOUT,100)(A(K),K=ISTART,IEND)
 10   CONTINUE
 100  FORMAT(15(1X,F7.5))
      RETURN
      END
```

FUNCTION DSNVAL

```
      FUNCTION DSNVAL (BIGOZ,NDAYS)
      DIMENSION BIGOZ(1)
C
C     RETURNS THE DESIGN VALUE OF ARRAY BIGOZ, WHICH CONTAINS THE 10
C     HIGHEST OBSERVATIONS FOR SOME A CELL. BIGOZ IS IN LOG UNITS,
C     DSNVAL IS RETURNED IN DATA UNITS. NDAYS MUST BE THE TOTAL
C     NUMBER OF DAYS (NOT THE TRUNCATED).
C     ALGORITHM IS TO DO A LEAST SQUARES LINEAR FIT TO THE LOGS OF
C     THE LARGEST VALUES VERSUS CORRESPONDING PERCENTILES.
C     MODIFIED FROM FRANK LUDWIG ISOPLETH PROGRAM.
C
      XDAYS=FLOAT(NDAYS)
      YDAYS=0.0
      SUMX=0.0
      SUMY=0.0
      SUMXY=0.0
      SUMX2=0.0
C
      DO 100 K=1,10
        IF(BIGOZ(K).LE.0.0)GO TO 101
        Y=BIGOZ(K)
        XN=K*100
        YDAYS=YDAYS+1.0
        X=XN/XDAYS
        SUMX=SUMX+X
        SUMY=SUMY+Y
        SUMXY=SUMXY+X*Y
        SUMX2=SUMX2+X*X
 100  CONTINUE
 101  IF(YDAYS.LE.0.0)GO TO 86
      XB=SUMX/YDAYS
      YB=SUMY/YDAYS
      B=(SUMXY-(YDAYS*XB*YB))/(SUMX2-(YDAYS*XB*XB))
      DSNVAL=YB+(B*(0.274-XB))
C     ANTILOG
      DSNVAL=EXP(DSNVAL)
      RETURN
 86   DSNVAL=-99.0
      RETURN
      END
```

PROGRAM MONTE

```
      PROGRAM MONTE(INPUT=404B,OUTPUT=404B,TAPE3=404B,TAPE7=404B,
     1     TAPE5=INPUT,TAPE6=OUTPUT)
      COMMON/LIMIT/THRESH,TRUNC
C      THRESH - THRESHHOLD - OVER THIS IS AN EXCEEDANCE(LOG UNITS)
C     TRUNC - DATA TRUNCATION VALUE - DAYS FOR WHICH AVERAGE IS BELOW
C             THIS ARE EXCLUDED FROM E M COMPUTATION
      COMMON /UNITS/IN,IOUT,ISIM,IPARM
      COMMON/DATA/X(200,50)
C     X - LOGS OF DAILY VALUES
      COMMON/STATS/XMU(50),XSD(50)
C
C     XMU - ESTIMATE OF CELL MEAN (OF LOG VALUES)
C     XSD - ESTIMATE OF CELL STANDARD DEVIATION
C
      COMMON/ANCOV/SIGMA(1300),CROSSP(1300)
C     SIGMA - ESTIMATE OF CORRELATION MATRIX, SYMMETRIC STORAGE MODE
C     CROSSP - CROSSPRODUCT MATRIX IN SYMMETRIC STORAGE MODE
      COMMON/SUMS/SUMX(50)
C     SUMX - SUM OF LOG DAILY VALUES
      COMMON/COUNTS/EXCED(50),OVR100(50),NDAYS,NCELLS,N1,N2,
     1   NTOTC,NTOTD,NFINAL,INITC,INDX(50),INSET(200)
C     EXCED(I) - NO. OF EXCEEDANCES AT CELL I
C     OVR100(I) - NO. OVER 100 PPB CELL I
C     NDAYS - FINAL NO. OF DAYS (ACTUAL NO. OF DAYS INPUT TO THIS PROG)
C     NCELLS - FINAL NO. OF CELLS
C     N1 - GROUP 1 CELLS (NO STATIONS, HIGH NO. OF EXCED OR DESIGN)
C     N2 - GROUP 2 CELLS (WITH STATIONS)
C     NTOTD - TOTAL NO OF DAYS
C     NFINAL - FINAL NO. OF DAYS AFTER TRUNCATION
C     NTOTC - TOTAL NO. OF CELLS
C     INITC - INITIAL NO. OF CELLS
C     INDX(50),INSET(200) - CELL INDEX ARRAY
      COMMON /PCTILE/ VAL,QTILE(50)
C     VAL - VALUE OF DESIRED PERCENTILE TO USE AS INTEGRATION LIMIT.
C           VAL IS BETWEEN .01 AND .99 (1 AND 99TH PCTILE)
C     QTILE(I) - THE VAL-TH QUANTILE FOR VARIABLE I
      COMMON/EST/TECH(2,3)
      INTEGER TECH
C     TECH - ALPHA NAME OF ESTIMATION TECHNIQUE (ISOPLETH, SIMPLE, OR EM)
C
      COMMON /RAND/DSEED
      DOUBLE PRECISION DSEED
C     DSEED - SEED FOR RANDOM NUMBER GENERATOR
      COMMON/WORK/SCRAT(3000)
      DATA TECH/4HISOP,4HLETH,4HSIMP,4HLE   ,4HE - ,4HM
C
C
C     MAIN CONTROL PROGRAM FOR OZONE GRID MONTE CARLO SIMULATION
C
C     IN - INPUT UNIT
C     IOUT - PRINTER OUTPUT UNIT
C     ISIM - SIMULATION OUTPUT IS WRITTEN HERE FOR FURTHER ANALYSIS
C     IPARM - UNIT FROM WHICH TO READ PARAMETER CARD
C
      IN=3
      IPARM=5
      IOUT=6
      ISIM=7
C
C     READ OPTIONS FROM PARAMETER CARD
      CALL CRACK(METH,NSIM,METHMU,METHEM,ITER,THRESH,TRUNC)
```

```
C
C     READ IN DAILY CELL VALUES
C
      CALL READ1(THRESH,TRUNC)
      CALL INFO
C
C     ESTIMATIONS AND SIMULATIONS BASED ON NAIVE ESTIMATES
      IF(METH.GT.0)GO TO 10
      WRITE(IOUT,49)
 49   FORMAT(1H1,9X,25HNAIVE PARAMETER ESTIMATES //)
      CALL NAIVES(SUMX,CROSSP,NDAYS,NCELLS,XMU,SIGMA)
C     CALL MEANS - SD EXTRAPOLATION IF METHMU=1
      IF(METHMU.EQ.1)CALL EXTRAP(X,XMU,XSD,SIGMA,NCELLS,NDAYS,
     1                           NTOTD,SCRAT)
      WRITE(IOUT,100)(XMU(I),I=1,NCELLS)
      WRITE(IOUT,202)
      CALL WRSYM(SIGMA,NCELLS)
C
C     BRANCH OUT TO EM IF NO SIMULATIONS REQUESTED
      IF(NSIM.LE.0)GO TO 10
      CALL CSEED(DSEED)
      CALL SIMULA(NSIM,NDAYS,TECH(1,2))
C
 10   CONTINUE
      IF(METH.LT.0)STOP
C     NOW CHANGE X TO REFLECT TRUNCATION
      NX=200
      CALL COMPRS(INSET,NFINAL,NCELLS)
      CALL NEWSUM(X,NX,NCELLS,NFINAL,SUMX,CROSSP)
C
C     INITIATE E M ALGORITHM - PERLIMINARY CALCULATIONS
      NO=NTOTD-NFINAL
C     COMPUTE SUBSCRIPTS OF VARIOUS REQUIRED MATRICES AND VECTORS
      ISIG=1
      INU=ISIG+(NCELLS*(NCELLS+1)/2)
      IWK1=INU+NCELLS
      IWK2=IWK1+NCELLS
C
      WRITE(IOUT,50)
 50   FORMAT (1H1,10X,36HE - M ALGORITHM PARAMETER ESTIMATION  //)
      NO=NTOTD-NFINAL
      CALL EMDRIV(SUMX,CROSSP,SIGMA,SCRAT(ISIG),XMU,NTOTD,NO,
     1SCRAT(IXNU),NCELLS,SCRAT(IWK1),SCRAT(IWK2),TRUNC,METHEM,ITER,IER)
      IF(IER.GT.0)GO TO 99
      WRITE(IOUT,100)(XMU(I),I=1,NCELLS)
 100  FORMAT(10X,22HESTIMATED MEAN OF LOGS/(1X,21(F5.3,1X)/))
      WRITE(IOUT,202)
 202  FORMAT(/10X,17HCOVARIANCE MATRIX/)
      CALL WRSYM(SIGMA,NCELLS)
      IF(NSIM.LE.0)STOP
C     NSIM IS THE NUMBER OF SIMULATIONS THAT WILL BE RUN
C     COMPUTE RANDOM NUMBER GENERATOR SEED BASED ON SYSTEM CLOCK
      CALL CSEED(DSEED)
C     CALL SIMULATION DRIVER
      CALL SIMULA(NSIM,NFINAL,TECH(1,3))
      ITRAIL=-1
      WRITE(ISIM,701)ITRAIL
 701  FORMAT(5I3)
      STOP
 99   WRITE(IOUT,199)
      WRITE(ISIM,701)ITRAIL
 199  FORMAT(//10X,32H*** ERROR IN E - M ALGORITHM ***)
      STOP
      END
```

SUBROUTINE EXTRAP

```
      SUDROUTINE EXTRAP(X,XMU,XSD,SIGMA,NCELLS,NDAYS,NTOTD,SCRAT)
      DIMENSION X(200,50),XMU(50),XSD(50),SIGMA(1300),SCRAT(3000)
C
C     COMPUTES LOG MEAN AND STANDARD DEVIATION BASED ON THE ESIMATES
C     OBTAINED BY FITTING A LINE TO THE VALUES VERSUS THEIR PERCENTILE
C     AND THEN EXTRAPOLATING MU AND SD. SUBROUTINE MUVAR IS USED
C     FOR THE BULK OF THE COMPUTATION, WITH THIS ROUTINE ACTING AS
C     POST PROCCESSOR TO RECOMPUTE THE NEW SIGMA FROM THE OLD AND THE
C     NEW SD VECTOR XSD.
      DO 10 I=1,NCELLS
        CALL MUVAR(X(1,I),XMU(I),XSD(I),SCRAT(1),NTOTD,NDAYS,IER)
        DO 10 J=1,I
          II=I*(I-1)/2+I
          JJ=J*(J-1)/2+J
          IJ=I*(I-1)/2+J
          SIGMA(IJ)=(SIGMA(IJ)/(SIGMA(II)*SIGMA(JJ)))*XSD(I)*XSD(J)
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE MUVAR

```
      SUBROUTINE MUVAR(X,XMU,XSD,WORK,NTOTD,NDAYS,IER)
C
C     COMPUTES ESTIMATE OF MEAN AND VARIANCE BY LEAST SQUARES GIVEN THE
C     UPPER TAIL OF THE DISTRIBUTION
C     X - VECTOR OF INTEREST
C     NDAYS - NUMBER OF ENTRIES OF X
C     NTOTD - TOTAL NUMBER OF DAYS (GE NDAYS)
C     XMU,XSD - PARAMETER ESTIMATES
C
      DIMENSION X(1),WORK(1)
      DO 10 I=1,NDAYS
        WORK(I)=X(I)
 10   CONTINUE
C     SORT WORK VECTOR (X IS LEFT UNCHANGED)
      CALL VSRTA(WORK,NDAYS)
C
      ISUB=SUB
      IF(SUB-ISUB.GE.0.5)ISUB=ISUB+1
      QTILE=WORK(ISUB)
C
C
C     COMPUTE EQUIVALENT Y ARRAY IF WE WERE TO PLOT THIS IN PROBABILITY
C     GRAPH PAPER
C     ALSO COMPUTE LEAST SQUARES LINE ON THE FLY
      NOVER=0
      SUMX=0
      SUMX2=0
      SUMY=0
      SUMXY=0
      DO 20 I=1,NDAYS
        ISUB=NDAYS-I+1
        NOVER=NOVER+1
        SUMX=SUMX+WORK(ISUB)
        SUMX2=SUMX2+WORK(ISUB)*WORK(ISUB)
        P=FLOAT(NTOTD-1)/FLOAT(NTOTD)
```

```
        CALL MDNRIS(P,Y,IER)
        SUMY=SUMY+Y
        SUMXY=SUMXY+Y*WORK(ISUB)
 20   CONTINUE
 30   CONTINUE
      DEN=FLOAT(NOVER)*SUMX2-SUMX*SUMX
      B=(SUMY*SUMX2-SUMX*SUMXY)/DEN
      A=(FLOAT(NOVER)*SUMXY-SUMX*SUMY)/DEN
C
C     NOW Y=AX+B, OR (Y-B)/A=X
C     Y=0 GIVES 50TH PERCENTILE OR MEAN EST.
C     Y=1.645 GIVES 95TH, WHICH AFTER SUBTRACTING THE MEAN AND
C     DIVIDING BY 1.645 GIVES SD.
      XMU=-B/A
      XSD=(1.645 - B)/A
      XSD=(XSD-XMU)/1.645
      RETURN
      END
```

SUBROUTINE CRACK

```
      SUBROUTINE CRACK(METH,NSIM,METHMU,METHEM,ITER,THRESH,TRUNC)
C
C     INTERPRETS THE USER PARAMETER OPTION CARD
C     IF NO CARD OR AN INCONSISTENT CARD IS PRESENTED, DEFAULTS ARE SET.
C
C     ***********    NOTE    ***********    NOTE     ********     NOTE    *
C     THIS FORM OF EOF CHECK IS CDC DEPENDENT. ON IBM TYPE COMPUTERS,
C     USE    READ(...,END=10)...
C
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
      READ(IPARM,100)METH,NSIM,METHMU,METHEM,ITER,THRESH,TRUNC
 100  FORMAT(5I3,2F10.6)
      IF(EOF(IPARM).NE.0.0)GO TO 10
C
C     A CARD WAS READ. SET NONSPECIFIED STUFF TO DEFAULT
      IF(METHMU.LT.0)METHMU=0
      IF(METHMU.GT.1)METHMU=1
      IF(NSIM.EQ.0)NSIM=20
      IF(THRESH.LE.0.0)THRESH=120.0
      IF(ITER.LE.0)ITER=10
      IF(TRUNC.LE.0.0)METH=-1
      IF(TRUNC.LE.0.0)TRUNC=-99.0
      METHMU=0
      GO TO 20
 10   CONTINUE
C
C     NO PARAMETER CARD. ASSUME THRESH OF 120 AND NO TRUNC, SO CANNOT DO EM
C     AND THUS FORCE METH LT 0
      METH=-1
      TRUNC=-99.0
      NSIM=20
      THRESH=120.0
 20   CONTINUE
      IF(TRUNC.GT.0.0.AND.METHEM.EQ.0)METH=0
C     LOG VALUES
      IF(TRUNC.GT.0.0)TRUNC=ALOG(TRUNC)
      THRESH=ALOG(THRESH)
      RETURN
      END
```

SUBROUTINE WRSYM

```
      SUBROUTINE WRSYM(A,NA)
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
      DIMENSION A(1)
C     PRINTS A MATRIX STORED IN SYMMETRIC STORAGE MODE - LOWER DIAGONAL
C     NA IS THE ORDER OF A
      DO 10 I=1,NA
        ISTART=I*(I-1)/2+1
        IEND=ISTART+I-1
        WRITE(IOUT,100)(A(K),K=ISTART,IEND)
 10   CONTINUE
 100  FORMAT(15(1X,F7.5))
      RETURN
      END
```

SUBROUTINE READ1

```
      SUBROUTINE READ1(THRESH,TRUNC)
C
C     READS THE INPUT DATA IN FORMATTED MODE FROM UNIT IN.
C     THE DATA FOR THE CELLS WITH STATIONS IS ASSUMED TO BE FIRST ON
C     THE INPUT LINE (THERE ARE NG2 STATIONS)
      COMMON /COUNTS/EXCED(50),OVR100(50),NDAYS,NCELLS,NG1,NG2,
     1 NTOTC,NTOTD,NFINAL,INITC,INDX(50),INSET(200)
C     EXCED - NUMBER OF EXCEEDANCES AT THIS CELL
C     OVR100 - UMBER OF OBSERVATIONS OVER 100 AT THIS CELL
C     NDAYS - NUMBER OF DAYS IN FINAL WORKING SET - LE 200
C     NG1 - NUMBER OF CELLS IN WORKSET WITHOUT MONITORING
C           STATIONS.
C     NG2 - NUMBER OF CELLS IN WORKSET WITH ONE OR MORE STATIONS
C     NCELLS - NUMBER OF CELLS IN WORKING SET = NG1+NG2
C     NTOTC - TOTAL NUMBER OF CELLS ON INPUT FILE
C     NTOTD - TOTAL NUMBER OF DAYS IN STUDY - REQUIRED FOR
C             SOME STATS
C     NFINAL - FINAL NO. OF DAYS AFTER TRUNCATION
      COMMON/GHOST/GCROSS(1300),GSUM(50)
C
C     /GHOST/ CONTAINS THE SUMS AND CROSSPRODUCTS OF THE REFLECTIONS
C     ALONG THE TRUNCATION AXIS. THIS IS USED TO YIELD HOPEFULLY
C     BETTER STARTING ESTIMATES OF MU AND SIGMA THAN WOULD BE
C     OBTAINED SIMPLY BY TAKING SUMX/NFINAL FOR MU AND
C     CROSSP(I,J)/NFINAL - MU(I)*MU(J) FOR SIGMA
C
C
      INTEGER EXCED,OVR100
      COMMON/DATA/X(200,50)
C     X - THE ACTUAL WORKING DATA SET
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
C     IN - INPUT UNIT
C     IOUT - OUTPUT UNIT
C     ISIM - OUTPUT OF SIMULATION
      COMMON/EST/TECH(2,3)
      INTEGER TECH
      COMMON/WORK/SCRAT(3000)
      INTEGER ISCRAT(1)
      EQUIVALENCE (SCRAT,ISCRAT)
C     SCRAT - UTILITY SCRATCH WORK AREA
      COMMON/SUMS/SUMX(50)
C     SUMX - SUM OF CELL VALUES FOR EACH CELL
```

```
C
      COMMON/JULDAY/JDAY(400)
C     JDAY(I) - JULIAN DAY OF OBS(I)
      COMMON/ANCOV/SIGMA(1300),CROSSP(1300)
C
C     CONTAINS THE SUMS, SUMS OF SQUARES, AND CROSSPRODUCTS
C     CROSSPRODUCTS ARE IN SYMMETRIC STORAGE MODE
C     THAT IS, THE CROSS PRODUCT OF CELLS I AND J
C     IS IN ENTRY I(I-1)/2 + J
C
      READ(IN,100)NTOTD,INITC,NDAYS,NTOTC,IHALF,NG2,NCELLS
      WRITE(ISIM,100)NTOTD,INITC,NDAYS,NTOTC,IHALF,NG2,NCELLS
 100  FORMAT(7I10)
C
C     AND NOW THE CELL INDEXES
      READ(IN,101)(ISCRAT(I),I=1,NCELLS)
      NG1=NCELLS-NG2
      DO 1 I=1,NCELLS
        OVR100(I)=0
        EXCED(I)=0
        SUMX(I)=0.0
        GSUM(I)=0.0
C       REARRANGE INDEX VECTOR - GROUP 2 ARE FIRST N2 ENTRIES, WE
C       WANT THESE AT THE END
        ISUB=NG2+I
        IF(I.GT.NG1)ISUB=I-NG1
        INDX(I)=ISCRAT(ISUB)
C
        DO 1 J=1,I
          JSUB=I*(I-1)/2+J
          CROSSP(JSUB)=0.0
          GCROSS(JSUB)=0.0
 1    CONTINUE
      WRITE(ISIM,101)(INDX(I),I=1,NCELLS)
 101  FORMAT(16I5)
C
C     WRITE NO. OF NETWORKS OF REAL DATA, I. E. , ONE
      NETS=1
      WRITE(ISIM,201)NETS,TECH(1,1),TECH(2,1)
 201  FORMAT(I3,2A4)
      WRITE(ISIM,202)NETS,NCELLS,NDAYS
 202  FORMAT(5I3)
C
C
C     READ IN X, Y COORDINATES OF EACH CELL
C     READ FIRST FOR CELLS 1 TO IHALF
      IX=1
      IXE=2*IHALF
      READ(IN,102)(SCRAT(J),J=IX,IXE)
 102  FORMAT(10F8.2)
      IX=IXE+1
C     NOW READ COORDINATES OF OTHER CELLS
      IXE=NTOTC-IHALF
      IXE=IX+(2*IXE)
      READ(IN,102)(SCRAT(J),J=IX,IXE)
      IX=IXE+1
C
      IXE=IXE+NTOTC
      IX1=IXE+1
      IXE1=IX1+NCELLS-1
C     READ IN DAILY VALUES
      NFINAL=0
      IGHS=IXE
      XLOW=TRUNC*FLOAT(NCELLS)
      AL100=ALOG(100.)
C
C     NOW READ IN DAILY VALUES. THEY ARE ASSUMED TO BE IN ASCENDING
C     ORDER OF CELL INDEX. WHEN THEY ARE FINALLY MOVED INTO THE X
```

```
C     MATRIX, THEIR ORDER REFLECTS THE ORDER OF THE CELL INDEXES
C     IN INDX, THAT IS, THE JTH COLUMN OF X HAS THE VALUES FOR
C     CELL NUMBER INDX(J).
C     NFINAL=NUMBER OF DAYS MEETING THE EM TRUNCATION CRITERION (I.E.,
C     THE AVERAGE VALUE INSTEAD OF THE MAX IS GT TRUNC FOR THAT DAY.
C     INSET(K),K=1,NFINAL EQUALS J WHENEVER J FORMS THE KTH ROW
C     OF THE FINAL SET AFTER EM COMPRESSES IT.
C
      ND=NDAYS
      I=0
      DO 10 ICOUNT=1,ND
C       READ IN A DAY OF VALUES
        READ(IN,200)IDAY,(SCRAT(J),J=IX,IXE)
C       THROW OUT THE DAY IF ANY VALUE IS LT 0
        DO 11 J=1,NCELLS
          IPOINT=INDX(J)
          ISCR=IX+IPOINT-1
        ISCR2=IX1+J-1
        SCRAT(ISCR2)=SCRAT(ISCR)
          IF(SCRAT(ISCR2).LE.0.0)GO TO 12
 11     CONTINUE
        GO TO 13
 12     NDAYS=NDAYS-1
        GO TO 10
 13     CONTINUE
        I=I+1
        SUMDAY=0.0
C       DECIDE WHETHER OR NOT TO KEEP THE DAY FOR EM CALCULATIONS
        DO 20 J=1,NCELLS
          ISCR2=IX1+J-1
          X(I,J)=ALOG(SCRAT(ISCR2))
          SUMDAY=SUMDAY+X(I,J)
 20     CONTINUE
C
C       WRITE OUT DATA AS READ TO UNIT ISIM - THIS IS SIMULATION O
C       OR THE "REAL" DATA
C
        WRITE(ISIM,200)IDAY,(SCRAT(J),J=IX1,IXE1)
C
        IF(SUMDAY.LT.XLOW)GO TO 22
C       THIS DAY IS TO BE INCLUDED
        NFINAL=NFINAL+1
        INSET(NFINAL)=I
 22     JDAY(I)=IDAY
C       MOVE DAY INTO X MATRIX AT ROW I
        DO 30 J=1,NCELLS
          IF(X(I,J).GT.AL100)OVR100(J)=OVR100(J)+1
          IF(X(I,J).GT.THRESH)EXCED(J)=EXCED(J)+1
C         IGJ=IGHS+J-1
C         TO ACCUMULATE GHOST SUMS, ACTIVATE THESE LINES
C         SCRAT(IGJ)=2*TRUNC-X(I,J)
C         GSUM(J)=GSUM(J)+SCRAT(IGJ)
C         ACCUMULATE REAL SUMS
          SUMX(J)=SUMX(J)+X(I,J)
C         ACCUMULATE REAL AND GHOST(IF ACTIVE) CROSSPRODUCTS
          DO 30 K=1,J
            KSUB=J*(J-1)/2+K
            CROSSP(KSUB)=CROSSP(KSUD)+X(I,J)*X(I,K)
C           IGK=IGHS+K-1
C         GCROSS(KSUB)=GCROSS(KSUB)+SCRAT(IGJ)*SCRAT(IGK)
 30   CONTINUE
 10   CONTINUE
 200  FORMAT(6X,I6,6X,(10F10.5))
C
C     IF WE WROTE LESS THAN ND DAYS TO TO ZERO VALUES, WRITE A DUMMY RECORD
      IF(NDAYS.EQ.ND)RETURN
      DO 40 I=IX1,IXE1
        SCRAT(I)=-99.0
```

```
 40   CONTINUE
      NDP1=NDAYS+1
      JDAY(NDP1)=-1
      WRITE(ISIM,200)JDAY(NDP1),(SCRAT(I),I=IX1,IXE1)
C
      RETURN
      END
```

SUBROUTINE CSEED

```
      SUBROUTINE CSEED(DSEED)
      DOUBLE PRECISION DSEED
C
C     COMPUTES DOUBLE PRECISION INTEGER VALUED SEED FOR RANDOM NUMBER
C     GENERATORS. IT COMPUTES THE SEED AS A FUNCTION OF THE CP TIME
C     THE PROGRAM HAS USED. BLOWING IT UP BY 10**3 GIVES EMPHASIS
C     TO THE CP MILISECONDS, WHICH WE CAN EXPECT TO BE MORE
C     RANDOM THAN THE WHOLE SECONDS.
C     THIS IS MACHINE DEPENDENT
      CTIM=SECOND(DUM)
      CTIM=CTIM*1000
C     GET INTEGER
      ITIM=CTIM
      DSEED=ITIM
      RETURN
      END
```

SUBROUTINE SIMULA

```
      SUBROUTINE SIMULA (NSIM,NDSIM,TECH)
C
C     THIS VERSION OF SIMULA UTILIZES IMSL ROUTINE GGNSM TO GENERATE
C     SIMULATED VALUES. IT USES A CHOLESKY DECOMP OF THE SIGMA MATRIX
C     INSTEAD OF THE SINGULAR VALUED DECOMPOSITION METHOD TO OBTAIN THE
C     SQUARE ROOT.
C
      INTEGER TECH(1)
      COMMON/DATA/X(200,50)
      COMMON/COUNTS/FILL(100),NDAYS,NCELLS,N1,N2,NTOTC,NTOTD,NFINAL,
     1      INITC,INDX(50),INSET(200)
      COMMON/JULDAY/JDAY(400)
      COMMON/STATS/XMU(50),XSD(50)
      COMMON/ANCOV/SIGMA(1300),CROSSP(1300)
      COMMON/WORK/SCRAT(3000)
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
      COMMON/RAND/DSEED
      DOUBLE PRECISION DSEED
C
C     MAIN CONTROL MODULE FOR SIMULATION PROCEDURE. INPUTS ARE
C     /ANCOV/ SIGMA HAS THE VARIANCE/COVARIANCE MATRIX (SYMMETRIC STORAGE)
C     XMU HAS THE MEANS ESTIMATE
C     PROCEDURE -
C     THE COVARIANCE MATRIX IS COMPUTED FROM THE CORRELATION
C     MATRIX IN COV.
C     THE COVARIANCE MATRIX IS SPLIT INTO COMPONENTS AS SPEC-
C     IFIED IN THE QUARTERLY REPORT OF 1/1 - 3/31 80.
```

```
C     THESE COMPONENTS ARE STORED IN A MODE COMPATIBLE WITH
C     2D ARRAYS, THOUGH SUBSCRIPTS ARE FIGURED COMPUTATIONALLY
C     AND WE ACTUALLY DEAL WITH THEM AS 1D ARRAYS. THIS ALLOWS
C     ADJACENT STORAGE WITH NO WASTE.
C     THE COMPONENTS OF THE COVARIANCE ARE MANIPULATED AS
C     NECCESSARY, AND THE COVARIANCE MATRIX OF THE CONDITIONAL
C     DISTRIBUTION IS COMPUTED. THIS IS DECOMPOSED INTO ITS
C     SQUARE ROOT MATRIX.
C     THEN FOR EACH SIMULATION (NSIM TOTAL), WE GENERATE A
C     MEAN VECTOR FOR EACH DAY, GENERATE A RANDOM VECTOR WITH
C     THE COMPUTED DISTRIBUTION, AND WRITE IT TO FILE ISIM FOR
C     LATER ANALYSIS.
C     SCRAT(1) CONTAINS THE C11 COMPONENT OF THE COVARIANCE
C     MATRIX, THEN IT CONTAINS THE SIGMA SQUARE ROOT MATRIX).
C
C     IF N1=0,ALL INFORMATION IS KNOWN AND NO SIMULATION IS REQUIRED.
      IF(N1.EQ.0)RETURN
C
      WRITE(IOUT,100)
 100  FORMAT(//10X,21HS I M U L A T I O N S//)
      I11 = 1
      I12 = I11 + N1*N1
      I12T = I12 + N1*N2
      I22 = I12T + N1*N2
      ITERM = I22 + N2*N2
      NMAX = MAX0(N1,N2)
      IWK = ITERM + NMAX*NMAX
C     IF WE HAVE A CORRELATION AND NOT A COVARIANCE MATRIX, COMPUTE
C     COVARIANCE VIA ROUTINE COV(CORRELATION,COVARIANCE,NCELLS,ST DEVS)
C     CALL COV(CORMAT,SIGMA,NCELLS,XSD)
C
C     WRITE OUT COMPUTED VARIANCE/COVARIANCE MATRIX
      WRITE(IOUT,102)
 102  FORMAT(/10X,*ORIGINAL VARIANCE/COVARIANCE MATRIX*)
      CALL WRSYM(SIGMA,NCELLS)
C
C     SPLIT THIS MATRIX, STORE COMPONENTS IN VARIOUS SCRAT LOCATIONS
      CALL SPLIT(SIGMA,SCRAT(I11),SCRAT(I12),SCRAT(I12T),
     1  SCRAT(I22),N1,N2)
C
C     INVERT C22 COMPONENT OF THE COVARIANCE MATRIX
C     NOTE WE MOVE IT INTO WHAT WILL BE SCRAT(ITERM) = C12*C22INV
      CALL LINV1F(SCRAT(I22),N2,N2,SCRAT(ITERM),3,SCRAT(IWK),IER)
      IF(IER.EQ.129)RETURN
      N22=N2*N2
C     MOVE THIS BACK INTO C22, FREEING UP TERM
      DO 10 I=1,N22
        I22S=I22+I-1
        I22T=ITERM+I-1
        SCRAT(I22S)=SCRAT(I22T)
 10   CONTINUE
C
C     COMPUTE C11 MATRIX = COVARIANCE MATRIX OF CONDITIONAL DISTN.
      CALL CSIGMA(SCRAT(I11),SCRAT(I12),SCRAT(I12T),SCRAT(I22),
     1 SCRAT(IWK),SCRAT(ITERM),N1,N2)
C
C     I11 IS START OF C11 COMPONENT (COVARIANCES OF GROUP 1 CELLS).
C     CONVERT TO SYMMETRIC STORAGE TO SAVE SPACE, WRITE INTO
C     CROSSP.
      CALL VCVTFS(SCRAT(I11),N1,N1,CROSSP)
C     WRITE OUT FINAL C11 MATRIX
      WRITE(IOUT,150)
 150  FORMAT(/1H1,9X,"FINAL COVARIANCE MATRIX (C11) "/)
      CALL WRSYM(CROSSP,N1)
C
C     NOW DECOMPOSE THIS INTO THE OLD C11 LOCATION
      IWK1=IWK
      IWK2=IWK+N1
```

```
C      CALL SQRM(CROSSP,N1,SCRAT(I11),N1,SCRAT(IWK1),SCRAT(IWK2),IER)
C      ABOVE MOD ALLOWS US TO TEST IMSL GGNSM ROUTINE
C
       IF(IER.GE.128)RETURN
C      WRITE(IOUT,175)
C      DO 20 I=1,N1
C        ISTART=I+I11-1
C        IEND=I+N1*(N1-1)+I11-1
C        WRITE(IOUT,155)(SCRAT(J),J=ISTART,IEND,N1)
C20    CONTINUE
C      CALL WRSYM(SCRAT(I11),N1)
 155   FORMAT((1X,10(F10.5,1X)))
 175   FORMAT(/10X,25HSQUARE ROOT MATRIX OF C11)
C
C      NOW WE ARE READY FOR THE ACTUAL SIMULATIONS. FIRST COMPUTE
C      TARGET LOCATIONS FOR DAILY MEAN AND DAILY SIMULATED VALUE
C      VECTOR.
C
       IMEAN=IWK
       IVECT=IMEAN+N1
       IWK1=IVECT+NMAX
       IX=IWK1+1
       SCRAT(IWK1)=0.0
C
C      WRITE RECORD CONTAINING NUMBER OF SIMULATIONS IN THIS BLOCK
C      A BLOCK OF NSIM SIMULATIONS IS WRITTEN UPON EACH CALL
       WRITE(ISIM,700)NSIM,TECH(1),TECH(2)
 700   FORMAT(I3,2A4)
C
       DO 30 I=1,NSIM
C      WRITE SIMULATION HEADER INFO
       IF(I.EQ.1)WRITE(IOUT,200)I,(INDX(J),J=1,NCELLS)
 200   FORMAT(/10X,*NO. *,I3,/1X,*DAY*,10X,*CELLS*/4X,20(2X,I3,1X)/)
       WRITE(ISIM,701)I,NCELLS,NDSIM
 701   FORMAT(5I3)
       DO 35 J=1,NDSIM
C      COMPUTE DAILY MEANS VECTOR
       CALL CMEAN(XMU,SCRAT(ITERM),SCRAT(IMEAN),SCRAT(IVECT),
      1   J,N1,N2)
C
C      NOW GENERATE VECTOR N1 LONG OF N(0,1)
C      WRITE VECTOR TO SCRAT(IVECT)
       CALL GGNSM(DSEED,1,N1,CROSSP,1,SCRAT(IVECT),SCRAT(IWK1),IER)
       IF (IER.GE.129)RETURN
C      ADD MEANS - WRITE OVER OLD MEANS VECTOR
       DO 40 K=1,N1
       ISUB1=IMEAN+K-1
       ISUB2=IVECT+K-1
       SCRAT(ISUB1)=SCRAT(ISUB1)+SCRAT(ISUB2)
C      ANTILOG
       SCRAT(ISUB1)=EXP(SCRAT(ISUB1))
 40    CONTINUE
       DO 50 K=1,N2
         ISX1=IX+K-1
         ISX2=K+N1
         SCRAT(ISX1)=EXP(X(J,ISX2))
 50    CONTINUE
C      WRITE OUT
C      TO SAVE SPACE, WRITE ONLY THE FIRST SIMULATION TO UNIT IOUT, BUT
C      WRITE ALL TO ISIM
       IF(I.EQ.1)WRITE(IOUT,300)JDAY(J),(SCRAT(K),K=IMEAN,ISUB1),
      1       (SCRAT(K),K=IX,ISX1)
 300   FORMAT(1X,I3,1X,(20(1X,F5.1)))
       WRITE(ISIM,702)JDAY(J),(SCRAT(K),K=IMEAN,ISUB1),
      1       (SCRAT(K),K=IX,ISX1)
 702   FORMAT(6X,I6,6X,(10F10.5))
 35    CONTINUE
       SCRAT(IWK1)=1.0
```

```
 30    CONTINUE
       WRITE(IOUT,888)NSIM
 888   FORMAT(//10X,I3,2X,"SIMULATIONS WRITTEN TO RAW OUTPUT FILE")
       RETURN
       END
```

SUBROUTINE COV

```
       SUBROUTINE COV(CORMAT,SIGMA,N,XSD)
       DIMENSION CORMAT(1),SIGMA(1),XSD(1)
C
C      GIVEN A CORRELATION MATRIX COMPUTED BY ANY OF THE TECHNIQUES WE
C      USE, AS WELL AS A VECTOR OF STANDARD DEVIATIONS, THIS MODULE
C      COMPUTES THE CORRESPONDING COVARIANCE MATRIX. THE MATRICES
C      MAY OR MAY NOT SHARE STORAGE. N IS THE DIMENSION OF THE MATRICES
C      AND ALSO OF XSD.
C      SIGMA(I,J)=CORMAT(I,J)*XSD(I)*XSD(J)
C      IF I=J, WE GET JUST THE VARIANCE (DIAGONALS)
C
       DO 10 I=1,N
       IM1=I-1
       IDIAG=I*(I-1)/2 + I
       SIGMA(IDIAG) = XSD(I)*XSD(I)
       IF (I.EQ.1)GO TO 10
       DO 20 J=1,IM1
         ISUB=I*(I-1)/2 + J
         SIGMA(ISUB)=CORMAT(ISUB)*XSD(I)*XSD(J)
 20      CONTINUE
 10    CONTINUE
       RETURN
       END
```

SUBROUTINE CSIGMA

```
       SUBROUTINE CSIGMA (C11,C12,C12T,C22INV,WORK,TERM,N1,N2)
       DIMENSION C11(1),C12(1),C12T(1),C22INV(1),WORK(1),TERM(1)
C
C      COMPUTES SIGMA = FINAL COVARIANCE MATRIX OF THE CONDITIONAL DIST -
C      RIBUTION. ALL MATRICES ARE STORED IN FULL MODE. DYNAMIC ARRAY
C      ALLOCATION IS DONE BY TREATING ALL MATRICES AS 1 DIMENSION ARRAYS
C      COMPUTING THE SUBSCRIPT CORRESPONDING TO THE I,J ENTRIES. THIS
C      ALLOWS THE STORAGE AREA OF THESE MATRICES TO BE BROKEN UP
C      AS NEEDED FOR EACH PROBLEM WITH NO WASTE. THE LOGICAL DIMENSIONS
C      OF THE MATRICES ARE AS FOLLOWS
C
C      C11  - N1 BY N1
C      C12  - N1 BY N2
C      C12T - N2 BY N1
C      C22INV - N2 BY N2
C      TERM - N1 BY N2
C      SIGMA IS WRITTEN OVER C11 AND IS N1 BY N1
C
C      TERM = C12 * C22INV IS COMPUTED HERE AND SAVED FOR USE BY
C      CMEAN, WHICH COMPUTES THE FINAL MEAN VECTOR.
C      IMSL VMULFF IS USED FOR ALL MATRIX MULTIPLICATIONS
C
```

```
C      COMPUTE TERM
       CALL VMULFF(C12,C22INV,N1,N2,N2,N1,N2,TERM,N1,IER)
C      TERM * C12T = WORK (N1 BY N1)
       CALL VMULFF(TERM,C12T,N1,N2,N1,N1,N2,WORK,N1,IER)
C      NOW FINAL SIGMA = C11 - CURRENT WORK
C      IT IS WRITTEN INTO C11
       N11=N1*N1
       DO 10 I=1,N11
         C11(I)=C11(I)-WORK(I)
 10    CONTINUE
       RETURN
       END
```

SUBROUTINE CMEAN

```
       SUBROUTINE CMEAN (XMU,TERM,XMEAN,WORK,NDAY,N1,N2)
       DIMENSION XMU(1),TERM(1),XMEAN(1),WORK(1)
       COMMON /DATA/ X(200,50)
C
C      COMPUTES THE DAILY MEAN VECTOR GIVEN THE CELL MEANS AND
C      TERM (=C12 * C22INV)
C      TERM IS COMPUTED IN THE MODULE THAT COMPUTES SIGMA - IT IS
C      INVARIANT FROM DAY TO DAY.
C      ALL THE MATRICES ARE LOGICALLY TREATED AS 2D ARRAYS BUT COMP-
C      UTATIONALLY AS 1D AARRAYS. THIS ALLOWS US TO DYNAMICALLY
C      ALLOCATE THE SCRATCH WORK AREA AS NEEDED IN EACH PROBLEM.
C
C      COMUTE FIRST N2 ELEMENTS OF WORK = X2 - MU2
C
       ISTART=N1
       DO 10 I=1,N2
         IX=ISTART+I
         WORK(I)=X(NDAY,IX) - XMU(IX)
 10    CONTINUE
       ISTART=N2
C      NEXT N1 ELEMENTS OF WORK ARE C12 * C22INV * (X2 - M2)
       CALL VMULFF(TERM,WORK,N1,N2,1,N1,N2,WORK(ISTART),N1,IER)
C
C      NOW MEAN VECTOR M = MU1 + WORK (J,J=ISTART,N1+ISTART)
       DO 20 I=1,N1
         IW=ISTART+I-1
         XMEAN(I) = XMU(I) + WORK (IW)
 20    CONTINUE
       RETURN
       END
```

SUBROUTINE SPLIT

```
       SUBROUTINE SPLIT(M,M11,M12,M12T,M22,N1,N2)
       REAL M(1),M11(1),M12(1),M12T(1),M22(1)
C
C      TAKES A SQUARE MATRIX IN SYMMETRIC STORAGE MODE AND SPLITS IT
C      INTO 4 PIECES AS BELOW
C
C       I-----I-----I
C       I M11 I M12 I
C       I-----I-----I
C       I M12TI M22 I
C       I-----I-----I
C
C      M11 IS SQUARE N1 BY N1
C      M22 IS SQUARE N2 BY N2
C      M12 AND M12T (ITS TRANSPOSE) ARE N1 BY N2 AND N2 BY N1
       NMAX=N1+N2
       DO 10 I=1,NMAX
       DO 10 J=1,NMAX
         INDEX=I*(I-1)/2+J
         IF(J.GT.I)INDEX=J*(J-1)/2+I
         IF(I.GT.N1)GO TO 5
C        M11 AND M12 BOTH HAVE I LE N1, SOLVE FOR THEM NOW
         IF(J.GT.N1)GO TO 1
         NEWSB=I+N1*(J-1)
         M11(NEWSB)=M(INDEX)
         GO TO 10
 1       CONTINUE
         NEWSB=I+N1*(J-N1-1)
         M12(NEWSB)=M(INDEX)
         GO TO 10
C
 5       CONTINUE
C        I GT N1 HERE, SOLVE FOR M12T AND M22
         IF(J.GT.N1)GO TO 6
         NEWSB=(I-N1)+N2*(J-1)
         M12T(NEWSB)=M(INDEX)
         GO TO 10
 6       CONTINUE
         NEWSB=(I-N1)+N2*(J-N1-1)
         M22(NEWSB)=M(INDEX)
 10    CONTINUE
       RETURN
       END
```

SUBROUTINE INFO

```
       SUBROUTINE INFO
       COMMON/WORK/SCRAT(3000)
       COMMON/UNITS/IN,IOUT,ISIM,IPARM
       COMMON/COUNTS/EXCED(50),OVR100(50),NDAYS,NCELLS,N1,N2,NTC,NTD,
      1    NFINAL,INITC,INDX(50),INSET(200)
       INTEGER EXCED,OVR100
C
C      PRINTS CELL COUNT INFORMATION
C
```

```
      WRITE(IOUT,100)NDAYS,NFINAL,NCELLS,N1,N2,NTD,INITC,NTC
 100  FORMAT(1H1,10X,20HNUMBER OF DAYS       ,I4/
     1           11X,20HFINAL NO. DAYS       ,I4/
     1            11X,20HANALYZED CELLS      ,I4/
     2            11X,20HNUMBER IN GROUP 1   ,I4/
     3            11X,20HNUMBER IN GROUP 2   ,I4/
     4            11X,20HINITIAL NO. DAYS    ,I4/
     5            11X,20HINITIAL NO. CELLS   ,I4/
     6            11X,20HFINAL NO. OF CELLS  ,I4///
     7      11X,4HCELL,6X,11HCOORDINATES,5X,5HEXCED,3X,8HOVER 100/)
      IX=1
      IY=NCELLS+1
      DO 10 I=1,NCELLS
        J=INDX(I)
        ISX=IX+2*(J-1)
        ISY=ISX+1
        WRITE(IOUT,200)INDX(I),SCRAT(ISX),SCRAT(ISY),EXCED(I),
     1   OVR100(I)
 10   CONTINUE
 200  FORMAT(11X,I3,4X,F8.2,1H, ,F8.2,3X,I3,7X,I3)
      RETURN
      END
```

SUBROUTINE EMDRIV

```
      SUBROUTINE EMDRIV(SUMX,CROSSP,SIGMA,SIG,XMU,N,NO,XNU,P,WORK1,
     1  WORK2,T,METHEM,ITLIM,IER)
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
      INTEGER P
      DIMENSION SUMX(1),CROSSP(1),SIGMA(1),XMU(1),XNU(1),WORK1(1),
     1   WORK2(1),SIG(1)
C
C     MAIN DRIVER ROUTINE FOR THE EM ALGORITHM. INPUTS -
C     SUMX - SUMS VECTOR
C     CROSSP - CROSSPRODUCTS VECTOR
C     N - NUMBER OF OBS
C     NO - NUMBER OF OBS AFTER TRUNCATION
C     P - DIMENSION OF THE RANDOM VECTOR
C     T - THRESHHOLD
C     METHEM - INITIAL ESTIMATE METHOD. IF ZERO, USE SIMPLE ESTIMATE.
C             IF NONZERO, USE SUFFICIENT STATISTIC ESTIMATE FROM EMO.
C     ITLIM - ITERATION LIMIT.
C
C     OUTPUTS -
C     SIGMA - ESTIMATE OF VARIANCE - COVARIANCE
C     XMU - ESTIMATES OF MEANS
C     INTERNALS -
C     SIG - CHOLESKY DECOMP MATRIX OF SIGMA AT NEXT TO LAST ITERATION
C     XNU - SUM OF COLUMNS (=SUM OF ROWS) OF SIG
C     WORK1, WORK2 - WORK AREAS OF LENGTH P AND P*P
C
C     INPUT CHECK - IF N = NO OR NO = 0 QUIT
      IF(N.NE.NO.AND.NO.GT.0)GO TO 2
      IER=1
      WRITE(IOUT,99)N,NO
 99   FORMAT(/10X,21HEM INPUT ERROR - N = ,I4,10X,5HNO = ,I4)
      RETURN
 2    CONTINUE
C     SUBROUTINE EMO WOULD GIVE AN INITIAL ESTIMATE OF XMU AND SIGMA
C     BASED ON THE COMPLETE STATISTICS FOR THE FINAL SET (AFTER BOTH
C     TUNCATIONS).
```

```
      IF(METHEM.NE.0)CALL EMO(SUMX,CROSSP,N,NO,P,XMU,SIGMA)
      IDUM=0
      WRITE(IOUT,100)IDUM
 100  FORMAT(//10X,10HITERATION ,I3/10X,4HMEAN)
 101  FORMAT(/10X,5HSIGMA)
      WRITE(IOUT,200)(XMU(J),J=1,P)
 200  FORMAT(10X,10F12.7)
      WRITE(IOUT,101)
      DO 5 J=1,P
        ISTART=J*(J-1)/2+1
        IEND=ISTART+J-1
        WRITE(IOUT,200)(SIGMA(K),K=ISTART,IEND)
 5    CONTINUE
C
C     ITERATION LOOP. WE DO THE FOLLOWING -
C     1. TAKE THE CHOLESKY DECOMP OF SIGMA (INTO SIG)
C     2. GET CONDITIONAL EXPECTATIONS E1 AND E2 FROM ROUTINE ET
C     3. CALL EMITER TO DO THE COMPUTATIONS FOR ONE ITERATION
C     4.PRINT RESULTS OF ITERATION
C
      DO 10 ITER=1,ITLIM
        CALL LUDECP(SIGMA,SIG,P,D1,D2,IER)
        IF(IER.NE.0)RETURN
C       GET RECIPROCALS OF DIAGONAL ELEMENTS
        DO 15 I=1,P
          II=I*(I-1)/2+I
          SIG(II)=1.0/SIG(II)
 15     CONTINUE
        CALL ET(E1,E2,SIG,XMU,T,P,XNU,IER)
        IF(IER.NE.0)RETURN
        CALL EMITER(XMU,SIGMA,SIG,XNU,E1,E2,N,NO,P,SUMX,CROSSP,WORK1,
     1   WORK2)
        WRITE(IOUT,100)ITER
      WRITE(IOUT,200)(XMU(J),J=1,P)
      WRITE(IOUT,101)
      DO 20 J=1,P
        ISTART=J*(J-1)/2+1
        IEND=ISTART+J-1
        WRITE(IOUT,200)(SIGMA(K),K=ISTART,IEND)
 20   CONTINUE
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE EMO

```
      SUBROUTINE EMO(SUMX,CROSSP,N,NO,P,XMU,SIGMA)
      INTEGER P
      DIMENSION SUMX(1),CROSSP(1),XMU(1),SIGMA(1)
C
C     COMPUTES INITIAL (ITERATION 0) ESTIMATES OF MUHAT AND SIGMA
C     SUMX - THE SUMS VECTOR
C     CROSSP - THE CROSSPRODUCT MATRIX IN SYMMETRIC STORAGE MODE
C     N - INITIAL NUMBER OF OBSERVATIONS
C     NO - NUMBER OF OBS IN TRUNCATED SET
C     P - DIMENSION OF OBSERVATION VECTOR (=LEN(SUMX)=ORDER(CROSSP)
C
      DEN=FLOAT(N-NO)
      DO 10 I=1,P
C       COMPUTE MEAN VECTOR ENTRY
        XMU(I)=SUMX(I)/DEN
        DO 10 J=1,I
```

```
C         SINCE J LE I, WE HAVE ALL NEEDED MEANS FOR THIS CALCULATION
          ISUB=I*(I-1)/2+J
          SIGMA(ISUB)=CROSSP(ISUB)/DEN-XMU(I)*XMU(J)
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE EMITER

```
      SUBROUTINE EMITER(XMU,SIGMA,SIG,XNU,E1,E2,N,NO,P,SUMX,CROSSP,
     1   WORK1,WORK2)
      INTEGER P
      DIMENSION XMU(1),SIGMA(1),SIG(1),XNU(1),SUMX(1),CROSSP(1),
     1   WORK1(1),WORK2(1)
C
C     GIVEN ESTIMATES XMU AND SIGMA, EMITER PERFORMS ONE ITERATION OF
C     THE EM ALGORITHM
C     XMU - CURRENT MEAN ESTIMATE REPLACED BY NEW
C     SIGMA - CURRENT VAR - COVAR MATRIX IN SYMMETRIC STORAGE MODE.
C             ON OUTPUT, SIGMA HAS THE NEW SIGMA MATRIX
C     SIG - THE SQUARE ROOT MATRIX OF SIGMA (SYM.STORAGE)
C     XNU - THE VECTOR SIG.(1) NORMALIZED TO HAVE NORM 1 (SUM OF COLUMNS)
C     E1, E2 - E1(TSTAR) AND E2(TSTAR)
C     N - NUMBER OF OBS IN WHOLE SET
C     NO - NUMBER OF OBS IN TRUNCATED SET
C     P - ORDER OF THE MEAN VECTOR, SIGMA MATRIX, ETC.
C     SUMX - SUM OF THE ENTRIES ACCROSS ALL OBS IN THE DATA VECTOR
C     CROSSP - CROSSPRODUCTS MATRIX
C     WORK1 - WORK AREA OF LENGTH GE P
C     WORK2 - WORKA AREA OF LENGTH P*(P+1)/2
C
C     ADVANTAGE IS TAKEN OF THE VARIOUS SYMMETRIC MATRICES BOTH FOR
C     STORAGE AND COMPUTATION, REDUCING BOTH BY A QUANTITY THAT
C     APPROACHES A FACTOR OF 2 AS P GETS LARGE
C
C     COMPUTE SIG(XNU)
      CALL MULTV(SIG,XNU,WORK1,P,P)
C
      FN=FLOAT(N)
      FNO=FLOAT(NO)
C
C     COMPUTE XNU*TRAN(XNU), MOVE INTO WORK2 (SYM STOR MODE)
      CALL XXT (XNU,P,WORK2)
C     PERFORM NECCESSARY TRANSFORM ON WORK SO IT BECOMES I - (1-E2)XNU*XNUT
      DO 20 I=1,P
        DO 20 J=1,I
          ISUB=I*(I-1)/2+J
          WORK2(ISUB)=(E2-1.0)*WORK2(ISUB)
          IF(J.EQ.I)WORK2(ISUB)=1+WORK2(ISUB)
 20   CONTINUE
C
C     NOW SIG*WORK2*SIGT
      CALL ABATT(SIG,WORK2,P,SIGMA)
C
      DO 30 I=1,P
C       COMPUTE UPDATED MEAN EST - STORE FOR NOW IN WORK2
         WORK2(I)=(SUMX(I)+FNO*(XMU(I)+WORK1(I)*E1))/FN
        DO 35 J=1,I
          ISUB=I*(I-1)/2+J
C         NOW ADD E1*(SIG*XNU*XMUT + XMU*XNUT*SIG) TO WHAT WAS IN SIGMA
          SIGMA(ISUB)=SIGMA(ISUB)+E1*(WORK1(I)*XMU(J)+WORK1(J)*XMU(I))
```

```
C         NOW WE ARE READY TO COMPUTE NEW SIGMA, OVERWRITE OLD
          SIGMA(ISUB)=(CROSSP(ISUB)+FNO*(XMU(I)*XMU(J)+SIGMA(ISUB)))/FN-
           WORK2(I)*WORK2(J)
              35   CONTINUE
              30   CONTINUE
             C
             C     UPDATE MEAN VECTOR
 60                DO 40 I=1,P
                     XMU(I)=WORK2(I)
              40   CONTINUE
                   RETURN
                   END
```

SUBROUTINE ET

```
      SUBROUTINE ET(E1,E2,SIG,XMU,T,P,XNU,IER)
      DIMENSION XMU(1),XNU(1),SIG(1)
      INTEGER P
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
      EXTERNAL NORM
      REAL NORM
C
C     COMPUTES CONDITIONAL EXPECTATIONS E1 AND E2
C
C     UNIVARIATE NORMAL DENSITY IS SMLPHI
      SMLPHI(X)=EXP(-X*X/2.0)/SQRT(2.0*3.1415927)
C
C     XMU - MEAN VECTOR EST
C     SIG - SQUARE ROOT OF VARIANCE - COVARIANCE
      CALL SUMT(SIG,XNU,P)
      XL=NORM(XNU,P)
      IF(XL.EQ.0)GO TO 99
      DO 10 I=1,P
        XNU(I)=XNU(I)/XL
 10   CONTINUE
      TSTAR=(T*FLOAT(P)-SUMX(XMU,P))/XL
C     CALL IMSL TO GET CORRESPONDING PROBABILITY
      CALL MDNOR(TSTAR,PT)
      IF(PT.LE.0.0.OR.PT.GE.1.0)GO TO 99
      E1=-SMLPHI(TSTAR)/PT
      E2=1.0+TSTAR*E1
      WRITE(IOUT,101)TSTAR,PT,E1,E2
 101  FORMAT(/10X,6HTSTAR ,F10.5,3X,14HP(Y.LE.TSTAR) ,F8.6,3X,
     1    3HE1 ,F10.5,3X,3HE2 ,F10.5)
      IER=0
      RETURN
 99   WRITE(IOUT,100)XL,TSTAR,PT
 100  FORMAT(/10X,44HERROR IN CONDITIONAL EXPECTATION CALCULATION/
     1  10X,9HNORS(NU) ,F10.5,5X,6HTSTAR ,F10.5,5X,
     2  14HP(Y.LE.TSTAR) ,F8.6)
      IER=1
      RETURN
      END
```

SUBROUTINE MULTV

```
      SUBROUTINE MULTV(A,X,Y,N,NA)
      DIMENSION A(1),X(1),Y(1)
C
C     MULTIPLIES LOWER TRIANGULAR A STORED IN LOWER TRIANGULAR MODE
C     BY VECCTOR X YIELDING VECTOR 1
      DO 20 I=1,N
        Y(I)=0.0
        DO 10 J=1,I
          IJ=I*(I-1)/2+J
          Y(I)=Y(I)+A(IJ)*X(J)
 10   CONTINUE
 20   CONTINUE
      RETURN
      END
```

SUBROUTINE ABATT

```
      SUBROUTINE ABATT(A,B,N,C)
      DIMENSION A(1),B(1),C(1)
C
C     COMPUTES A*B*A TRANSPOSE, WHERE A IS LOWER TRIANGULAR AND B IS
C     SYMMETRIC. ONLY LOWER TRIANGULAR PORTIONS ARE STORED.
C     RESULT C IS ALSO SYMMTRIC
      DO 40 I=1,N
        DO 30 J=1,I
          SUMK=0.0
          DO 20 K=1,J
            SUML=0.0
            DO 10 L=1,I
             IL=I*(I-1)/2+L
             LK=L*(L-1)/2+K
             IF(K.GT.L)LK=K*(K-1)/2+L
             SUML=SUML+A(IL)*B(LK)
 10         CONTINUE
            JK=J*(J-1)/2+K
            SUMK=SUMK+A(JK)*SUML
 20   CONTINUE
        IJ=I*(I-1)/2+J
        C(IJ)=SUMK
 30    CONTINUE
 40   CONTINUE
      RETURN
      END
```

SUBROUTINE SUMT

```
      SUBROUTINE SUMT(X,SX,N)
      DIMENSION X(1),SX(1)
C
C     SUMS ROWS OF LOWER TRIANGULAR MATRIX A STORED IN COMPRESED MODE AND
C     RETURNS RESULT IN SX
C
      DO 5 I=1,N
        SX(I)=0.0
 5    CONTINUE
C
      DO 20 I=1,N
        DO 10 J=1,I
          ISUB=I*(I-1)/2+J
          SX(J)=SX(J)+X(ISUB)
 10     CONTINUE
 20   CONTINUE
      RETURN
      END
```

SUBROUTINE XXT

```
      SUBROUTINE XXT(X,N,C)
      DIMENSION X(1),C(1)
C
C     MULTIPLIES VECTOR X TIMES ITS TRANSPOSE, STORES RESULT IN SYMMETRIC
C     STORAGE MODE MATRIX C
      DO 10 I=1,N
        DO 20 J=1,I
          ISUB=I*(I-1)/2+J
          C(ISUB)=X(I)*X(J)
 20     CONTINUE
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE SUMM

```
      SUBROUTINE SUMM(X,S,N)
      DIMENSION X(1),S(1)
C
C     SUMS THE ROWS OF A MATRIX IN SYMMETRIC STORAGE MODE, RETURNS RESULT
C     IN VECTOR S. N = ORDER OF X = LEN(S)
      DO 10 I=1,N
        S(I)=0.0
        ICON=I*(I-1)/2
        DO 10 J=1,N
          ISUB=J*(J-1)/2+I
          IF(J.LE.I)ISUB=ICON+J
          S(I)=S(I)+X(ISUB)
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE SQRM

```
      SUBROUTINE SQRM(X,NX,SQRX,NSQ,WORK1,WORK2,IER)
      DIMENSION X(1),SQRX(1),WORK1(1),WORK2(1)
      INTEGER MSG(3)
      COMMON/UNITS/IN,IOUT,ISIM,IPARM
C
C     COMPUTES THE SQUARE ROOT MATRIX OF X, THAT IS, THE MATRIX A
C     SUCH THAT A*A=X. ALGEBRAICALLY, A = T (E**.5) T TRANSPOSE,
C     WHERE T IS THE EIGENVECTOR MATRIX AND E IS THE MATRIX
C     WITH EIGENVALUES ALONG THE DIAGONAL AND ZERO ELSEWHERE
C     WE WILL USE IMSL EIGRS FOR COMPUTING THE VALUES, AND THEN
C     USE A MINIMUM STORAGE FORMULA TO MULTIPLY THE RESULTS.
C     ARRAY X IS IN SYMMETRIC STORAGE MODE, SQRX IS IN
C     FULL STORAGE MODE (NSQ,M) WHERE M GE NSQ GE NX. NX IS
C     THE ORDER OF X. WORK AREAS MUST BE NX AND NX BY NX LONG,
C     RESPECTIVELY. A MESSAGE INDICATING THE PERFORMANCE INDEX
C     OF EIGRS IS PRINTED TO UNIT IOUT.
C
      DATA MSG /4HGOOD,4HFAIR,4HPOOR/
C
C     USE SQRX INITIALLY AS A WORK AREA FOR THE IMSL, MOVING
C     EIGENVALUES INTO WORK1 AND EIGENVECTORS INTO WORK2
C
      JOBN=2
      CALL EIGRS(X,NX,JOBN,WORK1,WORK2,NX,SQRX,IER)
      IF(IER.GE.128)RETURN
C     PERFORMANCE INDEX OUTPUT
      IF(SQRX(1).LT.1.0)I=1
      IF(SQRX(1).GE.1.0.AND.SQRX(1).LE.100.0)I=2
      IF(SQRX(1).GT.100.0)I=3
      WRITE(IOUT,100)SQRX(1),MSG(I)
 100  FORMAT(/10X,*EIGENVALUE ROUTINE PERFORMANCE INDEX *,F10.5,5X,A4)
      WRITE(IOUT,101)(WORK1(I),I=1,NX)
 101  FORMAT(/10X,12HEIGENVALUES: / (10X,12F10.5))
      DO 10 I=1,NX
      DO 10 J=1,I
C       COMPUTE SUBSCRIPT OF SQUARE ROOT MATRIX AND ZERO IT
        ISQ=I*(I-1)/2+J
        SQRX(ISQ)=0.0
        DO 10 K=1,NX
          IEIG=I+NX*(K-1)
          IEIGT=J+NX*(K-1)
          IF(WORK1(K).LT.0)GO TO 500
          SQRX(ISQ)=SQRX(ISQ)+WORK2(IEIG)*SORT(WORK1(K))*WORK2(IEIGT)
 10   CONTINUE
      RETURN
 500  IER=129
      WRITE(IOUT,501)
 501  FORMAT(/10X,*NEGATIVE EIGENVALUE*)
      RETURN
      END
```

SUBROUTINE NAIVES

```
      SUBROUTINE NAIVES(SUMX,CROSSP,N,NCELLS,XMU,SIGMA)
      DIMENSION SUMX(1),CROSSP(1),XMU(1),SIGMA(1)
C
C     GIVEN SUMS VECTOR AND CROSSPRODUCTS MATRIX, THIS ROUTINE COMPUTES
C     ESTIMATES OF THE MEAN VECTOR AND THE VARIANCE COVARIANCE MATRIX.
C     CROSSP AND SIGMA ARE ASSUMED TO BE IN SYMMETRIC STORAGE MODE.
C
      FN=FLOAT(N)
      FNM1=FN-1.0
      DO 10 I=1,NCELLS
        XMU(I)=SUMX(I)/FN
        DO 10 J=1,I
          IJ=I*(I-1)/2+J
          SIGMA(IJ)=(CROSSP(IJ)-FN*XMU(I)*XMU(J))/FNM1
 10   CONTINUE
      RETURN
      END
```

SUBROUTINE NEWSUM

```
      SUBROUTINE NEWSUM(X,NX,NCOLS,NROWS,SUMX,CROSSP)
      DIMENSION X(1),SUMX(1),CROSSP(1)
C
C     COMPUTES SUMS AND CROSSPRODUCTS OF COLUMNS OF MATRIX X.
C     NX IS THE ROW DIMENSION OF X AS SPECIFIED IN THE CALLING PROGRAM.
C     NCOLS AND NROWS ARE THE NUMBER OF ROWS AND COLUMNS OF X ACTUALLY
C     IN USE.
C     SUMX AND CROSSP ARE THE OUTPUT SUM VECTOR AND CROSSPRODUCT MATRIX,
C     RESPECTIVELY.
      DO 10 I=1,NCOLS
        SUMX(I)=0.0
        DO 10 J=1,I
          IJ=I*(I-1)/2+J
          CROSSP(IJ)=0.0
 10   CONTINUE
C
      DO 20 I=1,NROWS
      DO 20 J=1,NCOLS
        IJX=I+NX*(J-1)
        SUMX(J)=SUMX(J)+X(IJX)
        DO 20 K=1,J
          JK=J*(J-1)/2+K
          IKX=I+NX*(K-1)
          CROSSP(JK)=CROSSP(JK)+X(IJX)*X(IKX)
 20   CONTINUE
      RETURN
      END
```

SUBROUTINE COMPRS

```
      SUBROUTINE COMPRS(INSET,NFINAL,NCELLS)
      DINENSION INSET(1)
      COMMON/DATA/X(200,50)
      COMMON/JULDAY/JDAY(400)
C
C     COMPRESSES THE X VECTOR TO CONTAIN ONLY THE FINAL TRUNCATED SET
      DO 10 I=1,NFINAL
        ISUB=INSET(I)
        JDAY(I)=JDAY(ISUB)
        DO 10 J=1,NCELLS
          X(I,J)=X(ISUB,J)
 10   CONTINUE
      RETURN
      END
```

FUNCTION DOT

```
      FUNCTION DOT(X,Y,N)
      DIMENSION X(1),Y(1)
C     DOT PRODUCT OF VECTORS X AND Y (X(TRANSPOSE)*Y)
      DOT=0.0
      DO 10 I=1,N
        DOT=DOT+X(I)*Y(I)
 10   CONTINUE
      RETURN
      END
```

FUNCTION NORM

```
      REAL FUNCTION NORM(X,N)
      DIMENSION X(1)
C     COMPUTES NORM OF VECTOR X
      NORM=0.0
      DO 10 I=1,N
        NORM=NORM+X(I)*X(I)
 10   CONTINUE
      NORM=SQRT(NORM)
      RETURN
      END
```

FUNCTION SUMX

```
      FUNCTION SUMX(X,N)
C     COMPUTES SUM OF ENTRIES OF A VECTOR (X DOT (1,1,1...1) )
      DIMENSION X(1)
      SUMX=0.0
      DO 10 I=1,N
        SUMX=SUMX+X(I)
 10   CONTINUE
      RETURN
      END
```

APPENDIX E

PROCEDURES FOR DETERMINING CONCENTRATIONS AND ORIGINS OF BACKGROUND AND TRANSPORTED OZONE

The design value is determined by conditions that prevailed on the ten days when the highest ozone concentrations were observed or were estimated from the simulation methodologies. These days are identified by the isopleth methodology. Practical applications require that the origins and amounts of ozone transported into the area on these critical days be identified, so that it will be possible to estimate how much of the ozone was locally generated, and how much was of natural origin or transported from elsewhere. This appendix describes procedures for identifying natural ozone sources and for estimating the amounts and origins of ozone (and precursors) from other regions. The overall rationale and approach were presented in the body of this report, but for the sake of completeness, some of that material is repeated here.

BACKGROUND

Any objective scheme for estimating the transported ozone is likely to be unreliable because of the vagueness of the process. However, a semi-objective method is feasible and will provide information that can be used to develop strategies to achieve compliance with ozone standards. It is not necessary to evaluate transported ozone for every day in a data set for this purpose; at most, it is necessary to determine

transported ozone on days where exceedences have occurred, and in fact, it will usually be sufficient to determine the transported ozone component only for those ten days used to derive the design value. The presumption in designing a control strategy is that if the control strategy can reduce the concentrations on which the design value is based to the point where the revised design value no longer exceeds the standard; there will then be general compliance with the standard. Thus, one need only examine the meteorological conditions that prevailed for the days from which the design value was estimated.

To develop an effective control strategy, the transported ozone entering the area must be known, so that the local contribution to the observed ozone concentrations can be evaluated. Some sort of modeling is then applied to evaluate the effectiveness of various proposed control strategies. Several basic approaches to modeling are possible, including:

- Rollback.
- The Empirical Kinetic Modeling Approach--EKMA (23).
- Langrangian (trajectory) photochemical modeling.
- Eulerian (airshed) photochemical modeling.

Each type of model has slightly different requirements for transported-ozone inputs; however, the first three are quite similar. The last (Eulerian) approach requires more information about the temporal and

spatial variability in transported ozone than do the other three.

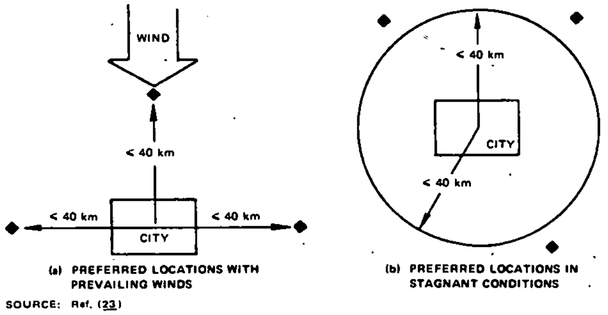
Because most regions of the United States will impose control strategies in the future, the background ozone concentration transported from an area where control measures have been imposed will change. Thus, the evaluation of transported ozone must provide for estimating effects of future control measures--upwind or in the area itself--by calculating historical air trajectories to determine if the source of transported ozone was in areas where control measures might be invoked in future years.

Finally, ozone concentrations reach relatively high values from natural causes in certain situations, e.g., intrusion of stratospheric ozone to relatively low altitudes. Methods are needed to recognize the meteorological situtations that might accompany high ozone concentrations from natural causes. The guidelines for interpreting the ozone standard (1) provide that days can be excluded from consideration by an EPA Regional Administrator; presumably, days when ozone concentrations exceeded the standard by virtue of natural causes could be excluded.

One of the first guides given for estimating transported ozone was included with the discussions of the EKMA (23), where it was stated that three-hour average surface ozone readings are most indicative of ozone concentrations through the mixing layer when they are made upwind (or at least outside the urban plume) after the nocturnal radiative inversion has broken, but early enough in the day that photochemical ozone will not have formed from residual precursors in the transported air. The

EKMA discussions suggest that measurements should be made between about 1100 and 1300 (local time). As will be shown later, there are often "signatures" in commonly available meteorological data that can be used to define the time of breakup of the nocturnal inversion so strict adherence to the above objective definition may not be desirable. Figure E-1, from EPA (23), shows the generalized locations considered acceptable for determining transported ozone according to the EKMA reports.

A somewhat different approach to the estimation of background ozone for use with EKMA has been taken by Maxwell and Martinez (24), who suggest selecting background ozone concentrations on the basis of vector-averaged 0600–1400 wind direction; the background ozone concentration is assumed to be the peak ozone concentration recorded during the day at the most distant station in the upwind direction. This method tacitly includes another component in the background or transported ozone concentration. It is assumed that the peak ozone concentration at an upwind site would have been observed whether or not the city existed. This is tantamount to a redefinition of background ozone to include not only the ozone that is transported as ozone, but also ozone that can be generated photochemically from transported precursors. The user may consider this to be a more valid definition for purposes of designing control strategies than the more limited definition included with the EKMA documentation (23). It is the approach underlying the procedure described below.



(a) PREFERRED LOCATIONS WITH PREVAILING WINDS

(b) PREFERRED LOCATIONS IN STAGNANT CONDITIONS

SOURCE: Ref. (23)

FIGURE E-1 EXAMPLE OF ACCEPTABLE MONITORING LOCATIONS FOR ESTIMATING TRANSPORTED OZONE

Figure E-2 is a schematic diagram of the procedure described below for identifying transported or background ozone components. As is seen in the figure, the process begins by identifying the days from which estimates of the design value are to be obtained. The next step is to acquire all the air quality and meterological information that is available for those days: The best estimate of background ozone will be obtained when all available meteorological and air quality data have been carefully considered. The data should be examined to determine whether the observed ozone concentrations might have arisen from natural causes. For example, areas to the south and west of troughs in the upper air circulation are candidates for intrusions of ozone-rich stratospheric air. If there is reason to believe that a natural source is implicated, the next step would be to contact the EPA Regional Administrator and request that the days in question be examined and exempted from consideration for purposes of determining the design value.

If no natural causes are evident, the next step is to determine whether air movement (and the direction of movement) is generally well defined or whether conditions are more stagnant. Figure E-1 indicated types of stations that might be used to define transported ozone concentrations when the wind direction is well-defined [Figure E-1(a)] and when it is stagnant [Figure E-1(b)]. In the latter case, the ozone concentration that is sought is not a transported ozone, but rather a background value. In either event, the next step is to determine whether the available measurements of ozone at the surface were representative of concentrations through the mixed layer. (In some special cases,

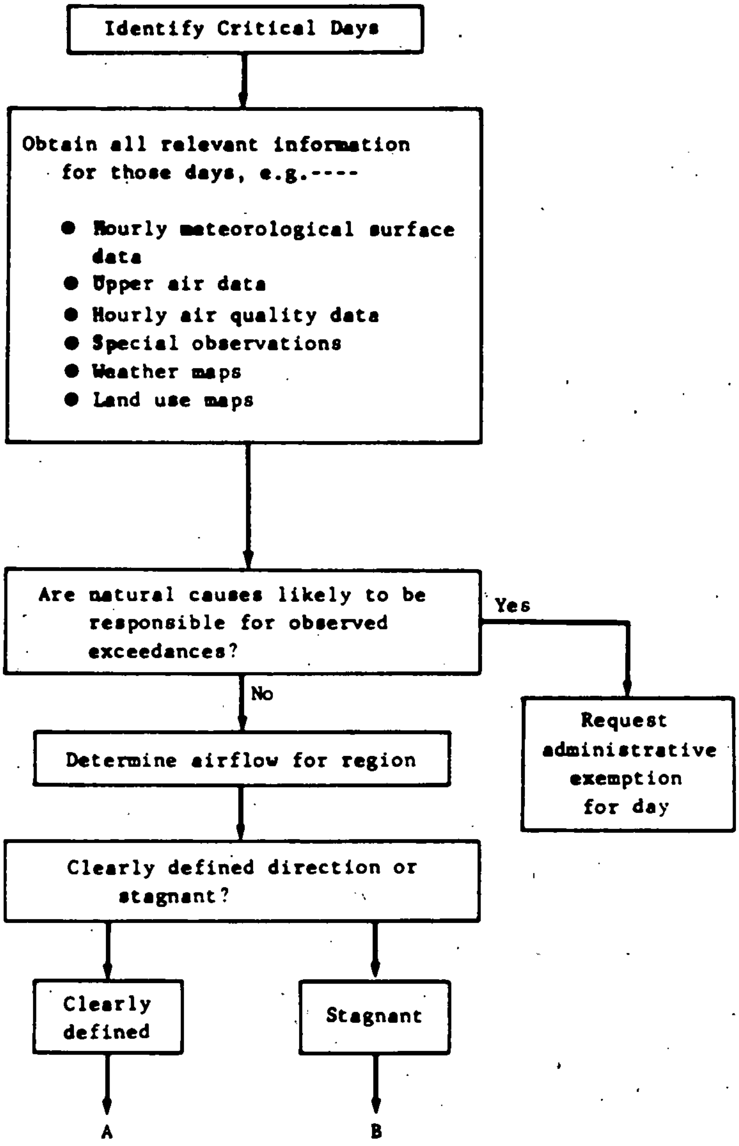


FIGURE E-2 SCHEMATIC DIAGRAM OF A PROCEDURE FOR ESTIMATING TRANSPORTED AND BACKGROUND OZONE

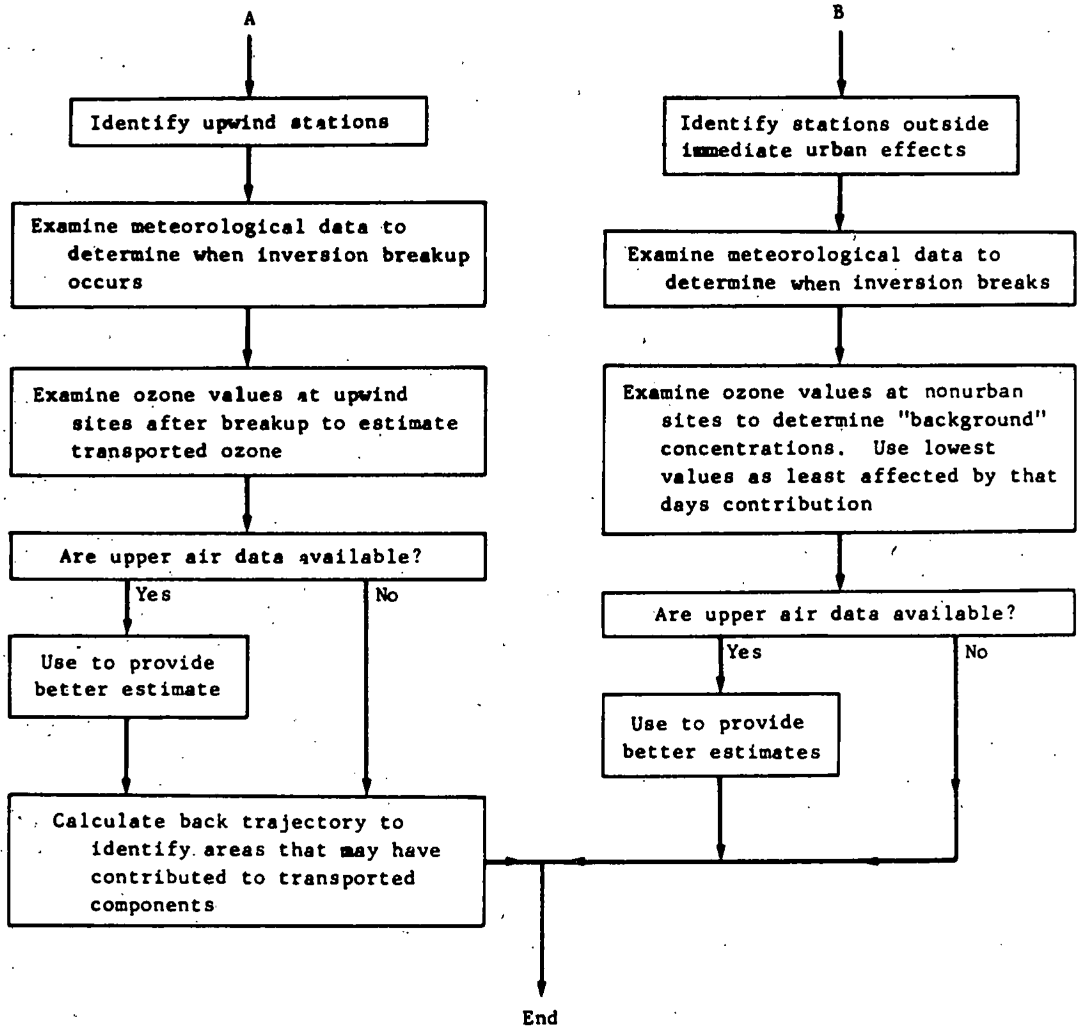


FIGURE E-2 SCHEMATIC DIAGRAM OF A PROCEDURE FOR ESTIMATING TRANSPORTED AND BACKGROUND OZONE (Concluded) aircraft measurements of ozone in the mixing layer may be available and should be used.) Finally, if the air has been transported from elsewhere, the source regions for that air are identified for the critical days, providing a basis for estimating how the transported background ozone might change when future control measures are implemented.

The following sections are based on Figure E-2. The various requirements and decisions are discussed with recommendations regarding how they may be accomplished. Examples are given where appropriate.

THE PROCEDURE

Identify Critical Days

As shown in Figure E-2, the process begins by defining critical days, i.e., those from which the estimates of the design value (1) have been obtained. These days are determined by the isopleth methodology, and are those days when the highest daily-maximum ozone concentrations have occurred. Because the object of the analysis is developing strategies to achieve compliance with the standard, it is not necessary to address days that are not involved in the design value process. Days that are used for the calculations of design value will be the ones for which transported ozone should be determined.

Obtain Relevant Data

Meteorological Data. These data are generally available from the National Climatic Center (NCC).* The NCC attempts to obtain a copy of all meteorological records collected in the United States. These data are available and can be ordered on microfilm, magnetic tape, hard copies, or as copies of raw data. A guide to available data is available from NCC (36). The Center answers inquiries and analyzes, evaluates, and interprets data. Routine letters or telephone inquiries are usually answered without charge; other services are provided at cost.

The bulk of the data at the Climatic Center is meteorological observations made at airfields by the National Weather Service, the Federal Aviation Administration, and the Defense Department. Figure E-3 shows an example of the kind of information to be found on a Climatological Data Monthly Summary form for one month at one station. More detailed meteorological data can be obtained from copies of National Weather Service WBAN Form 10A (an example is shown in Figure E-4). The WBAN form contains hourly data for sky cover, visibility, temperature, humidity, wind speed, and wind direction.

---

*Director, National Climatic Center, Federal Building, Asheville, North Carolina 28801, Telephone: (704) 258-2850.

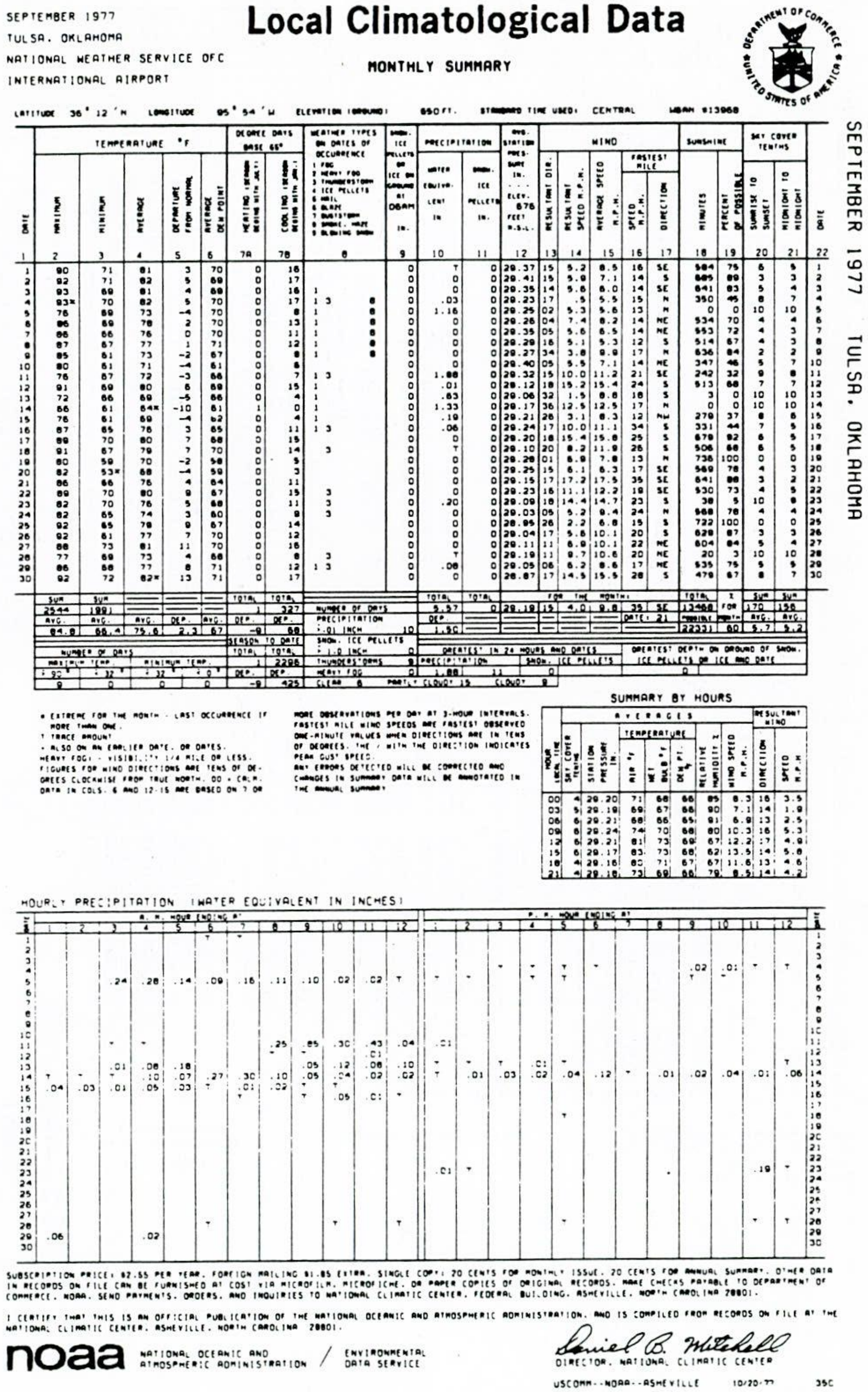
Twice-daily upper air maps and surface weather maps at three-hour intervals can also be obtained on microfilm from the NCC. Examples of these maps are shown in Figures E-5 and E-6. Daily surface weather maps for 0700 EST (Eastern Standard Time) for the United States are published routinely and mailed to subscribers once per week. Any agency responsible for interpretation of air quality data will probably find that a subscription to this series of weather maps* is very useful. Figure E-7 shows a map from this series.

Schools, radio and television stations, industrial complexes, fire stations, highway and transportation departments, universities and colleges, environmental studies groups, air pollution districts, and utility districts may have continuing meteorological records or special weather studies available. A direct call to these agencies may yield useful data not available elsewhere.

Air Quality Data. Data for the days of interest should be acquired. Records of the hourly concentration records of ozone, oxides of nitrogen, and nonmethane organic hydrocarbon (NMOC) stations in the study area should be reviewed. Data for the immediate study area should already be available, because these data are needed as input to the analysis procedure for determining the ozone design values and exceedances. Later, when the origins of the air entering the city have

---

*"The Daily Weather Map--Weekly Series" is for sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. At this writing, the annual subscription rate in the United States is $32.00.



SEPTEMBER 1977
TULSA, OKLAHOMA
NATIONAL WEATHER SERVICE OFC
INTERNATIONAL AIRPORT

Local Climatological Data

MONTHLY SUMMARY

LATITUDE 36° 12′ N LONGITUDE 95° 54′ W ELEVATION (GROUND) 650 FT. STANDARD TIME USED: CENTRAL WBAN #13968

SEPTEMBER 1977 TULSA, OKLAHOMA

DATE	TEMPERATURE °F MAXIMUM	MINIMUM	AVERAGE	DEPARTURE FROM NORMAL	AVERAGE DEW POINT	DEGREE DAYS BASE 65° HEATING (SEASON BEGINS WITH JULY)	COOLING (SEASON BEGINS WITH JAN.)	WEATHER TYPES ON DATES OF OCCURRENCE 1 FOG 2 HEAVY FOG 3 THUNDERSTORM 4 ICE PELLETS 5 HAIL 6 GLAZE 7 DUSTSTORM 8 SMOKE, HAZE 9 BLOWING SNOW	SNOW, ICE PELLETS OR ICE ON GROUND AT 06AM IN.	PRECIPITATION WATER EQUIVALENT IN	SNOW, ICE PELLETS IN.	AVG. STATION PRESSURE IN. ELEV. 676 FEET M.S.L.	WIND RESULTANT DIR.	RESULTANT SPEED M.P.H.	AVERAGE SPEED M.P.H.	FASTEST MILE SPEED M.P.H.	FASTEST MILE DIRECTION	SUNSHINE MINUTES	PERCENT OF POSSIBLE	SKY COVER TENTHS SUNRISE TO SUNSET	MIDNIGHT TO MIDNIGHT	DATE
1	2	3	4	5	6	7A	7B	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	90	71	81	3	70	0	16		0	T	0	29.37	15	5.2	8.5	16	SE	584	75	6	5	1
2	92	71	82	5	69	0	17		0	0	0	29.41	15	5.9	7.1	14	S	685	89	3	3	2
3	93	69	81	4	69	0	16	1	0	0	0	29.35	14	5.6	6.0	14	SE	641	83	5	4	3
4	93*	70	82	5	70	0	17	1 3 8	0	.03	0	29.23	17	.5	5.5	15	N	350	45	8	7	4
5	76	69	73	-4	70	0	8	1 8	0	1.16	0	29.25	02	5.3	5.6	13	N	0	0	10	10	5
6	86	69	78	2	70	0	13	1 8	0	0	0	29.26	04	7.4	8.2	14	NE	534	70	4	4	6
7	86	66	76	0	70	0	11	1 8	0	0	0	29.35	05	5.6	6.5	14	NE	553	72	4	3	7
8	87	67	77	1	71	0	12	1 8	0	0	0	29.29	16	5.1	5.3	12	S	514	67	4	3	8
9	85	61	73	-2	67	0	8	1 8	0	0	0	29.27	34	3.8	9.9	17	N	636	84	2	2	9
10	80	61	71	-4	61	0	6		0	0	0	29.40	05	5.5	7.1	14	NE	347	46	5	7	10
11	76	67	72	-3	66	0	7	1 3	0	1.88	0	29.32	15	10.0	11.2	21	SE	242	32	9	8	11
12	91	69	80	6	69	0	15	1	0	.01	0	29.12	18	15.2	15.4	24	S	513	68	7	7	12
13	72	66	69	-5	66	0	4	1	0	.63	0	29.06	32	1.5	8.8	18	S	3	0	10	10	13
14	66	61	64*	-10	61	1	0	1	0	1.33	0	29.17	36	12.5	12.5	17	N	0	0	10	10	14
15	76	61	69	-4	62	0	4	1	0	.19	0	29.21	28	3.1	8.3	12	NW	279	37	8	6	15
16	87	65	76	3	65	0	11	1 3	0	.06	0	29.24	17	10.0	11.1	34	S	331	44	7	5	16
17	89	70	80	7	68	0	15		0	0	0	29.20	18	15.4	15.8	25	S	679	92	6	5	17
18	91	67	79	7	70	0	14	3	0	T	0	29.10	20	8.2	11.9	26	S	506	68	6	5	18
19	80	59	70	-2	58	0	5		0	0	0	29.28	01	6.9	7.8	13	N	738	100	0	0	19
20	82	53*	68	-4	59	0	3		0	0	0	29.25	15	6.1	6.3	17	SE	569	78	4	3	20
21	86	66	76	4	64	0	11		0	0	0	29.15	17	17.2	17.5	35	SE	641	88	3	2	21
22	89	70	80	9	67	0	15	3	0	0	0	29.23	16	11.1	12.2	19	SE	530	73	4	5	22
23	82	70	76	5	68	0	11	3	0	.20	0	29.09	18	14.4	14.7	23	S	38	5	10	8	23
24	82	65	74	3	60	0	9	3	0	0	0	29.03	05	5.2	9.4	24	N	568	78	4	4	24
25	92	65	79	9	67	0	14		0	0	0	28.95	26	2.2	6.8	15	S	722	100	0	0	25
26	92	61	77	7	70	0	12		0	0	0	29.04	17	5.6	10.1	20	S	629	87	3	3	26
27	88	73	81	11	70	0	16		0	0	0	29.11	11	6.9	10.1	22	NE	604	84	5	4	27
28	77	69	73	4	68	0	8	3	0	T	0	29.19	11	9.7	10.6	20	NE	20	3	10	10	28
29	86	68	77	8	71	0	12	1 3	0	.08	0	29.05	06	6.2	8.6	17	NE	535	75	5	5	29
30	92	72	82*	13	71	0	17		0	0	0	28.87	17	14.5	15.5	28	S	479	67	8	7	30

SUM	SUM				TOTAL	TOTAL
2544	1991				1	327
AVG.	AVG.	AVG.	DEP.	AVG.	DEP.	DEP.
84.8	66.4	75.6	2.3	67	-9	68
					SEASON TO DATE	
NUMBER OF DAYS					TOTAL	TOTAL
MAXIMUM TEMP.		MINIMUM TEMP.			1	2298
≥ 90°	≤ 32°	≤ 32°	≤ 0°		DEP.	DEP.
9	0	0	0		-9	425

NUMBER OF DAYS
PRECIPITATION ≥ .01 INCH 10
SNOW, ICE PELLETS ≥ 1.0 INCH 0
THUNDERSTORMS 9
HEAVY FOG 0
CLEAR 6 PARTLY CLOUDY 15 CLOUDY 9

TOTAL	TOTAL	FOR THE MONTH:						TOTAL	% FOR MONTH		SUM	SUM
5.57	0	29.19	15	4.0	9.8	35	SE	13468			170	156
DEP.						DATE: 21		POSSIBLE MONTH			AVG.	AVG.
1.50								22331	60		5.7	5.2

GREATEST IN 24 HOURS AND DATES				GREATEST DEPTH ON GROUND OF SNOW, ICE PELLETS OR ICE AND DATE	
PRECIPITATION		SNOW, ICE PELLETS			
1.88	11	0		0	

* EXTREME FOR THE MONTH - LAST OCCURRENCE IF MORE THAN ONE.
T TRACE AMOUNT.
+ ALSO ON AN EARLIER DATE, OR DATES.
HEAVY FOG: - VISIBILITY 1/4 MILE OR LESS.
FIGURES FOR WIND DIRECTIONS ARE TENS OF DEGREES CLOCKWISE FROM TRUE NORTH. 00 = CALM.
DATA IN COLS. 6 AND 12-15 ARE BASED ON 7 OR MORE OBSERVATIONS PER DAY AT 3-HOUR INTERVALS.
FASTEST MILE WIND SPEEDS ARE FASTEST OBSERVED ONE-MINUTE VALUES WHEN DIRECTIONS ARE IN TENS OF DEGREES. THE / WITH THE DIRECTION INDICATES PEAK GUST SPEED.
ANY ERRORS DETECTED WILL BE CORRECTED AND CHANGES IN SUMMARY DATA WILL BE ANNOTATED IN THE ANNUAL SUMMARY.

SUMMARY BY HOURS

HOUR LOCAL TIME	AVERAGES SKY COVER TENTHS	STATION PRESSURE IN.	TEMPERATURE AIR °F	WET BULB °F	DEW PT. °F	RELATIVE HUMIDITY %	WIND SPEED M.P.H.	RESULTANT WIND DIRECTION	SPEED M.P.H.
00	4	29.20	71	68	66	85	8.3	16	3.5
03	5	29.19	69	67	66	90	7.1	14	1.9
06	6	29.21	68	66	65	91	6.9	13	2.5
09	6	29.24	74	70	68	80	10.3	16	5.3
12	6	29.21	81	73	69	67	12.2	17	4.9
15	6	29.17	83	73	68	62	13.5	14	5.8
18	4	29.16	80	71	67	67	11.6	13	4.6
21	4	29.18	73	69	66	79	8.5	14	4.2

HOURLY PRECIPITATION (WATER EQUIVALENT IN INCHES)

DATE	A.M. HOUR ENDING AT 1	2	3	4	5	6	7	8	9	10	11	12	P.M. HOUR ENDING AT 1	2	3	4	5	6	7	8	9	10	11	12	DATE
1						T	T																		1
2																									2
3																									3
4															T	T	T	T			.02	.01	T	T	4
5			.24	.28	.14	.09	.16	.11	.10	.02	.02	T	T	T		T	T				T	T			5
6																									6
7																									7
8																									8
9																									9
10																									10
11			T	T				.25	.85	.30	.43	.04	.01												11
12								T	T		.01														12
13			.01	.08	.18				.05	.12	.08	.10	T	T	T	.01	T							T	13
14	T	T	T	.10	.07	.27	.30	.10	.05	.04	.02	.02	T	.01	.03	.02	.04	.12	T	.01	.02	.04	.01	.06	14
15	.04	.03	.01	.05	.03	T	.01	.02	T	T															15
16							T		T	.05	.01	T													16
17																									17
18																	T								18
19																									19
20																									20
21																									21
22																									22
23													.01	T									.19	T	23
24																									24
25																									25
26																									26
27																									27
28						T				T		T					T						T	T	28
29	.06			.02																					29
30																									30

SUBSCRIPTION PRICE: $2.55 PER YEAR. FOREIGN MAILING $1.85 EXTRA. SINGLE COPY: 20 CENTS FOR MONTHLY ISSUE, 20 CENTS FOR ANNUAL SUMMARY. OTHER DATA IN RECORDS ON FILE CAN BE FURNISHED AT COST VIA MICROFILM, MICROFICHE, OR PAPER COPIES OF ORIGINAL RECORDS. MAKE CHECKS PAYABLE TO DEPARTMENT OF COMMERCE, NOAA. SEND PAYMENTS, ORDERS, AND INQUIRIES TO NATIONAL CLIMATIC CENTER, FEDERAL BUILDING, ASHEVILLE, NORTH CAROLINA 28801.

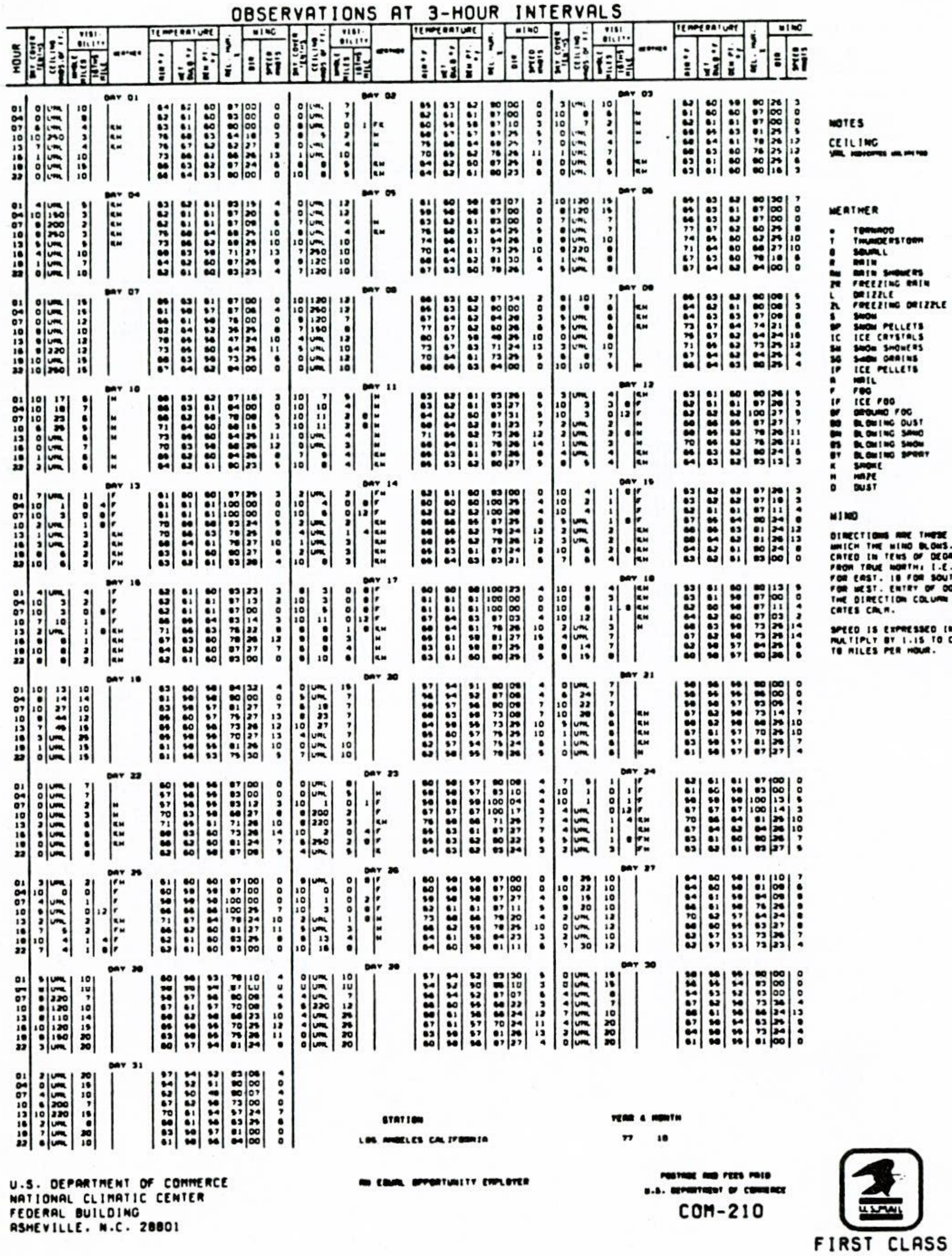
I CERTIFY THAT THIS IS AN OFFICIAL PUBLICATION OF THE NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, AND IS COMPILED FROM RECORDS ON FILE AT THE NATIONAL CLIMATIC CENTER, ASHEVILLE, NORTH CAROLINA 28801.

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / ENVIRONMENTAL DATA SERVICE

Daniel B. Mitchell
DIRECTOR, NATIONAL CLIMATIC CENTER

USCOMM--NOAA--ASHEVILLE 10/20/77 35C

FIGURE E-3 EXAMPLE OF MONTHLY METEOROLOGICAL SUMMARY



OBSERVATIONS AT 3-HOUR INTERVALS

NOTES

CEILING
UNL DENOTES UNLIMITED

WEATHER

•	TORNADO
T	THUNDERSTORM
Q	SQUALL
R	RAIN
RW	RAIN SHOWERS
ZR	FREEZING RAIN
L	DRIZZLE
ZL	FREEZING DRIZZLE
S	SNOW
SP	SNOW PELLETS
IC	ICE CRYSTALS
SW	SNOW SHOWERS
SG	SNOW GRAINS
IP	ICE PELLETS
A	HAIL
F	FOG
IF	ICE FOG
GF	GROUND FOG
BD	BLOWING DUST
BN	BLOWING SAND
BS	BLOWING SNOW
BY	BLOWING SPRAY
K	SMOKE
H	HAZE
D	DUST

WIND

DIRECTIONS ARE THOSE FROM WHICH THE WIND BLOWS, INDICATED IN TENS OF DEGREES FROM TRUE NORTH; I.E., 09 FOR EAST, 18 FOR SOUTH, 27 FOR WEST. ENTRY OF 00 IN THE DIRECTION COLUMN INDICATES CALM.

SPEED IS EXPRESSED IN KNOTS; MULTIPLY BY 1.15 TO CONVERT TO MILES PER HOUR.

STATION
LOS ANGELES CALIFORNIA

YEAR & MONTH
77 10

U.S. DEPARTMENT OF COMMERCE
NATIONAL CLIMATIC CENTER
FEDERAL BUILDING
ASHEVILLE, N.C. 28801

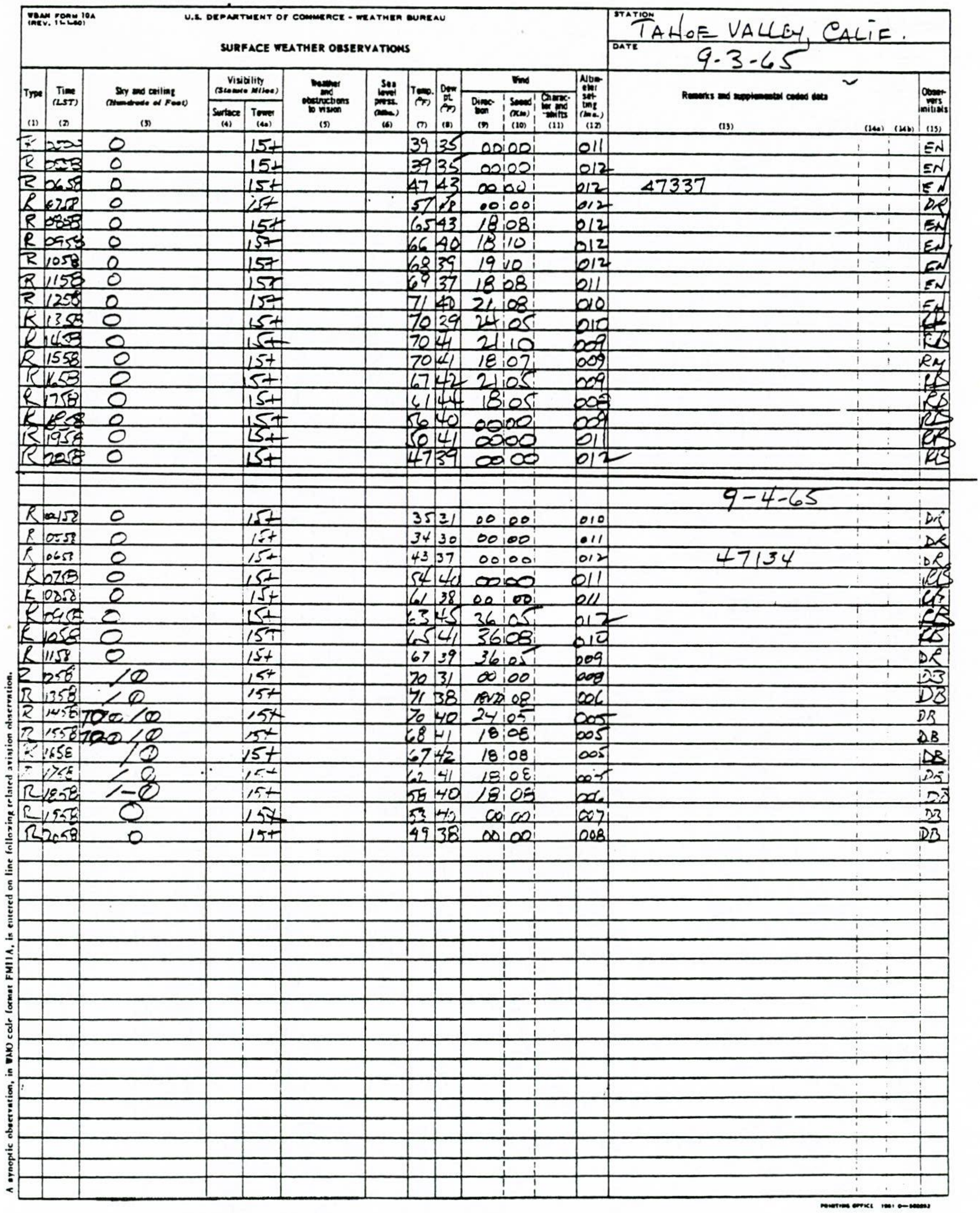
AN EQUAL OPPORTUNITY EMPLOYER

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE
COM-210

U.S.MAIL

FIRST CLASS

FIGURE E-3 EXAMPLE OF MONTHLY METEOROLOGICAL SUMMARY (Concluded)



WBAN FORM 10A
(REV. 1-1-60)

U.S. DEPARTMENT OF COMMERCE - WEATHER BUREAU

SURFACE WEATHER OBSERVATIONS

STATION TAHOE VALLEY, CALIF.

DATE 9-3-65

Type (1)	Time (LST) (2)	Sky and ceiling (Hundreds of Feet) (3)	Visibility (Statute Miles) Surface (4)	Visibility Tower (4a)	Weather and obstructions to vision (5)	Sea level press. (Mbs.) (6)	Temp. (°F) (7)	Dew pt. (°F) (8)	Wind Direction (9)	Wind Speed (Kts) (10)	Wind Character and shifts (11)	Altimeter setting (Ins.) (12)	Remarks and supplemental coded data (13)	(14a)	(14b)	Observers initials (15)
R	[illegible]	O		15+			39	35	00	00		011				EN
R	0558	O		15+			39	35	00	00		012				EN
R	0658	O		15+			47	43	00	00		012	47337			EN
R	0728	O		15+			57	48	00	00		012				DR
R	0858	O		15+			65	43	18	08		012				EN
R	0958	O		15+			66	40	18	10		012				EN
R	1058	O		15+			68	39	19	10		012				EN
R	1158	O		15+			69	37	18	08		011				EN
R	1258	O		15+			71	40	21	08		010				EN
R	1358	O		15+			70	39	24	05		010				RB
R	1458	O		15+			70	41	21	10		009				RB
R	1558	O		15+			70	41	18	07		009				RM
R	1658	O		15+			67	42	21	05		009				RB
R	1758	O		15+			61	44	18	05		008				RB
R	1858	O		15+			56	40	00	00		009				RB
R	1958	O		15+			50	41	00	00		011				RB
R	2058	O		15+			47	39	00	00		012				RB
													9-4-65			
R	0458	O		15+			35	31	00	00		010				DR
R	0558	O		15+			34	30	00	00		011				DR
R	0658	O		15+			43	37	00	00		012	47134			DR
R	0758	O		15+			54	40	00	00		011				RB
R	0858	O		15+			61	38	00	00		011				CB
R	0958	O		15+			63	45	36	05		012				RB
R	1058	O		15+			65	41	36	08		010				RB
R	1158	O		15+			67	39	36	05		009				DR
R	1258	/⊘		15+			70	31	00	00		008				DB
R	1358	/⊘		15+			71	38	16V20	08		006				DB
R	1458	100 /⊘		15+			70	40	24	05		005				DR
R	1558	100 /⊘		15+			68	41	18	08		005				DB
R	1658	/⊘		15+			67	42	18	08		005				DB
R	1758	/⊘		15+			62	41	18	08		005				DB
R	1858	/-⊘		15+			56	40	18	05		006				DB
R	1958	O		15+			53	40	00	00		007				DB
R	2058	O		15+			49	38	00	00		008				DB

A synoptic observation, in WMO code (format FM11A), is entered on line following related aviation observation.

PRINTING OFFICE 1961 O—[illegible]

FIGURE E-4 EXAMPLE OF WEATHER INFORMATION RECORDED ON NATIONAL WEATHER SERVICE WBAN FORM 10A

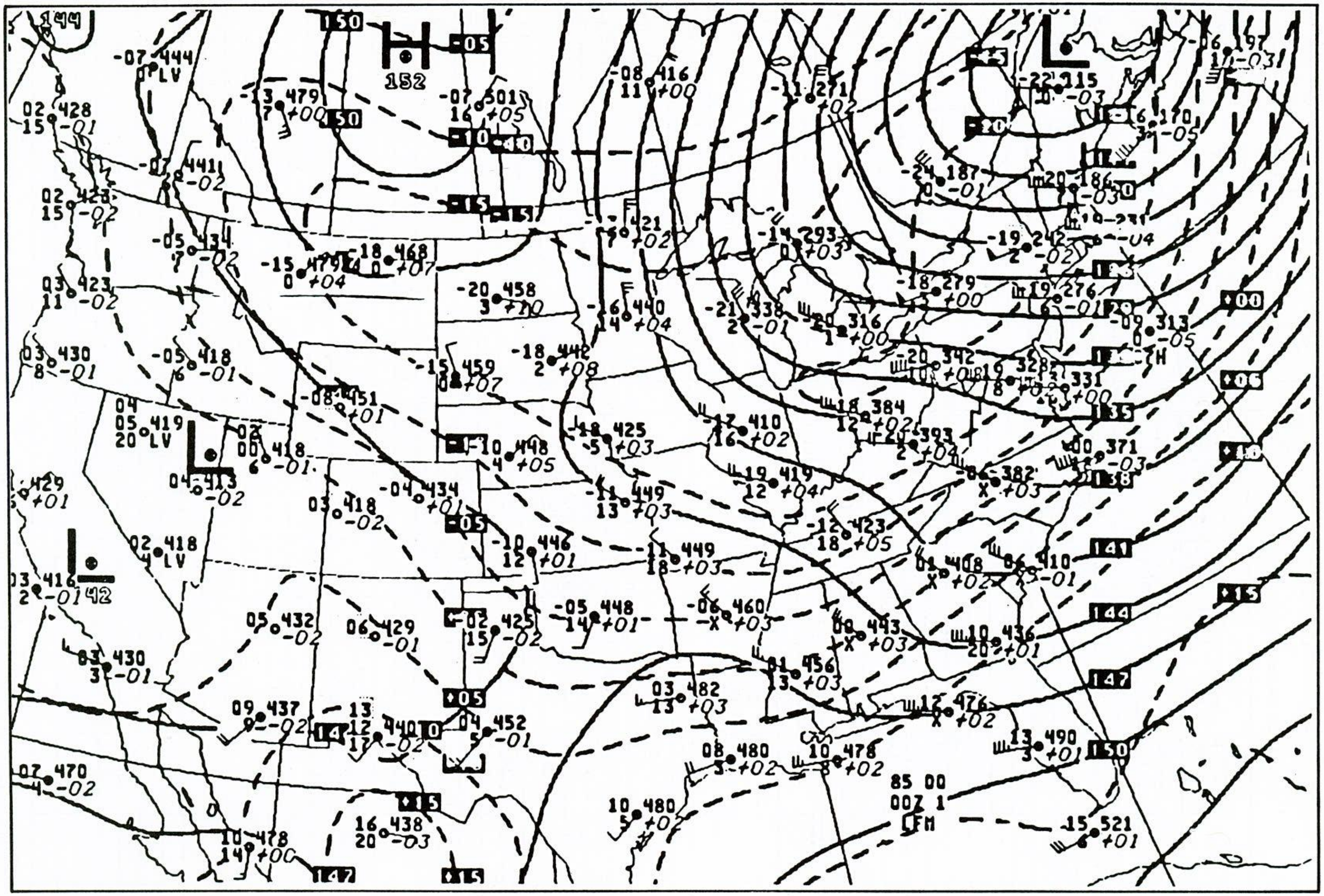


FIGURE E-5 EXAMPLE OF 850-mb (APPROXIMATELY 1500 m) CHART

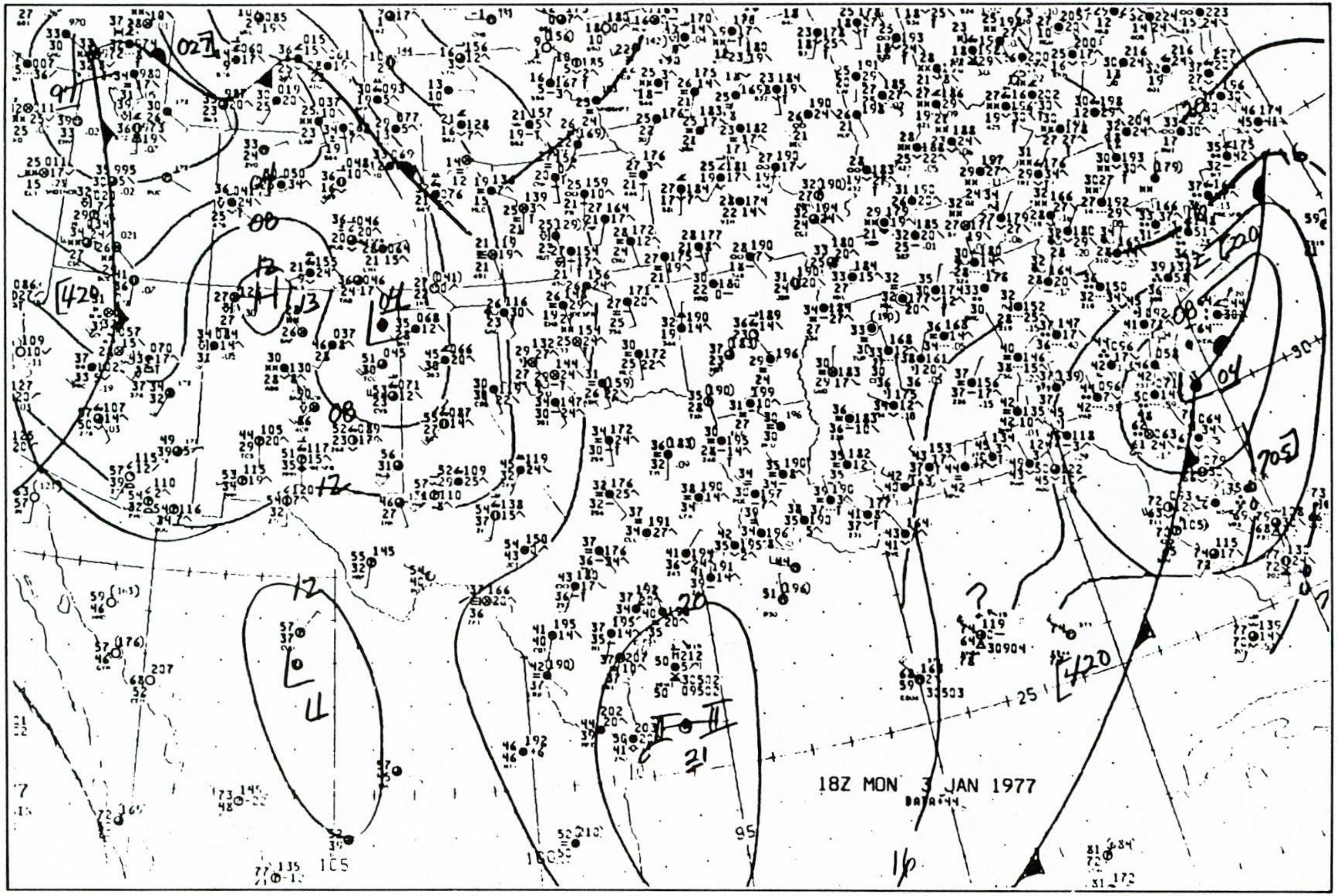


FIGURE E-6 EXAMPLE OF SURFACE WEATHER MAP AS AVAILABLE ON MICROFILM FROM NCC

104

WEDNESDAY, AUGUST 20, 1975

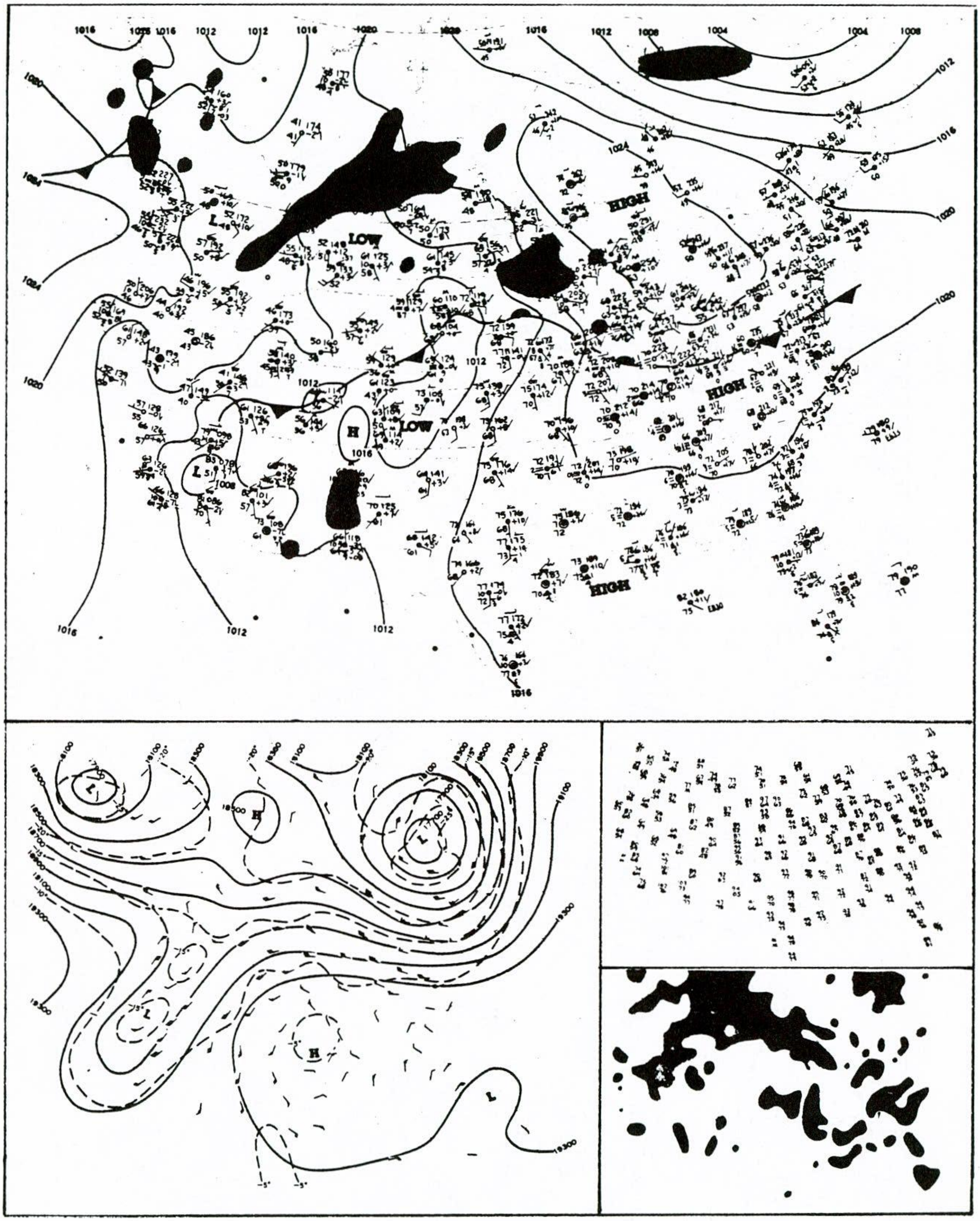


FIGURE E-7 EXAMPLE OF THE INFORMATION AVAILABLE IN "THE DAILY WEATHER MAP — WEEKLY SERIES" been ascertained, it may become important to obtain data from other areas that are upwind. In such instances, National Air Monitoring Stations (NAMS) data can be obtained from the EPA SAROAD data base by contacting EPA.*

The existence of other hourly data, such as that from State and Local Air Monitoring Stations (SLAMS) and Special Purpose Monitoring (SPM), can be ascertained from state or local air pollution control agencies. Names and telephone numbers can be obtained from the Directory, Governmental Air Pollution Agencies prepared by the Air Pollution Control Association.† This publication lists federal, state, regional, and county agencies conducting air pollution monitoring.

Maps. Maps will aid in identifying possible sources that contribute to the observed background concentrations. Topographic, census, traffic, and other maps of the surrounding areas (to about 300-km radius) should be obtained.

Topographic maps portray man-made and natural features and the shape and elevation of the terrain. They are readily available, economical, and usually supply all the detail that is required. These maps are classified according to scale: Figure E-8 is an example of three

---

*U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, National Air Data Branch (MD-14), Research Triangle Park, North Carolina 27711; Telephone (919) 541-5395.

†Editor: Directory, Governmental Air Pollution Agencies, Air Pollution Control Association, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213.

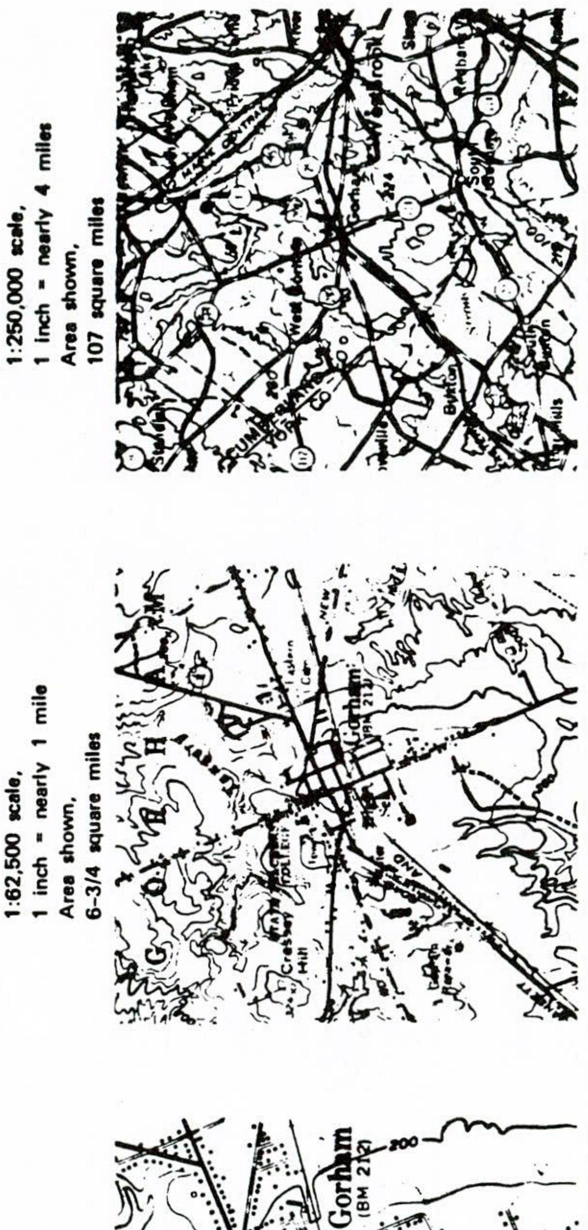
map scales of the same area showing the type of information that is available in large-, medium-, and small-scale maps. Table E-1 summarizes the principal maps and their essential characteristics. (Maps of the 250,000:1 scale are most appropriate for determining areas contributing importantly to transported ozone and precursors. They provide sufficient detail to identify major source areas while covering areas large enough for convenient assessment of contributions from potential sources within one- or two-day transport distances.)

To order maps of a specific area, obtain the "Index to Topographic Maps of [state]," which will include an order form and a list of local merchants that stock topographic maps. Indices and maps may be purchased by mail or over the counter from the U.S. Geological Survey (USGS).*

Emission inventories can be obtained on a county-by-county basis through EPA Regional Offices or from the State Implementation Plans (SIP). The emission inventories can be used to identify areas with major ozone precursor emissions along upwind trajectories.

---

*Areas west of the Missippi: Distribution Section, U.S. Geological Survey, Federal Center, Denver, Colorado, 80225. Areas east of the Mississippi: Distribution Section, U.S. Geological Survey, 1200 S. Eades Street, Arlington, Virginia 33303.



1:24,000 scale,
1 inch = 2000 feet
Area shown,
1 square mile

LARGE SCALE

1:62,500 scale,
1 inch = nearly 1 mile
Area shown,
6-3/4 square miles

MEDIUM SCALE

1:250,000 scale,
1 inch = nearly 4 miles
Area shown,
107 square miles

SMALL SCALE

FIGURE E-8 EXAMPLE OF THE INFORMATIONAL CONTENT OF LARGE-, MEDIUM-, AND SMALL-SCALE TOPOGRAPHIC MAPS

SOURCE: U.S. Geological Survey

SA-3515-21

Table E-1
NATIONAL TOPOGRAPHIC MAPS

Series	Scale	1 inch represents	Standard quadrangle size (latitude-longitude)	Quadrangle area (square miles)	Paper size E-W N-S width length (inches)
7½-minute	1:24,000	2,000 feet	7½ × 7½ min.	49 to 70	[1] 22 × 27
Puerto Rico 7½-minute	1:20,000	about 1,667 feet	7½ × 7½ min.	71	29½ × 32½
15-minute	1:62,500	nearly 1 mile	15 × 15 min.	197 to 282	[1] 17 × 21
Alaska 1:63,360	1:63,360	1 mile	15 × 20 to 36 min.	207 to 281	[2] 18 × 21
U.S. 1:250,000	1:250,000	nearly 4 miles	[3] 1° × 2°	4,580 to 8,669	[4] 34 × 22
U.S. 1:1,000,000	1:1,000,000	nearly 16 miles	[3] 4° × 6°	73,734 to 102,759	27 × 27

[1] South of latitude 31° 7½-minute sheets are 23 × 27 inches; 15-minute sheets are 18 × 21 inches.

[2] South of latitude 62° sheets are 17 × 21 inches.

[3] Maps of Alaska and Hawaii vary from these standards.

[4] North of latitude 42° sheets are 29 × 22 inches. Alaska sheets are 30 × 23 inches.

Source: U.S. Geological Survey (1969)

Determining Whether Natural Causes Are Responsible for Observed High-Concentration

Figure E-2 indicated that the first step following the accumulation of relevant background data is to determine for each day whether there is reason to believe that high ozone concentrations may have been the result of natural (rather than anthropogenic) causes. Two kinds of evidence can be used to make such a determination:

- Meteorological data and weather maps
- Temporal and spatial distributions of air quality.

The most common sources of natural concentrations of ambient ozone at the surface appear to be the downward transport of stratospheric ozone and the production of ozone from naturally occurring precursors within the troposphere. Synthesis from natural precursors, mainly of vegetative origin, is estimated to contribute a negligible increment. The concentration of ambient ozone of stratospheric origin is also generally well below the NAAQS, especially at monitors below about 1500 m above mean sea level (MSL). However, on rare occasions, abrupt intrusions from the stratosphere may cause localized high ozone concentrations that exceed the NAAQS. Since this cause is not anthropogenic, these periods should be excluded from the background ozone computation.

Much of the discussion of natural processes that follows has been extracted from a report by Singh et al. (37). The discussion is intended to provide some background for the understanding and

interpretation of the available data.

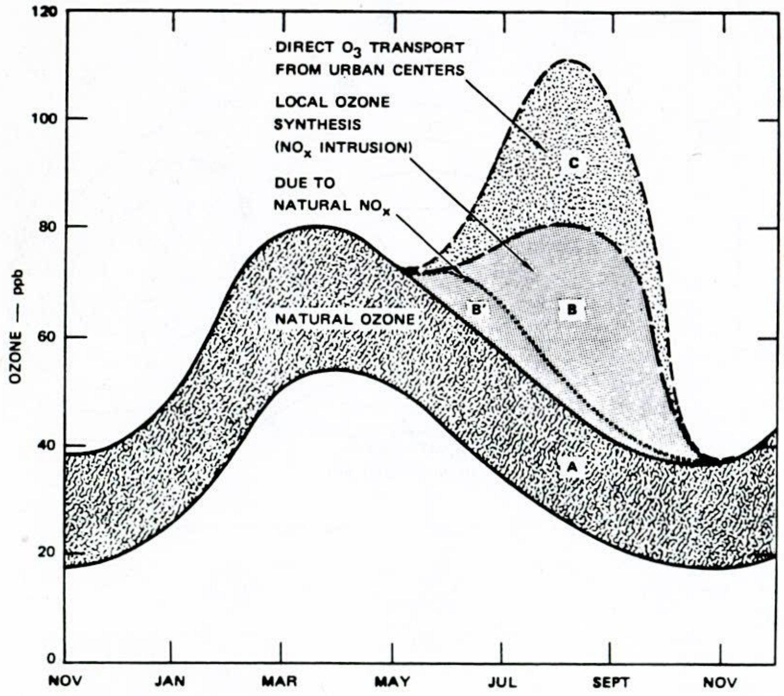
Tropospheric Synthesis. Tropospheric synthesis does not appear to be an important ozone source, although several mechanisms have been proposed for generating ozone from natural precursors. Among the earliest was Went's suggestion that tropospheric ozone might be synthesized photochemically from natural terpenes and natural NO_2 (38). Ripperton et al. tested this hypothesis under controlled conditions and confirmed that terpene and NO_x can result in ozone formation processes similar to those in polluted atmospheres (39). Although terpenoid compounds and NO_2 have been measured at relatively remote locations, the degree of their involvement in the tropospheric balance of ozone is uncertain because data bases are inadequate. Crutzen hypothesized another natural mechanism, entailing methane oxidation chains (40): This is an important proposition, because methane is ubiquitous and occurs at fairly high concentrations, viz., about 1.4 ppm. There is no consensus on the effectiveness of methane oxidation chains in producing ozone, because they can either produce or destroy ozone, depending on the NO_2 levels (41,42,43). Although natural reactive hydrocarbons (e.g., terpenes) and less reactive hydrocarbons (e.g., methane) are widespread in the atmosphere, their relation to the production of ozone appears to be critically controlled by the availability of oxides of nitrogen.

At this time, it appears very unlikely that either of the above natural mechanisms would produce ozone concentrations approaching the NAAQS. Furthermore, the hypothesized natural precursors to ozone are either omnipresent, as with methane, or at least have generally very widespread sources. For this reason, it usually will not be feasible to establish a strong connection between an NAAQS exceedance and a natural precursor source. At best, it should be possible to determine whether there has been any recent (within the past day or two) introduction of anthropogenic precursors. This determination should be sufficient for purposes of control-strategy development and evaluation. If the transported ozone has anthropogenic origins, then it is subject to some control and can be treated accordingly in modeling applications. If the ozone is not of man-made origin, the transported ozone concentrations are probably not subject to control and should be assumed to remain constant in the modeling exercises, regardless of what the specific origins might be.

Stratospheric Transport. Transport from the stratosphere has been suggested as a natural source of tropospheric ozone, because large amounts of ozone are known to be produced in the stratosphere and some of this stratospheric ozone is transferred to the troposphere by various meteorological processes. There are latitude dependent seasonal variations in the rate at which ozone is transferred to the troposphere at midlatitudes. The greatest rates of transfer occur in the late winter and spring (44). It appears that background ozone concentrations in the lower troposphere tend to lag one or two months behind the injection cycle from the stratosphere to the troposphere. The major sink for the tropospheric ozone is the destruction that takes place at the surface.

There is some uncertainty about the amount of ozone in the troposphere that can be attributed to stratospheric sources. Reiter (45) provides an estimate of 10 to 15 ppb as the average contribution of stratospheric ozone to the background at ground level. Singh et al., (46) and Danielsen and Mohnen (47) estimated the yearly mean tropospheric background ozone concentration to be about 30 ppb, nearly all of which can be attributed to a stratospheric source. Concentrations are likely to be higher than the annual average value in the spring at midlatitudes, and lower in the fall. In any event, there appears to be a natural background of ozone in the troposphere at a level of a few tens of parts per billion (ppb). This represents an appreciable fraction (20 to 40 percent) of the NAAQS for oxidant.

The annual variations in the natural tropospheric ozone burden at midlatitudes are shown schematically in Figure E-9 (46). At very remote sites, unaffected by anthropogenic emissions, the ozone concentrations reach their maximum in the early spring. In general, the natural ozone falls somewhere in the shaded area marked "A" in the figure. Natural concentrations reach their minimum in the late fall or early winter. The decline of ozone concentrations in these remote locations results in part from the decrease in stratospheric injection into the troposphere; it is also possible that photochemical processes destroy the natural ozone when NO_x is not present. If oxides of nitrogen are present, either from natural or anthropogenic sources, then the situation is quite different and photochemical reactions will cause a net increase in ozone.

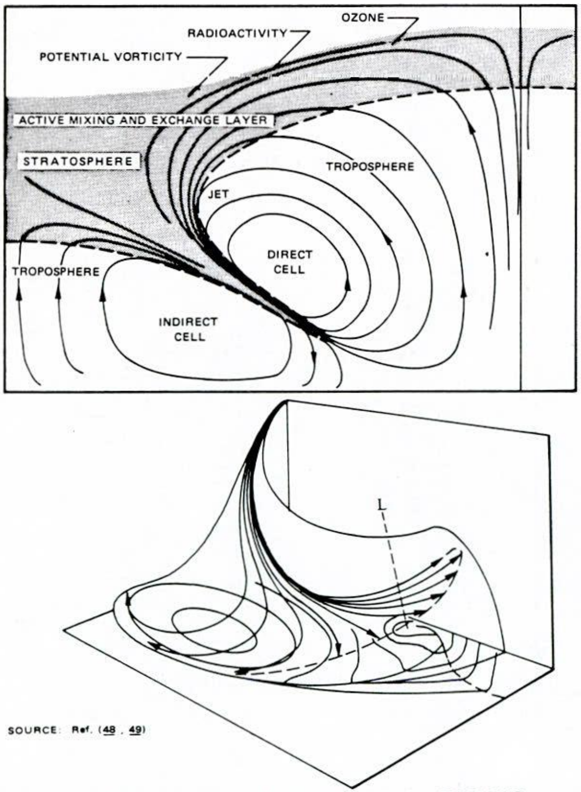


SOURCE: Ref. (46)

FIGURE E-9 IDEALIZED ANNUAL OZONE VARIATIONS AT REMOTE MIDLATITUDE LOCATIONS OF THE NORTHERN HEMISPHERE

Tropopause Folding. To this point, the stratospheric contribution has been discussed only in terms of averages. An important question is whether ozone rich stratospheric air ever reaches ground level before ozone concentrations have been diluted below the NAAQS. Danielsen has proposed a mechanism involving the folding of the tropopause that brings relatively undiluted stratospheric air deep into the troposphere, perhaps down to levels of 3000 m or so (48). Figure E-10 shows this tropopause folding. Stratospheric air containing ozone enters the troposphere in a large inclined, curved sheet. The shape of that sheet and the trajectories of the stratospheric air within it are illustrated schematically in the lower part of Figure E-10. The folded layer tends to enter the troposphere behind the surface cold front. The intruding stratospheric sheet moves in a descending clockwise path around the surface high pressure cell (48,49).

Recent data have shown the extent of the tropopause folding phenomenon and suggests that it occurs relatively frequently. Figure E-11 (from the rather extensive work of Johnson et al.) shows a cross section through a large stratospheric intrusion on 13 May 1978 as measured along the path shown in Figure E-12 (50). A nearly horizontal sheet of high ozone concentrations representative of the stratospheric air can be seen stretching over more than 100 km at an altitude of around 3.5 km (11,000 to 12,000 ft). It appeared in Figure E-10 that the stratospheric air would be swept around the low-altitude, high-pressure air in a clockwise direction. Reiter has attributed one instance where concentrations of nearly 200 ppb were observed at the



SOURCE: Ref. (48, 49)

FIGURE E-10 SCHEMATIC DIAGRAMS OF TROPOPAUSE FOLDING

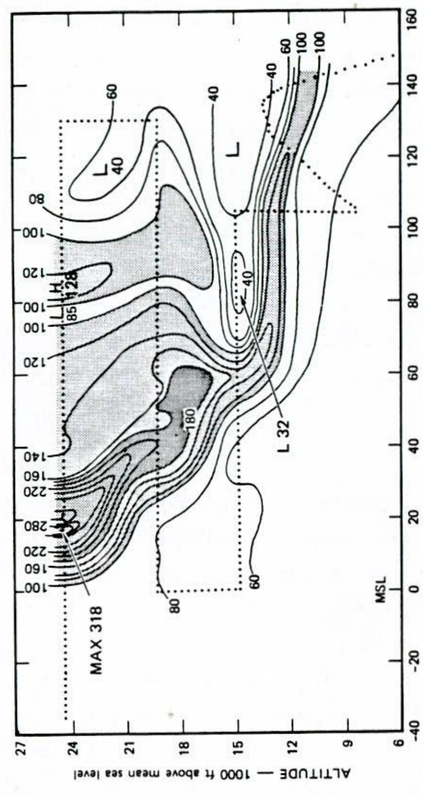
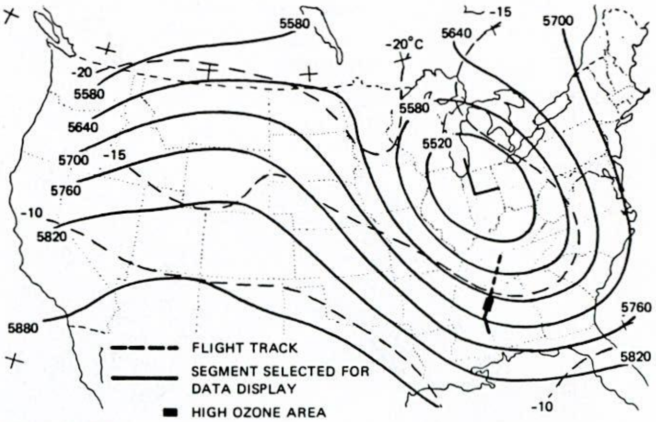


FIGURE E-11 VERTICAL CROSS SECTION OF OZONE CONCENTRATIONS (ppb) OBSERVED IN A TROPOPAUSE FOLD ON 13 MAY 1978

SOURCE: Ref. (50)



SOURCE: Ref. (50)

FIGURE E-12 FLIGHT TRACK FOR 13 MAY 1978

3000-m peak of the Zugsptize in Germany to an intrusion of stratospheric air (45). Singh et al. present another example from Mauna Loa, Hawaii, when concentrations of nearly 100 ppb were observed (37).

Both the above examples were observed at rather high altitudes; Lamb has conducted a detailed analysis of an incident near sea level in Santa Rosa, California (51). In this incident, hourly-averaged ozone concentrations of about 220 ppb were observed during the early morning hours of 19 November 1972. Lamb's analysis suggested that these high concentrations were the product of an unusual sequence of events. The ozone was brought into the troposphere from the stratosphere by large-scale circulations associated with the advance of the frontal zone. However, the ultimate transport to the ground resulted from small-scale air circulations around a shower cloud. Although the events produced very high concentrations, those concentrations were short-lived and affected only a relatively small area within a few tens of kilometers of the observation site.

From the evidence presented, it follows that the frequency of occurrence of stratospheric ozone intrusions is proportional to the frequency of cold front passages. Viezee and Singh present information verifying their conclusion (44). They show that monthly average beryllium-7 (^{7}Be) concentrations, an accepted tracer of stratospheric air, vary directly with the frequency of occurrence of 500-mb low-pressure troughs (Figure E-13). (Surface cold fronts are associated with elevated low-pressure troughs.) In addition, the frequency of low-pressure troughs is shown to vary with latitude and season.

Conclusions. In summary, the occurrence of high concentrations of ozone of stratospheric origin at low altitudes is not common, but does occur. Any observations of high ozone concentrations, especially during winter or spring, should be examined carefully. Their occurrence in the southwestern quadrant of a high-pressure system further increases the likelihood that natural stratospheric sources may be involved. If the air mass in which the ozone is observed has recently advanced rapidly from the north during a period of cyclogenesis, the probability of stratospheric involvement is yet stronger. Such cases should be recognized so that an EPA Regional Administrator can be requested to exempt them from the list of cases used to determine numbers of exceedances and design values. (Presumably, the EPA Administrator would have qualified meteorologists examine the data to determine whether stratospheric air was involved.)

Although violations of the air quality standards due solely to the introduction of stratospheric ozone into the lower atmosphere are rare, it is important not to dismiss them altogether, because violations induced by stratospheric air are associated with meteorological conditions that are very different from those producing photochemical ozone. Control strategies based on anomalous conditions are not likely be to very effective for reducing ozone concentrations of anthropogenic origin. For purposes of policy formulation, it is extremely important to differentiate between ozone of natural origin and that produced photochemically from anthropogenic emissions.

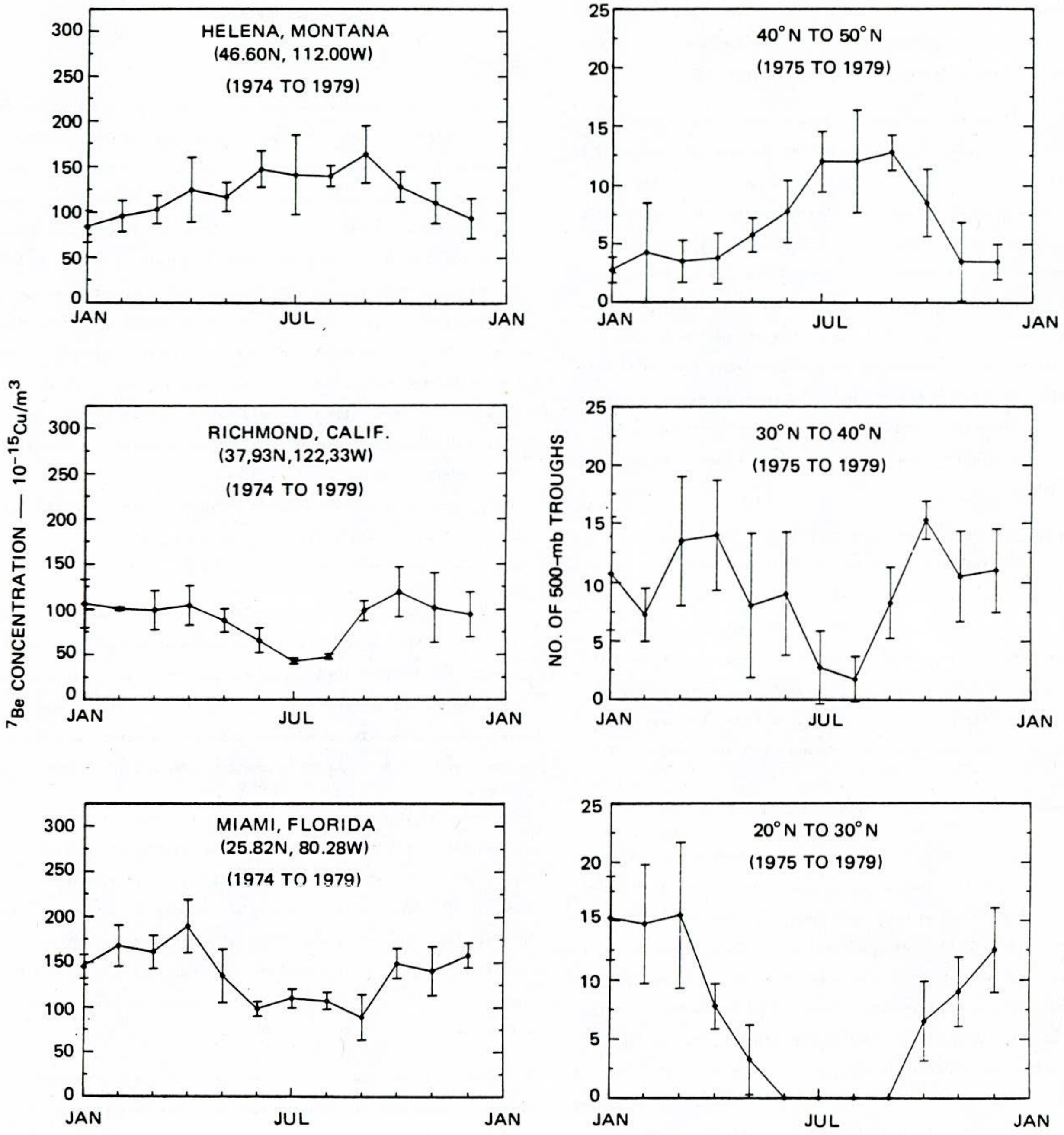
Determining Background Ozone Concentrations

When the contribution of background ozone concentrations to the observed high ozone concentrations is assessed, the contribution of transported ozone will be included as part of the background. The magnitude of the background/transported ozone concentrations will depend on the direction and speed of transport. Direction is important because the source depends on direction. Speed is important because it affects the degree to which the ozone is diluted. There are at least two methods for determining the magnitude of the background ozone concentrations. The first method is applied when there is no clearly defined airflow for the region, i.e., the air stagnation case: The second method applies when there is clearly defined airflow. To determine the background ozone concentration, it is first necessary to determine which method should be used. The discussion below first defines air stagnation episodes and how they differ from clearly defined airflow regimes and then describes the methods for determining the background/transported ozone concentrations.

Air stagnation episodes occur infrequently, but have historically been responsible for the worst air pollution episodes. Air stagnation or calm periods occur almost daily at the surface as a result of the formation of surface-based stable or inversion layers at night. However, there is usually well-defined transport occurring just above the surface stable layer. An air stagnation episode is therefore defined to be a case with little transport at the surface or aloft for an extended period, from one day to several days. The air stagnation episodes occur with strong, quasi-stationary, surface high-pressure systems. The surface high-pressure system is usually well supported in the upper atmosphere; that is, there is also a high pressure center at 850 mb (1.5 km) and usually a ridge of high pressure at 500 mb (5.5 km). Inversions that tend to trap pollutants in the lower layers are often associated with these synoptic conditions.

Several meteorological parameters and analyses from the NCC (see earlier discussions) will help to identify air stagnation cases. The climatological summaries may be more useful than the weather maps, because air stagnation periods do not always occur with the synoptic conditions described above. The central issue is the general lack of transport of air pollutants.

Climatological summaries from stations throughout the study area and its surroundings should be obtained, because some individual observation sites are located where they are influenced by local effects and are not representative of larger-scale airflow. For instance, the National Weather Service Office in Los Angeles is located at the Los Angeles International Airport, near the coast were the local sea breeze is important; so a day that should be considered an air stagnation case may be overlooked because a moderate sea breeze that penetrates no more than a few kilometers from shore may appear to be more prevalent than it really is. If weather observations from other sites, e.g., Burbank,



SOURCE: Ref. (44)

FIGURE E-13 SEASONAL AND LATITUDINAL VARIATION OF GROUND-LEVEL ^{7}Be (LEFT FRAMES) AND 500-mb LOW-PRESSURE TROUGH FREQUENCY

Ontario, and elsewhere were considered, a better picture would be obtained. Data should also be obtained from stations as far as 300 to 500 kilometers away, because these data may be needed to construct trajectories to determine the source regions of transported ozone. A discussion explaining the construction of back trajectories is presented later.

Observed weather parameters that identify air stagnation include:

- Wind speeds
- Wind directions
- Visibility
- Reported weather.

Wind Speeds. Reported hourly wind speeds that are less than 4 knots (2 m/s) with highly variable directions through most of the day usually indicate air stagnation. Wind speeds are not expected to be low all day, because mid- to late-afternoon convective motions usually increase the wind speeds at the surface unless winds aloft are also light. Visibilities less than 5 miles and reports of smoke or haze are indicative of pollutant accumulations because of stagnation and poor dispersion.

In summary, if wind speeds are consistently weak and the wind direction is highly variable, then air pollutants are not dispersed efficiently. If such wind conditions persist through most of the day, even during the afternoon hours, the day should be considered an air stagnation day.

Figure E-1 showed acceptable monitoring locations for estimating background ozone during stagnant conditions. Basically, the EPA suggests monitoring locations well away (40 km or more) from the urban area. The 40-km criterion ensures that the measurements are not heavily influenced by ozone that has been recently transported from the city by the light winds. Therefore, the first step in determining the background ozone concentration for the stagnant case is to identify the monitoring sites or sites that are at least 40 km from the urban center. If no sites are at least 40 km from the urban center, then the most distant and rural site should be used.

To estimate the background ozone concentration throughout the mixed layer, it is important to use a surface concentration observed after the morning inversion has been broken and vertical mixing has distributed the ozone uniformly through the mixed layer. Surface meteorological observations can be used to determine when the inversion breaks. One indicator is an increase in the surface wind speeds, caused by the downward flux of momentum from higher altitudes. Similarly, the surface wind directions should become more uniform. However, during stagnation periods, these effects are not likely to be as pronounced as they would otherwise be.

The appearance of cumulus or low clouds is another indicator of inversion breakup. When the inversion has broken, the vertical flux of moisture and heat often leads to the development of the fair weather cumulus clouds. Temperature observations can indicate that the inversion is still present. Before the inversion breaks, surface temperatures should rise faster than after the inversion breaks; insolation heats only the shallow layer below the inversion until that inversion has been broken. Then, the solar heating is distributed through a deeper layer, reducing the rate of temperature increase. In general, the nocturnal inversion should be expected to be erased by noon and often earlier.

When temperature soundings are available from a nearby site (Figure E-14 shows the U.S. locations where twice daily soundings are made), they should be used in conjunction with hourly surface temperature data to determine the time of inversion breakup. Typically, the inversion will break when surface temperature rises to the point where the potential temperature at the surface is greater than that at the top of the inversion. The appropriate soundings for this purpose are those taken in the predawn hours. (Soundings are made nominally at 0000 and 1200 GMT. The 1200 GMT sounding corresponds to a local standard time between 0400 on the Pacific Coast to 0700 in the east. Because the balloons are usually launched before the nominal time, they generally correspond to a time near dawn.) The best approach to determining the height of an inversion is to use a thermodynamic diagram and project the temperature at the top of the inversion to ground level along a dry adiabat to determine the surface temperature required to break the inversion. A simpler, rule-of-thumb alternative can be used. The sounding is examined to determine the height and temperature at the top of the inversion (the point where temperature resumes its decrease with height above a low layer where the temperature was constant or increased with height). The surface temperature needed to break the inversion will be about 1°C warmer than the temperature at the top of the inversion for every hundred meters that the inversion is above the surface. For example, if the top of the inversion is 350 m above the surface and the temperature there is 22°C, than the surface temperature required to break that inversion should be 25.5°C or more.

The hourly ozone data themselves can also serve to indicate that the inversion has broken. A sudden increase in ozone concentrations from near-zero to higher concentrations during the forenoon, is often caused by the sudden downward mixing of ozone from stable layers just above the surface. Such downward mixing is symptomatic of the erosion of the inversion. The ozone concentrations measured at the surface show how much background ozone has been left over from the preceding day.

If the data indicate that the nocturnal inversion layer has been eroded and an applicable background monitoring station (or stations) has been identified, the hourly averaged ozone concentrations observed at the site will be used to estimate the background ozone concentration, which will be the highest observed hourly averaged ozone concentration observed at the site during the day. If more than one applicable

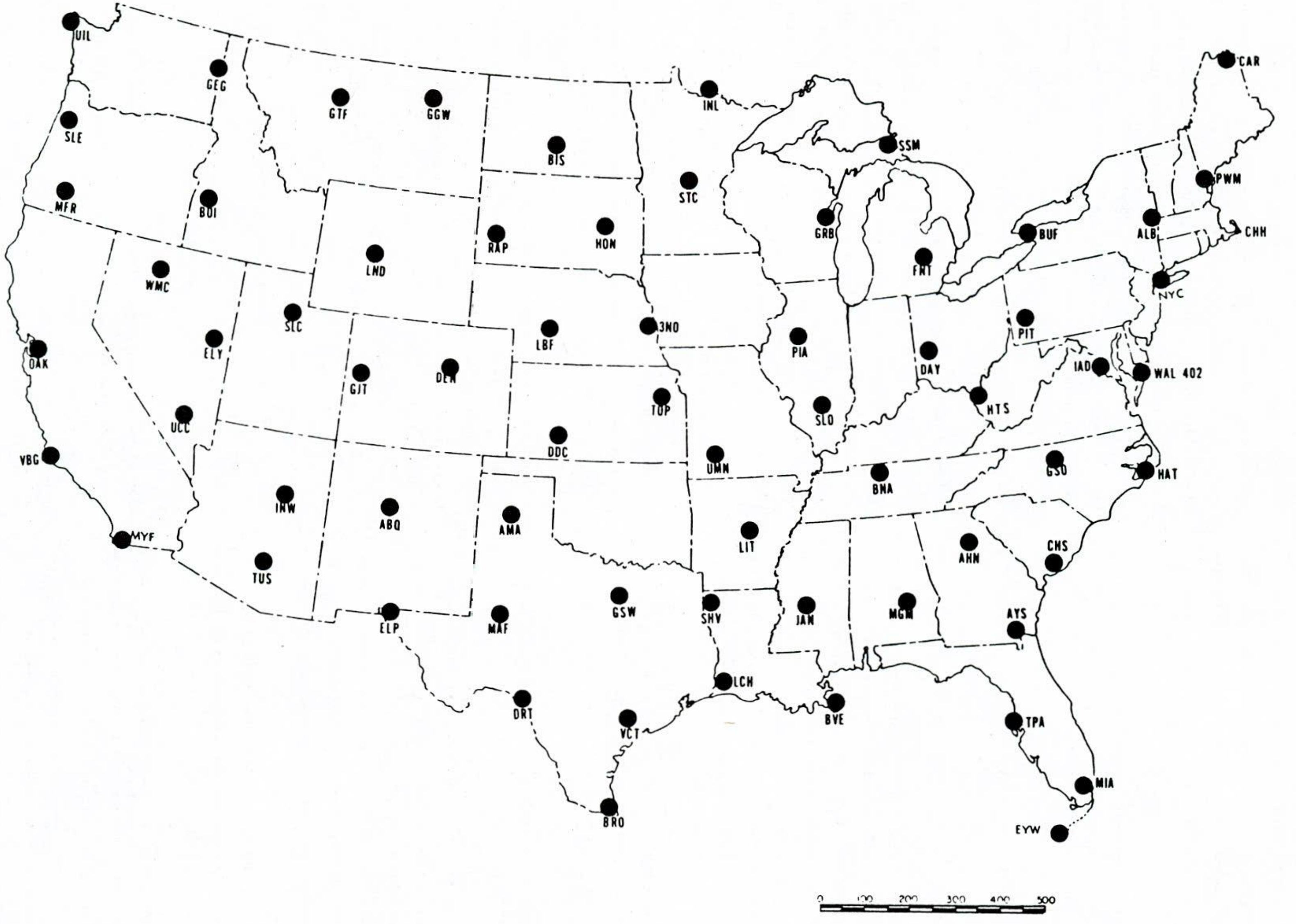


FIGURE E-14 LOCATIONS OF UPPER-AIR SOUNDING STATIONS IN THE UNITED STATES background site is available, the lowest observed peak-ozone concentration should be used to estimate the background ozone. The background peak-ozone concentration is expected to be observed between 1200 and 1700. A later peak indicates that some transport of ozone from the urban area may have reached the site.

The rationale for choosing the peak value observed at a background site was discussed earlier. Basically, the most reasonable definition of background ozone should be "the concentration that would have been observed in the absence of local precursors." The lowest peak-hour concentration observed at a site well outside the city should provide the best estimate (especially during stagnation conditions) of the ozone that would have been observed had the city not been present. However, if one chooses to estimate the ozone background with as little influence as possible from background precursors, the upwind concentration during the first two or three hours after nocturnal-inversion breakup provide the best estimates.

Finally, any ozone observations aloft in the afternoon mixed layer for the test day should be examined to determine whether those ozone concentrations were similar to the surface concentrations. If the assumption that the ozone is uniformly mixed is valid, then the concentrations at the surface and aloft will be similar. In general, measurements aloft are more representative of background values, so long as they are made within the mixed layer and are not anomalous peaks caused by the relatively short averaging times of most aircraft observations.

Wind Direction. It is necessary to know transported ozone concentration in order to assess control measures that will be needed in a given metropolitan area. If the contribution of transported ozone to the daily peak ozone concentration is large, then control strategies applied in upwind source areas will reduce the concentrations arriving in the urban area in question from upwind areas. This, in turn may substantially aid the achievement of air quality standards.

The selection of the monitoring site or sites to use in determining the transported background ozone concentration depends on wind direction, as was shown in Figure E-1. The background station must not be affected by ozone or precursors from the urban area itself. The climatological summaries and weather maps described earlier are used to determine the wind direction for the particular case day. The wind direction should be estimated for the hours from about 0600 (local time) to the time of the daily peak ozone concentration. In the early morning hours, the surface winds may be light and variable; thus, the 850-mb winds can often provide a better estimate of the transport direction. The 850-mb winds measured at 1200 GMT (0700 EST) should be used. During the later morning hours and the afternoon hours, the surface wind directions are usually more representative of the general direction of transport. Using the 0600 hour as the initial hour for determining the general transport direction ensures that the direction of transport of the early morning emissions of primary ozone precursors (NO_x and NMHC) are represented.

After the general direction of the airflow has been determined, appropriate monitoring sites can be identified for estimating transported background ozone concentrations. The EPA guidelines (1) suggest sites that are at least 40 km from the urban center in the upwind direction or to one side of the city as indicated in Figure E-1(a). If no sites are located beyond 40 km from the urban center, monitoring sites located directly upwind of the urban center to the extent possible (especially upwind of all NO_x and NMHC emission areas) will be acceptable, if they are free of local urban influences.

As in the case of stagnation, surface meteorological parameters are valuable for estimating the time of the upper-air-inversion breakup. The use of these parameters is the same as was discussed earlier.

The transported background ozone concentration is estimated from the hourly averaged ozone concentrations observed at the appropriate background monitoring stations. The peak hourly averaged ozone concentration observed after the breakup of the inversion should be used as the background estimate for the reasons cited earlier. When two or more background monitoring sites are available, the lowest peak-hour averaged ozone concentration observed among those sites should be used. The transported background ozone concentration should be determined from ozone data collected aloft during the afternoon within the mixed layer, if possible. Concentrations occurring during the evening are not likely to be representative of true background values, but are more likely to be from the plume downwind of the nearby city.

If the source of the transported ozone is nearby, or the trajectories (see next section) indicate that the source changed during the course of the day, then the peak afternoon values may not be appropriate because they will not be representative of the transported material that arrived in the morning. Material entering the area in the morning will augment the emissions from rush-hour traffic and contribute to the areas high ozone values. If modeling studies are planned, then the earliest measurements of ozone made after the breakup of the inversion (and NO_x and hydrocarbon measurements if available) should be used to provide upwind boundary values for modeling. If aircraft observations were made, they should be used of course. Ludwig (25) has discussed how surface ozone observations that are representative of conditions through the mixed layer can be recognized.

Back Trajectories. Construction of back trajectories will be necessary when the estimated transported background ozone concentration is large. Fairly complete records of observed winds in the region surrounding the urban areas being studied are needed to construct back trajectories. These data are obtained from the NCC for the observing sites within a radius of about 300 to 500 km. Probably the most convenient method for determining air flow that is representative of the lowest layers is by use of the 850-mb analyses from the NCC. An example of such an analysis was given earlier in Figure E-6.

Using the wind data, streamline maps for each observation period are drawn to depict the overall airflow in the region. In drawing a streamline map a short line depicting the wind direction reported at each observation site is first plotted on the map or tracing paper, as shown in Figure E-15(a). Wind directions are indicated by lines extending from the station circles. The lines extend from the circle toward the direction from which the air is moving. Thus, the station circle is at the end of the direction line that corresponds to the point of the air motion vector. The wind direction at points between the observed data is then interpolated between the observed wind directions by drawing additional line segments [Figure E-15(b)]. Finally, smooth curves are drawn parallel to the short line segments to produce a streamline map [Figure E-15(c)]. It should be noted that the streamlines at 850 mb will be very nearly parallel to the contours (of constant geopotential heights of the pressure surface) on the map.

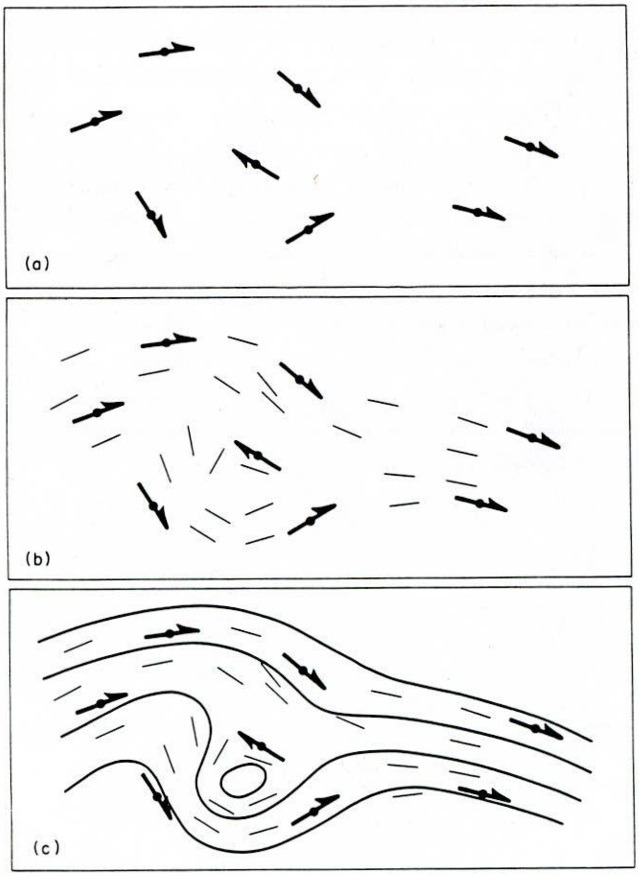


FIGURE E-15 CONSTRUCTION OF STREAMLINES

A streamline map estimating airflow is prepared for each observation time. For example, with wind observations reported every 12 hours at 0700 and 1900, the first streamline map drawn using the 0700 wind data is approximately applicable from about 0100 to about 1300. The next streamline map, based on the 1900 wind observations, represents the period from about 1300 to 0100.

The construction of the back trajectory starts at the time of the observed maximum daily peak ozone concentration. A wind speed appropriate to the study area must be interpolated. The wind speed can be an inverse-distance-weighted average of the two or three closest observation sites, or it can be estimated from the spacing of the contours on the 850-mb surface. Wind speeds are indicated by barbs and flags on the wind direction lines. Each whole barb indicates a wind speed of 10 knots (kt), a half-barb 5 kt. An unbarbed direction line with a hook at its end indicates a light wind of about 3 kt. Each triangular flag represents 50 kt. Thus, a flag and 1-1/2 barbs represents a 65-kt wind. It is sometimes convenient to recall that one knot is one nautical mile per hour, or one minute of latitude per hour. Thus, a 30-kt wind results in a movement of 0.5-latitude degree per hour, or 6 latitude degrees over a 12-hour period. It is convenient to use the degree of latitude as a unit of distance, even though the movements may be in directions other than north or south. In the following discussion, it is assumed that the hour of the peak concentration is also an hour with observed wind data.

To calculate the back trajectory, the distance an air parcel originating at the terminus would have traveled is estimated using interpolated wind speed. The distance calculated should be for a period equal to the time between wind observations. That is, for twice-daily wind observations, the distance traveled in 12 hours is used. The appropriate wind speed is the value estimate for a point about six-hours upwind of the terminus. Then, using the streamline map constructed for the hour during which the transported ozone arrived at the terminus, plot (on tracing paper) the distance traveled along a line parallel to the streamlines in a direction opposite the direction of the airflow. Next, place the tracing paper used to plot the distance traveled on a streamline map for the preceding wind observation time. Again, interpolate wind speeds (using the wind data from the earlier time) and plot the distance traveled, again starting backward from the terminus, parallel to the streamlines for the earlier hour. The estimated back trajectory for the first time step just before arrival comes from the point midway between the final points determined from the two sets of streamlines. Figure E-16 shows how the method is applied to two sets of trajectories. Note that the trajectory parallels the later streamlines at its terminus and the earlier streamlines at its point of origin.

To continue the back trajectory, repeat the steps described above using the new point as a new terminus. Of course, the streamline maps for the one and two time steps earlier are used. This backward stepping process should be continued for a day or two to determine whether the high transported background ozone concentration can be associated with a

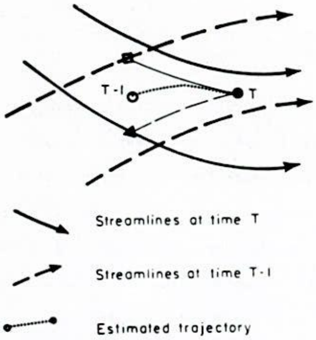


FIGURE E-16 SCHEMATIC DIAGRAM OF TRAJECTORY CONSTRUCTION METHOD E-50

specific upwind urban area that lies on or relatively close to the trajectory.

The preceding discussion describes a method by which trajectories can be manually constructed from readily available information. Automated methods are also available (8), but they often require data to be in a special format that is not readily available to every user. Figure E-17 shows examples of trajectories generated with such a program (8). The program generates trajectories that terminate at four different times of day based on winds that are interpolated both horizontally and temporally. The winds are averaged vertically through the afternoon mixing depth. In Figure E-17, the different trajectories are marked by different symbols ("A," "B") with positions at six-hour intervals upwind marked with numerals. If the user has access to an automated trajectory-calculating system, either directly or through the EPA, it should be used. However, lack of such access should not prevent the determination of the origins of transported ozone by the manual methods described above. E-51

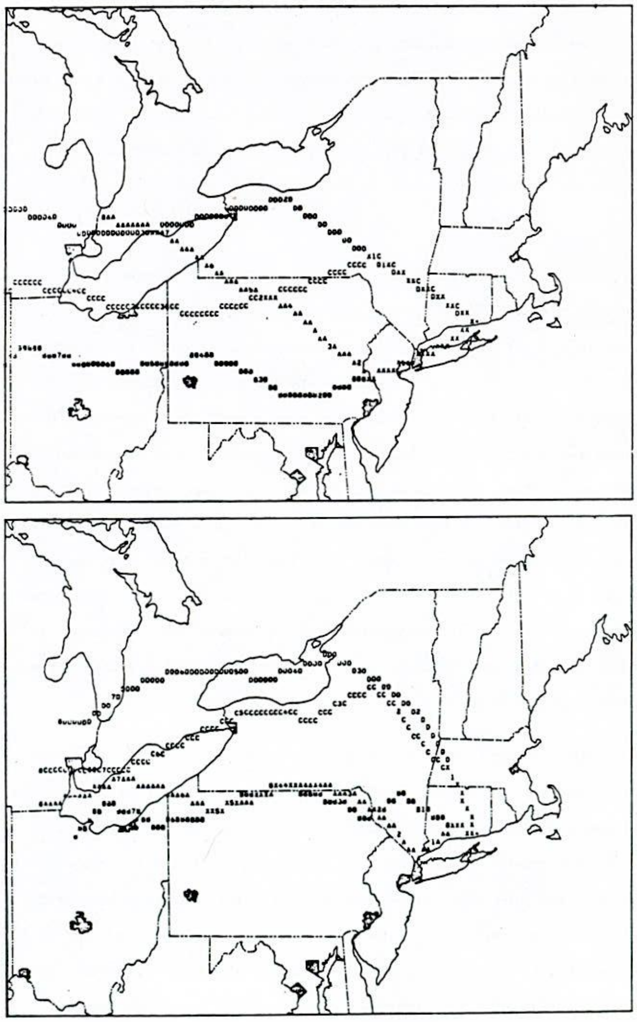


FIGURE E-17 EXAMPLES OF BACK TRAJECTORIES CALCULATED WITH A COMPUTER PROGRAM

APPENDIX F

SUPPLEMENTAL MONITORING PROGRAMS

GENERAL PRINCIPLES

Although it would be very desirable to have a permanent monitoring network with twenty or thirty permanently operating sites, the cost of such a network would be prohibitive. In principle, it should be sufficient to have only two permanent monitors if the objective of monitoring is to define the design value and the expected number of exceedances for the region: One monitor would be located where exceedances were most frequent and the other where the highest design value for the area occurs. If one carries this reasoning to even greater extremes, it would not be necessary to operate those two stations all the time. The design-value station would have to be operated only on those several days with the highest concentrations that are used to estimate design value. In principle, the exceedance station would have to be operated only on days when exceedances occurred; if the exceedances could be reliably predicted, then the station would not have to be operated at all.

The preceding paragraph presents two extremes that are impractical for different reasons. The first approach, that of using a large number of continually operating monitors, requires little outside information or judgement to determine design value or expected number of exceedances, but it is prohibitively expensive. The other extreme would require very little monitoring, but that monitoring would have to be done in exactly the right places at exactly the right times, which entails wholly unrealistic requirements for outside information and judgement. It should be possible to find a workable compromise between these two extremes. The best compromise appears to be a permanent network of stations that operate during the high ozone season, in locations where it is probable that they can define the most important features of the ozone concentration pattern. Specifically, a station is needed where it is most likely that the high ozone concentrations will occur, another where reliable measurements of background concentrations are possible and a few others to define the overall spatial distribution pattern in the area.

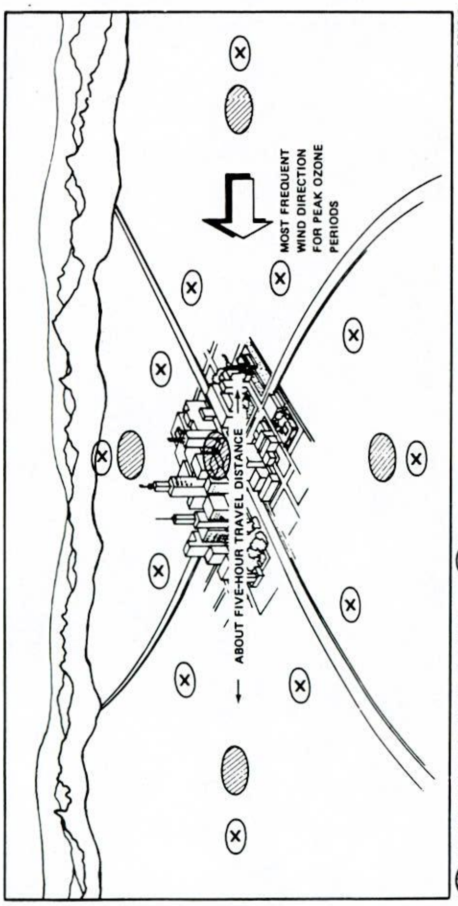
The permanent network can be supplemented on specific days when high concentrations are expected by deploying a few mobile monitors to previously identified locations. Before discussing the specifics of such an approach, we note that an approach using a limited number of fixed sites plus other sites that are temporarily deployed can lend itself to implementation by more than one agency. For example, the permanent monitoring can be done by a local or state environmental agency. It is not uncommon that transportation agencies and others specifically involved in project design and development conduct supplementary air quality monitoring studies. The approaches described below are quite consistent with such a division of effort.

OVERALL NETWORK

Permanent Stations

Large urban areas should have five or more permanent monitoring stations. For ozone, most of these should be outside the city. One or two should be in representative neighborhoods within the city to characterize the urban concentrations. Figure F-1 is a schematic diagram that shows 5 general areas for permanent monitoring sites. It also shows an arrow indicating the direction from which the wind blows most frequently during periods of peak ozone concentration. A computer program included at the end of this appendix can be used to determine the frequencies of wind directions and wind speeds for temperatures above 80°F, when ozone concentrations are high. An example of the output from that program is given in Table F-1. Alternate methods of estimating most frequent wind directions have been given by Ludwig and Shelar (14).

The most important locations for permanent monitors shown in Figure F-1 are those upwind and downwind of the urban area for the wind direction that is most frequent during periods of peak ozone. The upwind location meets requirements for estimating background concentrations. As noted in the figure, the distance from the upwind edge of the city to the monitoring location should be about five hours of travel. In the example shown in Table F-1, high-temperature (and hence high-ozone) winds are most frequently from the south at speeds of about 3 to 6 ms^{-1}. This wind speed corresponds to about 16 km/hr, or a distance of about 80 km from the upwind edge of the urbanized region. For an urban area that



▨ LOCATIONS FOR PERMANENT MONITORS ⊗ TEMPORARY LOCATION

FIGURE F-1 SCHEMATIC DIAGRAM OF APPROPRIATE AREAS FOR MONITORING SITES

Table F-1

TABULATED WIND STATISTICS FOR DAYTIME HOURS FOR ST. LOUIS, WHEN TEMPERATURE EXCEEDED 80°F

FREQUENCIES OF OCCURRENCES

DIRECT/CATEGORY	1.0-2.0	3.0-4.0	5.0-6.0	7.0-8.0	9.0-10.0	.GT.11.0	TOTAL
NNE	45.0	64.0	24.0	2.0	0.0	0.0	135.0
NE	38.0	63.0	29.0	2.0	0.0	0.0	132.0
ENE	49.0	76.0	14.0	1.0	0.0	0.0	140.0
E	54.0	99.0	29.0	2.0	0.0	0.0	184.0
ESE	75.0	125.0	41.0	2.0	0.0	0.0	243.0
SE	68.0	106.0	57.0	12.0	3.0	1.0	247.0
SSE	52.0	169.0	161.0	58.0	7.0	1.0	448.0
S	80.0	220.0	232.0	128.0	22.0	2.0	684.0
SSW	51.0	175.0	165.0	63.0	6.0	1.0	461.0
SW	74.0	220.0	136.0	38.0	10.0	2.0	480.0
WSW	80.0	198.0	162.0	35.0	5.0	0.0	480.0
W	87.0	122.0	91.0	32.0	6.0	2.0	340.0
WNW	65.0	88.0	52.0	26.0	8.0	1.0	240.0
NW	48.0	90.0	63.0	18.0	4.0	0.0	223.0
NNW	42.0	89.0	49.0	14.0	2.0	1.0	197.0
N	40.0	81.0	55.0	5.0	0.0	0.0	181.0
COLUMN TOTAL	948.0	1985.0	1360.0	438.0	73.0	11.0	4815.0

NO. OF CALM OBSERVATIONS: 46.0

PERCENTAGE OF OCCURRENCES

DIRECT/CATEGORY	1.0-2.0	3.0-4.0	5.0-6.0	7.0-8.0	9.0-10.0	.GT.11.0	TOTAL
NNE	.9	1.3	.5	.0	0.0	0.0	2.8
NE	.8	1.3	.6	.0	0.0	0.0	2.7
ENE	1.0	1.6	.3	.0	0.0	0.0	2.9
E	1.1	2.0	.6	.0	0.0	0.0	3.8
ESE	1.5	2.6	.8	.0	0.0	0.0	5.0
SE	1.4	2.2	1.2	.2	.1	.0	5.1
SSE	1.1	3.5	3.3	1.2	.1	.0	9.2
S	1.6	4.5	4.8	2.6	.5	.0	14.1
SSW	1.0	3.6	3.4	1.3	.1	.0	9.5
SW	1.5	4.5	2.8	.8	.2	.0	9.9
WSW	1.6	4.1	3.3	.7	.1	0.0	9.9
W	1.8	2.5	1.9	.7	.1	.0	7.0
WNW	1.3	1.8	1.1	.5	.2	.0	4.9
NW	1.0	1.9	1.3	.4	.1	0.0	4.6
NNW	.9	1.8	1.0	.3	.0	.0	4.1
N	.8	1.7	1.1	.1	0.0	0.0	3.7
COLUMN TOTAL	19.5	40.8	28.0	9.0	1.5	.2	99.1
						Calms	.9
						Total	100.0

is 40-km across, this would correspond to a site about 40-km downwind of the urban region. If the region were 80 or more kilometers across, then the location should be just beyond the downwind edge of the urban region. A station located in this way has the best chance of observing high ozone concentrations. A mirror-image site, upwind of the city, is needed to define background concentrations. The other two recommended locations outside the city provide for monitoring high ozone concentrations occurring with other wind directions. When combined with a location in the city, they also help to define general ozone patterns for the whole area. If it is absolutely impossible to operate permanent sites at five locations, then at least the two sites upwind and downwind of the city for the wind direction most frequently associated with high ozone and a site to characterize concentrations within the city should be operated.

It would be desirable to operate the permanent sites on a year-round basis, but compromises are possible. The EPA guidelines (1) require only that the data set be 75-percent complete for peak-potential-pollution seasons. Peak-potential-seasons are difficult to define precisely, especially if no historical monitoring data are available; historical data that are available should be used to define peak ozone seasons when the standards are expected to be exceeded once or more during the month. If historical data are not available, temperature records will provide reasonably good estimates. Where available, frequency distributions of daily maximum temperatures can be used; ozone exceedances are unlikely to occur when temperatures would not be expected to exceed 22°C (72°F) more than once a month. Compilations of the frequency distribution of daily maximum temperatures are not routinely available, but cooling-degree-day statistics have been published for most areas of the United States (22). The statistics should also be available from the National Climatic Center in Asheville, North Carolina. The meaning of the cooling degree-day concept is discussed in the body of this report. Any month with an average of fewer than 10 cooling degree days is unlikely to have occurrences of ozone concentrations in excess of the standard.

Temporary Stations

Temporary stations can be used to provide data at times and in places where high ozone concentrations are to be expected. In order to comply with the EPA guideline (1) definitions of valid monitoring days, temporary monitors should collect at least nine hours of data during the periods from 0900 to 2100 (local time). They should be deployed on days when the maximum temperature is predicted to be above 20°C (68°F). Suitable locations for monitoring should be identified ahead of time; Figure F-1 shows the general regions where specific locations should be selected. The characteristics of suitable locations are discussed in the next section.

If permanent monitoring is not being conducted at all five of the recommended areas shown in Figure F-1, then the first concern in any supplemental monitoring program will be to ensure that monitors are available for determining the concentrations near where a peak is expected, and at a location suitable for estimating background concentration, when high temperatures are anticipated. Forecast wind directions can be used to guide the deployment of temporary monitors. There should be at least one downwind station at the appropriate distance and, if necessary, an upwind station. A station well removed from the urban area to one side or the other will frequently serve to define background concentrations; an upwind station is not always necessary, but care should be exercised to make sure that the monitoring site is not influenced by nearby sources.

If the permanent network has five monitors, located generally as shown in Figure F-1, or if the wind direction is such that permanent monitors are located where peak and background can be monitored, then the objective of the supplemental monitoring will be to provide more details of the pattern in the vicinity of the expected maximum concentration. For the wind direction shown in Figure F-1, the three leftmost temporary locations would be the most suitable for this purpose, as they provide one station beyond where the maximum is expected and two that are closer to the city. Such an array provides a better estimate of the location and magnitude of the peak concentration. If more than three mobile monitors are available, then sites at about the same distance as the permanent monitor but 20 or 30 degrees to either side of it would be useful. These locations are not indicated in Figure F-1.

CRITERIA FOR SELECTING SPECIFIC LOCATIONS

Ludwig and Shelar (14) developed step-by-step procedures and guidelines for the selection of ozone monitoring sites. Their report contains considerable backup information regarding the criteria that they recommend, and it should be consulted for complete details; the material that follows has been extracted from that report.

Figure F-2 suggests that three somewhat different types of monitoring location will be required. It indicates one location in the urban region to characterize typical concentrations there, one location to characterize background concentrations; and finally, a location to measure the maximum or near-maximum concentrations. Figure F-2 summarizes the procedures used for selecting these three different kinds of sites. The figure shows that the procedure begins by assembling such relevant information as maps, emissions inventories, climatological data, and any historical monitoring data that are available. After the background information has been assembled, the next step is to determine which type of monitor is being located.

The right side of Figure F-2 shows steps in selecting background monitors. The process uses the most frequent wind directions associated with important photochemical activity to estimate the upwind direction; a good background site will be a few tens of kilometers from the city in that direction. The final step in selecting a specific background monitoring location is finding a site with characteristics like those shown in the box.

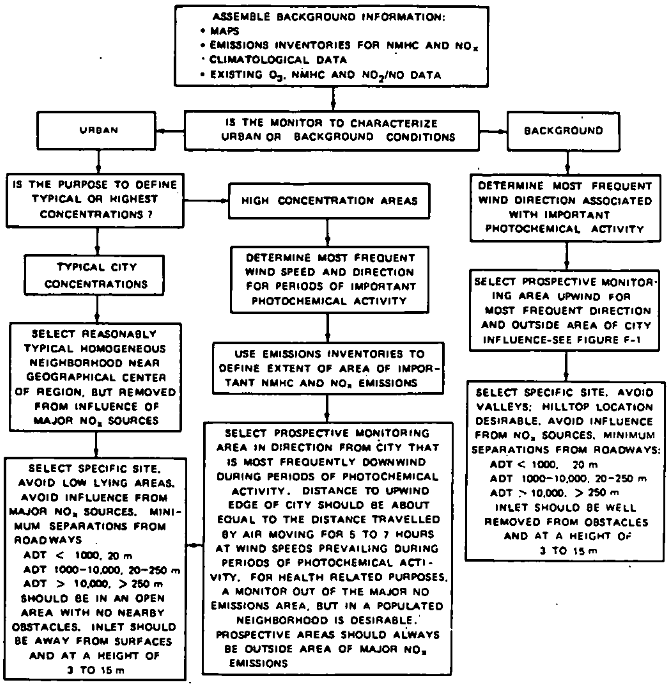


FIGURE F-2 SCHEMATIC DIAGRAM OF PROCEDURE FOR SELECTING OXIDANT MONITORING SITES

Two different types of location shown in Figure F-1 are influenced by the city, and hence cannot be considered background sites. One of these is in the city itself; the other is in a downwind location where maximum ozone concentrations are probable. Ludwig and Shelar (14) point out that it is not possible to identify the point of maximum ozone concentration with absolute certainty, which is one of the reasons for introducing supplemental monitoring. As Figure F-2 shows, wind directions are used in conjunction with the emissions inventories, which define the limits of the area within which most of the precursor emissions are found. If emissions inventories are not available, conventional topographic or street maps can be used to define the outer limits of the city; the edges need not be defined precisely. As was discussed earlier, locations several hours travel-time downwind of the city's upwind edge are appropriate.

The basic requirement for a site within a city is that the site be located within a reasonably homogeneous and typical neighborhood somewhere near the geographical center of the region. However, a city site, and all other sites, must be well removed from the influence of major sources of oxides of nitrogen. Figure F-2 gives some criteria for separation distances between highways and monitoring sites so that the influence of the local emissions will be reduced to acceptable levels. Consideration should also be given to the presence of any major elevated point sources whose nitric oxide emissions might impact the monitoring site. Such effects should be avoided.

COMPUTER PROGRAM FOR IDENTIFYING WINDS APT TO BE ASSOCIATED WITH OZONE PRODUCTION

As noted earlier, high ozone concentrations are most apt to be associated with relatively high temperatures. Thus joint-frequency distributions of wind direction and wind speed for hours when temperatures were high provide guidance for selecting monitoring locations. Ludwig and Shelar (14) developed a short computer program for generating such joint-frequency distributions.

Program WNDROS calculates the frequency distribution of wind direction and speed from standard National Climatic Center surface-observation data. The program includes only winds that accompany temperatures above 80°F (26.7°C) and occur during the daylight hours (0600 to 2000 LST). If the user chooses to select some other temperature cut-off or time period for inclusion, the program could be easily modified: it was written for use on a Control Data computer, but with a few modifications it can be used with other machines. The program reads WBAN/WMO hourly surface observations from tapes prepared by the National Climatic Center in Asheville, North Carolina. In addition to the data tape, the only other user-supplied input is a card image indicating the year/month/day of the dates when the user wants the processing to start and stop; the format for that card is 2F7.0. An example of the output of the program was given in Table F-1. For the convenience of the user, a listing of that program is included on the following pages.

Program WINDROSE

```
      PROGRAM WNDROSE (INPUT.OUTPUT.TAPE1.TAPE2)
C
C     WIND ROSE PROGRAM - - - CALCULATES FREQUENCY DISTRIBUTION OF WIND
C       SPEED VS WIND DIRECTION.  PROGRAM WILL READ WBAN/WMO HOURLY
C      SURFACE OBSERVATION CARD IMAGE TAPES PREPARED BY THE NATIONAL
C      CLIMATIC CENTER, NOAA.  USER MUST ENTER START AND STOP DATES OF
C      DATA THAT IS  TO BE PROCESSED.
C
      DIMENSION DAT(80).CAT(17.7).CLASS(6).ICLAS(12).WNDIR(17)
      DIMENSION WSWD(16.6).NOC(16)
      DIMENSION TOTSPD(7).TOTDIR(17)
      DATA (WNDIR=4HCALM.3HNNE.3H NE.3HENE.3H E .3HESE.3H SE.3HSSE.3H S
     2.3HSSW.3H SW.3HWSW.3H W .3HWNW.3H NW.3HNNW.3H N )
      DATA (ICLAS=1.2.2.3.3.4.4.5.5.6.6.7)
      DATA (CLASS=            7H1.0-2.0.7H3.0-4.0.7H5.0-6.0.7H7.0-8.0.
     18H9.0-10.0.8H.GT.11.0)
 1    FORMAT (1H1.25X*WIND ROSES FOR ST. LOUIS FOR DURING *F6.0* TO*F8.
     1/)
 2    FORMAT (/1H .*EOF. NO. =*I4)
 3    FORMAT (/1H .*P.E. NO. =*I4)
 4    FORMAT (/1H .*REC. NO. =*I4)
 5    FORMAT  (2F7.0)
  6    FORMAT (5X.F6.0.F2.0)
 7    FORMAT (8X.F2.0.F2.0.4X.A1.F2.0)
 8    FORMAT (// .30X*FREQUENCIES OF OCCURRENCES*//1H .*DIRECT/CATEGORY
     1*6A10.3X*TOTAL*/)
 9    FORMAT (1H .2XA4.9X.7(F8.1.2X))
 10   FORMAT (8A10)
 11   FORMAT (1H .8A10)
 12   FORMAT (// .30X*PERCENTAGE  OF OCCURRENCES*//1H .*DIRECT/CATEGORY
     1*6A10.3X.*TOTAL*/)
 13   FORMAT (/1H .I5.2F10.0.3F10.1.3I10.F10.0)
 14   FORMAT (/1H .*CALMA =*F10.2.5X*NO. OF OBS. =*I10/)
 15   FORMAT (1H .16I5)
 16   FORMAT (/1H .*NO. OF CALM OBSERVATIONS:*F7.1/)
18    FORMAT(/1X.*COLUMN TOTAL*3X.7(F8.1.2X))
      CALL MEMSETX (0.0.CAT.119) $ NOBS=0
C
C  READ START AND STOP DATES TO BE PROCESSED - FORMAT IS 2F7.0
C
      READ 5.BDATE.EDATE
      PRINT 1.BDATE.EDATE
C
C  READ SURFACE OBS TAPE - WRITTEN FOR CDC COMPUTER
C
 100  BUFFER IN (1.0) (DAT(1).DAT(80))
      IF (UNIT(1)) 130.110.120
110   NF=NF+1
      PRINT 2.NF
      GO TO 200
120   NP=NP+1
      PRINT 3.NP
130   NR=NR+1
      LEN=LENGTH(1)
       DO 180 I=1.LEN.8
       DECODE (13.6.DAT(I)) DATE.HOUR
      IF (DATE.LT.BDATE) GO TO 180
      IF (DATE.GT.EDATE) GO TO 200
C
C   CHECK FOR TIME OF DAY
C
      IF (HOUR.LT.6.0.OR.HOUR.GT.20.0) GO TO 180
      DECODE (19.7.DAT(I+3)) DIR.SPD.ITX.TT
      IF (ITX.EQ.1HX) GO TO 140
      IF (ITX.EQ.1HO) GO TO 150
      TT=100.+TT
      GO TO 150
 140  TT=-TT
150   CONTINUE
C
C    CHECK FOR TEMPERATURE LESS THAN 80 DEG F
C
      IF(TT.LT.80.) GO TO 180
C
C           CONVERT WIND SPEED UNITS (KTS TO MPS)
C
      WSPD=SPD*0.51479
      IF (WSPD.GE.11.0) GO TO 160
      IWS=WSPD+1.0 $ GO TO 170
```

```
 160   IWS=12
 170   IC=ICLAS(IWS)
C
C  CONVERT WBAN CODES INTO 16 WIND DIRECTIONS
C
C
       CALL WINDIR (DATE.DIR.WDIR.IWD)
       CAT(IWD.IC)=CAT(IWD.IC)+1.0 $ NOBS=NOBS+1
  180   CONTINUE
        GO TO 100
 200   PRINT 8.CLASS
       DO 201 I=2.17
        TOT=0.0
       DO 202 L=2.7
       TOT=CAT(I.L)+TOT
202    CONTINUE
       TOTDIR(I)=TOT
201      CONTINUE
       ATOT=0.
       DO 203 L=2.7
       TOT=0.0
       DO 204 I=2.17
       TOT=TOT+CAT(I.L)
204     CONTINUE
       TOTSPD(L)=TOT
       ATOT=ATOT+TOT
203    CONTINUE
       DO 210 I=2.17
       PRINT 9.WNDIR(I).(CAT(I.L).L=2.7).TOTDIR(I)
 210   CONTINUE
       PRINT 18.(TOTSPD(L).L=2.7).ATOT
       PRINT 16.CAT(1.1)
C
C           COMPUTE THE PERCENTAGE OF OCCURRENCES AT EACH DIRECTION IN EACH
C      WIND SPEED CLASS.
       DO 220 IW=2.17
       DO 220 IC=2.7
       WSWD(IW-1.IC-1)=CAT(IW.IC)*100.0/NOBS
       IF (CAT(IW.IC).GT.0.0) NOC(IW-1)=IC-1
 220   CONTINUE
       CALMA=CAT(1.1)*100.0/NOBS
       PRINT12.CLASS
       DO 205 I=2.17
205    TOTDIR(I)=TOTDIR(I)/NOBS*100.
       DO 206 I=2.7
206    TOTSPD(I)=TOTSPD(I)/NOBS*100.
       ATOT=ATOT/NOBS*100.
       DO 230 I=1.16
       PRINT 9.WNDIR(I+1).(WSWD(I.L).L=1.6).TOTDIR(I+1)
 230   CONTINUE
       PRINT 18.(TOTSPD(L).L=2.7).ATOT
       PRINT 14.CALMA.NOBS
       STOP200
        END
```

THE TRANSPORTATION RESEARCH BOARD is an agency of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's purpose is to stimulate research concerning the nature and performance of transportation systems, to disseminate information that the research produces, and to encourage the application of appropriate research findings. The Board's program is carried out by more than 250 committees, task forces, and panels composed of more than 3,100 administrators, engineers, social scientists, attorneys, educators, and others concerned with transportation; they serve without compensation. The program is supported by state transportation and highway departments, the modal administrations of the U.S. Department of Transportation, the Association of American Railroads, and other organizations and individuals interested in the development of transportation.

The Transportation Research Board operates within the Commission on Sociotechnical Systems of the National Research Council. The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine.

The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the furtherance of science and technology, required to advise the Federal Government upon request within its fields of competence. Under its corporate charter the Academy established the National Research Council in 1916, the National Academy of Engineering in 1964, and the Institute of Medicine in 1970.

TRANSPORTATION RESEARCH BOARD
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

ADDRESS CORRECTION REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
WASHINGTON, D.C.
PERMIT NO. 42970

000015M001
JAMES W HILL
RESEARCH SUPERVISOR

IDAHO TRANS DEPT DIV OF HWYS
P O BOX 7129 3311 W STATE ST
BOISE ID 83707