EPA-450/2-77-024a
October 1977
(OAQPS NO. 1.2-083)

GUIDELINE SERIES

GUIDELINE ON PROCEDURES FOR CONSTRUCTING AIR POLLUTION ISOPLETH PROFILES AND POPULATION EXPOSURE ANALYSIS

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

EPA-450/2-77-024a
(OAQPS NO. 1.2-083)

GUIDELINE ON PROCEDURES FOR CONSTRUCTING AIR POLLUTION ISOPLETH PROFILES AND POPULATION EXPOSURE ANALYSIS

Monitoring and Reports Branch
Monitoring and Data Analysis Division

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

October 1977

OAQPS GUIDELINE SERIES

The guideline series of reports is being issued by the Office of Air Quality Planning and Standards (OAQPS) to provide information to state and local air pollution control agencies; for example, to provide guidance on the acquisition and processing of air quality data and on the planning and analysis requisite for the maintenance of air quality. Reports published in this series will be available - as supplies permit - from the Library Services Office (MD-35), Research Triangle Park, North Carolina 27711; or, for a nominal fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

Publication No. EPA-450/2-77-024a

(OAQPS No. 1.2-083)

PREFACE

Development of air pollution isopleths and population exposure analysis are components of the "Air Monitoring Strategy for State Implementation Plans" (EPA-450/2-77-010). The guideline document is intended to provide assistance in performing these activities. A companion "Users Manual" describes existing computer software applicable to these tasks.

ABSTRACT

This guideline document provides an overview of methodologies that exist for constructing pollutant isopleth displays and for estimating population exposure to air pollutants from air monitoring data. Actual examples of the methodologies are presented for applications to data for the New York-New Jersey-Connecticut area and for the Los Angeles area. This report is to assist the EPA's Regional Offices and States in reviewing their data bases to determine feasibiltiy of performing isopleth/population exposure analyses, and to guide regional, state, and local air pollution control agencies in actually conducting such analyses.

Spatial presentation of air quality monitoring information in the form of isopleth maps is useful to air pollution control agencies and highly informative to the public. Steps that must be taken to obtain an isopleth map from air monitoring data measured at widely separated monitoring stations as well as from dispersion-model-simulation outputs are described. Currently available methods of developing an isopleth map are documented for three types: manually drawn maps, character-printed maps, and line-drawn maps. Problems associated with each type of isopleth map are discussed.

A population exposure analysis, which combines air quality data and demographic data to estimate population exposure to air pollution, is outlined. Data-set preparation and analysis methods are explained for estimating both long-term and short-term exposure of the population to air pollution. Exposure to long-term average concentrations is described by the percentage of the population exposed above an annual National Ambient Air Quality Standard (NAAQS). Exposure to short-term (e.g., 1 hr or 24 hr) concentrations is quantified by the distribution of the population who are exposed above an hourly or 24-hour NAAQS at various percentages of the time (hours or days).

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

ACKNOWLEDGMENTS

This guideline document has been prepared by Neil Frank, Monitoring and Reports Branch, Monitoring and Data Analysis Division, Office of Air Quality Planning and Standards and Yuji Horie, Data Sciences Division, Technology Service Corporation, 2811 Wilshire Boulevard, Santa Monica, California 90403, under contract No. 68-02-2318, project No. DU-76-C190.

CHAPTER 1. INTRODUCTION

This guideline document is to assist the U.S. EPA, States and local air pollution control agencies in displaying air quality information in a geographical context and reporting air quality progress to the public in a more meaningful and understandable form. This report provides guidelines on developing contour lines of air pollution levels (isopleth map) from air quality data and presents an outline of the methodology that has been developed for analyzing population exposure to air pollution. To assist air pollution control agencies in actually doing such analysis, examples of the methodology and the computer programs developed in recent applications to the New York-New Jersey-Connecticut area and to the Los Angeles area are presented.

Isopleth maps are often used to illustrate the spatial variation of air quality and other environmental variables. An isopleth map can be drawn manually, or it can be drawn mechanically by using a readily available computer program such as SYMAP [1]. Because environmental variables are measured at widely separated monitoring stations, the performance of the spatial-interpolation scheme being used may significantly affect the isopleth map. Therefore, numerous interpolation schemes employed in meteorological mapping and in other areas are reviewed to determine their suitability for air pollution applications. The state of the art of spatial interpolation is discussed in regard to preparing isopleth maps for air pollution studies.

Although several reports [2,3,4,14] have recently been published on population exposure, the methods used need to be thoroughly explained before they can find wide application in various geographical regions. This report outlines the population exposure methodology that we have developed and discusses problems that have arisen in two applications -- one for the New York-New Jersey-Connecticut area and one for the Los Angeles area. Because a population exposure analysis may be performed under a variety of situations, alternative methods, as well as the currently applied methods, are discussed.

1.1 ANOTHER WAY TO STUDY AIR QUALITY

The nation's air quality is routinely measured by some 4,000 air monitoring stations all over the United States. While most studies of air quality data have been made for individual stations, important knowledge can be gained by collectively studying air quality information provided by an entire network of monitors in a specific geographic area. Conventional tabulations and graphs of air quality summary statistics, given separately for individual stations, do not provide a convenient basis for studying these spatial patterns of air quality. When air-quality monitoring information is presented in a geographical context, say through use of an isopleth map, one can readily comprehend the spatial variation in air quality.

A few economists and geographers have combined air quality data with socio-economic data, studying air pollution in relation to economic and demographic variables [5,6,7]. Following their lead, air pollution researchers have developed methods to quantify ambient air quality in terms of population exposure to air pollution instead of concentration units [2,3,8,9,14]. The methodology to perform population exposure analyses is described in this report.

1.2 POTENTIAL BENEFITS

Geographically oriented air-quality presentations supplement conventional tabulations of air-quality summary statistics. Therefore, an isopleth map of air quality data is very effective for providing:

- A visual perspective of the spatial variation in air quality and patterns of human exposure levels;
- A meaningful reference for evaluating air quality trends in relation to population-growth patterns, and temporal and spatial emission trends;
- A spatially representative baseline of air quality for evaluating and modifying regulatory programs (e.g., State Implementation Plans and new source reviews).

National Ambient Air Quality Standards (NAAQSs) have been set to protect the public health and welfare. Ambient air quality data provide the basic means of quantifying whether the public benefits by trends in air pollution levels. To make this quantification directly relatable to public health, it is useful to report air quality in terms of population exposure levels. To do so requires synthesis of population and air quality data bases.

Air pollution concentrations alone are not indicative of population health risks, because they do not specify what fraction of the population is exposed to various concentration levels. For example, it may be more meaningful for residents of an urban area to learn that the percentage of the urban population living in areas above the primary national standard for particulates decreased from 60% in 1970 to 10% in 1975, than to learn that the annual mean pollutant concentrations decreased from 69.5 $\mu g/m^3$ to 64.3 $\mu g/m^3$ in the same period. In the latter statement, even the people who know the

NAAQS (75 μg/m^3 for the primary and 60 μg/m^3 for the secondary) may misinterpret the implications of meeting the primary NAAQS but failing to meet the secondary standard in both 1970 and 1975. The former statement tells people, regardless of their knowledge, that the particulate air pollution in the urban area improved from 1970 to 1975 and that the number of people exposed to health risks from particulate matter was less in 1975 than in 1970.

The Population Exposure Methodology discussed in this report focuses on exposure to air pollution above established air quality standards. Exposure to long-term average concentrations is characterized by that percentage of the population exposed to air pollution above the yearly average NAAQS. Exposure to short-term (e.g., one hr or 24 hr) concentrations is quantified by the average percentage of the time (hours or days) that people were exposed above the short-term NAAQS. A more complete characterization of population exposure is given by cumulative distributions of people's exposures to long-term average or short-term concentrations.

1.3 PERFORMING A SPATIAL ANALYSIS

To develop an isopleth map from air monitoring data, a network of air monitoring stations must adequately cover the area of interest. The required size may vary from one pollutant to another. If sufficient monitoring sites are not available, isopleth analysis might depend on the results of a calibrated dispersion model. As few as three stations may be needed for estimating a spatial concentration distribution of photochemical oxidant and other secondary pollutants. For the primary pollutants such as

CO, TSP, and SO_2, a larger number of stations are required to obtain a meaningful isopleth map over a study area. In addition, these stations must be clustered so that concentration in the neighborhood of the monitoring sites can be inferred from those measured by the stations.

Isopleth maps can be used in several ways to describe air quality. First, an isopleth map for a single year can be used to inform the public of the spatial variation in air quality. Second, a series of isopleth maps for selected years can be used to present the trends in the spatial patterns of air quality over time. Third, by computing the land area within each isopleth level, both the regional air quality level and the percentage of land area above the NAAQS can be determined. Fourth, by combing air quality data with demographic data, exposure of the population to air pollution can be determined.

In obtaining an accurate isopleth map or in determining population exposure to air pollution, a rather elaborate analysis method and some computer software are required. The analysis method for developing an isopleth map is discussed in Chapter 2; that for performing a population exposure analysis is described in Chapter 3. The computer programs for actually doing such analyses are presented in a separate volume of the User's Manual.

CHAPTER 2 SPATIAL DISPLAY OF AIR QUALITY INFORMATION

Ambient air quality in most major U.S. cities is routinely measured by a network of air monitoring stations. The display of air quality monitoring information in a spatial context is useful to air pollution control agencies and highly informative for the public. However, many factors have to be considered in displaying air quality data from a finite set of monitors on a spatial scale. Considerations may include the number of stations, configuration of monitoring sites, spatial gradient of pollutant concentrations, emission source inventory, geographical features, and local climatic characteristics.

When air monitoring stations are too few, too far apart or clustered in too small an area, a spatial display of air quality monitoring information cannot be made in a meaningful manner. A spatial analysis might depend on the results of a validated dispersion mode. The first section discusses items to be included in checking feasibility of a spatial display of monitoring information.

The simplest method of displaying monitoring information in a spatial context is to plot monitoring sites and air quality values on a map and to manuall draw an air-quality isopleth map. In drawing an isopleth line, one employs either consciously or unconsciously, an "eyeball estimate" of air quality at nonstation locations from air quality values measured at nearby monitoring stations. This eyeball interpolation involves all the considerations of pollutant-concentration dispersion characteristics that the person perceives. The importance of these subjective factors in drawing an isopleth map is discussed in Section 2.2.

Isopleth displays have been used more frequently for dispersion-model outputs than for air monitoring data. The main reason is that air-quality monitoring information was not adequate for developing an isopleth map in many urban areas. In recent years, however, the air monitoring network in metropolitan areas has expanded considerably and its data quality has improved measurably. Section 2.3 discusses two alternate methods of estimating air quality at nonstation locations by 1) spatially interpolating air-quality monitoring information, and 2) applying an air quality dispersion model. Procedures involved in each of the two methods are described in separate subsections. Advantages and disadvantages of each method under various situations are also stated.

Modern computer graphic techniques can be used for drawing an isopleth map after air quality values are estimated at sufficiently dense grid points. Section 2.4 presents an overview of currently available computer mapping techniques and several examples of computer-drawn isopleth maps of air quality parameters.

2.1 FIRST STEP IN SPATIAL ANALYSIS - FEASIBILITY

Given a study pollutant and a geographical area, the first thing to be considered is whether an isopleth analysis will be feasible. It is suggested that the isopleth analysis be based on monitoring data. This is the easiest way to establish trends over time and provides a basis for convenient updates. If this is not possible, then the isopleth analysis might be based on the believable results of a calibrated dispersion model.

Several items are useful for establishing the feasibility of an isopleth analysis based on monitoring data. These items include previously published isopleth maps, monitoring location maps and a check of data completeness.

As obvious as it may sound, a good starting point is to check if isopleth maps have been previously published. These could be based on a past year of monitoring data or the results of a modeling exercise. This basically indicates that an isopleth analysis <u>can</u> be done and will provide a basis for comparison of relative spatial patterns with new analysis. This comparison will also ensure that contradictory new results won't be produced.

Another important item is a map which shows the location of existing monitors plus those which previously produced historical data. This will show if monitoring data might serve as the basis for the isopleth analysis. This map will help to establish the spatial coverage of the monitors and also help to define the boundaries of the possible study area. Some hints on locating monitoring sites on a map is discussed in Appendix A.

Next, air quality data completeness must be checked for all candidate monitors. The check list must include:

a. Validity of air monitoring data over a time period of interest, say a year or a season, and

b. Historical continuity of data over several years for a trend analyses.

2.1.1 Spatial Coverage of Monitoring Data

Suppose that 5 stations report hourly CO concentrations with an acceptable data quality at three locations near highway and at two urban background sites. In this case, it may not be feasible to draw a concentration isopleth map from the CO monitoring data because each station reading of CO concentration is representative for such a small area (less than 1 mi around a station) that CO concentrations at a place apart from a station site may differ considerably from concentration readings at that station.

On the other hand, if the above example is for O_x, it is quite possible to develop an isopleth map by interpolating the station readings to places between stations. The reason is that the concentration gradients of O_x are generally so small that O_x concentration at any place within the urban area can be estimated reasonably well from concentration readings at the nearby stations.

A "representative area" of a monitoring station may be defined as a distance from the station at which a pollutant concentration remains almost constant, say it does not change more than ±20% from the reading at that station. Then, based on expected pollution gradient of long-term concentration statistics (e.g., annual mean and annual percentile concentration), the "representative area" presented in Table 1 may be used as a guide of how far from a station air monitoring data can be extrapolated for various pollutants. The representative areas specified in Table 1 are only a guide and are open to user modifications. For example, a site affected by local emission sources may have a much smaller representative area than the representive area listed in Table 1. Similarly, a monitoring site in a high pollution gradient area (e.g., due to complex terrain or due to high emission density) may have a smaller representative area. On the other hand, a representative area by a monitor located in a uniform land use zone (e.g., agricultural land) may be substantially larger than the typical representative area of Table 1. As a matter of fact, areas with a similar land use tend to have a similar level of air pollution. In many cases, a land use map will be quite useful for refining a representative area on a site-by-site basis.

Although a representative area for the primary pollutant is influenced strongly by a particular site characteristic, this may not be a problem with the secondary pollutants (e.g., O_x and $SO_4^=$) whose concentration gradients tend to be moderate. One exception is that O_x concentration gradient near a heavily traveled highway is quite sharp. Therefore, O_x monitoring site near a busy highway will have a much smaller representative area than the typical representative area for O_x monitors.

Using the representative area listed in Table 1, one can check the feasibility of an isopleth analysis for a given set of air monitoring stations and a given pollutant. Various configurations of air monitoring stations are given in Figure 1 to illustrate how to check the feasibility of an isopleth analysis and to determine the study area.

Suppose that, after quality and completeness of air monitoring data have been checked, the stations whose locations are designated by dots in Figure 1 have been selected for a spatial analysis. By drawing a circle around each station, with the radius being equal to the representative area, one can determine a potential study area for a given configuration of station sites. Figure 1-a shows that the stations are located so far apart from each other compared to the representative area for TSP that one cannot form a contiguous study area. On the other hand, Figure 1-b shows that although two stations are too distant from any other station, the rest of the stations form a contiguous study area.

Table 1. Typical "Representative Area" of a Monitoring Station Estimated for Each Major Pollutant from the Expected Pollution Gradients of Long-Term Concentration Statistics.

Pollutant	Representative Area Distance from a Station
CO	Less than 1 mi
SO_2	3 mi
TSP	3 mi
NO_2	5 mi
O_x	10 mi
$SO_4^=$	20 mi

Figures 1-c and 1-d are for O_x. Although the stations of Fig. 1-c are scattered as far apart from each other as those of Fig. 1-a, the large representative area for O_x monitoring sites makes it possible to form a contiguous study area. In Fig. 1-d, the three stations are so far apart from the rest of the stations that the study area is split into two areas.

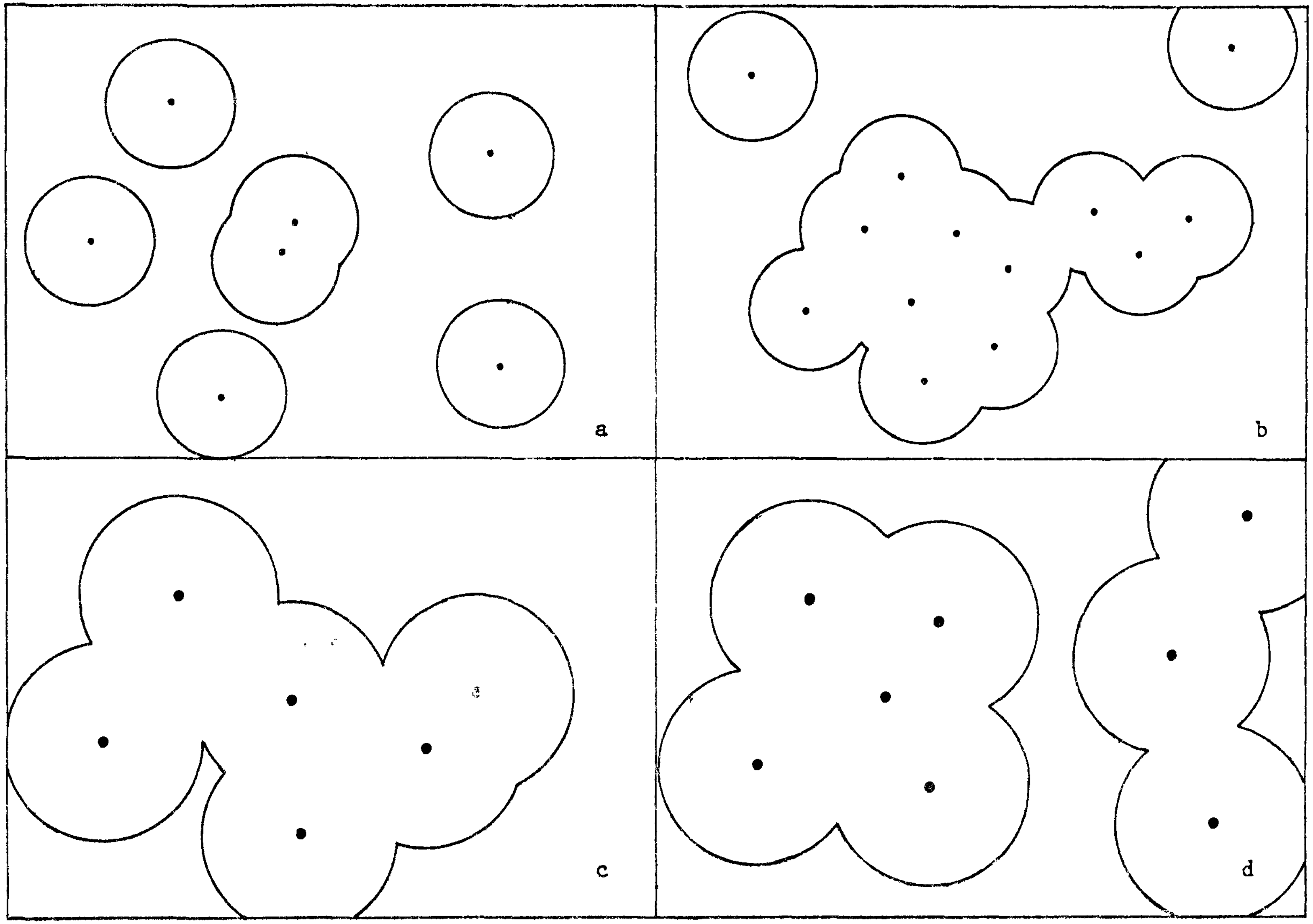


Figure 1. Determination of Study Area from a Set of Monitoring Stations by Using the Range of Influence (a and b for TSP, c and d for O_x).

When a set of monitoring stations cannot form a contiguous study area covering the region of interest, the spatial distribution of pollutant concentrations must be estimated by knowledge about pollution behavior in addition to the air monitoring data. If it is known that the pollution levels are intermediate between those of adjacent monitoring stations, then these areas can be estimated by a credible interpolation formula. Also a dispersion model may be helpful for depicting a spatial concentration distribution of a primary pollutant. In addition, pollution level in areas with no monitors may be guessed from the concentration readings of monitors in areas with the same or similar land use and similar adjacent emission sources. AQDM [10] and CDM [11] computer models are useful for SO_2 and TSP, while APRAC [12] and HIWAY [13] computer models are useful for CO. The use of these dispersion models for estimating a spatial distribution of concentration is discussed in Section 2.3.2.

2.1.2 Completeness of Air Quality Data

When air monitoring data are used for developing an air quality isopleth map, credibility of the monitoring data must be checked. The check list must include a) completeness of air monitoring data over a time period of interest, and b) historical continuity of data over several years for a trend analysis.

In reporting a seasonal or annual air quality, we have to check, for each station, the number of valid concentration measurements during the period of interest. In distinguishing valid stations from invalid stations, we need a specific criterion as to the minimum number of valid concentration measurements per station-year or station-season. The criterion for a valid station-year used by the U.S. Environmental Protection Agency's National Aerometric Data Bank (NADB) will serve as a preliminary data screening tool.

The NADB uses separate criteria for continuous and intermittent monitoring. For continuous monitoring, a year (or quarter) of data is valid if at least 75% of the total number of possible observations were recorded. For intermittent monitoring a year of data is valid if it has four valid quarters of data. For a valid quarter, a minimum of five observations is required. In addition, if one month has no measurements, the other two months must have at least two observations.

In order to establish air quality trends, spatial analysis of air monitoring data would be performed for a multi-year period rather than for a single year. This would entail at least two isopleth maps within a several year study period. A series of maps would show the change in the spatial pattern of air quality and can specifically show the change in the study area having air quality values above established air quality standards.

Ideally, only "trend" stations meeting the valid station-year criteria in every year during the study period should be selected for the analysis. In actual situations, however, such stations are usually too few to adequately depict the spatial variation in air quality. One solution to this problem is to use those stations that reported air quality data in several years. Then the air quality at stations in the missing years can be estimated from other data. If isopleth maps were developed only from the reporting stations in a given year some bias might be introduced into the trend analysis.

There are several possible approahces for estimating missing values. If the changes in air quality is gradual, then linear interpolation of air quality values between neighboring years might be used. Because of the problem of what to do with missing data in the start and end years, an alternate approach is to make use of the region-wide average change in air quality between successive years.

The average year-to-year change can be established from all sites with data for each successive two year period. This change can then be applied to estimate missing values. A third potential approach is to use a generalized linear model with class variables for years and sites. Unfortunately, this analysis of variance procedure will probably involve an unbalanced design due to the missing values and this may not be computationally convenient without the aid of a statistical analysis computer package. The missing values would be estimated from a term which describes the average air quality value for each year and from a term which describes the average air quality value for each site. For each of the above procedures, we can make use of emission density and/or land use maps to select an appropriate subset of sites that should be used for the estimation procedures.

2.1.3 Other Considerations

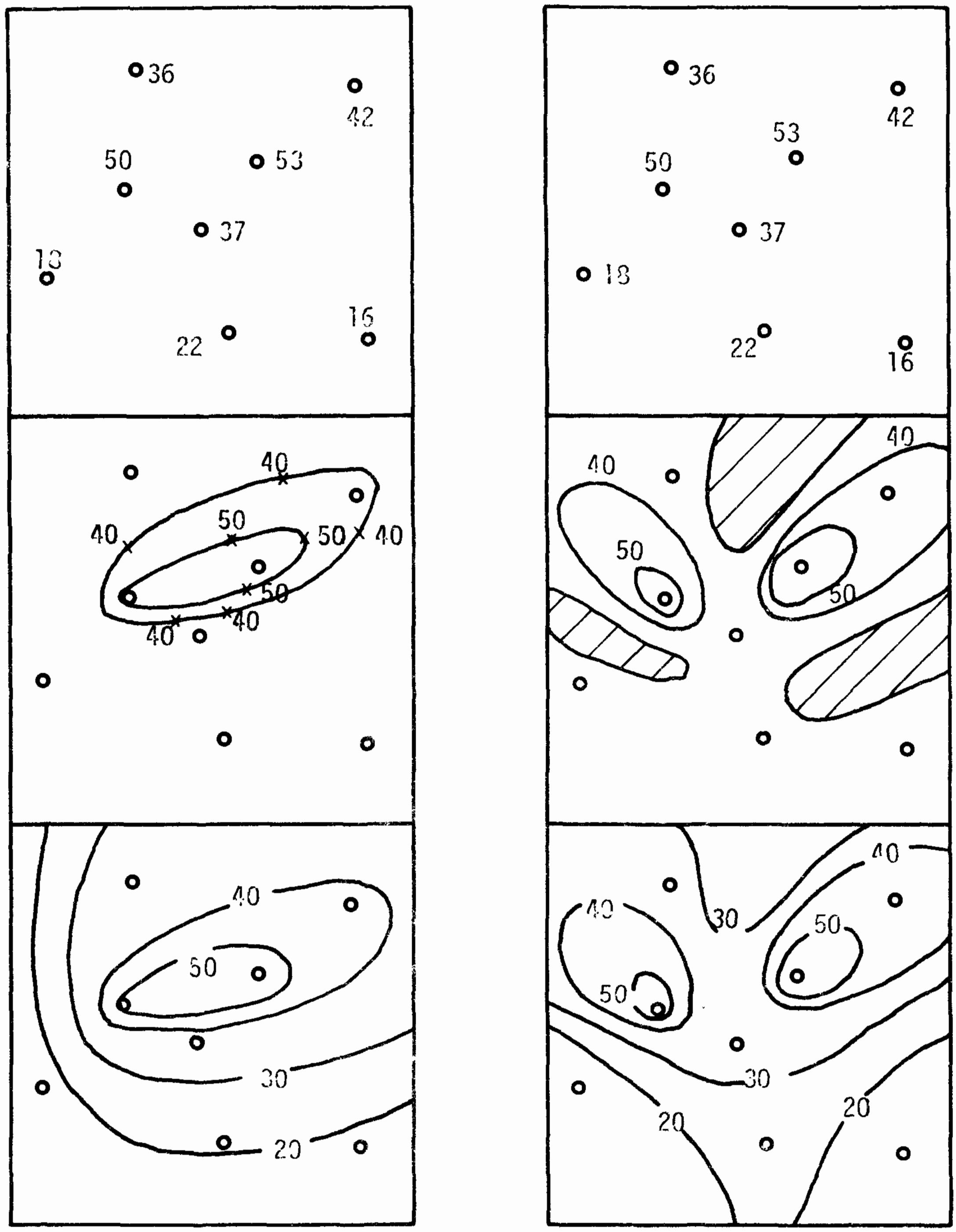
An adequate "area representativeness" of each monitoring site is essential to correctly estimate a spatial distribution of air quality from air monitoring data. If a monitoring station reports an extraordinary concentration caused by a local "hot spot", that station must be given special consideration in the analysis. The "hot spot" might be indicated on the isopleth maps as a localized condition. Because it is not representative of a wide area, it should not be included in a spatial analysis that estimates air quality by spatial interpolation from neighboring monitoring stations or in a population exposure analysis to be described in later sections of this guideline.

Site characteristics are another important factor. Since a pollutant concentration varies greatly with height above the ground, some adjustment of air monitoring data might be necessary to account for the height effects on concentration readings. Concentration readings at an elevated monitoring site, say, 50 ft above the ground, could be low because of the greater dilution at that elevation compared to the one at 5 ft above the ground. The height effect is great for CO and the other primary pollutants, but is not so great for O_x and the other secondary pollutants. Therefore, careful consideration of site characteristics such as height and local emission influence must be given for selecting a set of stations reporting concentrations of a primary pollutant.

2.2 MANUALLY DRAWN ISOPLETH MAPS

When air monitoring data are available at several station sites, the first step necessary for obtaining an isopleth map is to plot the station locations and concentration (or other air quality parameter) levels on a skeleton map as shown in Fig. 2.

Figure 2-a illustrates a standard procedure for drawing an isopleth map when one has concentration/station site information only. Starting with the stations where the highest concentration occurs, straight lines are drawn connecting adjacent station sites. When the concentration levels at two



a. Without Subjective Considerations

b. With Subjective Considerations

Figure 2. Illustration of the Stepwise Procedure for Preparing Isopleth Maps with and without Subjective Considerations.

adjacent station sites are considered, an intermediate grid point is assigned to one or more location on the edge where one or more isopleth levels fall. To find such an intermediate point between two stations, we use an eyeball interpolation of the two station values, which is usually similar to a linear interpolation. As seen in the second illustration of Figure 2-a, we can draw an isopleth line around the station with the highest concentration by connecting the intermediate points with the isopleth level of 50. Then, proceed to the next isopleth level, 40. Now, we consider all stations with concentrations near 40, plus their neighboring stations. By repeating the above procedure, an isopleth map is completed in the third illustration of Figure 2-a. Here, intermediate points and station values are removed to have a clear isopleth map.

The procedures described above are quite similar to those used in computer software for drawing an isopleth map. Figure 2-b illustrates a typical isopleth-drawing procedure employed by an air pollution expert and an experienced meteorologist. These experts also plot the station locations and concentration levels on a skeleton map. However, because they know the geographical feature of the area and its effect on the concentration field, they incorporate their subjective considerations in drawing an isopleth map.

In Figure 2-b, the locations of mountains are designated by hatched areas. Because an air pollution expert knows that a polluted air mass on one side of the mountains does not mix with the one on the other side of the mountains, he may draw an isopleth line separately for each of the two polluted valleys as shown in the second illustration of Figure 2-b. With his subjective consideration of the mountains, his final isopleth map, shown in the third illustration of Figure 2-b, is totally different from the isopleth map obtained without considering the mountains (Figure 2-a).

The above example considers only the effect of mountains on a concentration field. An air pollution expert may also consider potential effects on a concentration field of wind-flow patterns and configuration of major emission sources. Because the effects of mountains, emissions, and wind patterns on a pollutant concentration field are difficult to quantify, very little elaboration has been made in incorporating the subjective considerations mentioned above in a computer-drawn isopleth map. A technique to incorporate some of the subjective considerations in a spatial interpolation scheme is discussed in the next section.

2.3 SPATIAL REFINEMENT OF AIR QUALITY INFORMATION

The estimate of the spatial variation of air quality can be refined by two methods -- spatial interpolation of air monitoring data, and air quality simulation by a dispersion model. The former method is appropriate when air quality data are available at closely spaced network points. The latter method is more suitable when air quality data are not available at all (e.g., future air quality) or are available only at a few sites.

For most applications, a spatial distribution of pollutant concentration was obtained from dispersion-model outputs. The reason has been that, in early years of air quality surveillance, there were only a few air monitoring stations even in a large metropolitan area. Therefore, it was almost impossible to estimate a spatial distribution of pollutant concentration from concentration readings at a few widely separated monitoring sites.

The monitoring data have been used to validate the model calculations. However, in recent years, most major U.S. cities and metropolitan areas have a large and dense network of air monitoring stations. Because of this development of air monitoring networks, we can now develop an isopleth map of various air quality parameters from air monitoring data and monitoring site information.

2.3.1 Spatial Interpolation of Air Monitoring Data

To obtain a reproducible air quality isopleth map from air monitoring data, a credible spatial-interpolation scheme must be used to estimate concentrations at places other than monitoring sites.

The most commonly used interpolation scheme is linear interpolation. There are two types of linear interpolation formulas: true linear interpolation and pseudo linear interpolation. When applied to estimate a concentration at a place other than monitoring sites, the linear interpolation formulas may be written as

$$C_i = \left(\sum_{j-1}^{3} a_j C_j / d_j\right) \Bigg/ \left(\sum_{j=1}^{3} a_j / d_j\right) , \qquad (1)$$

where d_j is the distance between the i^{th} grid (or receptor) point and the j^{th} station site. C_j, $j=1,2,3$, is the observed air quality at each of the three stations nearest to the i^{th} grid point. The sign parameter a_j takes only

+1 inside the triangle formed by the three stations for both the true linear and pseudo linear interpolation formulas. However, outside the triangle, a_j of the true linear interpolation formula takes either +1 or -1 depending upon the position of the grid point, while a_j of the pseudo linear formula takes +1 only irrespective of the position of the grid point.

The performance of the linear interpolation formulas, Eq. (1), is shown in Figs. 3 and 4 for a hypothetical case where observed concentrations of 85, 15, and 60 (in arbitrary units) are assumed at points A, B, and C, respectively. In both figures the spatial distribution of the interpolated concentrations inside the triangle is generally in fair agreement with what we would expect from the concentrations at A, B, and C. Outside the triangle, however, the true interpolation formula tends to over-extrapolate the concentration values assigned at A, B, and C (Fig. 3). This over-extrapolation tendency of the true linear interpolation formula is very undesirable for air pollution applications. The pseudo linear interpolation formula does not show such over-extrapolation tendency (Fig. 4). In the resulting isopleth map, however, we miss both the highest observed concentration, 85, and the lowest, 15. Therefore, the pseudo linear interpolation formula is said to be over-smoothing particularly for its application to air quality mapping.

To avoid the problems of over-smoothing as well as over-extrapolation, the performance of various candidate interpolation formulas was examined. Among the interpolation formulas examined, the parabolic interpolation formula, given by Eq. (2), showed the most satisfactory performance [2].

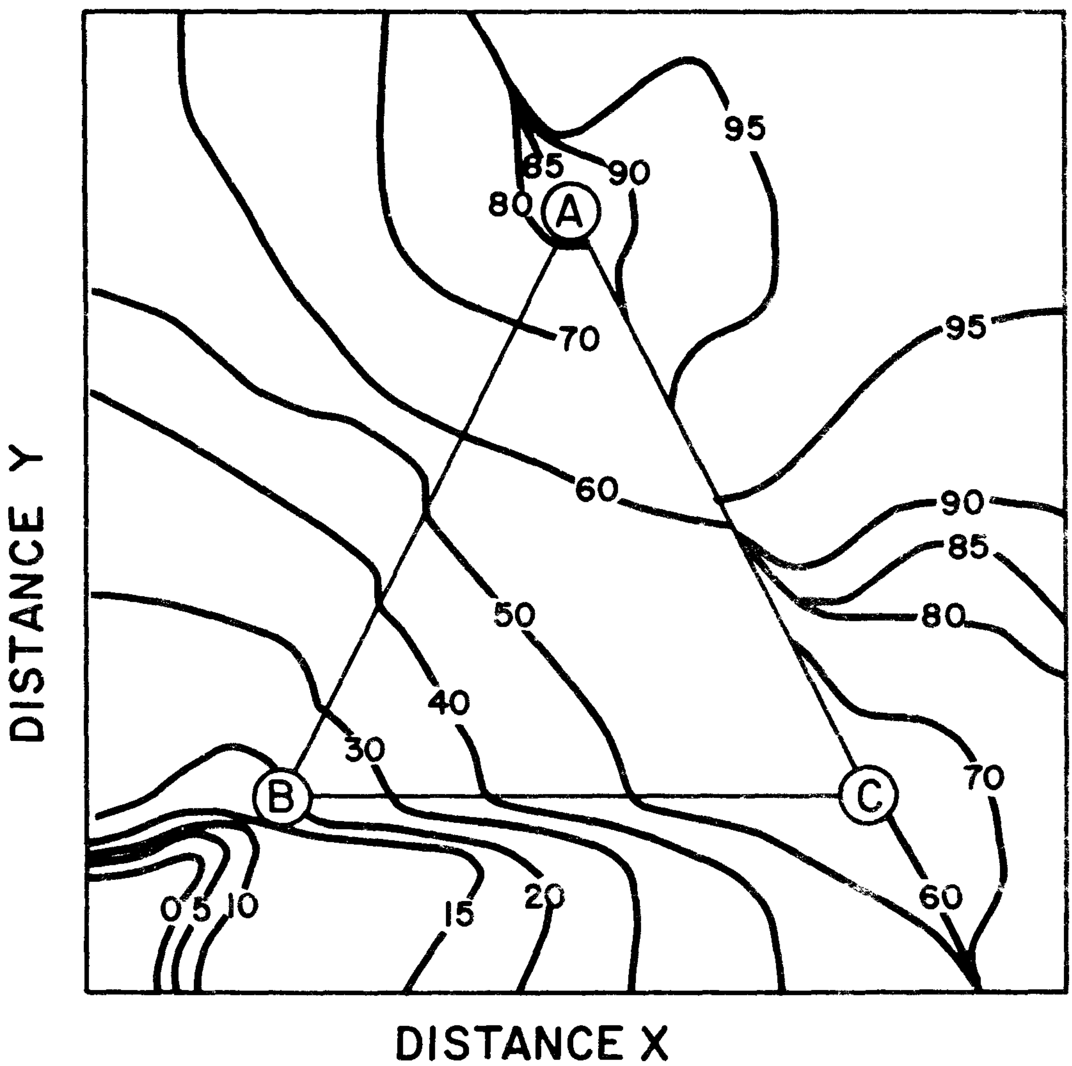


Figure 3. Performance of the True Linear Interpolation Formula [2].

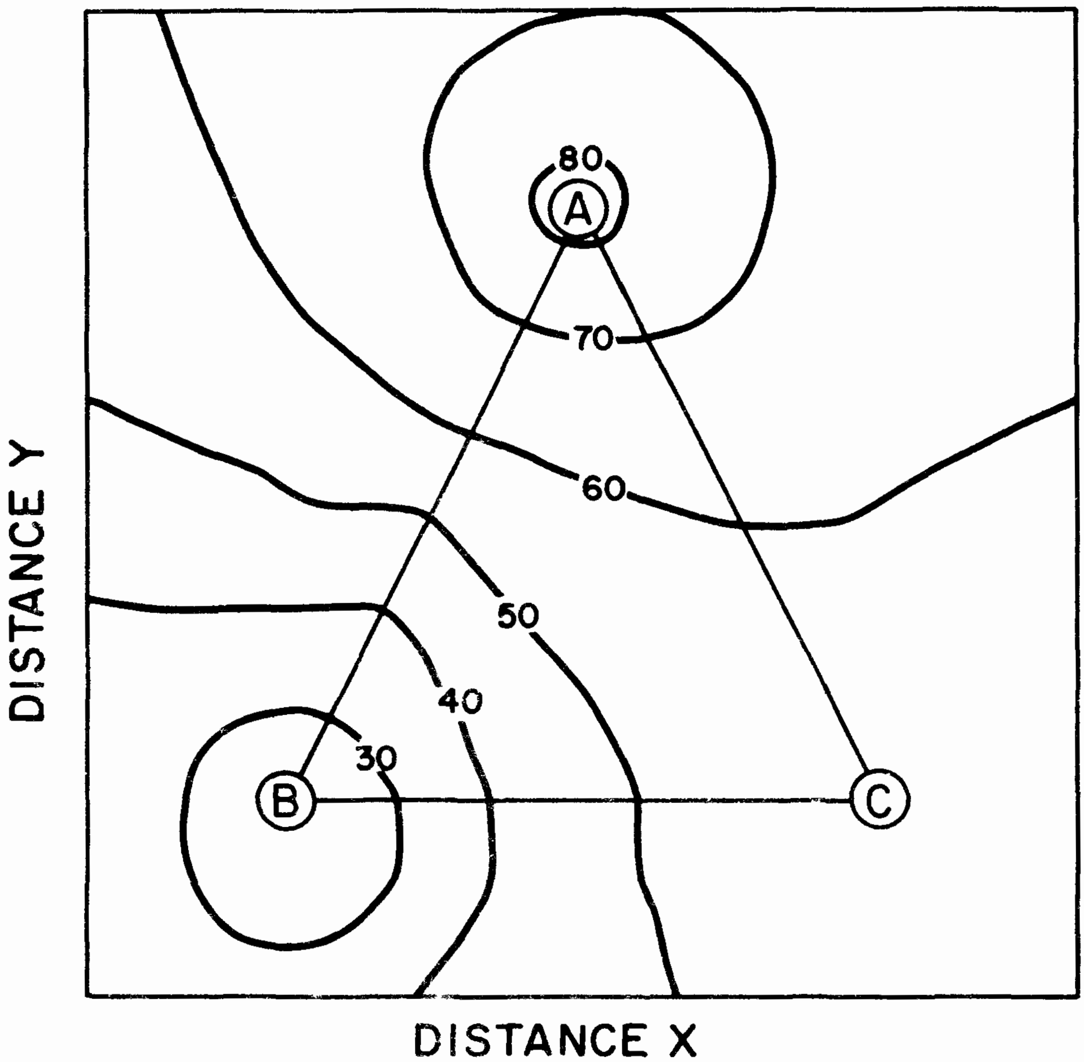


Figure 4. Performance of the Pseudo Linear Interpolation Formula [2].

$$C_i = \left(\sum_{j=1}^{3} C_j/d_j^2 \Bigg/ \sum_{j=1}^{3} 1/d_j^2 \right) . \tag{2}$$

The concentration isopleth map obtained by the parabolic interpolation formula is shown in Fig. 5 . It shows the isopleths of the highest, 85, and lowest, 15, observed concentrations. Although the parabolic interpolation formula is not perfect in its performance, it certainly minimizes the problem of over-smoothing as seen in the pseudo linear interpolation formula (Fig. 4). The conservative characteristic that no estimated value exceeds the maximum observed concentration nor is any estimated value less than the minimum observed concentration is also desirable for air pollution applications.

In Eq. (2), the summation can be extended to more than three stations. Although some interpolation schemes use more than three data points, we do not recommend to include more than the three nearest stations for interpolating the concentration at a non-station grid point. The reason is as follows: when more than three stations are used, the interpolated concentrations near the center of the triangle could become either higher or lower than any of the observed concentrations at the stations A, B, and C, depending upon the concentration levels at the other stations included in the summation.

Most computer mapping programs, e.g., SYMAP [1], employ linear interpolation or its derivatives. To avoid the aforementioned problems, the parabolic interpolation formula given by Eq. (2) should be used to interpolate air monitoring data to strategically located grid (or receptor) points before any computer mapping program is applied to the air monitoring data. The computer map drawn to air monitoring data at the station sites <u>and</u> estimated air quality at the

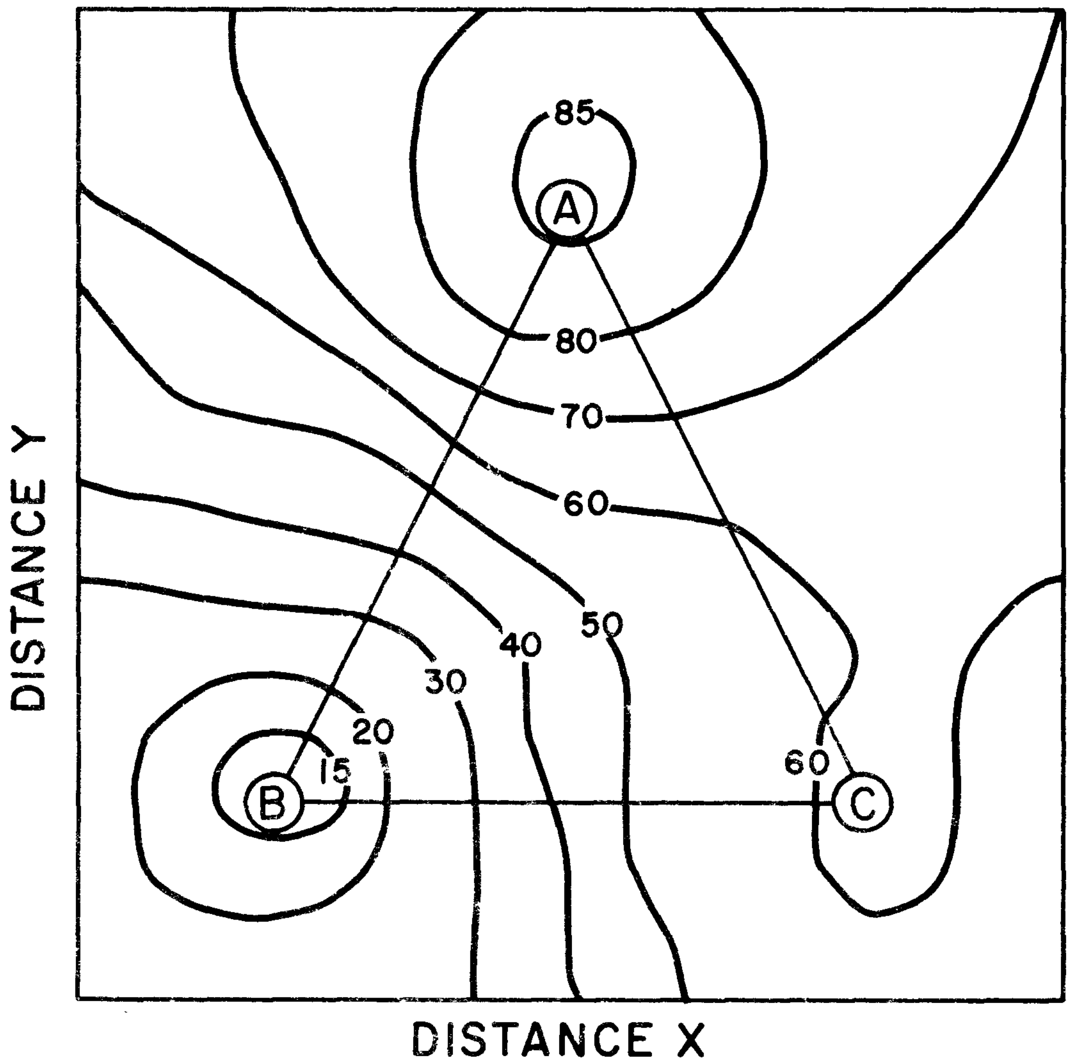


Figure 5. Performance of the Parabolic Interpolation Formula [2].

grid points should be closer to the real spatial distribution of air quality than a computer map drawn to the air monitoring data at the stations alone.

Any interpolation formula by itself cannot incorporate the effects of mountains, prevailing wind patterns, and configuration of major emission sources on the concentration pattern. Some computer algorithms (e.g., SYMAP, LPEM, and SPEM in User's Manual) can incorporate into air quality interpolation schemes the effects which geographical features exert on pollutant concentration fields. For example, when stations are located on both sides of mountains, the interpolated concentration at a receptor point will be determined primarily from the concentration readings at the stations located on the same side of the mountains as that receptor grid point. This effect can be accomplished by imposing a penalty distance to the stations on the opposite side of the mountain. This penalty distance simulates the difficulty of an air mass crossing the mountain. The location of mountains or other barriers would be approximated by a series of piecewise-linear penalty functions having a particular penalty value at each node. The technique can be applied to take into account the effects of the prevailing wind pattern, the configuration of large bodies of water and major emission sources. These techniques are incorporated into spatial interpolation schemes used in LPEM and SPEM computer models as well as the SYMAP graphics package (see User's Manual).

Spatial interpolation of air monitoring data should not be applied without constraints imposed by the physical reality of air pollution. The concentration reading at a station is representative only of a specified area about that station. As discussed in Section 2.1, the "representative area" depends on the type of pollutant and emission source pattern and is inversely associated with concentration gradient. A valid interpolated concentration will be limited to a receptor point lying within the representative area from the nearest station. Interpolated concentrations at receptor points outside the representative area of the nearest station should be denoted by some warning symbol (e.g., asterisks (*)). In this manner, the air quality in the study area will be estimated from concentration readings at appropriate monitoring stations. The study area may turn out to be one of the four cases illustrated in Fig. 1. In case (1-a), we would have to resort to a dispersion model or some other method than spatial interpolation of air monitoring data to estimate air quality at intermediate nonstation grid point.

2.3.2 Air Quality Simulation by Dispersion Model

Dispersion models are widely used to simulate the spatial distribution of air quality under various situations. The so-called climatological dispersion model--which uses, as input data, the emission inventory and the joint frequency distribution of wind speed, wind direction, and atmospheric stability--is considered useful for estimating the spatial variation of annual mean concentrations for nonreactive pollutants such as SO_2 and TSP.

A dispersion model would be more useful than spatial interpolation of air quality data when air monitoring data are scarce or when air quality data are not available. If a good emission inventory and meteorological data were available, it would also be useful for areas where there are high gradients in pollution concentration (e.g., point SO_2 source areas) and where the monitoring network might not be dense enough to measure these changes in pollution levels.

A dispersion model should first be calibrated by comparing the simulated values with the observed air quality at the monitoring station(s). Air quality at places including but not limited to the station sites would then be given by the predictions of the calibrated model. For additional details on modeling, refer to an upcoming OAQPS Guideline on Air Quality Models.

A dispersion model can be used to estimate the spatial distribution of future air quality when a proposed project is completed or when a proposed control strategy is implemented. First, the dispersion model is applied to the present emission inventory and present meteorology to simulate present air quality. Second, the simulated values are compared with the observed air qualities at several monitoring stations to calibrate the dispersion model. Then, the calibrated model is applied to the postulated emission inventory and multi-year average meteorology to simulate the air quality likely in the future.

A critical limitation of a climatological dispersion model is that it cannot be applied to complex terrain. When applied to flat or gentle terrain, a climatological dispersion model does generate useful spatial information of pollutant concentration, although it may not predict accurately a level of pollutant concentration. Therefore, there exists a possibility of obtaining a more accurate isopleth map by combining the spatial information from air monitoring data and from dispersion-model outputs than an isopleth map obtainable from air monitoring data or dispersion model outputs alone.

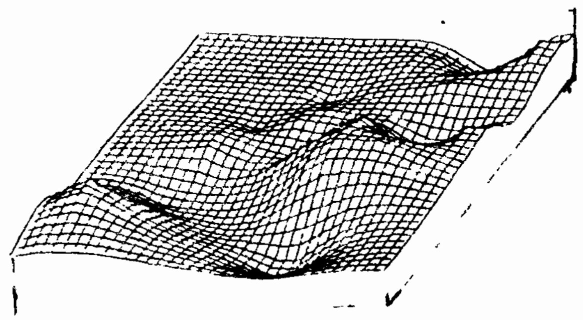
The following discussion presents a conceptual approach to accomplish the above objective. Specific guidelines are not yet available at this time.

Suppose that a predicted value overestimates a concentration level at one monitoring site in the south and underestimates at the other two monitoring sites in the north. Further, suppose that the emission inventory data used for a dispersion model are so complete that at least the spatial distribution pattern of predicted concentration is believable. One could then obtain a better isopleth map by superimposing the spatial distribution over an error field, as shown in Fig. 6. The error field is a plane that passes through the three points (x_m, y_m, e_m),m=1,2,3, where (x_m,y_m) is the x-y coordinates of the m^{th} monitoring station and e_m is the prediction error at that station, i.e., the observed concentration minus the predicted concentration.

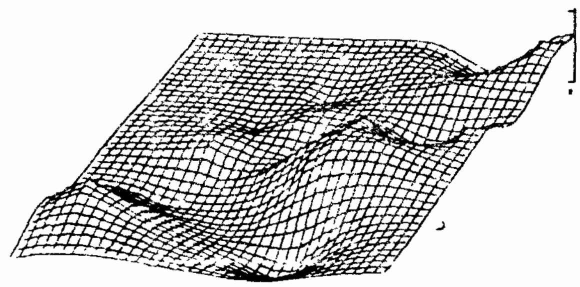
When there are more than three monitoring stations, the study area should be divided into several sub-areas formed by three adjacent stations. Then, the above adjustment can be made separately on the predicted spatial distribution for each sub-area. When a dispersion model is applied to complex terrain, the study area should be divided into several flat segments. Then, the adjustment can be made separately for each flat segment.

2.4 COMPUTER-DRAWN ISOPLETH MAPS

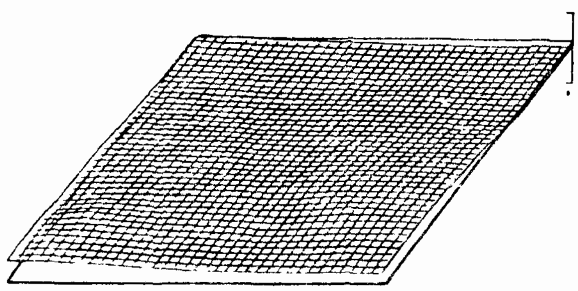
Before deciding to use a computer for drawing an air quality isopleth map, one has to consider advantages and disadvantages of employing a computer-graphics technique to his particular problems. Although a computer-drawn isopleth map is neat and objective, it requires some familiarity with computer graphics and a lot of data-base preparation. When prepared carelessly, a computer-drawn isopleth map can be quite erroneous. As mentioned in Section 2.2, one can incorporate into his hand-drawn isopleth all the considerations of pollutant-dispersion characteristics in a particular geographical area. These subjective factors often play a vital role in determining a pattern of isopleth lines, as demonstrated in Figure 2.



Adjusted Concentration Field



Simulated Concentration Field



Error Field

Figure 6. Adjustment of a Simulated Concentration Field According to Differences Between the Observed and Predicted Concentrations.

To draw an isopleth map by computer, all the data points and data values must be numerically specified in either computer cards or magnetic data tapes. In addition, isopleth levels and boundary of a study area must be numerically specified. This preparatory work requires substantial time and effort. Therefore, it may not be practical to use a computer for drawing a single isopleth map. Computer mapping is the most advantageous for drawing many isopleth maps repetitiously for similar applications. Here, a study area does not have to be the same as long as the type of isopleth mapping is similar.

Two types of computer-drawn maps are discussed in this report: character-printed maps and line-drawn maps. Character-printed maps are produced by a standard computer printer, which prints typewriter-like characters on standard computer-printout paper. Line-drawn maps are produced by either a pen plotter or a cathode ray tube (CRT). A brief description of the two types of computer mapping is presented in the first two subsections. In the third subsection, a hybrid method incorporating manually drawn and computer-drawn isopleth maps is explained. Computer software for these applications is presented in the User's Manual of this guideline document.

2.4.1 Character-Printed Maps

Character-printing--given access to a standard highspeed computer--is probably the most common means of drawing an isopleth map. SYMAP is a widely used computer program for drawing a variety of maps [1]. When SYMAP is used, the basic requirements are to specify the map boundaries numerically and to select the type of isopleth map from among three options: Conformal, Contour and Proximal [1].

In each of these maps, the study region is represented as a series of polygons. The polygons can be used to depict sub-areas such as counties. These polygons must be numerically specified and are used as basic input data of the package. Geographical features can also be incorporated into the resultant figure.

When the Conformal option is chosen, each polygonal area must be associated with a data value. Figure 7 is an example of the Conformal option. In this case, each polygonal area represents a sub-county census division. Solid lines, showing the boundaries of various counties, are overlayed over the Conformal SYMAP output. The SYMAP program simply compared each data value with the ranges of average earnings of male workers and printed the appropriate character in the polygonal area.

The Contour option is based on a concentration field established by linear interpolation among existing data points. With this option, the points at which data values are available must be specified by the X-Y coordinates. The contour intervals are pre-specified and interpolated values at intermediate data points are compared with the specified ranges of the variable. Then, the computer prints the appropriate symbol at each data point. The print positions at boundaries of one contour level and another can be left blank so that the shaded region of one contour level is clearly separated from the other shaded regions by a blank space. Due to the use of linear interpolation, spurious results can occur in areas of sharp pollutant gradients among existing data points.

Figure 8 is an example of the Contour option. In the figure, the geographical boundaries of the United States are featured by including Mexico

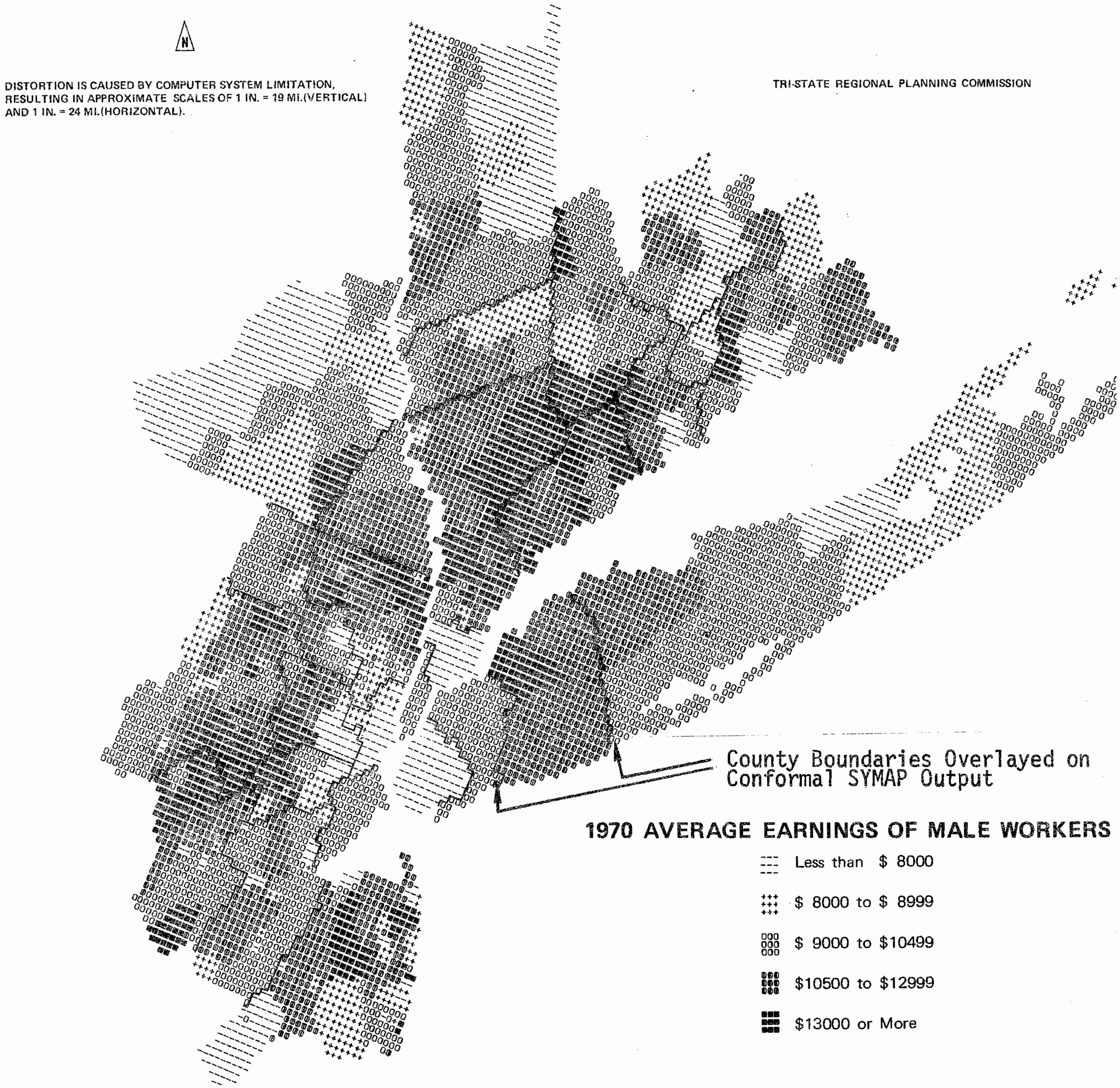


Figure 7. Example of Conformal Option, Showing Spatial Distribution of 1970 Average Earnings of Male Workers in the New York-New Jersey-Connecticut Tri-State Region [12].

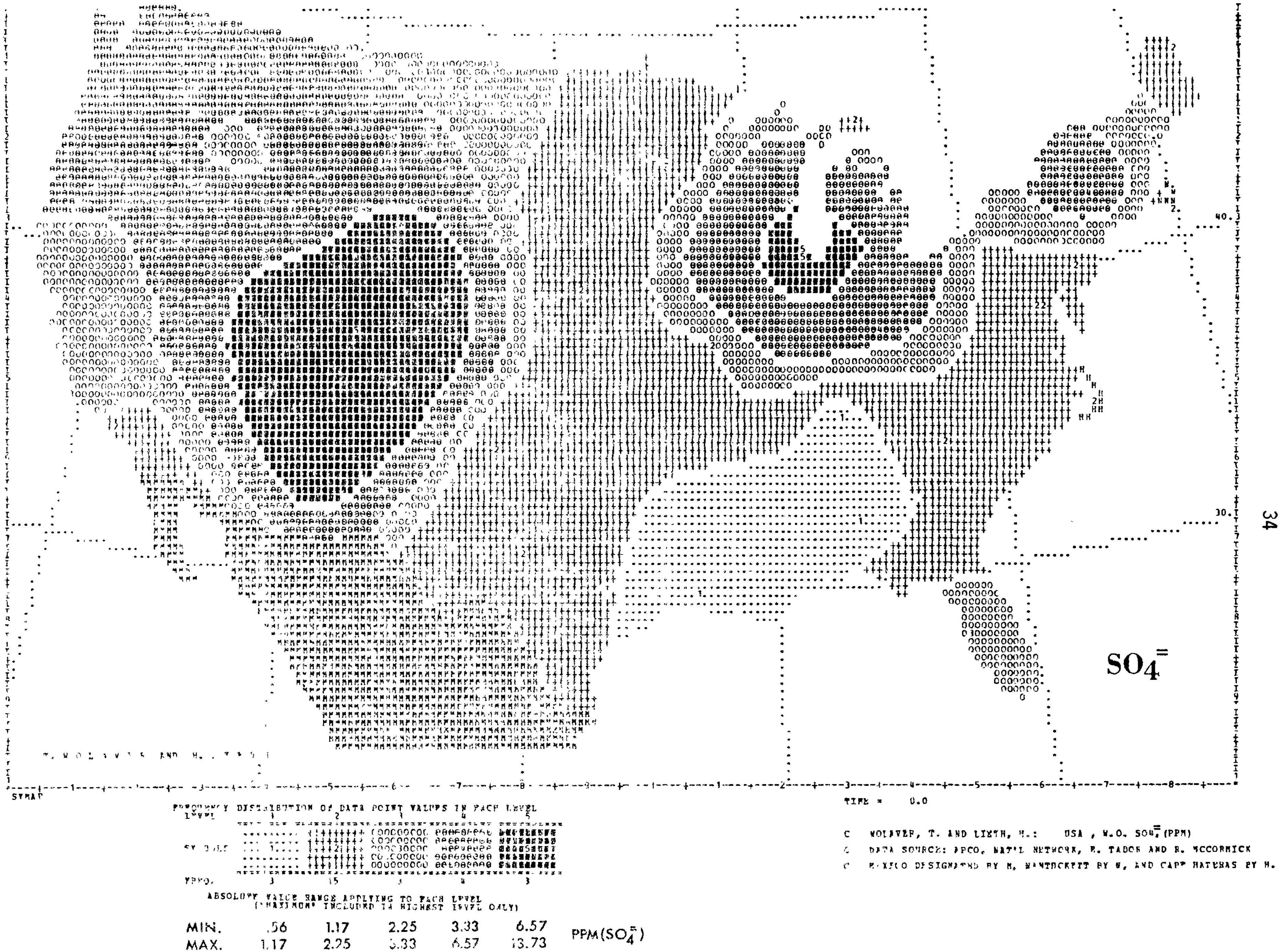


Figure 8. Example of Contour Option, Showing Spatial Variation of SO_4 = Washout Concentration Over the Contiguous United States [13]

(designated by M), the Nantucket Islands (designated by N), and Cape Hatteras (designated by H). The spatial variation of sulfate washout concentration over the United States is illustrated with the use of five different characters, each corresponding to a certain range of sulfate concentration.

In the Proximal option, the specifications of map boundaries, data points, and grid-point interval are made in the same manner as in the Contour option. When data values are provided at N data points, the program automatically divides the study region into N subregions, each of which surrounds one of the data points. The boundaries of subregions are given by an imaginary line equidistant between each pair of adjacent data points. The SYMAP program simply compares each data value with specified ranges and prints the appropriate character throughout the corresponding subregion.

In the Contour and Proximal options, one does not necessarily have to specify the actual boundary of a study area. Instead, one can overlay a transparent geographic map of the study area on a SYMAP isopleth map that is drawn over a rectangular area. By using a simple copying technique, one can obtain as nice an isopleth map as that obtained when the boundaries of a study area were meticulously defined numerically.

SYMAP employs two electives which can improve the spatial analysis. The first elective is a search radius applied to each monitor to determine its use in establishing interpolated values. This is analogous to the representative area discussed in Section 2.1. The second elective is barriers which can be used to simulate topographical features like mountains or known discontinuities in emission patterns.

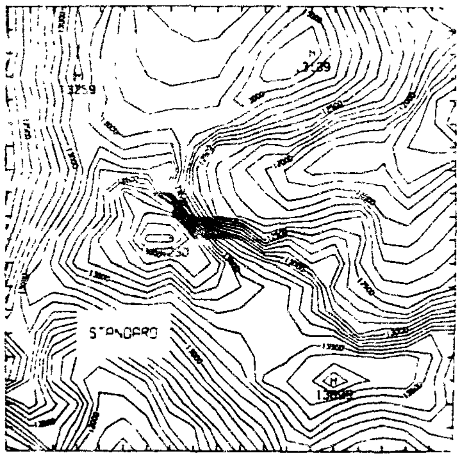
2.4.2 Line-Drawn Maps

Line-drawn isopleth maps can be generated by a number of plotting devices. These devices can be separated into two basic categories, Cathode Ray Tubes (CRTs) and Pen Plotters. CRTs are usually equipped with a photographic device so that the desired picture can also be printed on film or photo-sensitive paper.

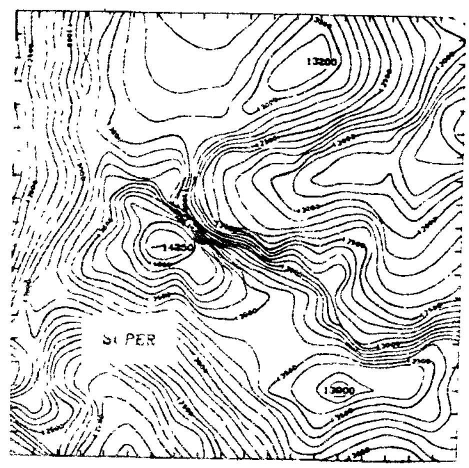
Whereas a character-printed map is obtained by using a simple procedural language described in the user's manual of SYMAP, a line-drawn map usually requires special programming. To generate a character-printed map, a computer simply calculates a variable value at every grid point, using a certain interpolation formula (e.g., linear interpolation), and prints a particular symbol according to the range in which the computed value falls. In contrast, to obtain a line-drawn map, the computer must command the plotting device to draw a contour line connecting the points at which the computed values are equal to the isopleth value.

Data points can be connected by a number of methods. From among the many methods, we recommend using either a piecewise linear fit, which is used in the computer program called TRICON listed in the User's Manual, or a splines-under-tension method. computer program for which is obtainable from the University of Colorado [17]. Under no circumstance should polynomial fit be used for drawing isopleth lines. Polynomial fit exhibits very poor performance with undesirable oscillations.

Figure 9 shows two weather maps produced by a CRT. In this figure, the first contour map was drawn using a piecewise linear fit; the second



a. Piecewise linear fit



b. Splines under tension

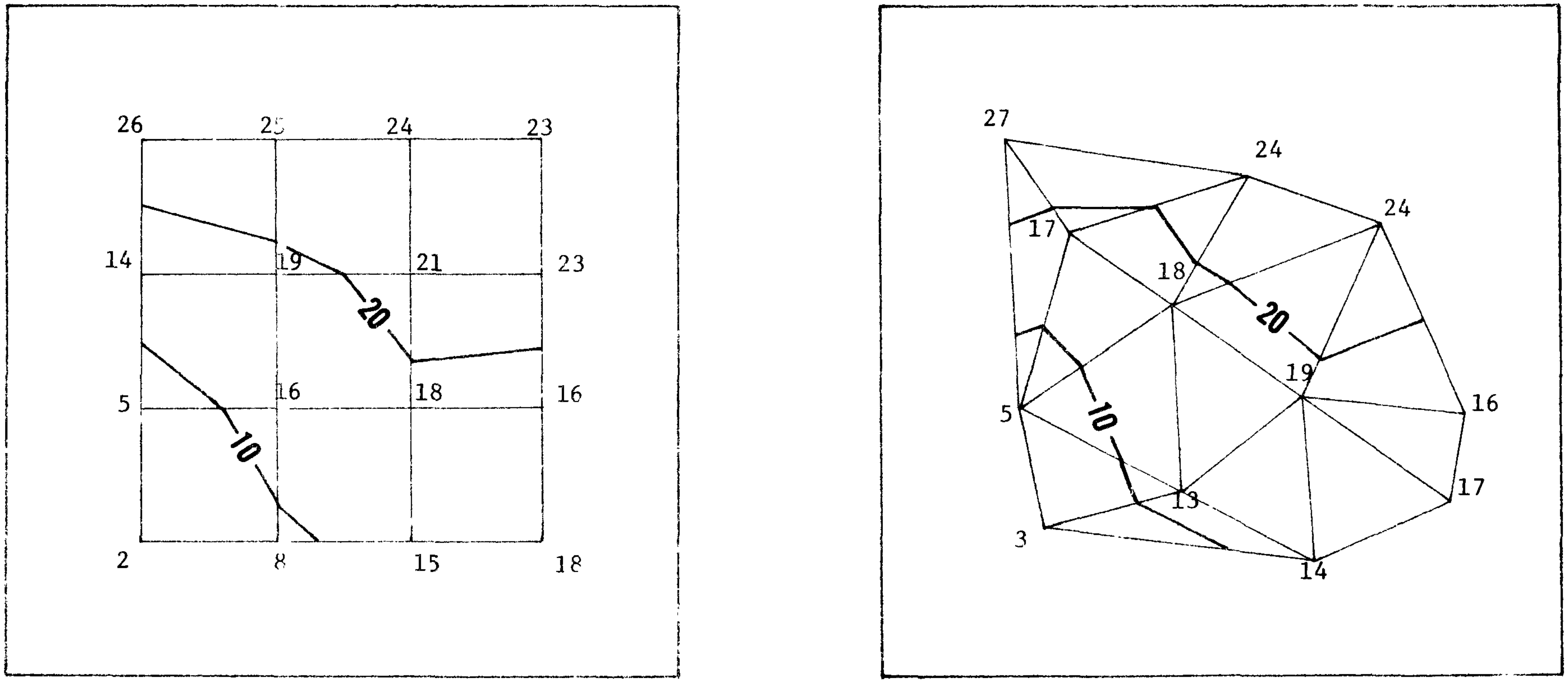
Figure 9. Contour Maps of Barometric Pressure Drawn by a Piecewise Linear Fit and by Splines under Tension [18].

contour map was drawn using splines under tension [18]. The former map was probably drawn from regularly spaced data points of barometric pressure by either writing a small computer program or simply using a procedural language of some preprogrammed package for computer graphics. Although the latter contour map has more aesthetic appeal than the former, it requires a large program and considerably longer computing time.

In the above examples, the data values were given at regularly spaced data points. The mechanism for determining an isopleth line for regularly spaced data points is illustrated in Figure 10a. For every edge whose two vertices have a data value greater than or equal to the isopleth level at one end and less than or equal to the isopleth level at the other, the isopleth point is determined by linearly interpolating the data values at the two vertices to that isopleth level. The isopleth line is then drawn by connecting all those isopleth points.

Air monitoring data (and population exposure variables to be discussed later) are available only at scattered data points, i.e., monitoring station sites and/or receptor points. A method of obtaining an isopleth map from such randomly spaced data points involves triangulation of data points as shown in Figure 10b. The triangulation can be performed in many ways. However, it is recommended that the outlying data points be connected first by straight lines to form a polygon, and the triangulation proceed inward. In forming a polygon by connecting the outlying data points, it is, to a certain extent, possible to make the polygonal area resemble the study region.

Once the triangulation is completed for randomly spaced data points, the steps for drawing an isopleth map are similar to those used for regularly



a. Regularly spaced data points

b. Randomly spaced data points

Figure 10. Concentration Isopleths Resulting from Regularly and Randomly Spaced Data Points.

spaced data points. An isopleth point is determined on each edge whose two vertices have a data value greater than or equal to the isopleth level at one end and less than or equal to the isopleth level at the other end. Then, the isopleth line is drawn by connecting all those isopleth points (Figure 10b). A computer program doing the above tasks has been developed by Martin Cohen at Technology Service Corporation. Listings of the computer program called TRICON are given in the User's Manual.

Figure 11 shows an example of a computer-drawn isopleth map for randomly spaced data points. In the figure, the boundary of the study area is drawn separately from the isopleth lines. The map boundary was drawn for the input data produced by a digitizer whose pencil-like probe traced the boundary of the original geographic map, starting from the bottom (note the small gap there). When such a convenient digitizing device is not available, the boundary can be numerically specified by reading off the X-Y coordinates of major vertices on the boundary line from an overlayed graph paper. A simpler approach, as mentioned before, is that of overlaying a transparent map of the study area on the computer-drawn isopleth lines.

The isopleth lines obtained from the randomly spaced data points cover the study area quite well. When data points sufficiently cover a study area, as seen in Figure 11, data values need not be computed for every regularly spaced data point; an isopleth map obtained from triangulation of randomly spaced data points is sufficient for most cases.

2.4.3 Hybrid Isopleth Map

Even when one cannot trust a computer-drawn isopleth map or computer software to draw an isopleth map is not readily available, a computer plotting device still can be a useful tool to draw an isopleth map. For

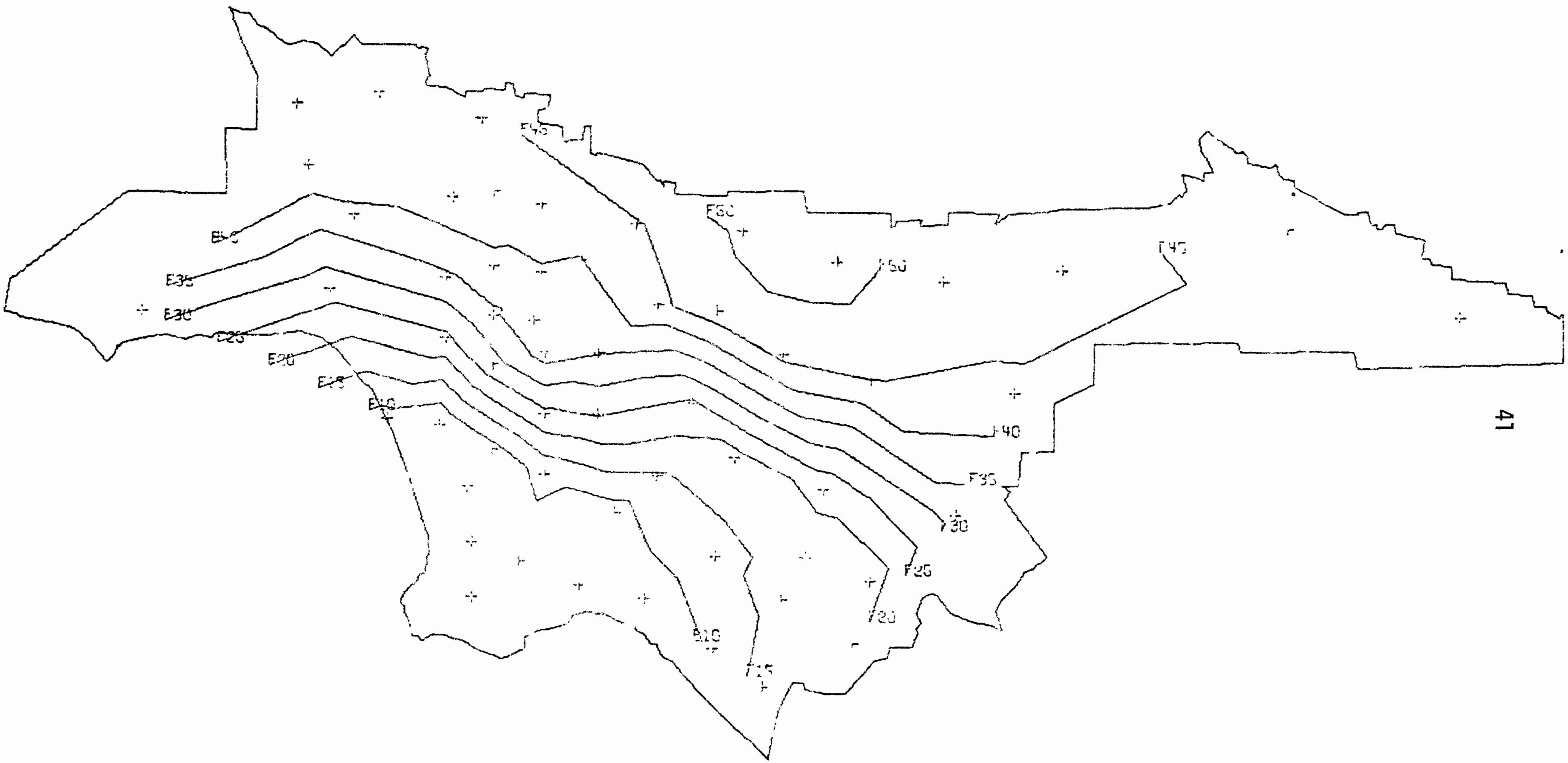


Figure 11. Computer-Drawn Isopleth Map, Showing Oxidant Air Quality over the Los Angeles Region in Percentage of Days on which the NAAQS was Exceeded During 1973/74 [3].

instance, a pen plotter will draw a specific symbol at an exact data point (monitoring station and/or receptor grid point) according to a range of data values. Then, one can easily draw an isopleth line by following the same symbols and procedures described in Section 2.2.

Figure 12 shows an example of a hybrid isopleth map. Three data sets are required:

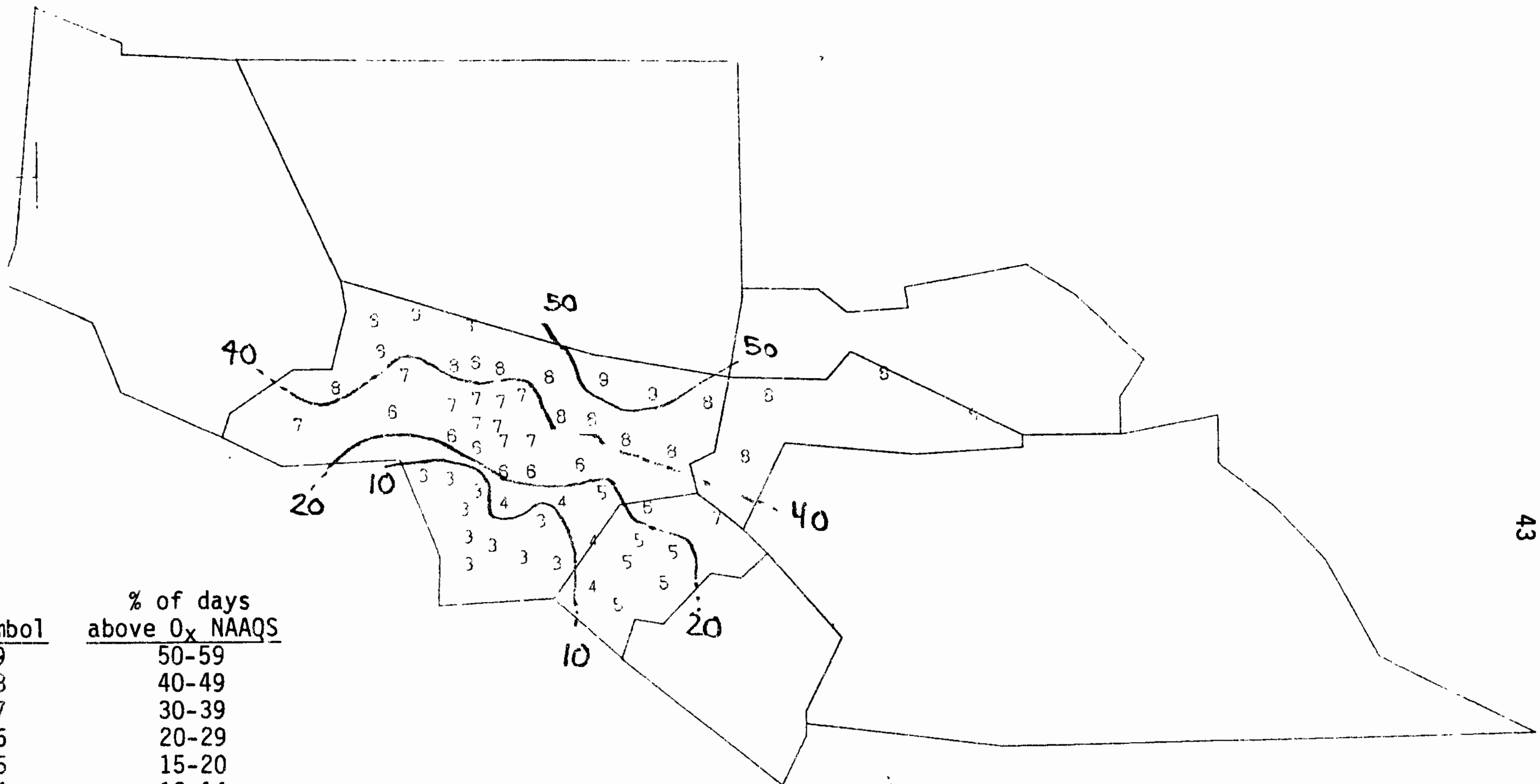
(1) A set of annual air quality statistics (here, percentage of days the O_x NAAQS was exceeded during 1973/74 period),

(2) X-Y coordinates of non-uniformaly spaced data points (here, receptor grid points at which the air quality was estimated from air monitoring data), and

(3) A set of X-Y coordinates describing polygonal areas that represent the study region.

The computer sorts each data value into an appropriate range of data values and commands the pen plotter to draw the corresponding symbol at the data point. Note that Figure 12 is drawn for the same data used to produce Figure 11.

With this hybrid approach, one can save some time by using a computer and at the same time can manually draw an isopleth map that the person perceives to be the most likely. The listing of the computer program to produce the hybrid map is given in the User's Manual.



Symbol	% of days above O_x NAAQS
9	50-59
8	40-49
7	30-39
6	20-29
5	15-20
4	10-14
3	5-9

Figure 12. Hybrid Isopleth Map, Showing Oxidant Air Quality over the Los Angeles Region in Percentage of Days on Which the NAAQS was Exceeded During 1973/74

CHAPTER 3. POPULATION EXPOSURE ANALYSIS

One can extend a spatial analysis of air quality information a step further by combining air quality data and demographic data. Ambient air quality can be quantified by the percentage of the population living in areas above an air quality standard instead of concentration units.

A population exposure analysis can report ambient air quality by a statement such as, "The percentage of the urban population living in areas above the primary national ambient air quality standard for particulates decreased from 60% in 1970 to 10% in 1975." This statement may be more informative to the public than telling them that the annual-mean-particulate concentrations decreased from 69.5 $\mu g/m^3$ to 64.3 $\mu g/m^3$ in the same period.

The following three different methods have been used to quantify population exposures from air monitoring information:

1. Associating populations to a monitoring site [5,6,7,14].
2. Transcribing air pollution isopleth patterns on a map of census tracts [8,9].
3. Interfacing air quality and population data at strategically located receptor grid points [2,3,14].

This guideline document describes the third method that provides the most objective estimates of population exposure and lends itself to computer processing.

The proposed population exposure analysis requires several steps. The first step is a search for a proper regional map and population data in a convenient form. The second step is development of a receptor network of artificial grid points that are used to approximate the spatially distributed population. The study area boundary is determined by the methods of Chapter 2. The location of the receptor points and the monitoring stations must be specified in a digital form, based on the coordinate system of the regional map. The third step involves the preparation of computer-ready data sets and computation of population exposure variables by computer software. The fourth step included analysis of the computed population exposure variables and presentation of the results.

The details of each step are discussed in the sections which follow.

3.1 DEVELOPMENT OF RECEPTOR NETWORK

Instead of employing the assumption that the population exposure level in various parts of a region is represented by the concentration reading at a single air monitoring station, we propose the use of a receptor network of artificial grid points. The development of this network is based on the spatial distribution of the population and is used to interface the population data and the air quality data at each receptor point. A receptor point is used to represent the local population in the areas where they reside. The location and density of receptor points are selected so that the distribution of receptor points reflects the spatial distribution of population density and adequately covers the land area throughout the study region. Air quality is estimated at each receptor point by spatial interpolation of air quality observed at neighboring monitoring sites.

To develop a receptor network, one has to have population data and a regional map. Population data are always obtainable from census population summaries. However, because it takes time and effort to reduce these data into a form that is practical for population exposure analysis, it may be worthwhile to contact a regional planning agency. The regional planning agency often has not only conveniently aggregated population data into properly aggregated statistical areas, but has also prepared a regional map showing boundaries of the statistical areas, cities, and counties.

When such a convenient source of information is not available, census population data must be used. Although every census summary contains maps showing tracts and statistical areas, a regional map covering the entire study area, say for an AQCR, may not be found in the census summary. Therefore, it is necessary either to look for other map sources (such as street and geographical maps) or to create a new regional map by combining several maps found in the census summaries.

The census population summarires range from detailed demographic data for each census tract to aggregated population data for each state and each standard metropolitan statistical area (SMSA). For a single city or small SMSA, the census tracts provide a basis for establishing a receptor network. The center of each census tract will serve as a receptor point. For a larger study area covering many cities and/or counties, the number of census tracts included in the study area may become too large. For the larger areas, the most convenient census statistics may be the population summaries for subcounties [19]. These summaries provide the population size of statistical areas that often correspond to administrative jurisdictions such as cities, towns, villages, or boroughs.

When a planning agency's statistical areas, census subcountry or census tract population statistics are found to be suitable for the analysis, receptor points are assigned to each "statistical" area according to the size of the population and the land area. Assuming that the population is uniformly distributed within each statistical area, the location of receptors (in most cases just one receptor) is determined such that the receptor points represent properly the spatial location and boundary of that statistical area. Empirically, we have found that the following method yields the proper number of receptor points to each statistical area:

1. Regardless of the size of the population and the land area, each statistical area is represented by at least one receptor point.
2. An additional receptor point is assigned for each 50 mi^2 increment land area or each 200,000-person increment of resident population.
3. Take the number of receptor points representing the larger increment (land or population).

A question arises here as to why the monitoring stations themselves are not included as receptor points in computing population exposure variables. The following two paragraphs provide some explanation.

Because the spatial interpolation scheme used in the population exposure methodology smoothes out the observed air qualities at the three nearest neighboring stations to each receptor point, the highest and lowest observed air qualities among the monitoring stations sometimes do not appear on the concentrations at the receptor points. Therefore, a population exposure analysis may indicate that the study population is exposed to concentrations either slightly above the lowest observed concentration or slightly below the highest observed concentration.

The problem of missing the highest and lowest concentrations, however, is already inherent in the limited number of monitoring sites. Although the spatial interpolation scheme used may further smooth out those highest and lowest concentrations, the conservative nature of population exposure methods should be a merit rather than a demerit. The population exposure analysis is less affected by the extreme values observed than is the air quality analysis that is applied directly to the air monitoring data. The smoothing effect of a spatial interpolation of air monitoring data is not necessarily undesirable for estimating population exposure because it is partially equivalent to people moving around and smoothing out their exposure levels.

When the receptor network is developed, both the receptor points and the monitoring stations have to be located on the same digitized regional map, e.g., a map on graph paper. To correctly locate a monitoring station on the map is not as easy as one might think; the regional map used may not contain any landmarks helpful in locating the station at the right place. The larger the study region, the more difficult the placement of the stations on the regional map. For a large, unmarked regional map, one's prior knowledge of the station sites does not work effectively. A mathematical method of correctly locating monitoring stations on a skeleton map, discussed in Appendix A should be used for such a case.

3.2 PREPARATION OF DATA SETS

The population exposure analysis requires careful preparation of several data sets. These include data on the resident population, receptor points, air quality, and monitoring stations. The data sets of air quality and monitoring stations similar to those required for performing a spatial analysis of air quality. Air quality data used for a population exposure analysis are commonly available statistical summaries: annual mean concentrations for analyzing long-term average population exposure, and percentage of the time above the air quality standard or percentile concentrations for analyzing short-term population exposure.

Population and receptor data sets to be prepared for performing a population exposure analysis are described in the following two subsections. These data sets, together with those of air quality and monitoring stations, are used as input data to one of the two computer software systems (LPEM and SPEM) described in the User's Manual.

3.2.1 Population Data

Sources of Population Data

National census data are the primary source for almost all demographic data. Detailed demographic data are available in both magnetic data tapes and published reports for some 35,000 census tracts in 241 Standard Metropolitan Statistical Areas (SMSAs) and the remaining unincorporated census statistical areas. However, it is rather difficult to transform these census data into a form useful for a population exposure analysis. Unless an in-house capability of handling the census data tapes exists, a search for pre-existing sources of conveniently aggregated population data is recommended.

Census statistical summaries issued from the Bureau of Census and local planning agencies are the two major sources of such aggregated population data. Among the census statistical summaries, the following two series provide useful population data for performing a population exposure analysis: PC(1)-A series [19], which summarizes the number of inhabitants for political jurisdictions (boroughs, towns, cities, counties, etc.), and PHC(1) series [20], which provides data on population characteristics (e.g., sub-populations by age group, nativity, place of work, worker and nonworker status) for each census tract. These two series also contain maps showing boundaries of political jurisdictions and/or census tracts. Some of these maps may be useful for a population exposure study.

Local planning agencies that may have useful population data include state highway departments, regional planning agencies, state housing administrations, and civil defense agencies. These agencies may have not only conveniently aggregated population data for regional statistical areas but also a high quality map showing boundaries of counties and regional statistical areas.

Sub-Population Data

Population exposure methodology discussed in this text assumes that people are locationally fixed at their residence location. This simplifying assumption may not be valid for working-age population, but may be good for sub-populations such as school-age and elderly who tend to stay near their residence location most of the time. Furthermore, these sub-populations are believed to be the most susceptible to air pollution. Therefore, population exposure analysis should be performed not only for total population but also for sub-populations including school-age children and elderly people.

Because all statistics of the sub-populations mentioned above are given for place of residence, the data for sub-populations should be converted to a percentage of the total resident population in each aggregated statistical area. Then, we can perform a population exposure analysis for various sub-populations by using the same receptor network developed for the total resident population.

For working-age population, its large diurnal mobility may invalidate the stationarity assumption employed in the population exposure methodology. However, a way exists to incorporate population mobility into the current population exposure methodology. It can be assumed that population exposure during working time occurs at work place, while population exposure during non-working time,occurs at residence location[3]. To compute exposure of the workers during working time requires employment statistics that provide the number of workers at their place of employment. This type of sub-population data should be given by the actual size of the sub-population in each aggregated statistical area, instead of the percentage of the total population.

Population Data for Off-Survey Years

When the analysis is made for many years, the population statistics should be computed for each year by interpolating known population statistics or population projections for other years to that year. For example, the size of resident population in New Y ork County (Manhattan borough) in 1965 is estimated as 1,618,757 by linearly interpolating the 1960 population, 1,698,281, and the 1970 population, 1,539,233, to the year 1965. The 1973 population may be approximated by the 1970 census population or may be estimated by interpolating the 1970 population and the population projected to 1980.

When the study period is only for a few years or when the spatial distribution of population over the study period remains almost the same, the population data of a single survey year can be used for population exposure analyses over the entire study period.

3.2.2 Receptor Point Data

In a population exposure analysis, a receptor point is used for several purposes. A receptor point represents both the size, the spatial location, and the spatial spread of the local population. The spatial location of the local populace is given by the x-y coordinates of the receptor point in the digitized regional map. Specifications for the formats are given in the User's Manual.

Air quality data measured at air monitoring stations are spatially interpolated to each receptor point and are merged with the population data at the receptor points to compute various population exposure variables. Thus, the receptor points can also be used to develop an isopleth map of air quality and population exposure variables from the values computed at each receptor point. For example, the isopleth map of long-term exposure to annual mean concentrations over the study region is developed from the values computed at each of the receptor points. Guidelines for producing isopleth maps of air quality and population exposure variables are given in Chapter 2.

3.3 CHARACTERIZATION OF POPULATION EXPOSURE TO AIR POLLUTION

Pollutant concentrations over an urban area are constantly changing in time and space. Exposure of people to these concentrations can be analyzed with respect to both long-term average exposure and statistical distribution of short-term exposures over a long period (usually a year). To make the analysis tractable, the resident population (the type of data available from the census summaries) is used, and individual members of the population are assumed to be locationally fixed at their residence location.

3.3.1 Analysis of Long-Term Population Exposure

An annual mean concentration is used to designate the long-term average pollution level to which people are exposed. The long-term average pollution level at each receptor point is estimated by spatially interpolating the annual mean concentrations observed at the three nearest neighboring stations to that receptor point by using Eq. (2). For a given receptor point, (x_i, y_i), the computer program called Long-Term Population Exposure Model (LPEM) automatically searches among all the stations used for the analysis for the three nearest stations, (x_j, y_j), $j = 1,2,3$, to that receptor point (see User's Manual).

Once the annual mean concentration, C_i, is determined for each receptor point, population exposure can be calculated by combining the air quality and population statistics. The LPEM computer program compares C_i with a given concentration threshold, C^*; enters the local population, P_i; and determines population exposure for the entire region, R. Mathematically, this can be expressed by

$$P(C^*) = \sum_{i\varepsilon R} P_i \; U(C_i - C^*) / \sum_{i\varepsilon R} P_i \quad , \tag{3}$$

where an indicator step function, $U(C_i - C^*)$, takes the value one for $C_i \geq C^*$ and zero for $C_i < C^*$. By changing the threshold level, C^*, from zero to a sufficiently large value, the LPEM program produces a cumulative distribution of the population exposed to various annual mean concentrations.

The annual mean concentrations obtained at both the receptor points and the station sites are used to draw an isopleth map of annual mean concentrations over the entire study region. Inclusion of the highest and lowest concentrations observed at the monitoring stations will make the resulting isopleth map more complete than that which would be produced from the receptor concentrations only. Such an isopleth map can be drawn either by hand or by the spatial computer program discussed in Section 2.4.

The long-term average exposure of the population may be expressed either by a pair of distribution curves as shown in Fig. 13 or by a bar graph as shown in Fig. 14. The pair of distribution curves is excellent for illustrating how much the population exposure to a given pollutant decreased (or increased) from one year to another and how the decrease (or increase) in population exposure is distributed over pollution levels. For instance, Fig. 13 shows that the percentage of the population exposed to total suspended particulate above the national primary standard (75 $\mu g/m^3$) dropped from 58% in 1971 to 17% in 1974.

The bar graph shown in Figure 14 is convenient for illustrating year-to-year variations in population exposure over a long period. The increase in population exposure during the middle years, 1967-72, is evidenced by the

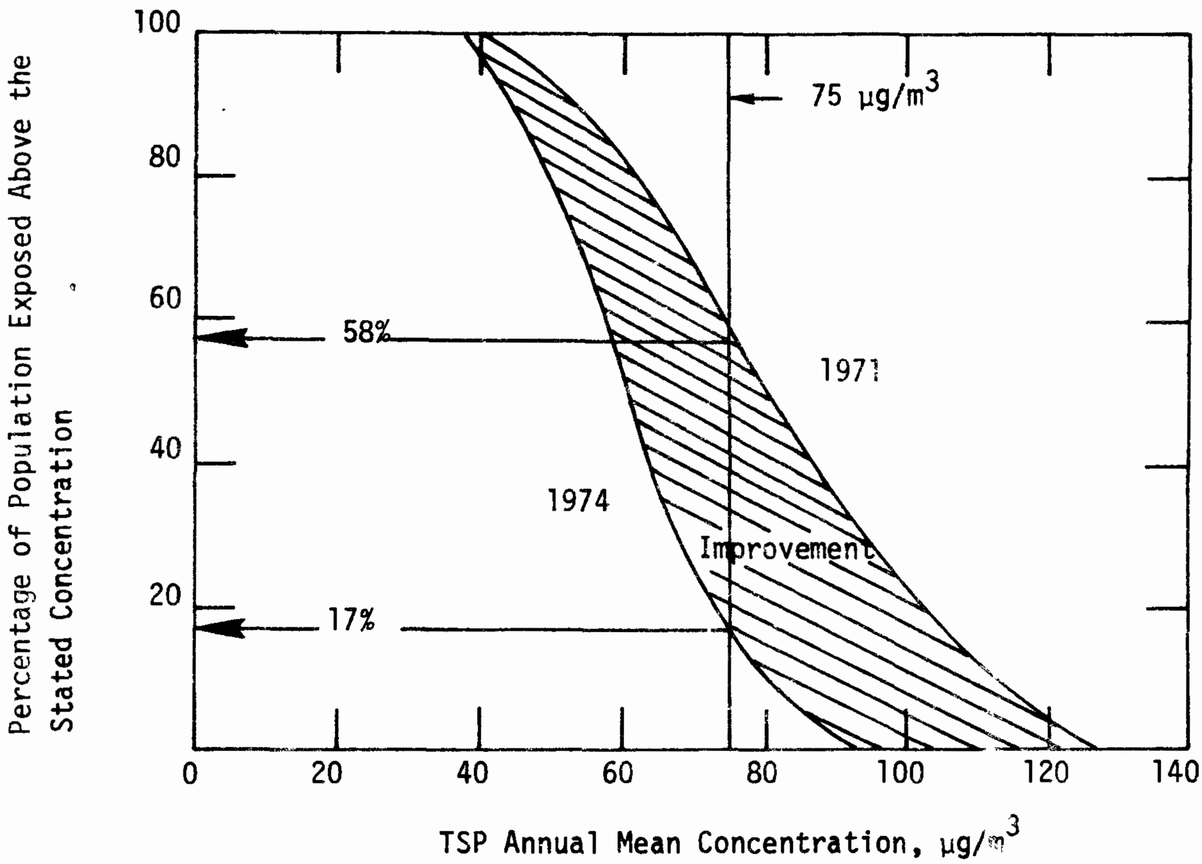


Figure 13. Decrease in Population Exposed to Total Suspended Particulate in New York-New Jersey-Connecticut Tri-State Region from 1971 to 1974 [4].

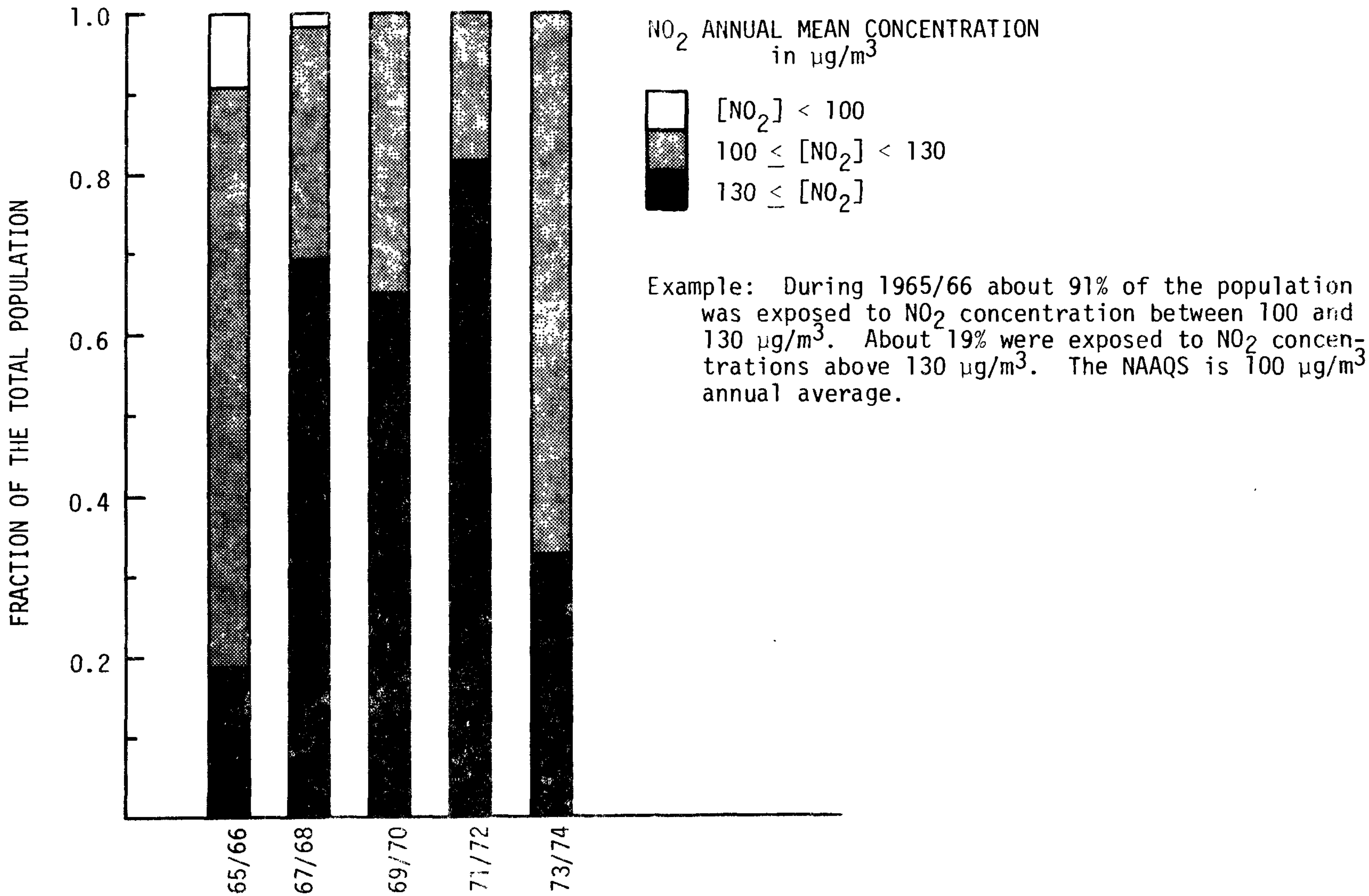


Figure 14. Changes in the Total Population Exposure to NO_2 During Five 2-Year Periods in the Los Angeles Region[3].

longer black bars indicating the percentage of the population exposed to annual mean NO_2 above 130 $\mu g/m^3$. The decrease in population exposure in 1973/74 is noted by the shorter black bar and the longer gray bar, which indicates the percentage of the population exposed to NO_2 between 100 and 130 $\mu g/m^3$.

The LPEM computer program also calculates three regional indices for characterizing long-term air quality and population exposure: station average concentration, space average concentration, and population average concentration. The station average concentration is given by an arithmetic mean of annual mean concentrations observed at individual stations. The space average concentration is given by an area weighted average of the concentrations at the receptor points. The population average concentration is given by a population weighted average of the receptor concentrations and indicates the air quality level most representative of the population. Mathematically, the space average concentration, AQ_s, and the population average concentration, AQ_p, are expressed respectively, by

$$AQ_s = \sum_{i \in R} S_i C_i / \sum_{i \in R} S_i \tag{4}$$

$$AQ_p = \sum_{i \in R} P_i C_i / \sum_{i \in R} P_i \tag{5}$$

where C_i is the annual mean concentration at the i^{th} receptor point, and S_i and P_i are, respectively, the land area and the local population size represented by the i^{th} receptor point.

3.3.2 Analysis of Short-Term Population Exposure

Analysis of population exposure to short-term peak concentrations is more difficult than the analysis for long-term average concentrations. The reason is that long-term average exposure involves only one pollution variable, i.e., concentration level, but short-term exposure involves at least two pollution variables, i.e., concentration level and frequency of occurrence or duration of that concentration level.

It is extremely costly to compute each instantaneous exposure level individually over a long time period. Instead of using each instantaneous concentration value, a percentage of the time that the short-term (1-hr or 24-hr) NAAQS was exceeded is used to characterize short-term exposures of the population over a long period. The quantity, "percentage of the time above the standard," is termed a "risk frequency" because it indicates how frequently people are exposed to a level of air pollution above the standard.

The risk frequency can be computed either from the number of standard violations or from the percentile concentrations. When a number of standard violations are available, the risk frequency at each monitoring station is used as input data to the LPEM computer program. Then, the LPEM spatially interpolates the risk frequencies observed at the monitoring stations to those at receptor points as if the risk frequencies were the annual mean concentrations (see User's Manual).

When percentile concentrations are available, the percentile concentrations at eight selected percentiles are used as input data to the computer program called Short Term Population Exposure Model (SPEM). First, the SPEM spatially interpolates the percentile concentrations at the monitoring stations to those at receptor points. Then, the SPEM compares each of the eight interpolated percentile concentrations with the air quality standard to determine the risk frequency at each receptor point (see User's Manual).

The risk frequency computed at each receptor point is used to draw an isopleth map of risk frequencies over the study region. The risk frequency map is an excellent aid in visualizing the spatial variation of short-term air quality levels.

When the risk frequencies are determined for all the receptor points, short-term exposure levels can be calculated for the entire study region. Both computer programs (LPEM and SPEM) compare the risk frequency, F_i, at each receptor point with a given frequency threshold, F^*; enter the local population, P_i; and scan the entire study region, R. Mathematically, this process can be expressed by

$$P(F^*) = \sum_{i \in R} P_i U(F_i - F^*) / \sum_{i \in R} P_i \quad , \tag{6}$$

where an indicator step function $U(F_i - F^*)$ takes the value one for $F_i \geq F^*$ and zero for $F_i < F^*$. By changing the threshold level, F^*, from 0% to 100%, both programs produce a cumulative distribution of the population exposed to air pollution above the standard at various frequencies.

The short-term exposures of the population over a long period may be expressed either by a pair of distribution curves as shown in Fig. 15 or by a bar graph as shown in Fig. 16. The pair of distribution curves illustrates how much the population exposure decreased (or increased) in a given time period, how the decrease (or increase) in population exposure was distributed over the population, and how the frequency of dangerous exposures changed. For instance, Fig. 15 shows the percentage of the population exposed to total suspended particulate above the national secondary 24-hr standard (150 $\mu g/m^3$) at least 10% of the days dropped from 33% in 1971 to 4% in 1974.

The bar graph shown in Fig. 16 is convenient for illustrating year-to-year variations in population exposure over a long period. The increase in population exposure during the middle years, 1967-72, is represented by the shorter white bars, indicating the percentage of the population exposed to hourly NO_2 above the California standard less than 6% of the days. The decrease in population exposure in 1973/74 is noted by the longer white bar and the shorter gray bar, indicating those who were exposed above the California standard between 6% and 12% of the days.

Both computer programs compute two regional indices for characterizing short-term air quality and population exposure: the percentage of time (hours or days) exceeding the standard, area weighted average over the study region; and the percentage of time exceeding the standard, population weighted average over the region. The latter quantity is the average percentage of time the standard is exceeded among all the people in the study region. Mathematically, the two regional indices, F_S and F_p, are expressed, respectively, by

$$F_S = \sum_{i \varepsilon R} S_i F_i / \sum_{i \varepsilon R} S_i \tag{7}$$

$$F_p = \sum_{i \varepsilon R} P_i F_i / \sum_{i \varepsilon R} P_i . \tag{8}$$

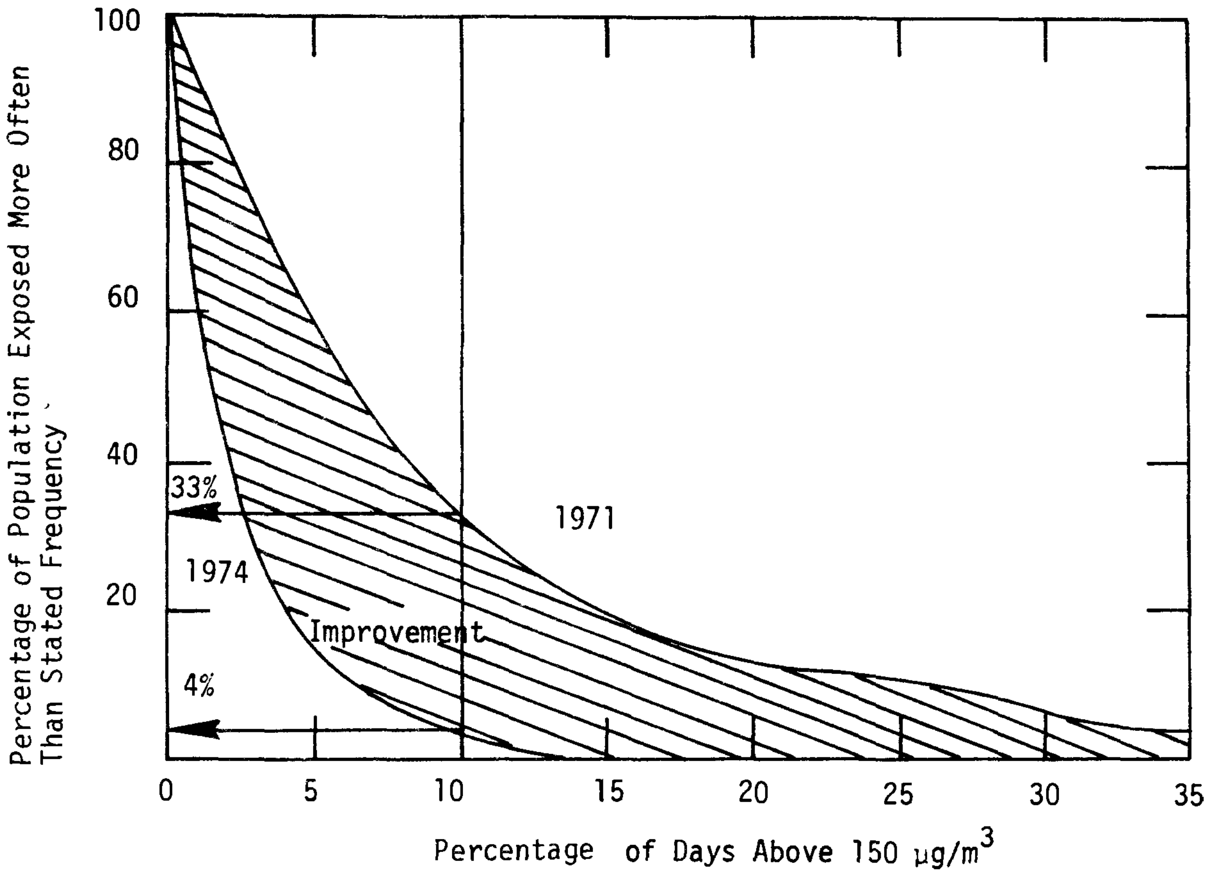


Figure 15. Population Exposed Daily to Total Suspended Particulate above 150 $\mu g/m^3$ in New York-New Jersey-Connecticut Tri-State Region[4].

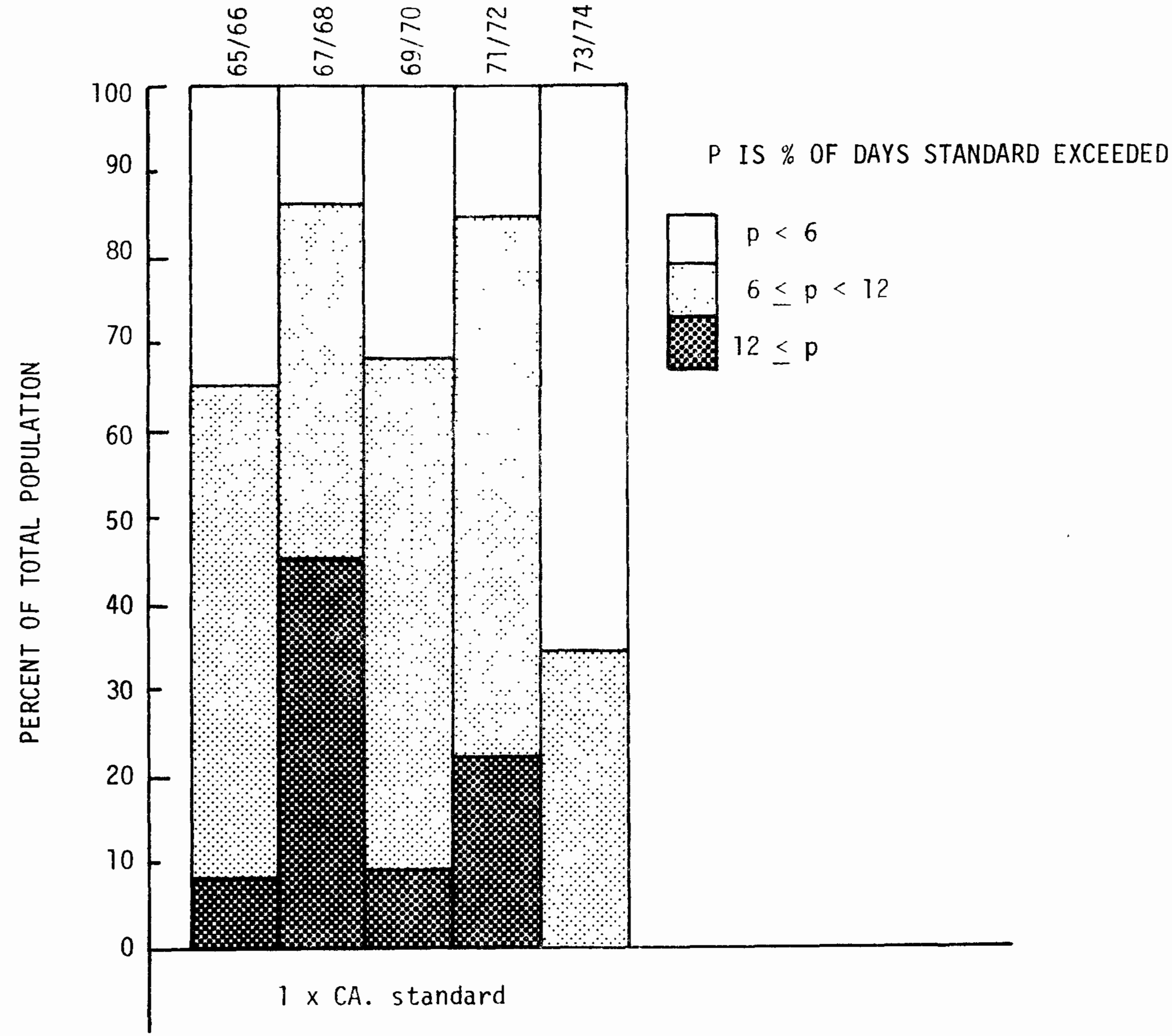


Figure 16. Changes in Population Exposure to NO_2 During Five 2-Year Periods in the Los Angeles Region [3].

REFERENCES

1. Harvard Laboratory for Computer Graphics, "User's Reference Manual for Synagraphic Computer Mapping 'SYMAP' Version V," Harvard University, Cambridge, MA, 1968.

2. Horie, Y., and A. C. Stern, "Analysis of Population Exposure to Air Pollution in New York-New Jersey-Connecticut Tri-State Region," U.S. EPA/OAQPS Publication No. EPA-450/3-76-027, University of North Carolina, Chapel Hill, NC, March 1976.

3. Horie,Y., et al., "Population Exposure to Oxidants and Nitrogen Dioxide in Los Angeles: Volumes I, II, and III," U.S. EPA/OAQPS Publication No. EPA-450/3-77-004a, b, and c, Technology Service Corporation, Santa Monica, CA, January 1977.

4. U.S. Environmental Protection Agency, "National Air Quality and Emissions Trends Report, 1975," U.S. EPA/OAQPS Publication No. EPA-450/1-76-002, Monitoring and Data Analysis Division, Office of Air Quality Planning and Standards, Research Triangle Park, NC, November 1976.

5. Zupan, J.M., The Distribution of Air Quality in the New York Region, Resources of the Future, Inc., Washington, DC, 1973.

6. McHarg, I. L., Design with Nature, Published for the American Museum of Natural History, Doubleday/Natural History Press, Garden City, NY, 1971.

7. Anderson, J. A. et al., "Correlation between Air Pollution and Socio Economic Factors in Los Angeles County," submitted to Urban Ecology, Department of Chemistry and Energy Center, University of California, San Diego, CA, January 1977.

8. Brian, J.L., Berry, et al. The Social Burdens of Environmental Pollution: A Comparative Metropolitan Data Source, Department of Geography, University of Chicago, 1976.

9. Istvan,Takacs and G. Bradford Shea. Estimations of Human Population-at-Risk to Existing Levels of Air Quality, EPA Contract No. 68-01-2820, Enviro Control, Inc., Rockville, Maryland, February 1975.

10. TRW Systems Group,"Air Quality Display Model," prepared for Department of Health, Education and Welfare, National Air Pollution Control Administration, Washington, DC, Contract No. Ph-22-68-60, and Available from NTIS, Springfield, VA, 22151 as PB-189-194, 1969.

11. Busse, A.D., and J. R. Zimmerman, "User's Guide for the Climatological Dispersion Model," U.S. EPA Publication No. EPA-R4-73-024, National Environmental Research Center, U.S. EPA, Research Triangle Park, NC, December 1973.

12. Ludwig, F. L., W. B. Johnson, A. E. Moon, and R. L. Mancuso, "A Practical Multipurpose Diffusion Model for Carbon Monoxide," Stanford Research Institute, Menlo Park, CA, Contracts CAPA-3-68 and CPA 22-69-64, 1970.

13. Zimmerman, J. R., and R. S. Thompson, "User's Guide for HIWAY, A Highway Air Pollution Model," U.S. EPA Publication No. EPA-650/4-74-008, National Environmental Research Center, U.S. EPA, Research Triangle Park, NC, February 1975.

14. Frank, N. H., W. F. Hunt, Jr., and W. M. Cox, "Population Exposure: An Indicator of Air Quality Improvement," Paper #77-44.2 presented at the 70th Annual Meeting of the Air Pollution Control Association, Toronto, Canada, U.S. Environmental Protection Agency, Research Triangle Park, NC, June 1977.

15. Tri-State Regional Planning Commission, Regional Profile--Regional Employment 1970, Vol. II, No.6, New York, NY, December 1973.

16. Wolaver, T. G., "The Dsstribution of Natural and Anthropogenic Elements and Compounds in Precipitation Across the U.S.: Theory and Quantitative Models," Botany Department, University of North Carolina, Chapel Hill, NC, October 1972. Copies available through Division of Ecological Research, U.S. EPA/NERC, Research Triangle Park, NC.

17. Cline, A. K., "Curve Fitting Using Splines Under Tension," Atmospheric Technology, National Center for Atmospheric Research, No. 3, pp 60-65, September 1973.

18. Wright, T. J., "Utility Plotting Programs," Atmospheric Technology, National Center for Atmospheric Research, No. 3, pp 51-57, September 1973.

19. Bureau of the Census, "Number of Inhabitants," PC(1)-A Series, U.S. Department of Commerce, Washington, DC, May 1972.

20. Bureau of the Census, "Census Tracts," PHC(1) Series, U.S. Department of Commerce, Washington, DC, May 1972.

APPENDIX A. USE OF COORDINATE TRANSFORMATION FOR LOCATING MONITORING STATIONS ON A SKELTON MAP

When the study region is large, a mathematical method of locating monitoring stations on a skeleton map of the study region often works more effectively and more accurately than heuristic methods, such as those based on the street address of station site or on knowledge of the relative location of one station to another. The mathematical method utilizes the UTM coordinates given in the SAROAD format and applies a coordinate transformation on the UTM coordinates of an individual station.

When the mathematical method is used, the first step is to locate the two most familiar stations on the skeleton map. The location of other stations on the map is then determined by a coordinate transformation of their UTM coordinates into the coordinates used in the map. The map coordinates (x, y) of a station whose UTM coordinates are (p, q) are given by the following equations:

$$\left.\begin{aligned} x &= m(p - p_1) + x_1 \\ y &= m(q - q_1) + y_1 \end{aligned}\right\} \qquad \text{(A-1)}$$

and the slope, m, is given by

$$m = \frac{x_1 y_2 - x_2 y_1}{(p_1 - p_2) y_2 - (q_1 - q_2) x_2} , \qquad \text{(A-2)}$$

where (x_1, y_1) and (x_2, y_2) are the map coordinates of the two known stations and (p_1, q_1) and (p_2, q_2) are their UTM coordinates.

ENVIRONMENTAL PROTECTION AGENCY
Technical Publications Branch
Office of Administration
Research Triangle Park, North Carolina 27711

OFFICIAL BUSINESS

AN EQUAL OPPORTUNITY EMPLOYER

POSTAGE AND FEES PAID
ENVIRONMENTAL PROTECTION AGENCY
EPA - 335

SPECIAL FOURTH-CLASS RATE
BOOK

U.S. MAIL

Return this sheet if you do NOT wish to receive this material ☐ or if change of address is needed ☐. (Indicate change, including ZIP code.)

PUBLICATION NO. EPA-450/2-77-024a
(OAQPS NO. 1.2-083)