Technical Report Documentation Page

1. REPORT No.

FHWA/CA/TL-84/01

2. GOVERNMENT ACCESSION No.

3. RECIPIENT'S CATALOG No.

4. TITLE AND SUBTITLE

Measurement and Analysis of Ambient Carbon Monoxide Concentrations for Project-Level Air Quality Impact Studies

5. REPORT DATE

February 1984

6. PERFORMING ORGANIZATION

7. AUTHOR(S)

P.E. Benson

8. PERFORMING ORGANIZATION REPORT No.

57328-604195

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Office of Transportation Laboratory
California Department of Transportation
Sacramento, California 95819

10. WORK UNIT No.

11. CONTRACT OR GRANT No.

E80TL13

12. SPONSORING AGENCY NAME AND ADDRESS

California Department of Transportation
Sacramento, California 95807

13. TYPE OF REPORT & PERIOD COVERED

Final

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

Recommended procedures are given for designing project-level ambient carbon monoxide (CO) monitoring studies and analyzing their results. Attention is focused on transportation related applications, although the procedures are general and can be applied to other types of air quality studies. A statistical model which uses the scheduling and duration of sampling as a means to minimize bias in estimating the second annual maximum CO concentration is developed. This model is verified using a data set comprised of 112 station-years of data collected at California air monitoring stations. The temporal distributions of seasonal maximums for this data set are also described.

Two types of sampling plans are discussed: Background sampling and corridor sampling. Specific criteria for choosing between these two plans are given. Recommendations are made regarding site selection, outlier analysis, rollback calculations and data handling. A computer program which estimates second annual maximum 1-hour and 8-hour ambient CO concentrations using observed maximums stratified by time of day periods is introduced. A complete set of user instructions for this program is given.

17. KEYWORDS

Air quality, carbon monoxide, sampling program, statistical model, ambient air

18. No. OF PAGES:

156

19. DRI WEBSITE LINK

http://www.dot.ca.gov/hq/research/researchreports/1981-1988/84-01.pdf

20. FILE NAME

84-01.pdf

STATE OF CALIFORNIA
DEPARTMENT OF TRANSPORTATION
DIVISION OF ENGINEERING SERVICES
OFFICE OF TRANSPORTATION LABORATORY

MEASUREMENT AND ANALYSIS OF AMBIENT CARBON MONOXIDE CONCENTRATIONS FOR PROJECT-LEVEL AIR QUALITY IMPACT STUDIES

Study Supervised by .... Earl C. Shirley, P.E.
Roy W. Bushey, P.E.

Principal Investigator .... Paul E. Benson, P.E.

Report Prepared by .... Paul E. Benson, P.E.

Assisted by .... Jim Quittmeyer
Kim Henry

RAYMOND A. FORSYTH, P.E.
Chief, Office of Transportation Laboratory

84-01

10-48

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
FHWA/CA/TL-84/01		

4. TITLE AND SUBTITLE	5. REPORT DATE
MEASUREMENT AND ANALYSIS OF AMBIENT CARBON MONOXIDE CONCENTRATIONS FOR PROJECT-LEVEL AIR QUALITY IMPACT STUDIES	February 1984
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
P. E. Benson	57328-604195
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Office of Transportation Laboratory California Department of Transportation Sacramento, California 95819	11. CONTRACT OR GRANT NO. E80TL13
	13. TYPE OF REPORT & PERIOD COVERED
12. SPONSORING AGENCY NAME AND ADDRESS	Final
California Department of Transportation Sacramento, California 95807	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

Recommended procedures are given for designing project-level ambient carbon monoxide (CO) monitoring studies and analyzing their results. Attention is focused on transportation related applications, although the procedures are general and can be applied to other types of air quality studies. A statistical model which uses the scheduling and duration of sampling as a means to minimize bias in estimating the second annual maximum CO concentration is developed. This model is verified using a data set comprised of 112 station-years of data collected at 12 California air monitoring stations. The temporal distributions of seasonal maximums for this data set are also described.

Two types of sampling plans are discussed: Background sampling and corridor sampling. Specific criteria for choosing between these two plans are given. Recommendations are made regarding site selection, outlier analysis, rollback calculations and data handling. A computer program which estimates second annual maximum 1-hour and 8-hour ambient CO concentrations using observed maximums stratified by time of day periods is introduced. A complete set of user instructions for this program is given.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Air Quality, Carbon Monoxide, Sampling Program, Statistical Model, Ambient Air	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.

19. SECURITY CLASSIF. (OF THIS REPORT)	20. SECURITY CLASSIF. (OF THIS PAGE)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified		

DS-TL-1242 (Rev.6/76)

NOTICE

The contents of this report reflect the views of the Office of Transportation Laboratory which is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Neither the State of California nor the United States Government endorse products or manufacturers. Trade or manufacturers' names appear herein only because they are considered essential to the object of this document.

CONVERSION FACTORS

English to Metric System (SI) of Measurement

Quantity	English unit	Multiply by	To get metric equivalent
Length	inches (in)or(")	25.40 .02540	millimetres (mmm) metres (m)
	feet (ft)or(')	.3048	metres (m)
	miles (mi)	1.609	kilometres (km)
Area	square inches (in^2)	6.432×10^{-4}	square metres (m^2)
	square feet (ft^2) acres	.09290 .4047	square metres (m^2) hectares (ha)
Volume	gallons (gal)	3.785	litres (l)
	cubic feet (ft^3)	.02832	cubic metres (m^3)
	cubic yards (yd^3)	.7646	cubic metres (m^3)
Volume/Time			
(Flow)	cubic feet per second (ft^3/s)	28.317	litres per second (l/s)
	gallons per minute (gal/min)	.06309	litres per second (l/s)
Mass	pounds (lb)	.4536	kilograms (kg)
Velocity	miles per hour (mph) feet per second (fps)	.4470 .3048	metres per second (m/s) metres per second (m/s)
Acceleration	feet per second squared (ft/s^2)	.3048	metres per second squared (m/s^2)
	acceleration due to force of gravity (G)	9.807	metres per second squared (m/s^2)
Weight Density	pounds per cubic (lb/ft^3)	16.02	kilograms per cubic metre (kg/m^3)
Force	pounds (lbs) kips (1000 lbs)	4.448 4448	newtons (N) newtons (N)
Thermal Energy	British thermal unit (BTU)	1055	joules (J)
Mechanical Energy	foot-pounds (ft-lb) foot-kips (ft-k)	1.356 1356	joules (J) joules (J)
Bending Moment or Torque	inch-pounds (ft-lbs) foot-pounds (ft-lbs)	.1130 1.356	newton-metres (Nm) newton-metres (Nm)
Pressure	pounds per square inch (psi)	6895	pascals (Pa)
	pounds per square foot (psf)	47.88	pascals (Pa)
Stress Intensity	kips per square inch square root inch (ksi $\sqrt{in}$)	1.0988	mega pascals $\sqrt{metre}$ (MPa $\sqrt{m}$)
	pounds per square inch square root inch (psi $\sqrt{in}$)	1.0988	kilo pascals $\sqrt{metre}$ (KPa $\sqrt{m}$)
Plane Angle	degrees (°)	0.0175	radians (rad)
Temperature	degrees fahrenheit (F)	$\frac{tF - 32}{1.8} = tC$	degrees celsius (°C)

ACKNOWLEDGEMENTS

The author would like to thank the following: Dick Wood and Ray Baishiki for providing the software needed to interface between the CMS time-share and PANVALET systems, and for furnishing technical support throughout development of the OBSMAX program. Mr. Peter Dirrim's initial work on outlier analyses is also appreciated. Mr. Gary Honcoop and Ms. Diana Doolin of the California Air Resources Board are thanked for providing Caltrans with the raw data from which the California data set was extracted.

Virtually all the programming and data analysis assignments for this study were carried out by Mr. Jim Quittmeyer and Ms. Kim Henry. They are commended for their professional contributions to the study and their long hours of work spent on the IBM 370.

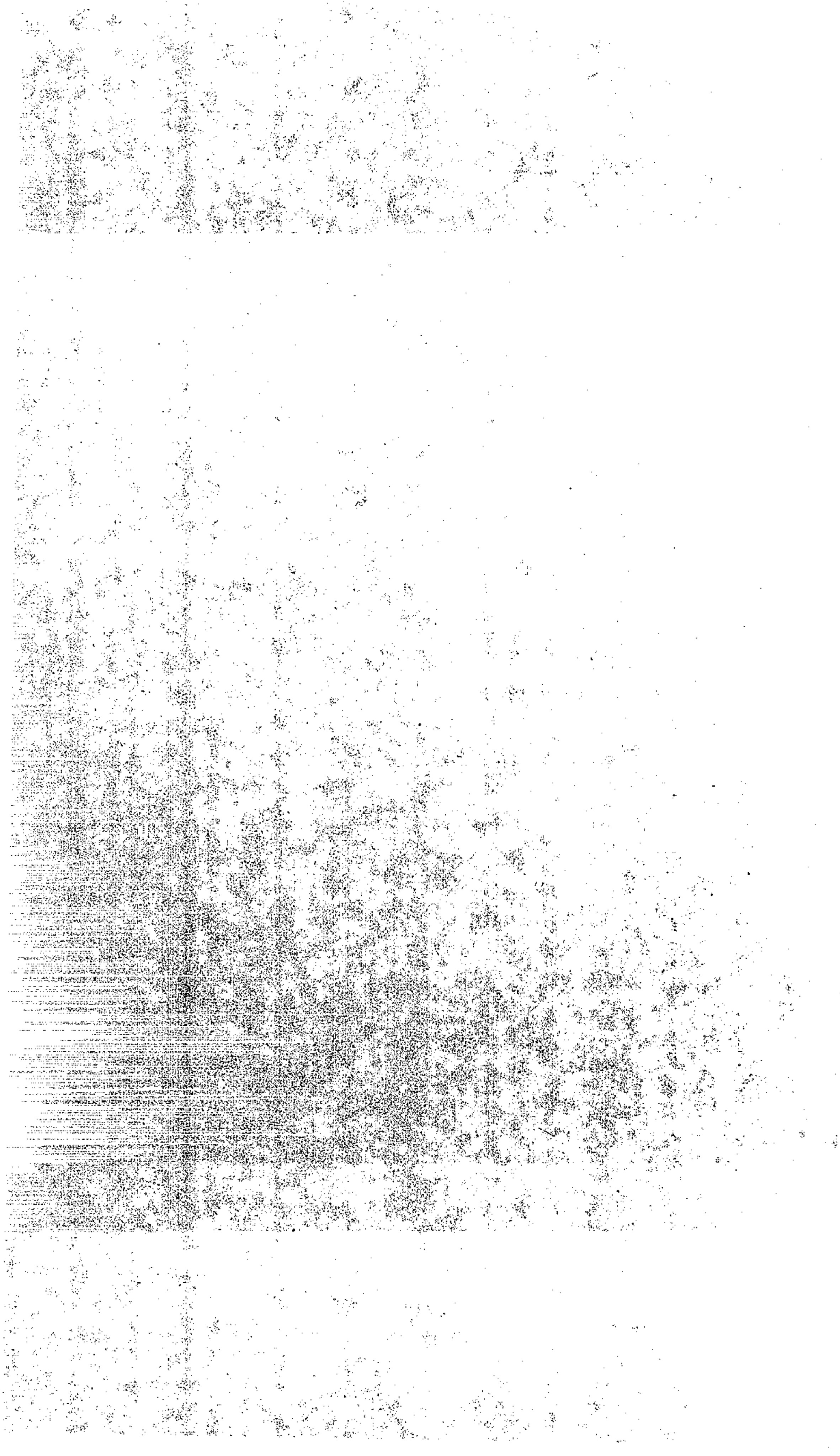


TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The determination of second annual maximum 1-hour and 8-hour ambient carbon monoxide (CO) concentrations from limited field monitoring data is an important component in transportation air quality impact studies. Since the significance of an air quality impact is judged on the basis of comparison to an absolute standard rather than between alternatives, accurate estimation of the ambient or background concentration is critical. It can often mean the difference between the finding of an acceptable or unacceptable impact. This is particularly true when project related impacts are small relative to background concentrations. Many highway improvement projects in urban areas fall into this category.

In this report, an examination of the methods used by the California Department of Transportation (Caltrans) for measuring and analyzing ambient CO data is made. Recommendations are made regarding the scheduling and duration of sampling, the selection of sampling locations, and the acquisition and storage of results. New methods for analyzing the data are also presented including an outlier analysis for editing aerometric data, a scheme for categorizing results into time of day periods, and an observed maximum analysis for estimating worst case ambient concentrations. A computer program implementing these new methods is included along with a complete set of user instructions.

The two types of sampling objectives for use in transportation air quality studies are introduced. The first, background sampling, is normally used in conjunction with a dispersion model for estimating microscale air quality impacts. The second, corridor sampling, can be used effectively for estimating impacts from transportation improvement projects along existing corridors without the use of a dispersion model. Specific criteria for choosing between these two plans are given in the report.

2. BACKGROUND

The method currently used by Caltrans to estimate second annual maximum 1-hour and 8-hour CO concentrations from field measurements was first introduced by R. I. Larsen in 1971(1). It was developed empirically from aerometric data collected at eight urban sites during the years 1962 through 1968. A two-parameter lognormal distribution was used by Larsen to extrapolate expected maximum values from random field measurements. In 1973, Singpurwalla interpreted Larsen's empirical relationships concerning averaging time in terms of extreme value theory(2). His findings were contingent upon the assumption that aerometric data are lognormally distributed. A computerized version of Larsen's model was developed by Caltrans in 1976(3).

Since its introduction in 1971, Larsen's two-parameter lognormal model has been studied and, in some ways, improved upon. The weaknesses of the original model primarily involve three areas: 1) the suitability of the two-parameter lognormal distribution, 2) the implicit assumption that sequential aerometric measurements are independent and evolve from a stationary process, and 3) the requirement that a random sampling scheme be followed.

Several authors, including Larsen, recognized that the two-parameter lognormal distribution was not appropriate for all cases. The original averaging time component of Larsen's model was modified in 1974 to an expanded five year base(4). In effect, this modification meant that the annual maximums extrapolated by the model would represent five year highs. The approach was designed as a safeguard against year-to-year sampling variability. In 1977, Larsen proposed a three-parameter lognormal distribution for use on data collected at urban and source-affected sites(5). The third parameter caused a positive shift in the conventional two-parameter distribution to account for the absence of very low concentrations at such sites. Mage and Ott recommended use of a censored three-parameter lognormal distribution in 1978(6). Barbalic examined the mutual dependence between parameters in Larsen's original two-parameter model in 1981, and developed a modified form of the model(7). Surman, Simpson and Stokoe evaluated the accuracy of Larsen's model on ozone and total suspended particulate concentrations obtained at a single monitoring station in Australia over a one-year time period(8). They reported overpredictions of maximum concentrations, especially for larger averaging times. In 1982, Holland and Fitz-Simons presented a computer program which fit a variety of statistical distributions to measured data, and evaluated the goodness-of-fit of each based on six criteria(9). The user could easily choose the best distribution for his particular data base using this program.

Earlier, in 1975, Curran and Frank suggested an alternative solution to the distribution problem(10). They saw no reason to attempt to fit a distribution to all of the data when the distribution of the highest values was the desired goal. In fact, using all available data for determining the explicit distribution gave far too much weight to the more abundant average-level concentrations. They suggested that a one or two-parameter exponential distribution fit exclusively to the higher values (say the upper fifth) would perform more reliably at predicting maximum values than a lognormal distribution fit to all the data. In more recent articles, Georgopoulos and Seinfeld have also drawn attention to this approach(11), while Roberts has demonstrated the value of using asymptotic distributions derived from extreme-value statistics for estimating pollutant concentration maximums(12).

In 1973, Patel objected to the implicit assumption of independence between sequential aerometric measurements contained in Larsen's model(13). He reasoned that air pollution concentrations measured at a fixed site would exhibit some degree of serial correlation. Neustadter and Sidik later showed that the assumption of independence was reasonable for successive measurements made three to six days apart(14). This was hardly practical for short averaging time sampling strategies, however. Larsen stated that the error generated by the admittedly incorrect assumption of sample independence was approximately balanced by a compensating arithmetic error (also detected by Patel) contained in the model(4).

More recently, Horowitz and Barakat discussed the effects of serial correlation and non-stationarity on conventional statistical applications to aerometric data such as Larsen's model(15). They concluded that serial correlation between sequential measurements would not seriously limit the usefulness of Larsen's model, but that deviations from the implicit assumption of stationarity could. A stationary process can be defined as a process whose statistical properties are invariant over time. Diurnal, weekly and seasonal periodicities in meteorology and pollutant source strength clearly make aerometric data a non-stationary process. Horowitz and Barakat solved this problem by functionally defining the aerometric non-stationary process using a least squares polynomial specification, then analyzing the residuals as a stationary autoregressive process. Another approach, used in this report, is to stratify sampling by season and day of week, and to stratify the analysis by time of day. In this way, the analysis is applied to discrete subsets of the data, each of which satisfies the stationarity assumption.

The final area in which the assumptions implicit in Larsen's model deviate from actual practice is the matter of random sampling. A survey conducted by Meisel and Dushane showed that continuous aerometric sampling over a period of three weeks to three months was the normal field practice(16). Random sampling by day or hour was characterized by respondents as inconsistent with efficient field operations. Larsen recommended collection of a 24-hour sample every third day as an example of a randomized sampling strategy for sulfur dioxide(1). Caltrans suggested moving CO samplers on a random schedule so that 20 to 25% (15% as an absolute minimum) of the complete seasonal sample was obtained at each site(3). Both these schemes proved impractical for Caltrans operations, however, because it was more efficient to site a sampler for a fixed block of time and sample on a 3 to 5-day weekly schedule than to move samplers on a daily basis. An analytical methodology consistent with this quasi-continuous type of sampling plan was needed.

In the late 1970s, the Transportation Research Board recognized this need by funding a National Cooperative Highway Research Program (NCHRP) research project aimed at developing and verifying a new approach to the prediction of ambient second annual maximum CO concentrations in urban areas. This project stressed the importance of minimizing costly field sampling by the use of auxiliary data and statistical methods. The final report by Meisel and Dushane, NCHRP 200, was published in 1979(16).

3. OVERVIEW OF NCHRP 200 METHODOLOGY

3.1 Definitions and Assumptions

The NCHRP 200 methodology is designed to amplify limited, project-specific CO data by the use of an auxiliary data set collected concurrently at a nearby, year-round monitoring station. The method assumes a significant temporal correlation between the two sets of data. The more extensive auxiliary station data set is used to characterize the pollution potential of the days sampled at the project site target station. Monitoring at the target station is assumed to run from one to two months during the high wintertime CO season. Data from the auxiliary station are then used to determine adverse days occurring during this period. An adverse day is defined by NCHRP 200 as one containing an 8-hour daily maximum which ranks in the upper 20% of readings for the year. An alternative method for determining adverse days from an area-specific meteorological index is also discussed in the report.

The overall methodology is based on an analysis of the 8-hour daily maximums recorded at the target station. These maximums are assumed to be independently distributed and correlated with the auxiliary data. The moving 8-hour average from which the daily maximum is chosen is not calculated through midnight, so that adjacent daily maximums cannot overlap. This serves to assure a degree of independence between the daily maximums.

For purposes of comparison to the national 8-hour standard, NCHRP 200 defines the second annual maximum as the second highest nonoverlapping 8-hour average occurring in a year. By assuming that the second annual maximum will not occur on the same day as the first annual maximum (an assumption supported by the data), the second annual daily maximum can be considered equivalent to the second annual maximum. Furthermore, when background levels are being measured, the difference between the first and second annual 8-hour maximum is assumed to be small. This implies that relatively little error is involved in using adverse day daily maximums as a predicator of the second annual maximum.

To support this approach, NCHRP 200 assumes that high background CO concentrations occur as a result of similar meteorological regimes, and that these regimes are quite different from those producing average or low pollutant levels. Measurements made on nonadverse days contain little significant information concerning the distribution of the peak annual maximums according to this argument. Therefore, nonadverse day daily maximums are not used in the NCHRP 200 methodology.

3.2 Method

The NCHRP 200 report recommends between one to two months of sampling at the target station. It further recommends that a 24-hour continuous or semi-continuous (3 to 5-day midweek) sampling schedule be followed. A full year of data from an auxiliary station, collected up to and during the target station monitoring period, are analyzed to determine the number and distribution of adverse days during the period. Prior to analysis, both the target and auxiliary station data are screened for missing values. Interpolation of missing values is permitted provided that no more than three total or two consecutive hours are missing in an eight-hour period.

After interpolation of the data and determination of the adverse days, the target station data may be analyzed in one of three ways: the distribution, observed maximum or combination methods. The latter is simply a weighted average of results from the first two methods.

In the distribution method, the 8-hour daily maximums measured at the target station during adverse days are plotted as a cumulative frequency function using,

$$f_i = 0.80 + 0.20 \left(\frac{i}{N+1}\right), \qquad (1)$$

where f_i is the cumulative frequency of the ith ranked daily 8-hour maximum, and N is the number of adverse days during the sampling period. NCHRP 200 then recommends fitting an exponential function to this distribution using least squares regression, and calling the value of this function at $f_i = 0.9945$, the second annual maximum.

In the observed maximum method, the number of adverse days occurring during the sampling period is used as a prequalification. If there are at least six adverse days in a one month period or 10 in a two month period, the highest 8-hour daily maximum observed during the period is used as an estimate of the second annual maximum. In cases where no auxiliary data are available, NCHRP 200 recommends sampling periods be prequalified on the basis of nationwide or statewide monthly distributions of adverse days. Based on five continuous monitoring stations with a combined total of 83 station-years of data, 30 day sampling periods conducted between October and January (February in California) inclusive are expected to contain at least six adverse days.

The NCHRP 200 methodology recommends that the 1-hour second annual maximum be determined as a function of the 8-hour maximum rather than directly from the 1-hour daily maximums. This is primarily due to the strong likelihood that the first and second 1-hour annual maximums will occur on the same day. Based on 1974 data from 67 continuous monitoring stations, NCHRP 200 recommends the following relationship:

$$CO_{1\ hr} = (1.26 \times CO_{8\ hr}) + 4.4, \qquad (2)$$

where concentrations are given in parts per million (ppm).

Year-to-year variability is not directly dealt with in the NCHRP 200 methodology. To compensate for it, the report recommends increasing the predicted second annual maximum by 15%. This conservative estimate is the average percent difference between 74 consecutive second annual maximums measured at four locations. The report also suggests extending the adverse day concept to multiple years in borderline cases if auxiliary data are available.

3.3 Verification

Five urban areas were chosen nationwide to provide the data needed to develop and verify the NCHRP 200 procedure. Each of the areas had a combination of extensive, long-standing monitoring programs and quality computerized data bases. Four to eight monitoring stataions were chosen in each area. These stations were classified as either background, intermediate (slightly source-affected) or heavily source-affected based on their proximity to major roadways, parking lots and other significant CO sources. Three stations classified as heavily source-affected were eliminated from consideration because their data were not consistent with the objective of the study to develop a method for estimating maximum background CO concentrations.

The urban areas of Los Angeles and New Jersey were designated primary locations based on the quality of their data. As such, they were used for initial development of the methodology. The areas of San Francisco, Denver and St. Louis were selected as secondary locations. Both primary and secondary data were used to verify the distribution, observed-maximum and combination methods for a variety of conditions. These conditions represented various combinations of sampling duration, station category and adverse day specification.

For each condition tested, the overall standard error and bias were determined. The bias represented the average signed difference between the actual second annual 8-hour maximum at the target station and the predicted value. It tended to be positive (overprediction) for the distribution method and negative (underprediction) for the observed-maximum method. Also, the bias tended to be greater for both methods at intermediate stations than at background stations. The standard error was used to develop confidence intervals for each of the conditions tested.

Since all three methods exhibited significant bias according to the verification study, an adjustment was incorporated into the NCHRP 200 procedure. Depending on the method, sampling duration, station category and adverse day specification, the corresponding bias determined in the verification study was subtracted from the initial prediction yielding an unbiased estimate of the 8-hour second annual maximum. In essence, results of the verification study were used to calibrate the methodology.

The 95% confidence limits for the observed maximum method applied to one month of sampling were approximately ±2.5 ppm for background stations and ±7.5 ppm for intermediate stations. These values were even higher for the distribution method. Clearly, the methodology performed unreliably for intermediate stations. However, for background stations, the confidence limits were within acceptable range.

3.4 Need for Modifications

During the course of implementing NCHRP 200 for use by Caltrans, a number of modifications were made to the original method. These modifications were made in the interest of improved program economy, flexibility and accuracy.

The first major element of NCHRP 200 requiring modification concerned the use of auxiliary data. Early in the project it became clear that the time lag in acquiring auxiliary data and the expense in processing were excessive. It was decided that the method for determining adverse days from statewide monthly distributions should be followed. However, this approach also presented a problem. The distribution of adverse days by month, given in NCHRP 200, contained data from only two regions in California. In addition, possible differences between the monthly distributions of adverse days and second annual maximums had not been investigated. Data from 12 California monitoring stations were collected and analyzed in this study as a substitute for the original NCHRP 200 work.

Another difficulty with the original method was the lack of temporal resolution in the final result. For example, an observed 8-hour maximum that occurred late in the evening might be used as representative of a morning commute hour ambient. Without organizing the analysis by time of day, similar examples of data mismatching could occur. Furthermore, by not crossing midnight in the computation of 8-hour averages, high nocturnal ambients such as reported by Remsburg, Buglia and Woodbury(17) would not be identified. These problems were addressed by extending 8-hour computations across midnight, and organizing the analysis by morning, midday, evening and nocturnal time-of-day categories.

The estimation of 1-hour maximums as a function of observed 8-hour maximums (Equation 2) was another area where the method needed modification. The intercept of 4.4 ppm contained in Equation 2 was based on 1974 data. Significant trends toward lower ambient levels in urban areas since 1974 made this constant inappropriate for current use. Use of Equation 2 also limited the applicabilty of the procedure to large urban areas similar to the ones studied in NCHRP 200. These same objections applied to the fixed adjustment values for bias used in the method. They were based primarily on 1974 data. In the modified procedure, the need for a fixed value bias adjustment was eliminated by using a sampling plan designed to minimize bias. Furthermore, estimation of the 1-hour second annual maximum was based directly on the 1-hour daily maximums, eliminating the need for a general expression to convert the 8-hour observed maximum to a 1-hour peak value. A reasonable degree of independence between the 1-hour daily maximums was guaranteed by the separation of the analysis into the four time-of-day categories.

A thorough outlier analysis was developed in this study to edit the data prior to application of the observed maximum method. This was not covered in the scope of the NCHRP 200 work, but was clearly needed if observed maximums were to be used as estimators of annual maximums.

The final need for modification to the NCHRP 200 methodology was brought about by potential changes in the National Ambient Air Quality Standards for CO(18). These proposals, still pending approval at this writing, would essentially change the definition of the worst case condition normally addressed in the project level analysis from the second to the sixth annual maximum. Therefore, the modifications made to the method were designed so that the overall sampling plan and analysis procedure could easily accommodate this possible future change.

4. ANALYSIS OF AMBIENT CARBON MONOXIDE DATA FROM SELECTED CALIFORNIA STATIONS

4.1 California CO Data Set

In order to help develop and verify the intended modifications to NCHRP 200, a large, representative data set was required. Fortunately, historical data from a comprehensive network of air quality monitoring stations throughout California were readily available from the California Air Resources Board in an edited, machine-readable form. Twelve stations with relatively complete records over a period of years were chosen from this data base. They represented a variety of geographic and demographic settings typical of California.

The selected data set was comprised of 1-hour averaged CO concentrations for each hour of the day. In cases where missing data were encountered, the NCHRP 200 interpolation method was used. If gaps within a 31 hour period (midnight to 7 a.m. of the following day) exceeded the size and frequency criteria set down in NCHRP 200, the entire day was dropped from the data set. After editing for missing values, the data set was stratified by CO season, starting July 1 and ending June 30 of the following year. The seasonal stratification was made in lieu of a calendar year division so that the monthly distribution of maximums would accurately represent the distribution encountered when sampling within a season. Seasons with more than 10 days missing in any single month from October through February, or more than 25 days missing over the entire five month wintertime period, were deleted from the data set. This editing left a total of 112 station-seasons in the data set, each comprised of an average of 349 days worth of 24 one-hour CO concentrations.

Peak 8-hour averages were determined for each day. Calculations were made crossing midnight with the start hour of the 8-hour period determining the date of the maximum. Overlapping 8-hour daily maximums were not permitted. The daily maximums within each station-season were then ranked. Dates and start hours of the top six ranks were retained in the final version of the data set. Tied values were assigned the same rank, making multiple occurrences of seasonal maximums possible. A similar treatment was given to 1-hour maximums, with the exception that multiple annual maximums within the same day were allowed. A description of the final data set is given in Table 1.

4.2 Distribution of Seasonal Maximums

The distributions of the seasonal high 8-hour daily maximums by month, day of week and hour of day were key elements in development of the modified sampling procedure. Instead of using the top 20% of the data, as in NCHRP 200, the modified method focused on the probabilities of encountering maximums within the top six ranks of the seasonal statistics. Study was limited to these seasonal maximums rather than adverse days because their temporal distributions were expected to more closely follow the distribution of second annual maximums.

The modified method assumes that the monthly distribution of seasonal maximums is independent of averaging time and rank. A categorical analysis of variance performed on the 12-station data set showed no significant difference between both monthly distributions of 1-hour and 8-hour

TABLE 1. CALIFORNIA AMBIENT CO DATA SET

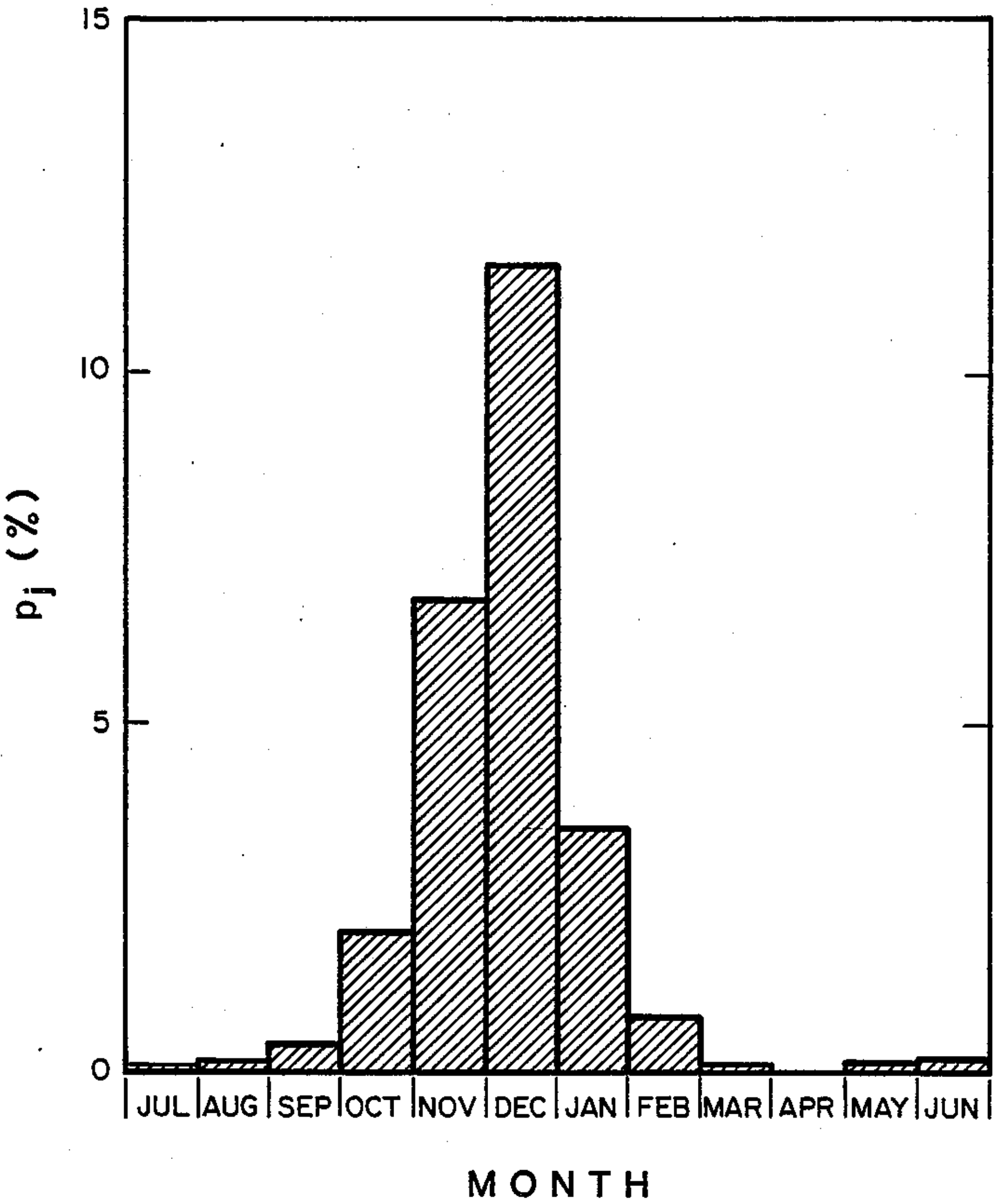
Metro Pop. (1980)	Station	Area/Site Code	Years Studied	Total Seasons	2nd Max. (81/82, PPM) 1-Hr	8-Hr
<100K	Pittsburg	700/430	1969-82	12	8	4.9
	Lancaster	7000/82	1971-82	10	9	4.9
	Escondido	8000/115	1975-82	7	12	8.0
	Santa Barbara	4200/355	1974-82	7	15	8.1
	Salinas	2700/544	1976-82	5	4	2.9
100K to 500K	Bakersfield	1500/203	1972,73,76-79,81-82	5	14	10.1
	Stockton	3900/252	1965-67,79-82	5	14	7.5
	Redwood City	4100/541	1968-82	13	10	5.5
	Sacramento	3400/282	1972-80	7	11[a]	7.4[a]
	Pomona	7000/75	1966-82	14	12	9.6
>500K	San Diego	8000/120	1973-82	8	12	8.6
	Burbank (L.A.)	7000/69	1963-82	19	25	20.1
			Total	112		

[a]79/80 Season

maximums, and first through sixth ranked seasonal high 8-hour maximums. There was a significant difference in the distribution of maximums by station. However, this was slight enough to justify the aggregation of results over the 12 stations as representative of a composite California location. The final sampling plan was specified using the aggregated monthly distribution of the proportion of days containing 8-hour seasonal maximums shown in Figure 1.

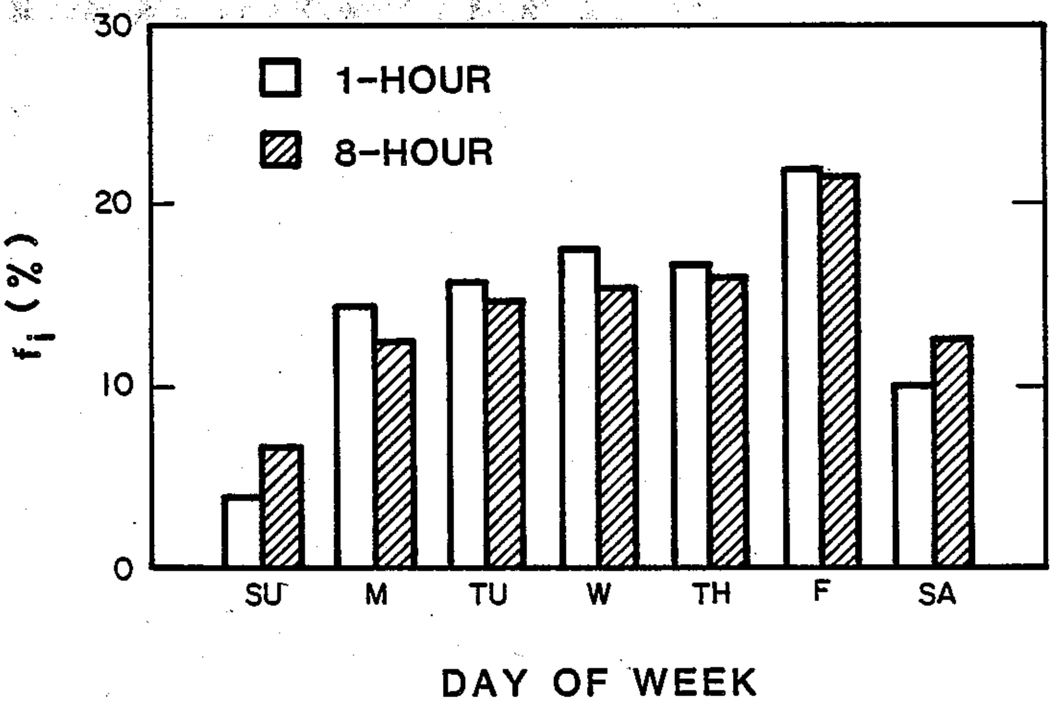
The distribution of seasonal maximums by day of week shown in Figure 2 was also important in the development of the modified method. The fraction of 1-hour seasonal maximums is somewhat greater than 8-hour maximums on weekdays, while the opposite is true on weekends. The relatively short duration of weekday traffic peaks and the broader temporal distribution of traffic volumes on weekends is consistent with this pattern. The day-to-day trends in Figure 2 are roughly similar for both 1-hour and 8-hour averaging times. There are gradually more seasonal maximums occurring through the week until a peak is reached on Friday. The number of seasonal maximums then drops significantly for Saturday and reaches a minimum on Sunday. Cross-stratification of the data by time of day and day of week revealed that the additional Friday occurrences and many of the Saturday occurrences took place in the late evening hours. The few occurrences of Sunday maximums also took place in the evening about one hour later than weekday commute peaks.

These temporal patterns exhibited by the seasonal maximums closely follow expected traffic distributions reported by Shirley(19). A composited version of the 1-hour and 8-hour day of week distributions was used to determine the probabilities of encountering seasonal maximums associated with different day of week sampling plans.



MONTHLY DISTRIBUTION OF THE PROPORTION OF DAYS, p_j, CONTAINING 8-HOUR DAILY MAXIMUMS WITHIN THE TOP SIX SEASONAL RANKS

FIGURE 1



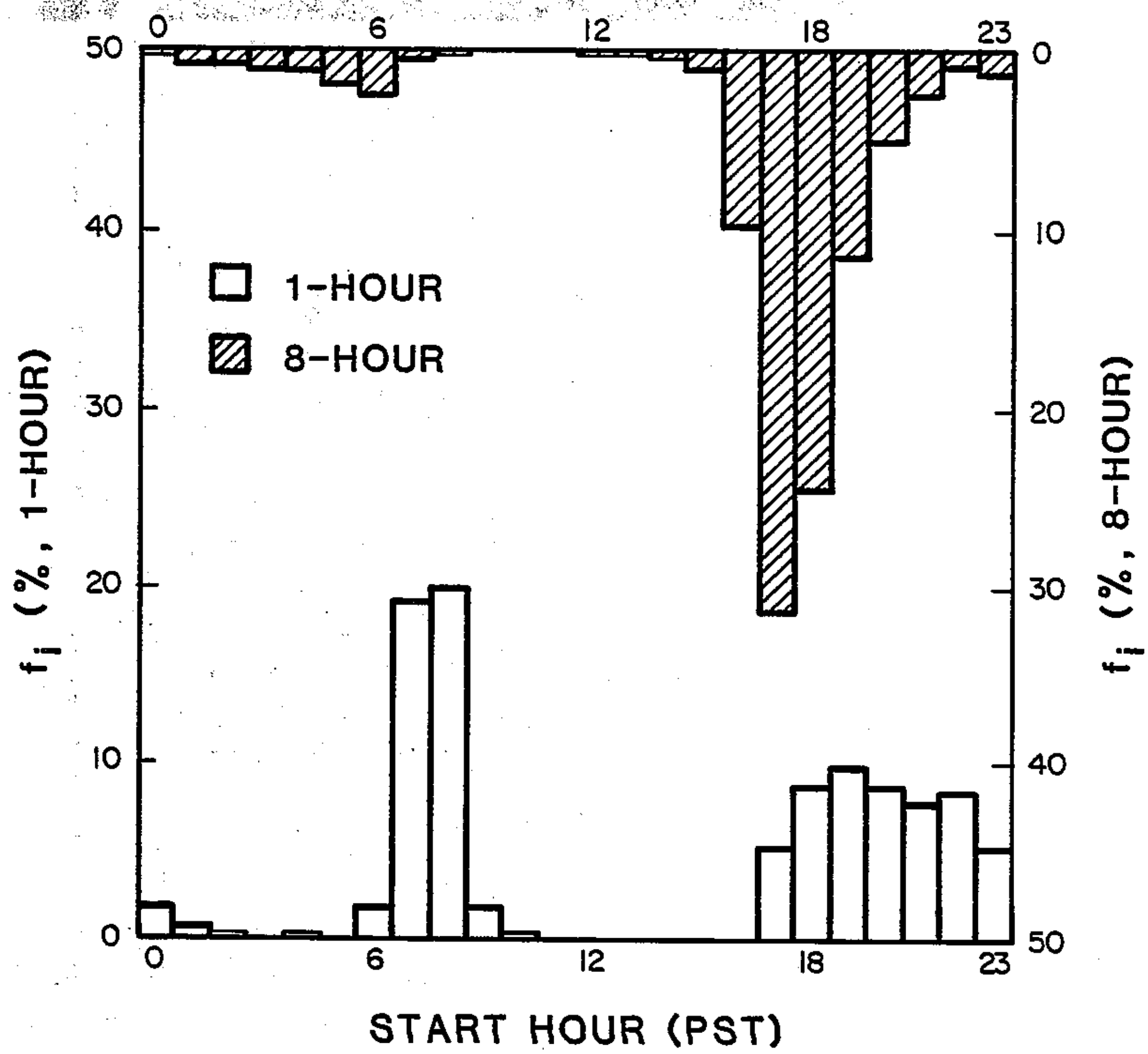
FRACTION OF 1-HOUR AND 8-HOUR SEASONAL MAXIMUMS, f_i, DISTRIBUTED BY DAY OF WEEK

FIGURE 2

The distributions of 1-hour and 8-hour seasonal maximums by start hour are shown in Figure 3. A start-hour is defined as the hour at which measurement begins. For example, a start hour of 18 could refer to a 1-hour average concentration measured from 6 to 7 p.m. or an 8-hour average from 6 p.m. to 2 a.m. of the next day. The times are given in Pacific Standard Time (PST).

The distributions are quite dissimilar, particularly regarding the occurrence of morning maximums. One-hour maximums occur most frequently between the morning commute hours of 7 to 9 a.m. Approximately 40% of the 1-hour maximums and 10% of the 8-hour maximums occur during this period. The explanation for this difference lies in the combined temporal distributions of traffic and meteorology. During the evening hours, these two factors combine over sufficiently long periods to yield the bulk of the 8-hour daily maximums. In the morning hours, a pronounced morning commute combined with very stable meteorological conditions lead to a substantial number of 1-hour seasonal maximums. However, the short duration of the morning commute peak, and the rapid shift to unstable meteorological conditions typically following this peak limit the number of morning 8-hour seasonal maximums to a very small percent of the total.

The time of day distributions were used to determine an appropriate division of the modified observed maximum analysis into four time periods: morning, midday, evening and nocturnal (Table 2). By dividing the analysis in this way, ambient maximums could easily be matched with traffic and meteorology from corresponding time periods. The monthly distribution of overall daily maximums contained in the California data set was assumed to accurately represent



FRACTION OF 1-HOUR AND 8-HOUR SEASONAL MAXIMUMS, f_i , DISTRIBUTED BY START HOUR

FIGURE 3

TABLE 2. TIME PERIODS FOR ANALYSIS OF AMBIENT CO CONCENTRATIONS

Time Period	Start Hour 1-Hr	Start Hour 8-Hr	Occurrence of Seasonal Maximum (%) 1-Hr	Occurrence of Seasonal Maximum (%) 8-Hr
Morning	0600	0100	43	10
Midday	1000	0800	1	>1
Evening	1700	1100	32	80
Nocturnal	2100	2000	24	10

similar distributions of daily maximums by time period. The sampling plan developed from the distribution of overall daily maximums was considered equally valid for estimating maximums by time period under this assumption.

5. A PROBABILISTIC MODEL FOR ESTIMATING THE DISTRIBUTION OF OBSERVED MAXIMUMS

5.1 Model Development and Verification

The fundamental difference between the NCHRP 200 method and the modified method developed in this report involves the procedure for determining the duration of sampling. Instead of sampling for a fixed 30 days with a minimum of 6 adverse days required, the modified procedure calls for a sampling program whose duration varies with the month or months sampled. The duration of sampling is chosen so as to yield an extremely high probability of attaining as an observed maximum, an unbiased estimate of the expected second annual maximum. This is achieved through the combined application of the binomial distribution and the combinatorial analysis. The specific sampling intervals recommended in this paper are based on the California data set. However, the principles can be extended to any comparable data set.

A distribution of randomly chosen, independent events characterized by two mutually exclusive outcomes can be described by the binomial expansion, $(q+p)^n$, where q and p represent the probability of occurrence attached to each outcome. The rth term of the expansion equals the probability that the outcome, whose underlying probability is denoted by p, will occur r times in n samples. This can be stated as

$$P(r|p,n) = \frac{n!}{r!(n-r)!} p^r q^{n-r} . \qquad (3)$$

The binomial expansion was used to generate expected monthly probabilities of encountering r seasonal maximums (defined as daily 8-hour maximums within the top six ranks for the season) in an n-day sampling period based on the underlying probabilities shown in Figure 1. Thus, for a full month sample taken during the jth month,

$$P(r|j) = P(r|p_j,n_j), \tag{4}$$

where p_j and n_j equal, respectively for the jth month, the probability of encountering seasonal maximums and the number of days in the month. Equation 4 was used to predict the distribution of occurrences of seasonal maximums for the full month sampling periods of October through February. These are compared in Table 3 to the observed distributions taken from the California data set. Probabilities are rounded off to the nearest whole percent so that totals may not exactly equal 100%.

Use of the binomial distribution assumes that the 8-hour daily maximums are randomly chosen, independent events. In fact, they are a set of sequentially sampled, autocorrelated events. The assumption of randomness is not seriously violated provided sampling is of sufficient duration to incorporate a majority of wintertime meteorological conditions. The assumption of independence between daily maximums presents a more serious problem, however. Examination of the California data set showed that clusters of consecutive seasonal maximums occur with significantly greater frequency than would be expected from a series of independent events. This was most evident for small clusters of two to three seasonal maximums, with 26% of paired values and 8% of groups of three occurring on successive days.

Clustering of seasonal maximums is caused by periods of calm, stable meteorological conditions between winter storms. The effect of clustering on the overall distribution of seasonal maximums is apparent in Table 3. There is a consistently higher percentage of months with no occurrences of seasonal maximums than would be expected from a truly independent distribution. This higher percentage is caused by the clustering of maximums in other months. By the same token, the overestimation of months with only one occurrence can be attributed to the likelihood that seasonal maximums will occur in clusters rather than as isolated events.

The binomial distribution does reasonably well at predicting the observed pattern for months having two or more occurrences. For November, it also does well in cases where there are less than two occurrences. Therefore, the assumptions of independence and randomness, while not entirely valid, were considered satisfied to the extent that the binomial distribution could be used as an approximation of the number of seasonal maximums, r, within a given sampling interval. A method was then needed to approximate the rank distribution of an observed maximum from a set of r seasonal maximums. Combinatorial analysis was used for this purpose.

The probability that an observed maximum equals the mth ranked seasonal maximum given a sample containing r seasonal maximums can be stated as

$$P(m|r) = \frac{_{(\ell+1)}C_r - {_\ell}C_r}{_6C_r} , \qquad (5)$$

where ℓ = number of ranks less than the mth rank (6-m) for the r seasonal maximums,

TABLE 3. OBSERVED AND (PREDICTED) PROBABILITIES OF ENCOUNTERING r SEASONAL MAXIMUMS BY MONTH

	Probability of Occurrence (%)				
r	Oct.	Nov.	Dec.	Jan.	Feb.
0	64 (52)	14 (12)	9 (2)	45 (33)	82 (79)
1	19 (34)	26 (27)	6 (9)	24 (37)	14 (19)
2	10 (11)	26 (28)	13 (18)	15 (20)	2 (2)
3	4 (2)	21 (19)	22 (22)	11 (7)	1 (0)
4	3 (0)	7 (9)	17 (20)	5 (2)	1 (0)
5	1 (0)	4 (3)	15 (14)	1 (0)	
6		2 (1)	10 (8)		
7		1 (0)	5 (4)		
8			2 (1)		
9			1 (0)		

and
$$ {}_nC_r = \begin{cases} \frac{n!}{r!(n-r)!}, & n \geq r \\ 0, & n<r \end{cases} $$

This formulation assumes that there are no multiple occurrences of seasonal maximums. For the California data set there were tied ranks, however, yielding an average of eight maximums within the top six ranks per season. Equation 5 was modified to account for this as follows:

$$ P(m|r) = \frac{1}{{}_8C_r}\Bigg[\frac{g^2}{36}\left({}_{(\ell+1)}C_r - {}_{\ell}C_r\right) + \frac{g}{18}\left({}_{(\ell+2)}C_r - {}_{\ell}C_r\right) $$

$$ + \frac{g\ell}{18}\left({}_{(\ell+2)}C_r - {}_{(\ell+1)}C_r\right) + \frac{1}{36}\left({}_{(\ell+3)}C_r - {}_{\ell}C_r\right) $$

$$ + \frac{\ell}{18}\left({}_{(\ell+3)}C_r - {}_{(\ell+1)}C_r\right) + \frac{\ell^2}{36}\left({}_{(\ell+3)}C_r - {}_{(\ell+2)}C_r\right)\Bigg] \qquad (6) $$

where g = number of ranks greater than the mth rank (m-1).

Equation 6 is based on the assumption that the two extra seasonal maximums are randomly distributed among the top six ranks (see Appendix C for the derivation of Equations 5 and 6). Overall, there is a slight increase with descending rank of the number of ties in the California data set. However, for months containing three or more seasonal maximums, the distribution of tied ranks is approximately uniform among the top six ranks.

Equation 6 was used to model the distribution of observed maximums among the top six ranks as a function of the number of occurrences of seasonal maximums. These values can be compared in Table 4 to the rank distribution of

TABLE 4. OBSERVED AND (PREDICTED) PROBABILITIES THAT THE OBSERVED MAXIMUM WILL EQUAL THE mth SEASONAL MAXIMUM GIVEN r OCCURRENCES

	Probability of Occurrence (%)							
m	r = 1	2	3	4	5	6	7	8+
1	14 (17)	23 (32)	44 (46)	56 (59)	74 (71)	77 (82)	100 (91)	100 (100)
2	14 (17)	20 (26)	25 (29)	25 (28)	17 (23)	15 (17)	0 (9)	
3	11 (17)	27 (20)	18 (16)	17 (10)	9 (5)	0 (2)		
4	23 (17)	13 (14)	9 (7)	3 (2)	0 (0)	8 (0)		
5	19 (17)	13 (7)	4 (2)	0 (0)				
6	19 (17)	4 (1)	0 (0)					

monthly observed maximums categorized by number of occurrences, r. The discrepancies between observed and predicted probabilities are primarily due to the non-uniform distribution of tied ranks, and the tendency for first and second seasonal maximums to be associated with clustered results.

To test the validity of the combined effects of the binomial distribution and Equation 6, the modeled distributions of observed maximums by rank for the months October through February were generated by

$$P(m|j) = \sum_{r=1}^{7} [P(r|j) \cdot P(m|r)] + P(r \geq 8|j) \cdot P(m|r=8), \tag{7}$$

where $P(m|j)$ equals the probability of the observed maximum during the jth month coming from the mth rank. The second term in Equation 7 relates the diminishing probabilities generated by the binomial distribution for $r \geq 8$ to the fixed probability for $r=8$ derived from the combinatorial analysis. These predictions generated by Equation 7 compare favorably to the observed distributions obtained from the California data set (Table 5).

5.2 Design Criteria for an Unbiased Sampling Plan

In general terms, the combined model developed in Section 5.1 can be expressed as follows:

$$P(m|p_j, n) = \sum_{r=1}^{n} P(r|p_j, n) \cdot P(m|r), \tag{8}$$

TABLE 5. OBSERVED AND (PREDICTED) PROBABILITIES BY MONTH THAT THE OBSERVED MAXIMUM WILL EQUAL THE mth SEASONAL MAXIMUM

	Probability of Occurrence (%)				
m	Oct.	Nov.	Dec.	Jan.	Feb.
1	9 (10)	27 (31)	51 (50)	16 (17)	3 (4)
2	4 (9)	25 (21)	15 (23)	7 (14)	3 (4)
3	5 (8)	13 (14)	13 (12)	11 (12)	3 (4)
4	8 (7)	9 (10)	8 (6)	9 (10)	3 (4)
5	6 (7)	6 (7)	4 (3)	8 (8)	3 (3)
6	4 (6)	6 (5)	1 (2)	5 (6)	5 (3)

where n is the number of days sampled in the jth month, and $P(m|r)=P(m|r=8)$ for all $r \geq 8$. Given known values for p_j and the average differences between concentrations by rank from the California data set, Equation 8 can be used to approximate an unbiased sampling program. The values for p_j are given in Figure 1, while the distribution of differences between the second seasonal maximum (m=2) and the first through sixth seasonal maximums for the California data set is summarized in Table 6. The maximum and overall average differences are also given in Table 6. Because there was no significant difference between the monthly distributions of 1-hour and 8-hour seasonal maximums, a sampling program based on 8-hour values was assumed equally valid for 1-hour estimates.

The sampling program was designed so that the duration of sampling would be sufficient to guarantee a fixed probability, P_c, of obtaining one or more seasonal maximums. To select a proper value for P_c, the probabilities described by Equation 8 were used in combination with the average differences, $\bar{d}_{m-2}$, given in Table 6. Since the binomial distribution turned out to be relatively insensitive to values of p_j ranging from 0.01 to 0.10 given a fixed value of P_c, an averaged value, p, of 0.05 was used in the following final design equation:

$$B=\sum_{r=1}^{7}\left\{P(r|\bar{p},P_c)\cdot\sum_{m=1}^{6}[P(m|r)\cdot\bar{d}_{m-2}]\right\}$$

$$+P(r\geq 8|\bar{p},P_c)\cdot\bar{d}_{1-2}+(1-P_c)\cdot\bar{d}_o \quad . \qquad (9)$$

TABLE 6. DISTRIBUTION OF DIFFERENCES BETWEEN 2nd AND mth SEASONAL MAXIMUMS

$\lvert d_{m-2} \rvert$ (ppm)	Probability of Occurrence (%) m = 1	3	4	5	6
$\leq$ 0.5	38	62	30	14	6
0.6 - 1.0	22	21	29	29	25
1.1 - 1.5	17	10	22	20	21
1.6 - 2.0	6	2	7	17	17
2.1 - 2.5	6	1	4	5	11
2.6 - 3.0	2	2	3	4	6
> 3.0	9	3	5	11	14
$\lvert d_{m-2} \rvert_{max}$ (ppm)	8.6	7.2	7.7	7.9	8.4
$\lvert \overline{d}_{m-2} \rvert$ (ppm)	1.27	0.69	1.15	1.52	1.85

In this equation, B equals the expected bias in ppm given P_c, and $\bar{d}_o$ represents an estimate of the average difference between the second seasonal maximum and observed maximums occurring outside the top six rank interval. In cases where a seasonal maximum is not encountered during the sampling period, there is still a very high probability, P_c', that a daily maximum within the top twelve ranks will be found. It can be shown that

$$P_c' = 1 - \exp\left(\frac{\ell n(1-P_c)\,\ell n(1-2\bar{p})}{\ell n(1-\bar{p})}\right), \tag{10}$$

assuming that the average underlying probability of encountering a daily maximum within the top twelve ranks is double the probability of encountering a maximum in the top six ranks. Thus, the selection of the design probability, P_c, was based on B approaching zero in Equation 7 and P_c approaching one in Equation 10. The value for $\bar{d}_o$ was determined by extrapolating values of d_{m-2} for m=7 to 12 using the average differences given in Table 6, and compositing these values as follows:

$$\bar{d}_o = \sum_{r=1}^{7} \left\{ \frac{P(r|\bar{p},P_c)}{P_c} \cdot \sum_{m=7}^{12} [P(m-6|r) \cdot \bar{d}_{m-2}] \right\} + \frac{P(r \geq 8|\bar{p},P_c)}{P_c} \cdot \bar{d}_{7-2} \ . \tag{11}$$

Equation 11 esentially deals with the small probability $P_c'-P_c$, that an observed maximum will fall outside of the top six ranks, but within the top twelve ranks for the season. The same combined probabilities used to model the distribution of the top six ranks are used to weight the values of $\bar{d}_{m-2}$ for m=7 to 12 when developing the composited result, $\bar{d}_o$.

By trial and error solution, a design value for P_c of 0.93 was determined. This yields a value of 0.995 for P_c. Simply stated, this means that, given a sampling period of sufficient duration to assure a 93% chance of encountering at least one seasonal maximum, one can be 99.5% confident that the observed maximum is an unbiased estimate of the expected second annual maximum. It can also be stated that there is approximately a 62% chance that the observed maximum will equal the expected first or second seasonal maximum under such conditions.

The average differences between ranks were based on measurements made at a variety of stations. Seasonal 8-hour maximums for the Burbank station were routinely in the 30 to 40 ppm range during the 1960s, while values for Salinas in the 1970s were usually below 5 ppm. Thus, the distribution of differences given in Table 6 represents a wide range of exposures. The average differences were used in this paper strictly for the purpose of approximating an unbiased sampling program. It was assumed that

$$\frac{\bar{d}_{1-2}}{\bar{d}_{m-2}} = \text{CONSTANT} \qquad (12)$$

for each value of m regardless of location, time period, season or averaging time. The absolute random error that one can expect in terms of ppm for any given location will depend on the magnitude of the seasonal maximums at that location, not the average differences derived from the California data set.

6. OUTLIER ANALYSIS

6.1 Objectives

Occasionally, erroneous data may become imbedded in a data set as the result of equipment malfunction or human error. The consequences of this kind of error are magnified by the nature of the observed maximum method. If a very high, erroneous measurement is included in an ambient CO data set, it will likely be identified as an observed maximum and thus cause a significant overprediction of the second annual maximum. To avoid this, a thorough outlier analysis is performed prior to selection of the observed maximums. The analysis checks the data set for internal consistency, and identifies suspected outliers for inspection.

Three separate analyses for outliers are incorporated into the observed maximum computer program described in Section 9. Each analysis checks the internal consistency of the data for a different temporal stratification. The day-to-day consistency is checked by both period of the day (i.e., midday, evening, etc.) and by hour of the day. The within-day consistency is checked by analyzing deviations from a 5-hour moving average. Underlying distributions for each stratified subset of the data are assumed for purposes of comparison. If the data do not fit the assumed distributions, or contain an insufficient number of measurements for comparison, then the outlier analyses will not be effective in detecting anomalous values. However, for most CO data sets, the three analyses should provide adequate protection against erroneous data.

6.2 Gap Test

The first procedure used to screen the data is a modified form of the Gap Test(20). This procedure is strictly concerned with identifying anomalous gaps in the upper 1% of the data. Data values stratified by time period are rounded off to the nearest ppm and ranked in descending order. The upper 1% of the data is examined for gaps of 1 ppm or greater starting at the highest ordered value. If a gap is found, a two-parameter exponential frequency distribution is fit by least squares to the upper 20% of the data, excluding values above the gap. This serves as the expected distribution to which the observed gap will be compared. Before proceeding, a Chi-Squared Test for goodness-of-fit is performed at the 95% significance level on the exponential distribution. If this test indicates a significant lack of fit to the observed frequency distribution, the program moves on to examine the next gap, if any. Otherwise, the gap being considered is tested using the following Chi-Squared statistic:

$$\chi^2 = E(x_{i-1}+1) - E(x_i) + \frac{[O(x_i)-E(x_{i-1}+1)]^2}{E(x_{i-1}+1)}, \qquad (13)$$

where x_i = first value after gap
x_{i-1} = last value before gap
$E(x_i)$ = expected frequency of values greater than or equal to x_i
$O(x_i)$ = observed frequency of values greater than or equal to x_i

This statistic is tested at the 99% confidence level for a significant deviation from the expected frequency distribution. If the gap proves too large to be explained by the expected distribution, the value(s) above the gap are identified as suspected outliers and processed for later inspection. The test then proceeds to examine the next gap, if any. If the first gap tested is not significant, however, the program moves on to examine the next time period.

6.3 Grubbs Test

The second procedure used to test for outliers is called Grubbs Test(21,22). As with the Gap Test, Grubbs Test is used to identify anomalous high values. The test is performed on the highest values encountered in the data set for each hour of the day. The data are assumed to be lognormally distributed. The test statistic, T_j, is computed for the jth hour as follows:

$$T_j = \frac{\mathrm{Max}(z_{ij}) - \bar{z}_j}{s_j} \qquad (14)$$

where,

$$z_{ij} = \ell n x_{ij}$$

$\mathrm{Max}(z_{ij})$ = Maximum value of z_{ij} for the jth hour over n_j days.

$$\bar{z}_j = \frac{\sum_{i=1}^{n_j} z_{ij}}{n_j}$$

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (z_{ij})^2 - \frac{\left(\sum_{i=1}^{n_j} z_{ij}\right)^2}{n_j}}{n_j - 1}$$

This statistic is tested at the 1% significance level (one-sided) for the number of observations, n, contained in the data set for the hour being tested. Any value identified as a suspect outlier by this method is expected to occur as a valid member of the sample only one time out of every 100 samples of size n. In the case of 30 days of monitoring, this would amount to only one occurrence for every 3,000 values. Measurements identified as exceeding this probability should be examined closely.

6.4 Time Sequence Test

Only the highest 1% of hourly averages contained in a data set are likely to be examined by either the Gap or Grubbs Tests. While this furnishes the observed maximum method with reasonable protection for hourly peak values, further screening of the full range of data is needed to assure against anomalous 8-hour maximums. A Time Sequence Test using a 5-hour moving average was devised to meet this need.

In order to employ the Time Sequence Test, two hours of data before and after the value being tested must be available. Using these four values, an expected range for the center value, x_i, is developed as follows:

$$x_{max} = \mathrm{Max}(x_{i-1}+\Delta_{(-)},\ x_{i+1}+\Delta_{(+)}) \qquad (15)$$

$$x_{min} = \mathrm{Min}(x_{i-1}-\Delta_{(-)},\ x_{i+1}-\Delta_{(+)}) \qquad (16)$$

where,

$$\Delta_{(-)} = |x_{i-1}-x_{i-2}|$$

$$\Delta_{(+)} = |x_{i+1}-x_{i+2}| \ .$$

This range encompasses eight possible interpolative modes: four direct and reverse trends (sawtooth) using the pairs of values preceding and following x_i, and four interpolations between x_{i-j}, x_{i+k} for (j=1,2 [k=1,2]). When a value of $\Delta_{(-)}$ or $\Delta_{(+)}$ is less than 1 ppm, the value is raised to equal to one. This recognizes the instrument, sampling and round-off error inherent in each value of the data set. If x_i falls within the expected range, no further test is made. If not, the expected range is used as a yardstick to judge the compatibility of x_i with its neighboring values. A Dixon Ratio(23) is computed as follows:

$$R = \frac{x_i - x_{max}}{x_i - x_{min}}, \quad x_i > x_{max} \tag{17}$$

or

$$R = \frac{x_{min} - x_i}{x_{max} - x_i}, \quad x_i > x_{min} . \tag{18}$$

The ratio is tested against a critical value of 0.5. If R is greater than 0.5, then x_i is identified as a suspect outlier.

The Time Sequence Test is meant to act as a filter to screen the data for possible aberrations. It is not a rigorous statistical test and has no significance level attached to it. The critical value of 0.5 was chosen to identify values which experience indicates deserve inspection.

7. RECOMMENDED PROCEDURES

7.1 Sampling Plan Selection

The following procedures address two types of ambient CO sampling plans: corridor sampling and background sampling. Corridor sampling is conducted near existing highway facilities, and represents ambient CO levels directly influenced by vehicle emissions within the transportation corridor. Background sampling is carried out at sufficient distances from significant sources such as freeways, shopping centers or major arterials so as to represent a measure of ambient CO concentrations characteristic of well distributed area-wide sources (e.g., low volume surface streets, residences, etc.). The choice of which sampling program to use depends upon the air quality study objectives and the nature of the proposed transportation project.

Corridor sampling is recommended for projects where no substantial change is made to the distribution of emissions within an established corridor. Examples of this include widening within existing right-of-way, HOV lane conversion or construction of a median transitway. For these types of projects, a sufficiently accurate assessment of microscale air quality impacts can be made by adjusting on-site measurements for anticipated changes in traffic volume and vehicle emissions using rollback procedures.

Another possible use of corridor sampling is for projects where post-construction monitoring is planned. Such monitoring is called for when anticipated impacts border on the air quality standards and authorities have agreed to make implementation of mitigation measures contingent upon post-construction air quality.

Background sampling is recommended for projects that are expected to produce significant changes in the distribution of emissions. Projects involving new alignment, major widening (requiring additional right-of-way), signalization or parking facilities fall into this category. Changes in the distribution of emissions necessitate the use of a line source dispersion model such as CALINE3(24) to estimate post-construction concentrations. Background ambient CO levels are needed to add to the modeled contributions in order to estimate overall microscale impacts.

In the past, attempts to measure background concentrations were often carried out too close to existing facilities. In such cases, the resulting estimates of 1-hour and 8-hour background CO levels were positively biased by corridor emissions. This led to overestimates of impacts since corridor emissions were, in effect, counted twice. The recommended siting criteria for background sampling are designed to eliminate this problem.

7.2 Siting Criteria

Use of the observed maximum method recommended in NCHRP 200 makes siting decisions more critical than in the past. The outlier analysis provides some protection against isolated, extreme events, but proper siting is needed to keep the probability of encountering such events to a minimum.

Factors to consider when siting ambient monitoring stations include:

1. Topography
2. Land Use
3. Proximity of significant emission sources
4. Exposure
5. Serviceability
6. Security

The first four of these factors are evaluated in terms of the representativeness of the site. The last two are practical constraints best left to the judgment of the field supervisor.

For corridor sampling plans, sites representative of public or private access locations near the right-of-way should be chosen. Strictly speaking, the National Ambient Air Quality Standards (NAAQS) pertain to all areas of public access. However, judgment should be exercised to assure that the ambient levels are measured at locations compatible with the averaging time of the standard. In the case of the 1-hour and 8-hour NAAQS for CO, schoolyards, parks, private yards, etc., are locations consistent with the above criteria — parking lots, plowed fields, pedestrian walkways, etc., are not.

Since peak concentrations within the corridor are heavily influenced by emissions from the highway, significant changes in highway geometry or traffic volume will usually dictate the overall distribution of monitoring sites for the project. Whenever possible, these sites should be chosen in locations where public concern is expected to focus (e.g., schools, hospitals, etc.). Changes in land use along the corridor are of secondary concern because of the overriding influence of the highway. However, major changes in topography should be represented by separate monitoring sites (e.g., canyon to valley or opposite sides of a coastal ridge).

In designing a background sampling plan, the first consideration is the existing or projected land-use zones traversed by the project. These zones need only be broadly defined, and must predominate along one or both sides of the corridor for at least one mile to be considered for sampling. Broad categories of land use extensive enough to meet this criterion include:

Urban Residential
Urban Business
Industrial (Light or Heavy)
Recreational
Rural

Site selection criteria for background sampling plans differs considerably from corridor sampling criteria. Concentrations at a background site must represent emissions from many diverse activities and locations. No single emissions source can dominate. The site must characterize a single, broad land-use category, and be so situated that measurements are representative of background levels only.

In cases where a new alignment is being studied, background monitoring sites can be established within the proposed transportation corridor. However, if the new alignment is expected to be accompanied by changes in land use, monitoring should take place in the nearest area where existing conditions most closely match the projected land-use surrounding the proposed corridor. For example, background monitoring sites for a new alignment near an expanding urban or suburban region should be located within the nearest developed zones. Monitoring should not be conducted in vacant fields that will soon become housing tracts. On the other hand, monitoring sites should not be located so far from the project location that significant differences in meteorology can be expected.

In cases where background measurements are needed for a corridor containing an existing highway, sampling sites must be located so that they are representative of air quality along the corridor minus the effects of the existing highway. If this provision is not made, the background measurements will be influenced by the existing highway emissions, and thereby contribute to an overprediction of the modeled air quality impacts. Minimum setback distances from existing facilities should be maintained to avoid this problem. Recommended setback distances from streets and highways as reported in the literature vary widely. A summary of these distances is given in Table 7.

The recommended minimum setback distance (d) in meters from the edge of shoulder or curb for Caltrans background monitoring programs is as follows:

$$d = ADT/100 \qquad (19)$$

where ADT = average daily traffic in vehicles/day. Equation 19 was developed using a revised version of the CALINE3 line source dispersion model. The model results were used to determine minimum setback distances required to hold CO concentrations below 1 ppm under stable atmospheric conditions for varying ADT. Recent tracer gas release experiments conducted by Caltrans have helped confirm these model results out to a distance of 200 meters.

TABLE 7

Minimum Setback Distances from Streets and Highways for Background and Neighborhood Monitoring Stations (Various Sources)

Minimum Setback Distance (m)	Traffic Volume or Roadway Type (1000 veh./day)	Reference
35	8-50	(25)
2500	50+	
60	Local Street	(3)
120	Freeway	
100	0.5+	(26)
10	<10	(27)
25	10-15	
45	15-20	
80	20-30	
115	30-40	
135	40-50	
150	50-60	

The setback distances given by Equation 19 are meant to be applied to major existing sources whose projected contributions in future years will be obtained through the use of a dispersion model. Setback distances from local streets and other minor CO sources should be roughly equivalent to the distribution of similarly sized sources along the corridor being modeled. An absolute minimum setback distance of 10 m should be maintained in all cases(27). This provides some degree of protection against an unrepresentative emissions event occurring close enough to a sampling inlet to significantly bias results.

In cases where the recommended minimum setback given in Equation 19 cannot be achieved, the observed maximum 1-hour and 8-hour background concentrations can be adjusted by deducting modeled contributions from the existing facility. Care should be taken to collect wind speed, wind direction and traffic volume data concurrently if this adjustment is to be made. Sampling on both sides of the existing corridor and filtering the data for upwind concentrations is not recommended. This is because observed maximums will normally occur during stagnant conditions when emissions tend to drift to both sides of the corridor.

Proper exposure of the sampling inlet is the final factor to consider when establishing either a background or corridor sampling site. The inlet should be exposed to free air flow for an arc of at least 270°(27). Walls, dense foliage and solid fences are considered obstructions when making this determination. Chain link fence and similarly porous objects are not. Where possible, setback distances from significant obstructions should be at least three times the lesser of the height or width of the obstruction. In cases where a sampler is located downwind of an obstruction during recurring morning or evening wind regimes (e.g., upslope/downslope winds or land/sea breezes), this factor should be increased to 10. As mentioned earlier, the minimum setback distance from any CO source is 10 m. Sources such as motorcycles, emergency generators, idling lawnmowers, etc., should be considered in addition to automobiles when making this judgment. Sampling inlet height should be compatible with the "nose height" of 1.5 to 2 m normally used in air quality impact analyses. Exceptions to this can be made based upon the objectives of the monitoring program.

7.3 Quality Assurance Program

Nondispersive infrared analyzers are used by Caltrans to determine bag sample CO concentrations. Careful site selection and sample scheduling will yield reliable results only when these instruments are maintained and operated properly. To accomplish this, TransLab has instituted a rigorous quality assurance program. This program recommends procedures and schedules for instrument calibration, operator training and record keeping. It also includes an annual interlaboratory audit using "unknown" air samples, and an annual field review of procedures and equipment. TransLab's Air Quality Unit administers the program, but the districts have the ultimate responsibility for making use of the services offered. Details of the program are given in Reference 28.

7.4 Scheduling and Duration of Sampling

Scheduling and duration of sampling are the key elements in the modified observed maximum method. They are used to minimize bias and to assure a reasonable probability of encountering a maximum value equal or near the second seasonal maximum. Sample scheduling determines the probability, p_{jk}, of encountering a seasonal maximum given the jth month and the kth day-of-week sampling plan. Sampling duration determines the probability of encountering one or more seasonal maximums given p_{jk}. If the probability, $P(r \geq 1 | p_{jk})$, equals the design probability, P_c, then the observed maximum represents an unbiased estimate of the expected second seasonal maximum.

To facilitate selection and design of sampling plans, a table listing values of $P(r=0|p_{jk})=1-P(r \geq 1|p_{jk})$ for one week periods as a function of month and days sampled during the week was constructed. The following simplified form of the binomial distribution for r=0 was used to compute these probabilities:

$$P(r=0|p_{jk},n_k) = (1-p_{jk})^{n_k} , \qquad (20)$$

where n_k = number of days sampled/week.

Values of p_j taken from Figure 1 were modified according to

$$p_{jk} = p_j \cdot \frac{7 \sum_{i=1}^{7} D_{ik} \cdot \bar{f}_i}{\sum_{i=1}^{7} D_{ik}} , \qquad (21)$$

where $\bar{f}_i$ equals the average probability (from Figure 2) of encountering a seasonal maximum on the ith day, and $D_{ik}=1$ if the ith day is included in the kth day-of-week sampling plan or $D_{ik}=0$ if it is not. Equation 21 is simply a means for accounting for the significant difference in the distribution of seasonal maximums by day-of-week shown in Figure 2. Eight day-of-week sampling plans were considered. The results are given in Table 8.

To use Table 8, one simply selects the entry or entries for the month(s) and day-of-week sampling plan(s) being considered. Treating each probability as independent, the combined probability of encountering zero seasonal maximums over a given sampling period will be the product of the individual probabilities taken from Table 8. For instance, if a proposed sampling plan calls for three weeks of M-F sampling in December followed by two weeks of Tu-F sampling and one week of M-Sa sampling in January, the combined probability of encountering zero seasonal maximums would be given by

$$P(r=0) = (0.49)^3 \cdot (0.84)^2 \cdot (0.79) = 0.07.$$

The criteria for accepting a proposed sampling plan is

$$\prod_{j,k} P(r=0|p_{jk})^{w_{jk}} \simeq 1-P_c \ , \qquad (22)$$

where w_{jk} equals the number of weeks the kth day of week sampling plan will be repeated in the jth month. Since P_c=0.93, the example plan cited above meets the criteria. If fewer weeks had been proposed, the observed maximum would have been a negatively biased estimate of the second annual maximum. On the other hand, if the M-F sampling had been expanded to M-Sa, then the observed maximum would have been a positively biased estimate. Using Table 8, a field supervisor can choose a sampling plan that will yield as an observed maximum a relatively unbiased estimate of the expected second seasonal maximum. If the need arises, a prearranged plan can even be changed midstream and still

TABLE 8. PROBABILITY OF ENCOUNTERING ZERO SEASONAL MAXIMUMS IN A ONE WEEK SAMPLING PERIOD BY MONTH AND DAY OF WEEK SAMPLING PLAN

Days Sampled	$P(r=0 \mid p_{jk})$ Oct.	Nov.	Dec.	Jan.	Feb.
M-W	0.94	0.80	0.68	0.89	0.97
Tu-Th	0.93	0.79	0.66	0.89	0.97
W-F	0.92	0.76	0.62	0.87	0.97
M-Th	0.91	0.74	0.59	0.86	0.96
Tu-F	0.90	0.71	0.55	0.84	0.96
M-F	0.88	0.66	0.49	0.81	0.95
M-Sa	0.87	0.63	0.44	0.79	0.94
M-Su	0.86	0.61	0.43	0.78	0.94

meet the criteria stated in Equation 22. After sampling is concluded and the data checked for outliers, the 1-hour and 8-hour observed maximums by time period can be considered accurate estimates of their respective second annual maximums.

7.5 Rejection of Outliers

When a value has been identified as an outlier, a decision must be made by the project manager to either retain the value in the data set, or to replace it with an interpolated or "best judgment" value. Identification of an outlier by one or more of the tests described in Section 6 is not alone sufficient cause for rejection of the value from the analysis. The Gap, Grubbs and Time Sequence Tests are simply meant to point out anomalous values to the analyst. Final rejection of an outlier should come only after careful consideration of the possible, likely and sometimes provable reasons for rejection.

An anomalous reading should be examined in terms of the following four sources of systematic error: transient source, contaminated sample, instrument malfunction and transcription error. Any other sources of error capable of generating the observed inconsistency in the data should also be examined. The personnel responsible for the field sampling and laboratory measurements should be consulted for their input regarding possible sources of error. Comparison with concurrent measurements from companion samplers or nearby air monitoring stations can also be made, if available.

Absolute proof is not needed to reject an outlier. A likely explanation combined with the statistical improbability is sufficient cause for rejection. In either case, a basis for rejection should be established and documented in the working files of the project.

7.6 Rollback Calculations

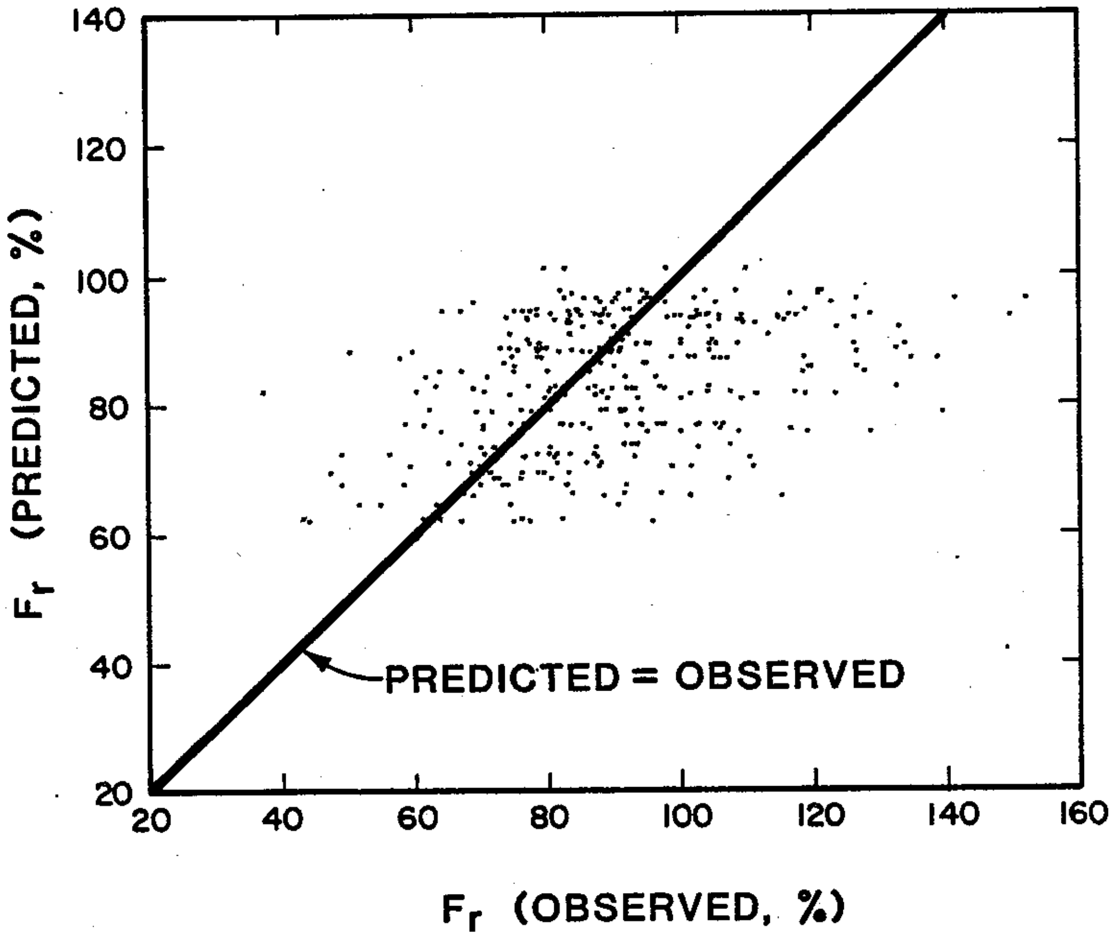
Estimates of second seasonal maximums obtained from either background or corridor sampling programs are valid only for the season sampled. Random variability in meteorology, long-term trends in emissions and other factors will produce different results from one year to the next. Normally, values obtained for a base year must be projected forward to some future year for purposes of comparison to standards over the life of a project. While random year-to-year variations can only be assumed to average out over time, predictable trends can be accounted for. This is done by linearly scaling base year concentrations forward to future years by means of a rollback factor, F_r.

The rollback factor is normally taken as a ratio of expected future year emissions to estimated base year emissions. In cases where detailed emission inventories and projections are available, F_r can be determined directly from these data files. Since mobile emissions are by far the most important source of CO in urban areas, anticipated increases in vehicle miles traveled (VMT) and decreases in vehicle emission factors (EF) are often used to estimate F_r in the absence of areawide emissions inventories.

Use of the linear rollback method assumes that spatial distribution of emissions and average meteorological conditions remain constant from year to year. This is a reasonable assumption to apply to average annual concentrations. However, Georgopoulos and Sienfeld have pointed out that this type of linear scaling is valid for maximum values only if the base year and future year distributions are lognormal with equal geometric standard deviations(11). Furthermore, because seasonal maximums are the result of infrequent combinations of peak emission rates and stagnant meteorology, they must be considered unique events. The magnitude and incidence of such events is quite variable from year-to-year. Rollback factors estimated from averaged projections of VMT and EF may not accurately characterize the ratio between base and future year emissions under the unique conditions likely to produce a seasonal maximum.

The California data set was used to investigate the validity of applying linear rollback to seasonal maximums. Observed values for F_r were computed using 8-hour second seasonal maximums from the 1971/72 through 1981/82 seasons. All possible combinations for F_r were computed for each station (55 values for a station with all 11 second seasonal maximums). Expected values were determined using the California emission factor program, EMFAC6D(29), and measured trends in state highway travel(30). Figure 4 is a graph of the expected versus observed values of F_r.

The considerable scatter of results in Figure 4 can, in part, be attributed to season-to-season fluctuations in meteorology. The fact that nearly half of the observed values for F_r fall above 100% indicates that these fluctuations tend to overshadow any underlying trends in average emissions. A station-by-station examination of seasonal maximums for the 11 season period corroborated this finding by failing to turn up any clear-cut trends in the data.



PREDICTED VERSUS OBSERVED ROLLBACK FACTORS, F_r, FOR THE CALIFORNIA DATA SET

FIGURE 4

Seasonal maximums are the result of a complex combination of numerous, sometimes interrelated factors. While trends in seasonal maximums can be found in the data over long periods of time, especially over the period from the 1960s to the 1970s, the erratic season-to-season pattern of results contained in the California data set indicates that most subtle changes in VMT and EF will not significantly change a seasonal maximum. This is particularly true given the leveling off of mandated vehicle emissions reduction programs and the recommended siting of samplers in areas representative of projected VMT.

In light of these findings, it is recommended that the use of rollback factors be discontinued for observed maximums from background monitoring sites. For corridor sampling programs, observed maximums may continue to be adjusted using the rollback method, provided that base and future year estimates of EF and VMT are consistent with both corridor-specific projections and worst case meteorological conditions. This would normally mean separate estimates of F_r for 1-hour and 8-hour applications.

7.7 Data Storage and Retrieval

The data collected during a background or corridor sampling program may well have application beyond its immediate use as a project-level air quality study. Other agencies may want to make use of the data, or Caltrans may need to reexamine it in the future. Since the data are in their most usable form after being entered into the computer and edited, the most logical place to store them is in the computer. Hard copy stored in file cabinets or boxes is appropriate only as a backup to the computerized files.

The Air Quality Data Handling System (AQDHS) is designed for the long-term storage of aerometric data. The data are stored in a format developed by the Environmental Protection Agency(31) in which uniform codes are used to specify project, location, type of measurement, test method and units. Each record contains data for a 12-hour period. A series of service programs is available to edit and update the AQDHS files and to provide monthly summary reports and a statistical analysis of the data(32).

It is strongly recommended that data from background or corridor sampling programs be entered into AQDHS. Instructions on how to do this are contained in Section 9 of this report.

8. IMPLEMENTATION

The recommended procedures contained in Section 7 will be implemented within Caltrans by the distribution of this report, the availability of an accompanying computer program, and the scheduling of a training course for district personnel. The computer program, OBSMAX, is available on the CMS (Conversational Monitoring System) time-share system. A complete set of user instructions is contained in Section 9 of this report. A training course covering this material and other updated microscale air quality modeling topics is tentatively scheduled for 1984.

Development of the sampling criteria specified by Table 8 and Equation 22 was presented in general form so that it might be implemented by other organizations as well. For locations where the monthly distribution of seasonal maximums differs significantly from the 12-station California data set, a more appropriate version of Table 8 can be constructed from local aerometric data by using the same design equations. In cases where the duration of sampling falls short of the recommended period because of time, funding or staff limitations, the probability of encountering the second or higher seasonal maximum can still be determined using Equation 8, and the observed maximum can be adjusted for bias by using Equation 9. Proposed revisions to the NAAQS for CO increasing the allowable number of exceedances to five per year(18) can also be accommodated by shifting the reference rank in Equation 9 from the second to the sixth seasonal maximum, and modifying Equation 6 for use with either the second or third observed maximum. If these proposed revisions are adopted, appropriate modifications to Table 8 and the OBSMAX program will be made and distributed to recipients of this report.

9. USER INSTRUCTIONS: "OBSMAX"

9.1 Introduction

OBSMAX analyzes ambient CO data according to the modified NCHRP 200 methodology described in this report. Concentrations measured from hourly bag samples taken during the CO season (October through February) serve as input to the program. The input data file is examined for outliers and missing data. Interpolated values are provided by the program for gaps of two hours or less.

The input data file is divided into four time periods: morning, midday, evening and nocturnal. The program reports the highest 1-hour and 8-hour concentrations contained in the input file for each time period, in addition to an overall daily maximum.

Output from the program consists of:

1. A chart summarizing the 1-hour and 8-hour observed maximums for each time period,

2. A brief summary of values identified as outliers,

3. A set of graphs for each time period that show ranked daily 1-hour and 8-hour maximums and detailed temporal plots of days on which observed maximums occur.

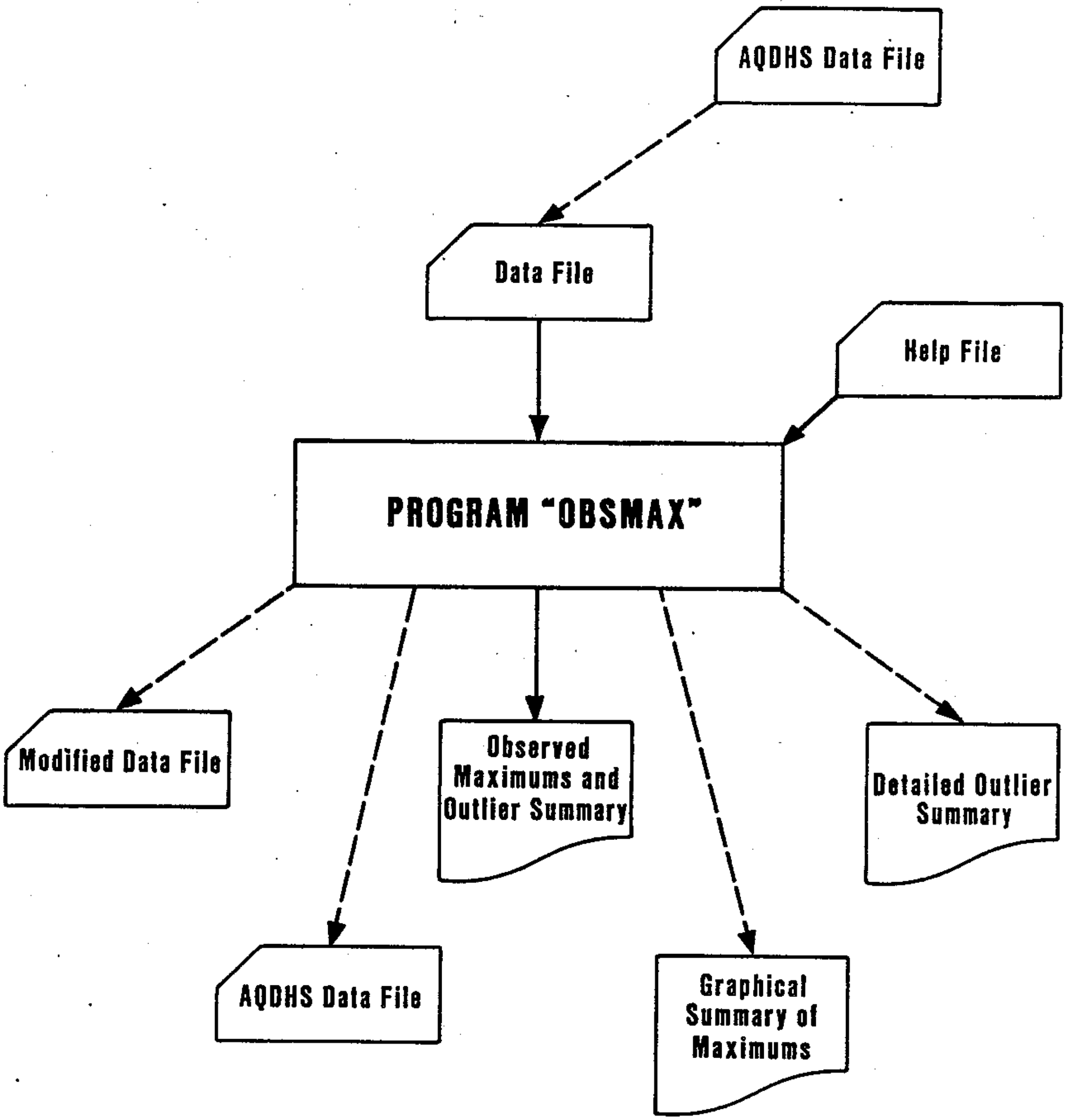
The program can accept data files from either a CMS (timeshare) account, or AQDHS. After an initial outlier analysis, the user may substitute estimated values for the indicated outliers or let the program supply interpolated values. The edited data file may then be saved in the user's CMS account. It may also be sent to AQDHS for permanent storage.

The program is interactive. The user will be asked to supply information and to make choices while the program is running. A prompt (>) will be printed each time the program is ready to receive input. Each prompt is preceded by a question or a statement. A general flowchart for the program is shown in Figure 5.

9.2 Input Files

Input files should consist of hourly CO readings in ppm chronologically ordered by day. The program can handle files containing up to 90 days of CO data. Single hour designations used for input or output will always refer to the start hour of the sample in terms of a 24-hour clock. For example, a 1700-hour reading will refer to a sample collected between 5 and 6 p.m. Times are to be recorded in Local Standard Time. Individual measurements may be input as either whole numbers or decimals rounded to the nearest .5 ppm. Values are rounded when they are graphed so that additional accuracy in the input file is not preserved in the output.

There are two input file formats acceptable to the program. Format A should be used for 24-hour data. Format B is better suited for partial blocks of data as might be collected by a single 12- or 16-hour bag sampler.



FLOW CHART FOR THE "OBSMAX" COMPUTER PROGRAM (DASHED LINE REPRESENTS OPTIONAL ACTIVITY)

FIGURE 5

Format A:

field	information
1	month number
2	day number
3	2-digit year number
4-27	24-hourly measurements

Each field should be separated from the next by a comma. The first value should be for the 12 to 1 a.m. sample, and so on for the 24 hours. Hours for which there are no readings should be entered as -1's.

Example of record Format A:

```
1,3,82,4,3,2,1,1,1,2,3,4,5,7,9,6,4,3,-1,-1,3,2,1,1,1,1,1
month | year
   day                 24 hourly readings
                   (-1's are missing hours)
```

Format B:

field	information
1	month number
2	day number
3	2-digit year number
4	*
5	start hour of first sample
6	start hour of last sample
7-30	values taken during

Each field should be separated from the next by a comma. The asterisk (*) in the fourth field indicates partial 24-hour data. An asterisk in any other field will generate an error message.

The hour delimiters of the monitoring period are the start hours of the first and last samples. Times should be based on a 24-hour clock with start hours ranging from 0 to 23. For instance, if monitoring began at 7 a.m., the start hour will be 7. If the last reading for the day was taken for the hour 4-5 p.m., the final start hour would be 16. Within the designated monitoring period, missing readings are entered as -1's.

Example of record Format B:

1, 2, 82, *, 7, 14, 1, 3, 2, 5, 4, 3, 3, 1

(1, 3, 2, 5, 4, 3, 3, 1: readings for the hours between 0700 and 1500)

9.3 Running the Program

OBSMAX is available through the Caltrans CMS Library. In order to run it, one must first logon to CMS with the proper account code, password, etc. The following step-by-step instructions assume the logon procedure has been successfully completed. Note that words in capital letters represent program initiated messages unless otherwise indicated.

Step	Instruction

1. Type RUN OBSMAX and press return.

   The system will load the program. This may take awhile, especially if there are numerous users on CMS.

2. MANUALLY SCROLL PAPER TO TOP OF PAGE AND PRESS RETURN.

   Use the carriage control knob to roll the paper so that the top of the printhead is about one-half inch below a perforation. Press return.

3. ENTER DATAFILE NAME (OR HELP OR QUIT).

   Type the name of the data file you wish to have analyzed and press return. If the response HELP is made, a brief description of the required data file format is listed. A response of QUIT will stop program execution.

4. VERIFY: YOUR DATA FILE IS __________? YES/NO.

5. ENTER UP TO 30 CHARACTERS OF JOB DESCRIPTION

   Enter any information you would like to have included in the output page heading. Press return.

Step	Instruction

6. ENTER 3 DIGIT SITE CODE.

   If the data file is to be transferred to AQDHS, a site code properly assigned by your AQDHS district coordinator is entered. If not, any 3 digit number may be entered.

7. CHECKING FOR OUTLIERS...

   At this point the program is checking for outliers. This may take as long as a minute. If no outliers are found, skip to Step 13 of these instructions.

8. THERE ARE ____ OUTLIERS. DO YOU WISH TO EXAMINE THEM? YES/NO.

   If YES is entered, charts showing the neighboring values for each outlier will be printed (see Section 9.4 for more details). A NO response will bypass printing of the outlier charts. Press return.

9. DO YOU WISH TO SUBSTITUTE VALUES FOR THE INTERPOLATED VALUES? YES/NO/QUIT.

   Enter QUIT if you would like to cancel the program run and study the outlier charts acquired in Step 8. Enter NO and proceed to Step 11 of these instructions if you agree with the interpolated values already substituted for the outliers. A YES response will lead to Step 10. Press return.

Step	Instruction

10. ENTER NUMBER OF MONTH, DAY, HOUR, AND SUBSTITUTE VALUE. PRESS RETURN TO STOP.

    Enter the requested information for any value in the data file you wish changed and press return. The Step 10 program message will be repeated. Continue to substitute values one at a time until your substitutions are complete. To deactivate the substitution option, press return without making an entry.

11. DO YOU WISH TO HAVE THE MODIFIED DATA FILE SAVED? YES/NO.

    Enter YES or NO and press return. If you enter NO, go to Step 13.

12. ENTER NEW FILE NAME. IT MAY BE UP TO 8 CHARACTERS LONG.

    Enter the name under which you want the new file saved. If you enter the original file name, the modifications will be saved as part of the original file. Press return.

Step	I n s t r u c t i o n

13. DO YOU WISH TO HAVE THIS DATA SENT TO AQDHS? ENTER YES OR NO.

    Enter YES or NO and press return. If you enter NO, go to Step 17. Details of the AQDHS option are discussed in Section 9.4.

14. ENTER 4 DIGIT AREA CODE.

    Enter area code of the sampling site and press return. AQDHS area codes are given in Appendix A.

15. ENTER 2 DIGIT PROJECT CODE.

    Enter project code obtained from district AQDHS coordinator and press return.

16. CAN'T FIND YOUR JOBCARD FILE.
    PLEASE ENTER THE FOLLOWING INFORMATION:

    TMS CODE?
    MSP CODE?
    LABOR COST CODE?
    YOUR NAME?

    When data are sent to AQDHS, the program uses a file called JOBCARD MVS to write the necessary JCL. If the jobcard file has not previously been created, the program does so interactively. Obtain the TMS, MSP and Labor Cost codes from your EDP manager or AQDHS coordinator and enter them accordingly. Your name is used to label the output and should be no more than 20 alphanumeric characters.

17. WORKING...

The program is determining the observed maximums. This may take up to 2 minutes during peak CMS usage periods.

The program will page and print a summary of the observed maximums. It will then page again and print response codes for various graphing options.

18. GRAPHS.

Enter the code number(s) of the graphs you want printed separated by commas (further details are given in Section 9.5).

The program will print the requested graphs.

Program execution ends.

9.4 Program Options

9.4.1 Outlier Correction Option

Since the accuracy of the information generated by this program depends on the validity of the input data, the input file is checked for anomalous values. The methods used to identify these outliers are discussed in Section 6. Occasionally you may want to replace interpolated values substituted by the program for identified outliers with your own estimate or with the outlier itself. Consequently, you have the option of substituting values into the data file in place of the interpolated values.

If any outliers are found, they are replaced by interpolated values at the beginning of the program. The program then stops and asks if you would like to examine the outliers. If you enter yes, charts similar to the following will be printed.

AMBIENT AIR QUALITY SUMMARY PAGE: 1
OBSERVED MAXIMUM ANALYSIS
POLLUTANT: CARBON MONOXIDE

JOB DESCRIPTION: DOCUMENTATION
SITE CODE: 123
RUN DATE: 83/11/17 DATA FILE: CODATA

SUMMARY OF OUTLIERS:

	12/5/82			12/15/82	
TIME	INPUT VALUES	INTERPOLATED VALUE	TIME	INPUT VALUES	INTERPOLATED VALUE
1900	3		0200	4	
2000	2		0300	4	
2100	3		0400	4	
2200	3		0500	2	
2300	2		0600	1	
0000	*18	1	0700	*17	1
0100	1		0800	1	
0200	1		0900	1	
0300	1		1000	1	
0400	1		1100	1	
0500	1		1200	1	

The asterisked value is the outlier. The interpolated value which has replaced it in the data set is shown next to the outlier. The date of the outlier is printed above the chart.

You have the option of substituting your own estimate for an interpolated value. If you do not substitute, the interpolated value will be used in the observed maximum determinations. If you do wish to substitute values, answer 'yes' to the question DO YOU WISH TO SUBSTITUTE VALUES FOR THE INTERPOLATED VALUES? The program will prompt you to enter the number of the month, day, hour and the substitute value. If you wish to have the program run with the original data value, you will have to substitute the original value for the interpolated value.

When you are offered the option to substitute values, you will also have the option to quit. This provides you with an opportunity to study the outliers at your leisure. When you return to the program, run it from the beginning. When you are asked if you wish to see the outlier charts respond 'no'. You will then be offered the chance to substitute values for the interpolated values. If you decide not to substitute, respond 'no' and continue with the program. If you decide to substitute values, respond 'yes' and you will be given the opportunity to do so.

9.4.2 Modified Data File Option

The modified data file that includes the interpolated or substituted values may be saved in your account. If you wish to save the modified file, you should enter 'yes' in response to 'DO YOU WISH TO HAVE THE MODIFIED DATA FILE SAVED?'. You will then be asked to enter a name for the new file. If you enter a name which is unique in your account, the new file will be saved under that name with a file type of WDATA. If you enter a name which already belongs to a file in your account, the modified file will replace the existing file. Consequently, if you want to have both the existing file and the modified file, you should take care to save the modified file with a unique name.

9.4.3 Retrieval of AQDHS Files

The Air Quality Data Handling System (AQDHS) is a data base designed to facilitate organization, storage and retrieval of air quality and meteorological data. It provides a data base from which researchers can access a large volume of data collected from a variety of sources. The OBSMAX program may be run on files that have been retrieved from AQDHS, or may be used to store newly entered data into AQDHS.

To retrieve CO data from AQDHS, run the program GETCO. The GETCO program has been designed to retrieve existing CO data from AQDHS in a format which can be readily handled by the OBSMAX program. GETCO requires the following information:

1. Either:

   A file in your CMS account called 'JOBCARD MVS' which contains a valid MVS jobcard.

   Or the necessary information for the GETCO program to create a jobcard file for you. You will need a TMS code, MSP code and a Labor Cost code.

2. Retrieval information.
You will need the area code, site code, and starting and ending dates (year, month and day) for the requested data.

To run GETCO:

a.) Logon to your CMS account.

b.) From CMS type RUN GETCO.

c.) The program will first check for a JOBCARD MVS file. If the file is found, the program will read in the jobcard information, display it, and ask if the information is OK. If you answer 'YES', the program will proceed to the retrieval information.

Example:

```
run getco
                    *** GETCO ***
                     Version 1.Ø
     For retrieving carbon monoxide data (CO) from
     the Air Quality Data Handling System (AQDHS).
------------------------------------------------------------
83/12/Ø5          12:14:14          TMLDICK
------------------------------------------------------------
Enter ABORT at any time.
------------------------------------------------------------
Reading your jobcard file.
------------------------------------------------------------
JOBCARD:
//TMLDICK  JOB (TM1ØØP,TMMW,TMPJR,V27ØL),'DICK WOOD'
------------------------------------------------------------
OK (Y/N) ? Y
```

If the program does not find a JOBCARD MVS file or if you answer 'NO' when asked if the jobcard information is OK, the program will prompt you for the necessary information and write a JOBCARD MVS file in your account.

Example:

```
run getco
                    *** GETCO ***
                     Version 1.Ø
     For retrieving carbon monoxide data (CO) from
     the Air Quality Data Handling System (AQDHS).
-----------------------------------------------------------
83/12/Ø6          Ø8:Ø8:56          TMLNOISE
-----------------------------------------------------------
Enter ABORT at any time.
-----------------------------------------------------------
Can't find your jobcard file.

*** JOBCARD DATA ***
Enter the following information:

TMS code ? tmmw
MSP code ? tmpjr
LABOR COST code ? v27Ø1
Your NAME ? dick wood
-----------------------------------------------------------
JOBCARD:
//TMLNOISE JOB (TM1ØØP,TMPJR,V27ØL),'DICK WOOD'
-----------------------------------------------------------
OK (Y/N) ? Y
  'JOBCARD MVS A'
  Has been written to your account.
```

d.) After the jobcard information has been supplied, the program will proceed to ask for retrieval parameters. You will be prompted for the area code, site code and starting and ending dates. The area code must be less than 10000, the site code less than 1000, and both codes greater than 0 (see Appendix A for area code designations). The starting and ending dates are input as 6 digit numbers in the form yymmdd. The starting and ending dates do not need to fall on valid AQDHS data points, and requested data may contain gaps. The AQDHS file is simply searched for any data which fall between the starting and ending dates, inclusive. The program will check to insure that the ending date is greater than the starting date. Once the retrieval parameters are entered, the program will display the data and ask if it's OK.

Example:

```
*** RETRIEVAL PARAMETERS ***
Enter AREA code ? 52Ø
Enter SITE code ? 24
Enter START DATE (yymmdd) ? 77Ø215
Enter  END  DATE (yymmdd) ? 77Ø311

RETRIEVAL PARAMETERS
--------------------
AREA code = Ø52Ø            SITE code = Ø24
-------------------------------------------------------
From FEB 15, 1977        To    MAR 11, 1977
-------------------------------------------------------
OK (Y/N) ? Y
```

e.) Once the retrieval parameters have been entered, the program will submit the necessary JCL file to the MVS system and return you to CMS. You should now get two messages from the system; one telling you that the file has been sent to MVS, and one telling you that the file has been received. Then after a few minutes wait, you should receive a series of messages telling you that two files have been 'spooled' to your account. If you have logged off the system and returned, the status of these files can be checked by entering a query files command from CMS (Q FILES).

Example:

```
        JOB SUBMITTED.

        Two files will be sent to your account.
        The 'rdr' file will contain your CO data.
        The 'listing' file will contain the system messages.
        Please do not delete the listing file until you are
        sure your job has run successfully.

        BYE-BYE

R;
 DMTNCM147I SENT FILE 6245 (6245) ON LINK TEALE TO TEALE JOB
 DMTRGX170I FROM TEALE: IAT9140 JOB RSCS6245 ADDED TO JES3 JOB QUEUE

 DMTRGX170I FROM TEALE: TMLDICK          OUTPUT QUEUED FOR TRANSMISSION
PRT FILE 6246 FROM RSCS    COPY 001   NOHOLD
 DMTAXM104I FILE (6245) SPOOLED TO TMLDICK -- ORG TEALE (TMLDICK) 12/05/83
PUN FILE 6247 FROM RSCS    COPY 001   NOHOLD
 CMTAXM104I FILE (6245) SPOOLED TO TMLDICK -- ORG TEALE (TMLDICK) 12/05/83
```

f.) After the two files have been returned to your account from MVS they may be loaded onto your 'A' disk with a recfiles command. The first file is a 'listing' file which contains the JCL and system messages from MVS. This file should be deleted after you have ascertained that the proper CO data have been retrieved from the AQDHS files. The second file is a cardreader ('rdr') file which should contain your CO data.

Example:

```
recfiles
THE TAG FOR PRINTER FILE 6246 WAS:
FILE (6245) ORIGIN TEALE    TMLDICK  12/Ø5/83 12:15:57 P.S.T.
TEMP1 LISTING A1
TEMP2 RDR A1
READER IS EMPY
2 FILES LOADED ON DISK 'A'
R;
```

9.4.4 AQDHS Transfer Option

The purpose of AQDHS is to store aerometric data efficiently and economically for long periods of time. A data file that was initially created on CMS for direct input to OBSMAX can be automatically transferred to AQDHS by using the AQDHS Transfer Option. This option is activated by answering YES to the question 'DO YOU WISH TO HAVE THIS DATA SENT TO AQDHS?'. If you type YES and press return, you will be asked to specify the area and project code for the sampling site. It is important that these codes, as well as the site code, be unique. They are the parameters by which data will be cataloged. A table of California area codes can be found in Appendix A. Project and site codes should be obtained from an AQDHS district coordinator responsible for maintaining a record of project and site codes used in your district. With the entry of the proper codes, OBSMAX will transfer a copy of your modified input file (including interpolated or substituted values) directly to AQDHS.

9.5 Program Output

The following is an example of the summary page of output from OBSMAX:

AMBIENT AIR QUALITY SUMMARY PAGE: 2
OBSERVED MAXIMUM ANALYSIS
POLLUTANT: CARBON MONOXIDE

JOB DESCRIPTION: DOCUMENTATION
SITE CODE: 123
RUN DATE: 83/11/17 DATA FILE: CODATA

OBSERVED MAXIMUMS:

TYPE	AVG.TIME (HRS.)	DAYS USED	VALUE (PPM)		DATE	TIME
DAILY	1	30	11.0		12/12/82	1800-1900
	8	30	8.6		12/21/82	1600-0000
MORNING	1	29	10.0		12/09/82	0700-0800
	8	23 $	7.0	(I)	11/27/82	0500-1300
MIDDAY	1	29	8.0		12/12/82	1600-1700
	8	29	5.5		12/12/82	1000-1800
EVENING	1	24	11.0		12/12/82	1800-1900
	8	23	8.0		12/12/82	1500-2300
NOCTURNAL	1	23	10.0		12/01/82	0000-0100
	8	22 $	6.6		12/08/82	2000-0400

(I) CONTAINS INTERPOLATED OR SUBSTITUTED VALUE
$=INSUFFICIENT NUMBER OF DAYS

OUTLIER SUMMARY:

VALUE (PPM)	DATE	TIME
18	12/05/82	0000-0100
17	12/15/82	0700-0800

This page gives a summary of both observed maximums by time period, and outliers encountered. Detailed definitions for each output parameter are given in the following list.

TYPE: Refers to the time-of-day period of the start hour of the observed maximum (see Table 2).

AVG. TIME: Indicates whether observed maximum is 1-hour or 8-hour average.

DAYS USED: Refers to the number of qualified days from which the corresponding observed maximum was chosen.

For a day to qualify, all hourly measurements (or properly interpolated values) for the time period must exist. A $ indicates that the number of days are insufficient to meet the design criteria given in Section 7.4.

VALUE: The maximum 1-hour and 8-hour CO concentrations in ppm for the indicated time periods are given in this column. An asterisk signifies the observed maximum over all time periods. An (I) indicates that there was at least one interpolated value during the time period in which the observed maximum occurred. This interpolated value may or may not have been used in the determination of the observed maximum.

DATE: Indicates the date during which the maximum occurred. For 8-hour maximums, the date of the start hour is given.

TIME: Indicates the range of time during which the observed maximum occurred.

OUTLIER SUMMARY: The start hour, date and range of time of any outliers detected are given here. More detailed information on outliers may be obtained prior to printout of the observed maximum summary by exercising the Outlier Correction Option (see Section 9.4.1).

An example of the next page of output is shown below.

```
AMBIENT AIR QUALITY SUMMARY             PAGE:  3
OBSERVED MAXIMUM ANALYSIS
POLLUTANT: CARBON MONOXIDE

JOB DESCRIPTION: DOCUMENTATION
SITE CODE: 123
RUN DATE: 83/11/17           DATA FILE: CODATA

WHICH OUTPUT GRAPH(S) WOULD YOU LIKE TO SEE?
ENTER NUMBER(S) SEPARATED BY COMMAS.

0 = NONE
1 = DAILY
2 = MORNING
3 = MIDDAY
4 = EVENING
5 = NOCTURNAL
6 = ALL

GRAPHS?      >6
```

The code numbers on this page are used to obtain detailed graphs of the observed maximums. A brief description of each choice is listed below.

NONE: Ends program execution

DAILY: Graphs 1) ranked daily maximums for the complete data set, and 2) readings for the 24-hour period within which the 1-hour and 8-hour observed maximums occurred.

MORNING
MIDDAY
EVENING
NOCTURNAL: For the respective time period, graphs 1) ranked daily maximums, and 2) readings for the 24-hour period within which the 1-hour and 8-hour observed maximums occurred.

ALL: Prints all graphs previously described.

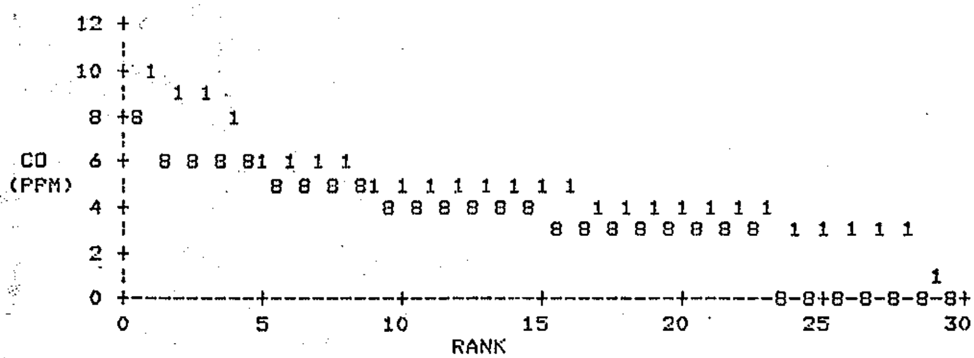
An example of the graphical output available is shown below.

AMBIENT AIR QUALITY SUMMARY   PAGE: 5
OBSERVED MAXIMUM ANALYSIS
POLLUTANT: CARBON MONOXIDE

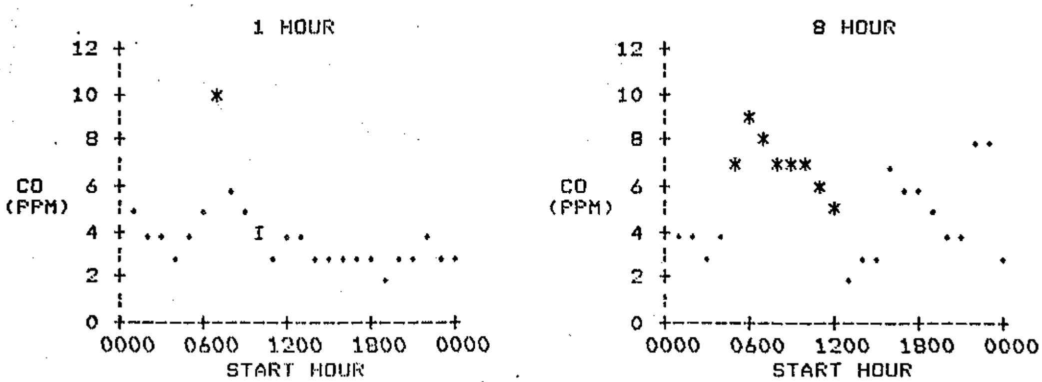
JOB DESCRIPTION: DOCUMENTATION
SITE CODE: 123
RUN DATE: 83/11/17   DATA FILE: CODATA

GRAPHICAL SUMMARY OF MORNING MAXIMUMS

RANKED RESULTS:



OBSERVED MAXIMUMS (*)
(I= INTERPOLATED VALUE)



The graph of 1-hour and 8-hour ranked results is printed first. Only the top 30 daily maximums will appear on the graph. Values are rounded to the nearest whole ppm. The "1" and "8" symbols are offset one space to avoid overlap.

The lower two graphs give the detailed distribution of hourly measurements made in the same 24-hour period as the 1-hour and 8-hour observed maximum. Again, results are rounded to the nearest ppm. The actual observed maximums are denoted by asterisks. Interpolated values are indicated by an "I". All other hourly measurements are denoted by a ".". The time axis will automatically center on either noon (1200) or midnight (0000) to best center the observed maximum measurements.

REFERENCES

1. R. I. Larsen. A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards. U.S. Environmental Protection Agency, AP-89, November 1971.

2. N. D. Singpurwalla. Extreme Values from a Lognormal Law With Applications to Air Pollution Problems. Technometrics, Vol. 14, August 1972, pp. 703-711.

3. P. D. Allen, A. J. Ranzieri and E. C. Shirley. Application of Larsen's Mathematical Model For Air Quality Studies, Transportation Laboratory, California DOT, TL-7082-12-76-25, 1976.

4. R. I. Larsen. An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions - Part 2. J. Air Pollution Control Association, Vol. 24, June 1974, pp. 551-558.

5. R. I. Larsen. An Air Quality Data Analysis System for Interrelating Effects, Standards, and Needed Source Reductions: Part 4. A Three-Parameter Averaging-Time Model. J. Air Pollution Control Association, Vol. 27, May 1977, pp. 454-459.

6. D. T. Mage and W. R. Ott. Refinements of the Lognormal Model for Analysis of Aerometric Data. J. Air Pollution Control Association, Vol. 28, August 1978, pp. 796-798.

7. N. Barbalic. Mutual Dependencies Between the Characteristic Parameters of the Lognormal Mathematical Model of Air Pollution Concentration. Atmospheric Environment, Vol. 15, 1981, pp. 1691-1698.

8. P. G. Surman, R. W. Simpson and J. Stokoe. Application of Larsen's Model to Brisbane, Australia, Atmospheric Environment, Vol. 16, 1982, pp. 2609-2614.

9. D. M. Holland and T. Fitz-Simons. Fitting Statistical Distributions to Air Quality Data By The Maximum Likelihood Method, Atmospheric Environment, Vol. 16, 1982, pp. 1071-1076.

10. T. C. Curran, and N. H. Frank. Assessing The Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations, 68th Annual Meeting of the Air Pollution Control Association, Boston, 1975.

11. P. G. Georgopoulos and J. H. Seinfeld. Statistical Distributions of Air Pollutant Concentrations, Environmental Science and Technology, Vol. 16, July 1982, pp. 401A-416A.

12. E. M. Roberts. Review of Statistics of Extreme Values with Applications to Air Quality Data - Parts 1 and 2. J. Air Pollution Control Association, Vol. 29, June/July 1979, pp. 632-637/733-740.

13. N. R. Patel. Comment on a New Mathematical Model of Air Pollution Concentration. J. Air Pollution Control Association, Vol. 23, April 1973, pp. 291-292.

14. H. E. Neustadter and S. M. Sidik. On Evaluating Compliance With Air Pollution Levels Not to be Exceeded More than Once a Year. J. Air Pollution Control Association, Vol. 24, June 1974, pp. 559-563.

15. J. Horowitz and S. Barakat. Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Non-Stationarity. Atmospheric Environment, Vol. 13, 1979, pp. 811-818.

16. W. S. Meisel and T. E. Dushane. Monitoring Carbon Monoxide Concentrations in Urban Areas. TRB, NCHRP 200, April 1979.

17. E. E. Remsberg, J. J. Buglia and G. E. Woodbury. The Nocturnal Inversion and Its Effect on the Dispersion of Carbon Monoxide at Ground Level in Hampton, Virginia. Atmospheric Environment, Vol. 13, 1979, pp. 443-447.

18. A. Samberg and J. Elston. The Proposed Revision to the NAAQS for Carbon Monoxide and Its Effect on State Designation. J. Air Pollution Control Association, Vol. 32, December 1982, pp. 1228-1230.

19. E. C. Shirley. Traffic Information Requirements for Estimates of Highway Impact on Air Quality. California Division of Highways, MR657082S(3)-72-09, 1972.

20. Screening Procedures for Ambient Air Quality Data. U.S. Environmental Protection Agency, EPA-450/2-78-037, July 1978.

21. F. E. Grubbs. Sample Criteria for Testing Outlying Observations. Annals of Mathematical Statistics, Vol. 21, March 1950, pp. 27-58.

22. F. E. Grubbs and G. Beck. Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations. Technometrics, Vol. 14, November 1972, pp. 847-854.

23. V. Barnett and T. Lewis. Outliers in Statistical Data. John Wiley and Sons, New York, 1978.

24. P. E. Benson. CALINE3 - A Versatile Dispersion Model for Predicting Air Pollutant Levels Near Highways and Arterial Streets. Transportation Laboratory, California DOT, FHWA/CA/TL-79/23, November 1979.

25. F. L. Ludwig and J. H. S. Kealoha. Selecting Sites for Carbon Monoxide Monitoring. U.S. Environmental Protection Agency, EPA-450/3-75-077, September 1975.

26. W. R. Ott. Development of Criteria for Siting Air Monitoring Stations. J. Air Pollution Control Association, Vol. 27, June 1977, pp. 543-547.

27. Ambient Air Monitoring and Data Reporting Requirements. U.S. Environmental Protection Agency, Federal Register, Vol. 44, May 10, 1979, p. 27594.

28. K. O. Pinkerman, R. W. Bushey and E. C. Shirley. Air Monitoring Quality Assurance Program. Transportation Laboratory, California DOT, Vols. 1 and 2, FHWA/CA/TL-81/10, June 1982.

29. D. M. Coats, R. W. Bushey and E. C. Shirley. Mobile Source Emission Factors. Transportation Laboratory, California DOT, CA-TL-80/15, May 1980.

30. Traffic Volumes on the California State Highway System. Division of Traffic Engineering, California DOT, 1971-1982.

31. D. H. Fair, J. B. Clements and G. B. Morgan. Storage and Retrieval of Aerometric Data - Parameter Coding Manual. U.S. Environmental Protection Agency, July 1971.

32. J. Faulkinbury and P. E. Benson. User Manual - Air Quality Data Handling System. Transportation Laboratory, California DOT, CA-TL-7288-77-32, November 1977.

APPENDIX A - AQDHS Area Codes for California

05 CALIFORNIA

0040 ALAMEDA
0060 *ALAMEDA
0070 ALAMO-DANVILLE
0080 ALBANY
0100 ALHAMBRA

0130 ALONDRA PARK
0140 *ALPINE
0160 ALTADENA
0180 ALTURAS
0200 ALUM ROCK

0220 *AMADOR
0230 ANAHEIM
0240 ANDERSON
0250 ANGWIN
0260 ANTIOCH

0270 APPLE VALLEY
0275 APTOS
0280 ARCADIA
0300 ARCATA
0320 ARDEN-ARCADE

0340 ARROYO GRANDE
0360 ARTESIA
0380 ARVIN
0390 ASHLAND
0400 ATASCADERO

0420 ATHERTON
0440 ATWATER
0460 AUBURN
0470 AUGUST SCHOOL AREA
0480 AVENAL

0490 AVOCADO HEIGHTS
0500 AZUSA
0520 BAKERSFIELD
0540 BALDWIN PARK
0560 BANNING

0580 BARSTOW
0600 BAYWOOD-LOS OSOS
0610 BEALE EAST
0620 BEAUMONT
0640 BELL

0660 BELLFLOWER
0680 BELL GARDENS
0700 BELMONT
0710 BELVEDERE
0720 BENICIA

0730 BEN LOMOND
0740 BERKELEY
0760 BEVERLY HILLS
0770 BIG BEAR
0780 BISHOP

0790 BLOOMINGTON
0800 BLYTHE
0820 BONNYVIEW
0830 BOYES HOT SPRINGS
0840 BRAWLEY

0860 BREA
0865 BRENTWOOD
0870 BRISBANE
0875 BRODERICK-BRYTE
0880 BUENA PARK

0900 BURBANK
0920 BURLINGAME
0960 *BUTTE
0980 *CALAVERAS
1000 CALEXICO

1020 CALIPATRIA
1030 CAMARILLO
1040 CAMPBELL
1050 CAPITOLA
1060 CARDIFF-BY-THE-SEA

1080 CARLSBAD
1100 CARMEL-BY-THE-SEA
1110 CARMEL VALLEY
1120 CARMICHAEL
1140 CARPINTERIA

1160 CARSON
1180 CASTRO VALLEY
1200 CASTROVILLE
1210 CATHEDRAL
1220 CENTRAL VALLEY

1240 CERES
1245 CERRITOS
1250 CHERRYLAND
1255 CHERRY VALLEY
1260 CHICO

1280 CHICO NORTH
1290 CHICO WEST
1295 CHINA LAKE
1300 CHINO
1320 CHOWCHILLA

1360 CHULA VISTA
1370 CITRUS HEIGHTS
1380 CLAREMONT
1390 CLEARLAKE HIGHLANDS
1400 CLOVERDALE

1420 CLOVIS
1440 COACHELLA
1460 COALINGA
1500 COLTON
1520 COLUSA

*COUNTY DESIGNATION

05 CALIFORNIA

1540 *COLUSA
1560 COMMERCE
1580 COMPTON
1600 CONCORD
1620 *CONTRA COSTA

1640 CORCORAN
1660 CORNING
1680 CORONA
1700 CORONADO
1720 CORTE MADERA

1740 COSTA MESA
1760 COVINA
1780 CRESCENT CITY
1800 CRESENT NORTH
1805 CREST FOREST

1810 CUCAMONGA
1830 CUDAHY
1840 CULVER CITY
1860 CUPERTINO
1865 CUTLER

1870 CYPRESS
1900 DALY CITY
1910 DANA POINT
1940 DAVIS
1950 DEL AIRE

1960 DELANO
1980 DEL MAR
1990 DESERT HOT SPRINGS
2000 *DEL NORTE
2010 DIAMOND BAR

2040 DINUBA
2060 DIXON
2065 DOMINGUEZ
2070 DOS PALOS
2080 DOWNEY

2100 DUARTE
2110 DUBLIN
2120 DUNSMUIR
2140 EARLIMART
2150 EAST COMPTON

2155 EAST LA MIRADA
2160 EAST LOS ANGELES
2170 EAST PALO ALTO
2180 EAST PORTERVILLE
2190 EDWARDS

2220 EL CAJON
2240 EL CENTRO
2260 EL CERRITO
2270 EL ENCANTO HEIGHTS
2290 *EL DORADO

2290 ELK GROVE
2320 EL MONTE
2340 EL PASO DE ROBLES
2360 EL RIO
2380 EL SEGUNDO

2385 ELSINORE
2390 EL TORO
2395 EL TORO STATION
2400 EMERYVILLE
2420 ENCINITAS

2440 ENTERPRISE
2460 ESCONDIDO
2480 EUREKA
2500 EXETER
2520 FAIRFAX

2540 FAIRFIELD
2550 FAIR OAKS
2580 FALLBROOK
2600 FARMERSVILLE
2620 FILLMORE

2630 FIREBAUGH
2640 FLORENCE-GRAHAM
2650 FLORIN
2660 FOLSOM
2680 FONTANA

2700 FORD CITY
2720 FORT BRAGG
2730 FORT IRWIN
2740 FORTUNA
2745 FOSTER

2750 FOUNTAIN VALLEY
2760 FREEDOM
2780 FREMONT
2800 FRESNO
2820 *FRESNO

2840 FULLERTON
2850 GALT
2860 GARDENA
2870 GARDEN ACRES
2880 GARDEN GROVE

2890 GEORGE
2900 GILROY
2920 GLEN AVON
2940 GLENDALE
2960 GLENDORA

2970 GONZALES
2980 *GLENN
2990 GRAND TERRACE
3000 GRASS VALLEY
3010 GREENFIELD

*COUNTY DESIGNATION

3020 GRIDLEY
3030 GROSSMONT-MT HELIX
3040 GROVER CITY
3060 GUADALUPE
3065 GUSTINE

3070 HACIENDA HEIGHTS
3090 HALF MOON BAY
3100 HANFORD
3110 HAWAIIAN GARDENS

3140 HAYWARD
3160 HEALDSBURG
3180 HEMET
3190 HEMET EAST
3200 HERMOSA BEACH

3210 HESPERIA
3230 HIGHLAND
3240 HILLSBOROUGH
3260 HOLLISTER
3280 HOLTVILLE

3290 HOME GARDENS
3300 *HUMBOLDT
3320 HUNTINGTON BEACH
3340 HUNTINGTON PARK
3360 IMPERIAL

3380 *IMPERIAL
3400 IMPERIAL BEACH
3420 INDIO
3440 INGLEWOOD
3460 *INYO

3464 ISLA VISTA
3468 KENSINGTON
3470 KERMAN
3480 *KERN
3500 KING CITY

3520 *KINGS
3530 KINGS CANYON NAT PARK
3540 KINGSBURG
3560 LA CANADA-FLINTRIDGE
3565 LA CRESCENTA-MONTROSE

3570 LADERA HEIGHTS
3580 LAFAYETTE
3600 LAGUNA BEACH
3605 LAGUNA HILLS
3610 LAGUNA NIGUEL

3620 LA HABRA
3640 *LAKE
3645 LAKE ARROWHEAD
3660 LAKELAND VILLAGE
3665 LAKEPORT

3670 LAKESIDE
3680 LAKEWOOD
3700 LA MESA
3710 LA MIRADA
3720 LAMONT

3740 LANCASTER
3750 LA PALMA
3760 LA PUENTE
3780 LARKSPUR
3800 *LASSEN

3810 LASSEN VOLCANIC NAT PARK
3820 LA VERNE
3840 LAWNDALE
3860 LEMON GROVE
3880 LEMOORE

3890 LEMOORE STATION
3900 LENNOX
3910 LENWOOD
3940 LINCOLN
3950 LINCOLN VILLAGE

3960 LINDA
3980 LINDSAY
4000 LIVE OAK
4010 LIVE OAK CITY
4020 LIVERMORE

4030 LIVINGSTON
4040 LODI
4050 LOMA LINDA
4060 LOMITA
4080 LOMPOC

4085 LOMPOC NORTH
4090 LOMPOC NORTHWEST
4100 LONG BEACH
4120 LOS ALAMITOS
4140 LOS ALTOS

4160 LOS ALTOS HILLS

4200 *LOS ANGELES
4220 LOS BANOS
4240 LOS GATOS

4260 LYNWOOD
4280 MC FARLAND
4300 MADERA
4320 *MADERA
4340 MANHATTAN BEACH

4360 MANTECA
4380 MARINA
4400 *MARIN
4420 *MARIPOSA
4440 MARTINEZ

*COUNTY DESIGNATION

05 CALIFORNIA

4480	MARYSVILLE
4490	MATHER
4500	MAYWOOD
4520	MEINERS OAKS-MIRA MONTE
4540	*MENDOCINO
4550	MENDOTA
4560	MENLO PARK
4580	MERCED
4600	*MERCED
4620	MILLBRAE
4640	MILL VALLEY
4660	MILPITAS
4700	MIRA LOMA
4710	MISSION VIEJO
4720	MODESTO
4730	MOJAVE
4740	*MODOC
4760	*MONO
4780	MONROVIA
4800	MONTCLAIR
4820	MONTEBELLO
4840	MONTEREY
4860	*MONTEREY
4880	MONTEREY PARK
4890	MONTE SERENO
4900	MOORPARK
4905	MORADA
4910	MORAGA
4920	MORGAN HILL
4940	MORRO BAY
4960	MOUNTAIN VIEW
4970	MUSCOY
4980	MULBERRY
5000	NAPA
5020	*NAPA
5040	NATIONAL CITY
5060	NEEDLES
5080	*NEVADA
5100	NEWARK
5120	NEWHALL
5130	NEWMAN
5140	NEWPORT BEACH
5150	NIPOMO
5160	NORCO
5170	NORTH FAIR OAKS
5180	NORTH HIGHLANDS
5185	NORTH ISLAND
5240	NORWALK
5260	NOVATO
5280	OAKDALE
5300	OAKLAND
5305	OAK VIEW
5310	OCEANO
5320	OCEANSIDE
5330	OILDALE
5340	OJAI
5360	OLIVEHURST
5380	ONTARIO
5400	OPAL CLIFFS
5420	ORANGE
5440	*ORANGE
5460	ORANGE COVE
5465	ORANGEVALE
5470	ORCUTT
5480	ORINDA
5520	ORLAND
5530	OROSI
5540	OROVILLE
5550	OTAY-CASTLE PARK
5560	OXNARD
5580	PACIFICA
5600	PACIFIC GROVE
5620	PALMDALE
5625	PALMDALE EAST
5630	PALM DESERT
5640	PALM SPRINGS
5660	PALO ALTO
5680	PALOS VERDES ESTATES
5685	PALOS VERDES PENINSULA
5700	PARADISE
5740	PARAMOUNT
5750	PARKWAY-SACRAMENTO SOUTH
5760	PASADENA
5770	PATTERSON
5774	PENDLETON NORTH
5776	PENDLETON SOUTH
5780	PERRIS
5800	PETALUMA
5820	PICO RIVERA
5840	PIEDMONT
5860	PINOLE
5870	PISMO BEACH
5880	PITTSBURG
5920	PLACENTIA
5940	*PLACER
5960	PLACERVILLE
5980	PLEASANT HILL
6000	PLEASANTON
6020	*PLUMAS
6030	POINT MUGU

*COUNTY DESIGNATION

6040 POMONA
6060 PORTERVILLE
6065 PORTERVILLE NORTHWEST
6070 PORTERVILLE WEST
6080 PORT HUENEME

6085 PORTOLA VALLEY
6090 POWAY
6100 QUARTZ HILL
6120 QUINCY
6130 RAMONA

6140 RANCHO CORDOVA
6145 RANCHO RINCONADA
6150 RANCHO SANTA CLARITA
6160 RED BLUFF
6180 REDDING

6200 REDLANDS
6220 REDONDO BEACH
6240 REDWOOD CITY
6260 REEDLEY
6280 RIALTO

6300 RICHMOND
6320 RIDGECREST
6340 RIO DELL
6350 RIO LINDA
6360 RIO VISTA

6370 RIPON
6380 RIVERBANK
6400 RIVERSIDE
6405 ROCKLIN
6410 RODEO

6415 ROHNERT PARK
6416 ROHNERVILLE
6420 *RIVERSIDE
6440 ROLLING HILLS ESTATES
6460 ROSELAND

6480 ROSEMEAD
6500 ROSEVILLE
6520 ROSS
6525 ROSSMOOR
6530 ROWLAND HEIGHTS

6535 RUBIDOUX
6540 RYANS SLOUGH
6560 ST HELENA
6580 SACRAMENTO
6600 *SACRAMENTO

6620 SALINAS
6640 SAN ANSELMO
6660 *SAN BENITO
6680 SAN BERNARDINO
6700 *SAN BERNARDINO

6720 SAN BRUNO
6740 SAN BUENAVENTURA
6760 SAN CARLOS
6780 SAN CLEMENTE
6800 SAN DIEGO

6820 *SAN DIEGO
6830 SAN DIMAS
6840 SAN FERNANDO
6860 SAN FRANCISCO
6880 *SAN FRANCISCO

6900 SAN GABRIEL
6920 SANGER
6940 SAN JACINTO
6960 *SAN JOAQUIN
6980 SAN JOSE

6990 SAN JUAN CAPISTRANO
7000 SAN LEANDRO
7020 SAN LORENZO
7040 SAN LUIS OBISPO
7060 *SAN LUIS OBISPO

7070 SAN MARCOS
7080 SAN MARINO
7100 SAN MATEO
7120 *SAN MATEO
7140 SAN PABLO

7160 SAN RAFAEL
7170 SAN RAMON
7180 SANTA ANA
7200 SANTA BARBARA
7220 *SANTA BARBARA

7240 SANTA CLARA
7260 *SANTA CLARA
7280 SANTA CRUZ
7300 *SANTA CRUZ
7320 SANTA FE SPRINGS

7340 SANTA MARIA
7345 SANTA MARIA SOUTH
7360 SANTA MONICA
7380 SANTA PAULA
7400 SANTA ROSA

7410 SANTEE
7420 SARANAP
7440 SARATOGA
7460 SAUSALITO
7470 SCOTTS VALLEY

7480 SEAL BEACH
7490 SEARLES VALLEY
7500 SEASIDE
7520 SEBASTOPOL
7540 SELMA

*COUNTY DESIGNATION

7550 SEQUOIA NAT PARK
7560 SHAFTER
7580 *SHASTA
7620 *SIERRA
7640 SIERRA MADRE

7660 SIGNAL HILL
7670 SIMI VALLEY
7680 *SISKIYOU
7690 SOLANA BEACH
7700 *SOLANO

7720 SOLEDAD
7740 SONOMA
7760 *SONOMA
7780 SONORA
7790 SOQUEL

7800 SOUTH EL MONTE
7820 SOUTH GATE
7825 SOUTH LAGUNA
7830 SOUTH LAKE TAHOE
7840 SOUTH MODESTO

7860 SOUTH OROVILLE
7900 SOUTH PASADENA
7940 SOUTH SAN FRANCISCO
7960 SOUTH SAN GABRIEL
7965 SOUTH SAN JOSE HILLS

7970 SOUTH WHITTIER
7980 SOUTH YUBA
7985 SPRING VALLEY
7990 STANFORD
8000 STANTON

8020 *STANISLAUS
8040 STOCKTON
8050 SUISUN CITY
8060 SUNNYMEAD
8080 SUNNYVALE

8100 SUSANVILLE
8120 *SUTTER
8140 TAFT
8160 TAFT HEIGHTS
8180 TEHACHAPI

8200 *TEHAMA
8220 TEMPLE CITY
8230 THERMALITO
8240 THOUSAND OAKS
8250 TIBURON

8260 TORRANCE
8280 TRACY
8300 *TRINITY
8320 TULARE
8340 *TULARE

8360 *TUOLUMNE
8380 TURLOCK
8384 TUSTIN
8385 TUSTIN FOOTHILLS
8390 TWENTYNINE PALMS

8392 TWENTYNINE PALMS BASE
8395 TWIN LAKES
8400 UKIAH
8420 UNION CITY
8440 UPLAND

8460 VACAVILLE
8465 VALENCIA
8470 VALINDA
8480 VALLEJO
8485 VANDERBURG

8490 VENTURA
8500 *VENTURA
8510 VICTORVILLE
8514 VIEW PARK-WINDSOR HILLS
8517 VILLA PARK

8520 VISALIA
8540 VISTA
8555 WALNUT CITY
8560 WALNUT CREEK
8570 WALNUT CREEK WEST

8580 WALNUT HEIGHTS
8590 WALNUT PARK
8600 WASCO
8620 WATSONVILLE
8640 WEED

8645 WEST ATHENS
8650 WEST CARSON
8655 WEST COMPTON
8660 WEST COVINA
8680 WEST HOLLYWOOD

8700 WESTMINSTER
8703 WEST MODESTO
8706 WESTMONT
8709 WEST PITTSBURG
8712 WEST PUENTE VALLEY

8715 WEST SACRAMENTO
8718 WEST WHITTIER-LOS NIETOS
8720 WHITTIER
8740 WILLITS
8750 WILLOWBROOK

8760 WILLOWS
8770 WINTON
8780 WOODLAKE
8800 WOODLAND
8820 WOODSIDE

*COUNTY DESIGNATION

05 CALIFORNIA

8840	*YOLO
8845	YORBA LINDA
8850	YOSEMITE NAT PARK
8860	YREKA CITY
8880	*YUBA
8900	YUBA CITY
8920	YUCAIPA
8940	YUCCA VALLEY

*COUNTY DESIGNATION

APPENDIX B - OBSMAX

(Program Listing)

```
00010
00020
00030
00040 !****************  CARBON MONOXIDE TEST  ********************************
00050
00060 ! WRITTEN BY  KIM HENRY
00070 ! CALTRANS TRANSPORTATION LABORATORY, SACRAMENTO
00080
00090 ! REVISED BY  JIM QUITTMEYER
00100 ! REVISION DATE: 11/30/83
00110
00120
00130 !*********************************************************************
00140
00150 !  THE DATA FILE FOR THIS PROGRAM CONSISTS OF UP TO 90- 26 ITEM RECORDS.
00160 !  EACH RECORD CONSISTS OF A MONTH, A DAY, AND 24 DATA VALUES.
00170 !  AN ALTERNATIVE FILE ORGANIZATION IS DISCUSSED IN THE HELP MESSAGE
00180 !  AND IN THE USER DOCUMENTATION.
00190
00200 !*********************************************************************
00210
00220 !**************  INITIALIZATIONS  ************************************
00230
00240 . OPTION BASE 1   %  OPTION NOPROMPT
00250
00260 !****  ARRAYS
00270
00280  DIM  AXIS$ (50)
00290  DIM  CONSEQ_DAYS$ (90)
00300  DIM  DAY_FACTOR(7),                     DAYS_IN_A_MONTH (12)
00310  DIM  GRAPH_SWITCH$(7),                  JDATE(90)
00320  DIM  FLAGS$ (90,28),                    HOLD (90,27)
00330  DIM  HORIZ_AXIS$(28)
00340  DIM  HORIZ_AXIS_1$ (64),                HORIZ_AXIS_2$(28)
00350  DIM  HORIZ_AXIS_3$ (28),                LABEL$ (50)
00360  DIM  MAXIMUM_1 (90,5),                  MAXIMUM_8 (90,5)
00370  DIM  MAX_TIME_1 (90,5),                 MAX_TIME_8 (90,5)
00380  DIM  MONTH_FACTOR(12)
00390  DIM  OUTLIER$ (11,6),                   OVER_MIDNIGHT (32)
00400  DIM  PERIOD_COUNT (5),                  PERIOD_COUNT_8(5)
00410  DIM  PS(5),                             PE(5)
00420  DIM  PS8(5),                            PE8(5)
00430  DIM  PERIOD_NAME$ (5)
00440  DIM  POSITION_1 (90,5),                 POSITION_8 (90,5)
00450  DIM  PROJECT_DESCRIPTION$ (30),         PROB(90,5,2)
00460  DIM  PLOT$ (50,30,2)
00470  DIM  SKIP_SWITCH$ (90),                 SORTING (90)
00480  DIM  SRT (90),                          STD_DEV (90)
00490  DIM  STRING$ (6),                       TCHECK (13)
00500  DIM  TEMP_HORIZ$(28)
00510  DIM  WORK (90,28),                      XBAR(90)
00520
00530 ! *** ARRAY DEFINITIONS
00540
00550 ! AXIS$ - VERTICAL AXIS, USED IN ALL SUMMARY GRAPHS
```

```
00560 ! CONSEQ_DAYS$ - INDICATES WHETHER NEXT RECORD IS FOR THE NEXT
00570 !     CONSEQUTIVE DAY
00580 ! FLAGS$ - INDICATES WHETHER THE VALUE WITHIN A RECORD IS AN
00590 !     INTERPOLATED VALUE (I), AN OUTLIER (O), AN INTERPOLATED
00600 !     OUTLIER (Q), A SUBSTITUTED VALUE (C), OR ONE OF THREE OR MORE
00610 !     CONSEQUTIVE MISSING VALUES (T)
00620 ! HOLD - HOLDS ORIGINAL DATA READ IN FROM DATA FILE, INCLUDING
00630 !      MONTH, DAY AND YEAR
00640 ! HORIZ_AXIS$ - USED TO FILL HORIZ_AXIS_2$ AND HORIZ_AXIS_3$
00650 ! HORIZ_AXIS_1$ - HORIZONTAL AXIS FOR RANK GRAPH
00660 ! HORIZ_AXIS_2$ - HORIZONTAL AXIS FOR 1-HOUR GRAPH
00670 ! HORIZ_AXIS_3$ - HORIZONTAL AXIS FOR 8-HOUR GRAPH
00680 ! MAXIMUM_1 (R,P) - HOLDS 1-HOUR MAXIMUM FOR RECORD R, PERIOD P
00690 !     (P: 1=DAILY, 2=MORNING, 3=MIDDAY, 4=EVENING, 5=NOCTURNAL)
00700 ! MAX_TIME_1 (R,P) - HOLDS START HOUR + 1 OF 1-HOUR MAXIMUM OF
00710 !     RECORD R, PERIOD P
00720 ! MAXIMUM_8 (R,P) - HOLDS 8-HOUR MAXIMUM FOR RECORD R, PERIOD P
00730 ! MAX_TIME_8 (R,P) - HOLDS START HOUR + 1 OF 8-HOUR MAXIMUM OF
00740 !     RECORD R, PERIOD P
00750 ! OVER_MIDNIGHT - TEMPORARILY HOLDS DATA FROM START HOUR 0000 TO
00760 !     0700 OF THE NEXT DAY TO CALCULATE MAXIMUMX OF 1 AND 8 HOUR
00770 !     INTERVALS FOR ALL TIME PERIODS
00780 ! PS - CONTAINS START HOUR + 1 OF 1-HOUR TIME PERIODS (USED IN
00790 !     CONJUNCTION WITH OVER_MIDNIGHT ARRAY)
00800 ! PE - CONTAINS END_HOUR + 1 OF 1-HOUR TIME PERIODS
00810 ! PS8 - CONTAINS START HOUR + 1 OF 8-HOUR TIME PERIODS
00820 ! PE8 - CONTAINS END HOUR + 1 OF 8-HOUR TIME PERIODS
00830 ! PLOT$ - USED FOR BOTH RANK AND MAXIMUM GRAPHS FOR BOTH 1 AND 8
00840 !     HOUR INTERVALS (PLOT$(V,HZ,HR), V=VERTICAL, HZ=HORIZONTAL,
00850 !     HR: 1=1-HOUR, 2=8-HOUR)
00860 ! WORK - HOLDS READINGS FOR A DAY PLUS TWO HOURS ON BOTH SIDES,
00870 !     AND IS USED IN ALL CALCULATIONS.  CONTAINS INTERPOLATED AND
00880 !     SUBSTITUTED VALUES
00890
00900 !***** INITIALIZE ARRAYS
00910
00920 MAT CONSEQ_DAYS$=('NO')         % MAT DAYS_IN_A_MONTH = (0)
00930 MAT EVENING_MAX_1=(-.5)         % MAT EVENING_MAX_8=(-.5)
00940 MAT GRAPH_SWITCH$ = ('OFF')     % MAT JDATE=(-99)
00950 MAT FLAGS$=('')                 % MAT HOLD=(-99)
00960 MAT PERIOD_COUNT = (0)          % MAT PERIOD_COUNT_8 = (0)
00970 MAT SORTING=(-99)
00980 MAT SRT=(-99)                   % MAT STD_DEV = (-99)
00990 MAT STRING$=('')                % MAT TCHECK=(0)
01000 MAT WORK=(-99)
01010 MAT MAXIMUM_1 =(-.5)            % MAT MAXIMUM_8 =(-.5)
01020 MAT LABEL$=(' ')                % MAT AXIS$ =('|')
01030 MAT MAX_TIME_1 =(-.5)           % MAT MAX_TIME_8 =(-.5)
01040 MAT POSITION_1 = (-99)          % MAT POSITION_8 = (-99)
01050 MAT OVER_MIDNIGHT = (-99)       % MAT PLOT$ = (' ')
01060 MAT PROB = (0)
01070 MAT SKIP_SWITCH$ = ('OFF')
01080  !***** CARRIAGE CONTROL STRINGS
01090
01100  H$ = CHR$(15)
```

```
01110  IN$ = RPT$ (' ',12)
01120  LF$ = CHR$(37) + CHR$(13)
01130  TOP$ = CHR$(37) + CHR$(13) + RPT$(CHR$(0),30)
01140  TF$ = CHR$(12) + RPT$(CHR$(0),65)
01150
01160  !***** HEADING STRINGS
01170  HED10$ = '                 AMBIENT AIR QUALITY SUMMARY'
01180  HED11$ = '                PAGE: '
01190  HED12$ = '                 OBSERVED MAXIMUM ANALYSIS'
01200  HED13$ = '                 POLLUTANT: CARBON MONOXIDE'
01210  HED14$ = '                 JOB DESCRIPTION: '
01220  HED15$ = '                 SITE CODE: '
01230  HED16$ = '                 RUN DATE: '
01240  HED17$ = '                DATA FILE: '
01250  OUTLIER_COLUMN_1$ ='                  INPUT    INTERPOLATED'
01260  OUTLIER_COLUMN_2$ ='        TIME     VALUES       VALUE'
01270
01280  !***** FORMAT STRINGS
01290
01300 FORMB$='             <##########    #      >## #  ##.# <####    @##@/@##@/@##@'
01310  FORMA$='      @##@00-@##@00'
01320  FORM1$= FORMB$ + FORMA$
01330  FORM2$='              >##            @##@/@##@/@##@        @##@00-@##@00'
01340  PLOT_FORM$ ='    #####>####'
01350  EMPTY$ = '             <##########    #            <#################'
01360  OUTLIER_FORM_1$='       ##00     >###          >##'
01370  OUTLIER_FORM_2$ = OUTLIER_FORM_1$ + RPT$(' ',12)    + OUTLIER_FORM_1$
01380  !***** OTHER VARIABLES
01390
01400  AQDHS_SWITCH$ = 'OFF'
01410  AREA_CODE =0
01420  CHART_COUNT = 7
01430  CHART_TOTAL=0
01440  INTP_FLAG$= 'OFF'
01450  JDATE = 0
01460  LINE_COUNT = 0
01470  NI_FLAG$ = 'OFF'
01480  NUMBER_OF_RECORDS = 0
01490  OUTLIER_COUNT = 0
01500  PAGE = 1
01510  PROJECT_CODE = 0
01520  PROJECT_DESCRIPTION$ = ''
01530  SITE_CODE = 0
01540  SIX_HOUR_SWITCH$ = 'OFF'
01550  STRING$ = ' '
01560  SUFF_FLAG$ = 'OFF'
01570
01580  !**************************************************************
01590  MAT READ MONTH_FACTOR
01600  DATA .0351,.0085,.0003,0.0,.0006,.0009,.0006,.0009,.0045,.0207
01610  DATA .0676,.1149
01620  MAT READ DAY_FACTOR
01630  DATA .371,.945,1.06,1.16,1.15,1.53,.795
01640  MAT READ PERIOD_NAMES$
01650  DATA 'DAILY','MORNING','MIDDAY','EVENING','NOCTURNAL'
```

```
01660  MAT READ DAYS_IN_A_MONTH
01670  DATA 31,28,31,30,31,30,31,31,30,31,30,31
01680  MAT READ PS
01690  DATA 1,7,11,18,22
01700  MAT READ PS8
01710  DATA 1,2,9,12,21
01720  MAT READ PE8
01730  DATA 24,8,11,20,25
01740  MAT READ PE
01750  DATA 24,10,17,21,30
01760  DATA 2.55,2.61,2.66,2.71,2.75,2.79,2.82,2.85,2.88,2.91,2.94,2.96,2.99
01770  MAT READ TCHECK
01780  REM  INPUT FILE
01790
01800   PRINT 'MANUALLY SCROLL PAPER TO TOP OF PAGE AND PRESS RETURN.'
01810   LINPUT TOP_OF_PAGE$
01820   PRINT CHR$(39);'C66'
01830   PRINT CHR$(39);'V'
01840     PRINT
01850     PRINT 'ENTER DATA FILE NAME (OR HELP OR QUIT).    ';
01860     INPUT FIL$
01870     IF FIL$='HELP' THEN GO TO 9160
01880     PRINT % PRINT 'VERIFY:YOUR DATA FILE IS ';FIL$;'? YES/NO.';
01890     INPUT FCHCK$
01900     FCHCK$=STR$(FCHCK$,1,1)
01910     IF FCHCK$='Y'
01920       GOTO 2000
01930     ELSEIF FCHCK$='N'
01940       GOTO 1850
01950     ELSE
01960       PRINT 'ANSWER YES OR NO PLEASE! ';
01970       GO TO 1880
01980     ENDIF
01990     ON CONV GO TO 11050
02000     OPEN #3, FIL$, INPUT
02010  PRINT
02020  PRINT 'ENTER UP TO 30 CHARACTERS OF JOB DESCRIPTION.    ';
02030   INPUT PROJECT_DESCRIPTION$
02040  IF LEN(PROJECT_DESCRIPTION$) > 30
02050      PRINT 'JOB DESCRIPTION MAY NOT BE LONGER THAN 30 CHARACTERS.'
02060.     GOTO 2020
02070  ENDIF
02080  PRINT
02090  PRINT 'ENTER 3 DIGIT SITE CODE.    ';  % INPUT SITE_CODE$
02100     IF LEN(SITE_CODE$)<>3
02110         PRINT 'CODE MUST BE 3 DIGITS. PLEASE RIGHT JUSTIFY AND ZERO FILL.'
02120          GOTO 2090
02130     ENDIF
02140 ! ******************** CALCULATIONS **********************************
02150
02160  REM FILL HOLD MATRIX.  THIS WILL HOLD THE ORIGINAL DATA FROM FIL$.
02170    D = 1
02180    LOOP
02190      MAT INPUT #3, STRING$;
02200       ON EOF GO TO 2570
```

```
02210       IF STRING$(4) = '*' THEN 2330       ! * FLAGS VARIABLE LENGTH DATA
02220
02230   !  DATA IS IN 24 READINGS FORM
02240       FOR H = 1 TO 6
02250         HOLD(D,H) = VALUE(STRING$(H))
02260       NEXT H
02270       FOR H = 7 TO 27
02280          INPUT #3, M;
02290          HOLD (D,H) =M
02300       NEXT H
02310       GO TO 2470
02320
02330   !  DATA IS IN VARIABLE LENGTH FORM
02340       HS = VALUE(STRING$(5))  %    HE = VALUE(STRING$(6))
02350       HOLD(D,1) = VALUE(STRING$(1))  %   HOLD(D,2) = VALUE(STRING$(2))
02360       HOLD(D,3) = VALUE(STRING$(3))
02370       FOR H = HS+1 TO HE+1
02380         INPUT #3,M;
02390         HOLD(D,H+3) =M
02400       NEXT H
02410       FOR H = 1 TO HS
02420         HOLD(D,H+3) = -1
02430       NEXT H
02440       FOR H = HE+2 TO 24
02450         HOLD(D,H+3) = -1
02460       NEXT H
02470       D = D+ 1
02480       IF D > 90
02490         PRINT % PRINT
02500         PRINT ' YOUR DATA FILE HAS DATA FOR MORE THAN 90 DAYS,'
02510         PRINT ' WHICH IS THE MAXIMUM THE PROGRAM CAN HANDLE.  '
02520         PRINT ' PROGRAM EXECUTION HAS BEEN HALTED.            '
02530         CLOSE #3
02540         GOTO 16500
02550       ENDIF
02560  ENDLOOP
02570      CLOSE #3
02580
02590    NUMBER_OF_RECORDS = D-1
02600
02610     REM CALCULATE JULIAN DATE
02620
02630         FOR D = 1 TO NUMBER_OF_RECORDS
02640           JDATE=0
02650           FOR J = 1 TO HOLD (D,1) -1
02660             JDATE = JDATE + DAYS_IN_A_MONTH (J)
02670           NEXT J
02680           JDATE(D) = JDATE + HOLD (D,2)
02690         NEXT D
02700
02710   REM CHECK JULIAN DATE FOR CONSEQ. DAYS
02720      FOR D = 1 TO NUMBER_OF_RECORDS-1
02730        IF JDATE (D+1) = 1 AND JDATE (D) > 364 THEN CONSEQ_DAYS$(D)='YES'
02740        IF JDATE (D+1) = JDATE (D)+1 THEN CONSEQ_DAYS$(D) = 'YES'
02750      NEXT D
```

```
02760
02770   REM FILL WORK MATRIX
02780     REM  THIS MATRIX HOLDS READINGS FOR A DAY PLUS 2 HRS ON BOTH SIDES
02790     WORK(1,1) = -1  %    WORK(1,2) = -1
02800     FOR D = 1 TO NUMBER_OF_RECORDS
02810       FOR H  = 3 TO 26
02820         WORK (D,H) = HOLD (D,H+1)
02830         IF WORK(D,H) = 0 THEN WORK(D,H) = .25   ! FOR OUTLIER ANAL
02840       NEXT H
02850     NEXT D
02860
02870     FOR D = 1 TO NUMBER_OF_RECORDS -1
02880       IF CONSEQ_DAYS$(D) = 'NO' THEN GO TO  2950
02890       REM  CONSEQ DAYS
02900        WORK (D+1,1) = WORK (D,25)
02910        WORK (D+1,2) = WORK (D,26)
02920        WORK (D,27)  = WORK (D+1,3)
02930        WORK (D,28)  = WORK (D+1,4)
02940        GO TO 3000
02950       REM NON-CONSEQ DAYS
02960        WORK (D+1,1) = -1
02970        WORK (D+1,2) = -1
02980        WORK (D,27)  = -1
02990        WORK (D,28)  = -1
03000      NEXT D
03010      WORK(NUMBER_OF_RECORDS,27)=-1 % WORK(NUMBER_OF_RECORDS,28)=-1
03020
03030  PRINT
03040  PRINT 'CHECKING FOR OUTLIERS...'
03050  PRINT
03060   GOSUB 11190     ! CHECK FOR OUTLIERS
03070
03080   FOR D = 1 TO NUMBER_OF_RECORDS
03090   FOR H = 1 TO 28
03100     IF WORK(D,H) = .25 THEN WORK(D,H) = 0
03110   NEXT H
03120  NEXT D
03130   REM CHECK FOR MISSING READINGS
03140     FOR D = 1 TO NUMBER_OF_RECORDS
03150       FOR H = 3 TO 26
03160         IF WORK (D,H) >= 0 THEN GO TO 3180
03170         GOSUB 9690      !INTERPOLATE VALUES
03180       NEXT H
03190
03200     NEXT D
03210  IF OUTLIER_COUNT >0
03220   PRINT 'THERE ARE ';OUTLIER_COUNT;' OUTLIERS.  DO YOU WISH TO EXAMINE ';
03230    PRINT 'THEM? YES/NO.  ';
03240      POSITION_1(1,1)=D
03250    INPUT A$
03260    A$=STR$(A$,1,1)
03270   IF A$= 'Y'
03280     FOR D = 1 TO NUMBER_OF_RECORDS
03290       FOR H = 3 TO 26
03300         IF FLAGS$(D,H)='O' OR FLAGS$(D,H)='Q' THEN GOSUB 10070 ! OUTLIERS
```

```
03310        NEXT H
03320      NEXT D
03330      IF NI_FLAG$='ON' THEN PRINT '(NI= NOT INTERPOLATED)'
03340    ELSE
03350       IF A$ <> 'N'
03360         PRINT 'PLEASE RESPOND WITH YES OR NO.'
03370        GOTO 3250
03380      ENDIF
03390    ENDIF
03400   IF  A$='Y'  THEN PRINT CHR$(12); RPT$(CHR$(0),65)
03410    PRINT
03420    PRINT 'DO YOU WISH TO SUBSTITUTE VALUES FOR THE INTERPOLATED VALUES? ';
03430    PRINT 'YES/NO/QUIT. ';
03440    INPUT A$
03450      A$=STR$(A$,1,1)
03460     IF A$ = 'Q' THEN GOTO 16500
03470     IF A$ = 'Y'
03480      LOOP
03490        B$= ' '
03500        PRINT 'ENTER NUMBER OF MONTH,DAY,START HOUR, AND SUBSTITUTE VALUE.'
03510        PRINT 'PRESS RETURN TO STOP.'
03520        LINPUT B$
03530        IF LEN(B$) = 0 THEN GOTO 3800
03540        FOR P = 1 TO 4
03550          COMA = IDX(B$,',')
03560          V(P) = VALUE (STR$(B$,1,COMA-1))
03570          B$ = STR$ (B$,COMA+1,80)
03580        NEXT P
03590        V(4) = VALUE(B$)
03600        FOR D = 1 TO NUMBER_OF_RECORDS
03610          IF HOLD(D,1) = V(1) AND HOLD(D,2) = V(2)
03620            HOLD(D,V(3)+4) = V(4)   ! SUBSTITUTE VALUE
03630            WORK(D,V(3)+3) = V(4)
03640            FLAGS$(D,V(3) +3) = 'C'
03650            IF V(3)+3>24 AND CONSEQ_DAYS$(D)='YES'
03660               WORK(D+1,V(3)-21)=V(4)
03670               FLAGS$(D+1,V(3)-21)='C'
03680            ELSEIF V(3)+3<5 AND CONSEQ_DAYS$(D-1)='YES'
03690               WORK(D-1,V(3)+27)=V(4)
03700               FLAGS$(D-1,V(3)+27)='C'
03710            ENDIF
03720          ENDIF
03730        NEXT D
03740       ENDLOOP
03750    ELSE
03760       IF A$ <> 'N'
03770        PRINT 'PLEASE RESPOND WITH YES, NO OR QUIT.'
03780        GOTO 3440
03790      ENDIF
03800    ENDIF
03810    PRINT
03820    PRINT 'DO YOU WISH TO HAVE THE MODIFIED DATA FILE SAVED? YES/NO.  ';
03830    INPUT A$
03840    A$=STR$(A$,1,1)
03850    NEW_FILE_FLAG$ = 'NO'
```

```
03860   IF A$ = 'Y'
03870     PRINT 'ENTER NEW FILE NAME. IT MAY BE UP TO 8 CHARACTERS LONG.  ';
03880     INPUT N$
03890     OPEN #5, N$, OUTPUT
03900     FOR D = 1 TO NUMBER_OF_RECORDS
03910       FOR H = 1 TO 3
03920         PRINT #5 USING '##,', HOLD(D,H);
03930       NEXT H
03940        FOR H = 3 TO  25
03950           PRINT #5 USING '##,',WORK(D,H);
03960        NEXT H
03970        PRINT #5 USING '##',WORK(D,26)
03980     NEXT D
03990     CLOSE #5
04000     NEW_FILE_FLAG$ = 'YES'
04010   ELSE
04020    IF A$<> 'N'
04030      PRINT 'PLEASE RESPOND WITH YES OR NO.'
04040      GOTO 3830
04050    ENDIF
04060  ENDIF
04070  ENDIF
04080   REM  INFORMATION FOR TRANSFERING DATA TO AQDHS
04090     PRINT
04100     PRINT'DO YOU WISH TO HAVE THIS DATA SENT TO AQDHS? ENTER YES OR NO. ';
04110
04120     INPUT AQDHS$
04130     AQDHS$=STR$(AQDHS$,1,1)
04140     IF AQDHS$ ='Y'  THEN GOTO 4200
04150     IF AQDHS$ ='N'  THEN GO TO 4210
04160      IF  AQDHS$<>'Y' AND AQDHS$<>'N'
04170          PRINT 'PLEASE RESPOND WITH EITHER YES OR NO.'
04180          GO TO 4120
04190      ENDIF
04200     GOSUB 14360  ! SEND DATA TO AQDHS
04210
04220  PRINT % PRINT 'WORKING...'
04230
04240 FOR D = 1 TO NUMBER_OF_RECORDS
04250   GOSUB 14210  ! DAY OF WEEK (DOW)
04260
04270
04280 REM FILL ARRAY OVER_MIDNIGHT WITH READINGS FROM START HOUR
04290 REM 0000 TO 0700 OF THE NEXT DAY IN ORDER TO HANDLE ALL TIME
04300 REM INTERVALS FOR BOTH 1 AND 8 HOUR CALCULATIONS
04310
04320   MAT OVER_MIDNIGHT = (-1)
04330   FOR H=1 TO 24
04340     OVER_MIDNIGHT(H)=WORK(D,H+2)
04350   NEXT H
04360   IF CONSEQ_DAYS$(D)='YES'
04370     FOR H=1 TO 8
04380     OVER_MIDNIGHT(H+24)=WORK(D+1,H+2)
04390     NEXT H
04400   ENDIF
```

```
04410
04420 REM 1-HOUR PERIODS
04430
04440 FOR PP=1 TO 5
04450 ! CHECK PERIOD FOR MISSING VALUES UNLESS DAILY
04460 IF PP=1 THEN 4590
04470 MISS=0
04480   FOR H=PS(PP) TO PE(PP)
04490     IF OVER_MIDNIGHT(H)<0 OR (CONSEQ_DAYS$(D)='NO' AND H>19)  &
04500  &              THEN 4710  ! SKIP PERIOD
04510     IF H>19
04520       IF HOLD(D+1,H-16)<0 THEN MISS=MISS+1
04530     ELSE
04540       IF HOLD(D,H+8)<0 THEN MISS=MISS+1
04550     ENDIF
04560     IF MISS>INT((5/8)*(PE(PP)-PS(PP))) THEN 4710  ! SKIP PERIOD
04570   NEXT H
04580
04590 ! FIND 1-HOUR MAXIMUMS
04600   PERIOD_COUNT(PP) = PERIOD_COUNT(PP) +1
04610   PROB(D,PP,1)=DAY_FACTOR(DOW)*MONTH_FACTOR(HOLD(D,1))
04620   FOR H=PS(PP) TO PE(PP)
04630     IF OVER_MIDNIGHT(H)>MAXIMUM_1(D,PP)
04640       MAXIMUM_1(D,PP)=OVER_MIDNIGHT(H)
04650       MAX_TIME_1(D,PP)=H    !H IS START HOUR + 1
04660     ENDIF
04670   NEXT H
04680
04690   REM 8-HOUR PERIODS
04700
04710   ! CHECK PERIOD FOR MISSING VALUES UNLESS DAILY
04720   IF PP=1 THEN 4910
04730     FOR H = PS8(PP) TO PE8(PP)+7
04740       IF OVER_MIDNIGHT(H)<0 THEN 5060   ! SKIP PERIOD
04750     NEXT H
04760     FOR H = PS8(PP) TO PE8(PP)
04770       MISS = 0
04780       FOR A = 0 TO 7
04790         IF H+A>19 AND CONSEQ_DAYS$(D) = 'NO'
04800            GOTO 5060 !  SKIP PERIOD
04810         ELSEIF H+A>19
04820            IF HOLD(D+1,H+A-16)<0 THEN MISS=MISS+1
04830          ELSE
04840            IF HOLD(D,H+A+8)<0 THEN MISS=MISS+1
04850         ENDIF
04860           IF MISS>3 THEN 5060  ! SKIP PERIOD
04870       NEXT A
04880     NEXT H
04890
04900 REM COMPUTE 8-HOUR AVERAGES AND FIND 8-HOUR MAXIMUMS
04910
04920       PERIOD_COUNT_8(PP)=PERIOD_COUNT_8(PP) + 1
04930       PROB(D,PP,2)=DAY_FACTOR(DOW)*MONTH_FACTOR(HOLD(D,1))
04940       FOR H = PS8(PP) TO PE8(PP)
04950         TOTAL = 0
```

```
04960         FOR A = 0 TO 7
04970           TOTAL = TOTAL + OVER_MIDNIGHT(H+A)
04980         NEXT A
04990         AVERAGE = TOTAL / 8
05000         IF AVERAGE > MAXIMUM_8 (D,PP)   ! PERIOD MAX
05010           MAXIMUM_8(D,PP) = AVERAGE
05020           MAX_TIME_8(D,PP) = H    ! H IS START HOUR + 1
05030         ENDIF
05040
05050       NEXT H
05060     NEXT PP
05070   NEXT D
05080
05090 REM CREATE POSITION ARRAY
05100 FOR J=1 TO 90
05110   FOR K=1 TO 5
05120     POSITION_1(J,K) = J
05130     POSITION_8(J,K) = J
05140   NEXT K
05150 NEXT J
05160
05170   ! SORT MAXS
05180   FOR PERIOD = 1 TO 5
05190     FOR I = 1 TO NUMBER_OF_RECORDS
05200       FOR D = 1 TO NUMBER_OF_RECORDS - 1
05210         IF MAXIMUM_1(D,PERIOD) < MAXIMUM_1(D+1,PERIOD)
05220           TEMP = MAXIMUM_1 (D,PERIOD)
05230           MAXIMUM_I (D,PERIOD) = MAXIMUM_1(D+1,PERIOD)
05240           MAXIMUM_I(D+1,PERIOD) = TEMP
05250           TEMP = POSITION_1(D,PERIOD)
05260           POSITION_1(D,PERIOD) = POSITION_1(D+1, PERIOD)
05270           POSITION_1(D+1,PERIOD) = TEMP
05280         ENDIF
05290         IF MAXIMUM_8 (D,PERIOD) < MAXIMUM_8 (D+1,PERIOD)
05300           TEMP = MAXIMUM_8(D,PERIOD)
05310           MAXIMUM_8(D,PERIOD) = MAXIMUM_8(D+1,PERIOD)
05320           MAXIMUM_8(D+1,PERIOD) = TEMP
05330           TEMP = POSITION_8(D,PERIOD)
05340           POSITION_8(D,PERIOD) = POSITION_8(D+1,PERIOD)
05350           POSITION_8(D+1,PERIOD) = TEMP
05360         ENDIF
05370       NEXT D
05380     NEXT I
05390   NEXT PERIOD
05400
05410
05420
05430 !********************** OUTPUT  **********************************
05440
05450   REM PRINT FIRST PAGE
05460     I = FN_PAGE_HEADING_BLOCK (PAGE)
05470     PAGE = PAGE + 1
05480     PRINT '               OBSERVED MAXIMUMS:'
05490     PRINT  % PRINT
05500     PRINT'                       AVG.TIME  DAYS  VALUE'
```

```
05510     PRINT'               TYPE    (HRS.)   USED  (PPM)         DATE';
05520     PRINT '            TIME'
05530     PRINT RPT$ (' ',12); RPT$ ('-',58)
05540     PRINT
05550
05560   REM PRINT OBSERVED MAXIMUM CHART
05570   FOR PP = 1 TO 5
05580     DAYS_8 = PERIOD_COUNT_8(PP)
05590     DAYS_1 = PERIOD_COUNT(PP)
05600
05610     NOCTURNAL_SWITCH$ = 'OFF'
05620     PRINT_DAY = POSITION_1(1,PP)
05630     RVALUE = MAXIMUM_1(1,PP)
05640     IF RVALUE < 0 THEN RVALUE = 0
05650     PERIOD$ = PERIOD_NAME$(PP)
05660     HOUR = 1
05670     TT=1
05680     GOSUB 14100  ! SUFFICIENT NUMBER OF DAYS?
05690     START_HOUR = MAX_TIME_1(PRINT_DAY,PP) -1
05700     IF START_HOUR>23 THEN START_HOUR=START_HOUR-24
05710     IF PP=5 THEN NOCTURNAL_SWITCH$ = 'ON'
05720     I = FN_FILL_PRINT_CONSTANTS
05730     I = FN_PRINT_LINE
05740
05750     PRINT_DAY = POSITION_8(1,PP)
05760     RVALUE = MAXIMUM_8(1,PP)
05770     IF RVALUE < 0 THEN RVALUE = 0
05780     PERIOD$ = '          '
05790     HOUR = 8
05800     TT=2
05810     GOSUB 14100  ! SUFFICIENT NUMBER OF DAYS?
05820     START_HOUR = MAX_TIME_8(PRINT_DAY,PP) -1
05830     IF START_HOUR>23 THEN START_HOUR=START_HOUR-24
05840     I = FN_FILL_PRINT_CONSTANTS
05850     I = FN_PRINT_LINE
05860     PRINT
05870   NEXT PP
05880       PRINT
05890       PRINT IN$; RPT$('-',58)
05900  IF INTP_FLAG$ = 'ON'
05910       PRINT IN$;'(I) CONTAINS INTERPOLATED OR SUBSTITUTED VALUE'
05920  ENDIF
05930  IF SUFF_FLAG$ = 'ON'
05940 PRINT IN$;'$=INSUFFICIENT NUMBER OF DAYS'
05950  ENDIF
05960       PRINT  % PRINT  % PRINT
05970
05980    REM PRINT OUTLIER SUMMARY
05990      IF OUTLIER_COUNT >0 THEN GO TO 6020
06000      PRINT  IN$;'OUTLIER SUMMARY: NO OUTLIERS FOUND'
06010      GO TO 6230
06020   REM THERE ARE OUTLIERS TO BE PRINTED
06030      IF OUTLIER_COUNT < 17 THEN GO TO 6060
06040      I = FN_PAGE_HEADING_BLOCK (PAGE)
06050      PAGE = PAGE +1
```

```
06060       I = FN_OUTLIER_HEADING_BLOCK
06070       LINE_COUNT = 13
06080        FOR D = 1 TO NUMBER_OF_RECORDS
06090          FOR H = 3 TO 26
06100               IF FLAGS$(D,H)<>'O' AND FLAGS$(D,H)<>'Q'  &
06110                & AND FLAGS$(D,H) <>'C' THEN GOTO  6200
06120              PRINT USING FORM2$, HOLD(D,H+1), HOLD(D,1), HOLD(D,2)&
06130               &   ,HOLD(D,3), H-3 ,H-2
06140              LINE_COUNT = LINE_COUNT + 1
06150              IF LINE_COUNT < 60 THEN GO TO 6200
06160              I = FN_PAGE_HEADING_BLOCK (PAGE)
06170              I = FN_OUTLIER_HEADING_BLOCK
06180              LINE_COUNT = 13
06190              PAGE = PAGE +1
06200          NEXT H
06210       NEXT D
06220
06230    REM FIND OUT WHICH GRAPHS USER WANTS TO SEE
06240     I = FN_PAGE_HEADING_BLOCK(PAGE)
06250     PAGE = PAGE + 1
06260     PRINT IN$;'WHICH OUTPUT GRAPH(S) WOULD YOU LIKE TO SEE?'
06270     PRINT IN$;'ENTER NUMBER(S) SEPARATED BY COMMAS.'
06280     PRINT
06290     PRINT IN$;'0 = NONE'
06300     PRINT IN$;'1 = DAILY'
06310     PRINT IN$;'2 = MORNING'
06320     PRINT IN$;'3 = MIDDAY'
06330     PRINT IN$;'4 = EVENING'
06340     PRINT IN$;'5 = NOCTURNAL'
06350     PRINT IN$;'6 = ALL'
06360     PRINT
06370     PRINT
06380     PRINT IN$; 'GRAPHS?       ';
06390
06400     LINPUT GRAPH_CHOICE$
06410
06420     REM SEPARATE NUMBERS FROM WITHIN STRING
06430     COMMA = 0
06440     LOOP
06450        COMMA = IDX (GRAPH_CHOICE$, ',')
06460        IF COMMA = 0 THEN 6570
06470        GRAPH$ =  (STR$ (GRAPH_CHOICE$, 1, COMMA -1))
06480       REM  VALIDATE GRAPH INPUT
06490       A = ORD (GRAPH$)
06500       IF A < 248 AND A > 239 THEN GO TO 6530
06510       PRINT GRAPH$;' IS NOT AN ACCEPTABLE VALUE. PLEASE REENTER LIST.'
06520       GO TO  6400
06530         GRAPH = VALUE (GRAPH$)
06540        GRAPH_SWITCH$(GRAPH) = 'ON'
06550        GRAPH_CHOICE$ = STR$ (GRAPH_CHOICE$, COMMA +1, 80)
06560   ENDLOOP
06570    REM  NO MORE COMMAS
06580       GRAPH = VALUE(GRAPH_CHOICE$)
06590       IF GRAPH = 0 THEN GO TO 16500   ! INPUT WAS 'NONE'
06600       IF GRAPH = 6
```

```
06610         FOR M = 1 TO 5
06620           GRAPH_SWITCH$(M) = 'ON'
06630         NEXT M
06640        ENDIF
06650     GRAPH_SWITCH$ (VALUE(GRAPH_CHOICE$)) = 'ON'
06660     REM PRINT REQUESTED GRAPHS
06670       FOR Z = 1 TO 5
06680          IF GRAPH_SWITCH$(Z) = 'OFF' THEN GOTO 8570
06690          PERIOD$ = PERIOD_NAME$(Z)
06700
06710       I= FN_PAGE_HEADING_BLOCK (PAGE)
06720       PAGE = PAGE + 1
06730       LINE_COUNT = 13
06740
06750
06760      REM CREATE VERTICAL AXIS
06770         LONG_AXIS_SW$ = 'OFF'
06780       IF FP(MAXIMUM_1(1,1)/2) = 0    ! TOP VALUE IS EVEN
06790         FOR I = 1 TO MAXIMUM_1(1,1)
06800           IF FP (I/2) = 0 THEN AXIS$(I) = VALUE$(I) + ' +'
06810         NEXT I
06820      ELSE                    ! TOP VALUE IS ODD
06830         FOR I = 1 TO MAXIMUM_1(1,1) +1
06840           IF FP(I/2) = 0 THEN AXIS$(I) = VALUE$(I) + ' +'
06850         NEXT I
06860         LONG_AXIS_SW$ = 'ON'
06870      ENDIF
06880        LABEL$ (INT(MAXIMUM_1(1,1)/2)+1) = ' CO '
06890        LABEL$ (INT(MAXIMUM_1(1,1)/2)) = '(PPM)'
06900
06910      REM CREATE HORIZONTAL AXIS
06920        HORIZ_AXIS_1$(1) = '0'
06930        HORIZ_AXIS_1$(2) = ' '
06940        HORIZ_AXIS_1$(3) = '+'
06950        FOR A = 4 TO 63
06960           IF FP((A-3)/10)=0 THEN HORIZ_AXIS_1$(A) = '+' &
06970         & ELSE HORIZ_AXIS_1$(A) = '-'
06980       NEXT A
06990
07000     REM FILL PLOTTING MATRIX FOR RANK GRAPH
07010    MAT PLOT$=(' ')
07020    HR1_GRAPH_SW$ = 'ON'    % HR8_GRAPH_SW$ = 'ON'
07030    IF MAXIMUM_1(1,Z) <= 0 THEN HR1_GRAPH_SW$ = 'OFF'
07040    IF MAXIMUM_8(1,Z) <= 0 THEN HR8_GRAPH_SW$ = 'OFF'
07050    IF HR1_GRAPH_SW$ = 'OFF' AND HR8_GRAPH_SW$= 'OFF'
07060      PRINT 'THERE ARE NO READINGS FOR THE ';PERIOD$;' PERIOD.'
07070      GOTO 8570
07080    ENDIF
07090    IF NUMBER_OF_RECORDS>30 THEN RR=30 ELSE RR=NUMBER_OF_RECORDS
07100    FOR J = 1 TO RR
07110       V1= INT(MAXIMUM_1 (J,Z) +.5)
07120       V8= MAXIMUM_8 (J,Z) +.5
07130       IF V1 > .5 THEN PLOT$(V1,J,1) = '1'
07140       IF V8 > .5 THEN PLOT$(V8,J,2) = '8'
07150       IF V8>= 0 AND V8<.5 THEN HORIZ_AXIS_1$((J*2)+2) = '8'
```

```
07160 NEXT J
07170
07180  REM PRINT TITLE
07190    PRINT IN$; 'GRAPHICAL SUMMARY OF ';PERIOD$;' MAXIMUMS'
07200    PRINT  % PRINT
07210    PRINT '    RANKED RESULTS:'
07220    PRINT
07230
07240  REM PRINT PLOT MATRIX
07250  IF LONG_AXIS_SW$ ='ON' THEN TOP = MAXIMUM_1(1,1)+1 ELSE &
07260    & TOP = MAXIMUM_1(1,1)
07270  FOR I = TOP TO 1 STEP -1
07280       PRINT USING PLOT_FORM$, LABEL$ (I), AXIS$(I);
07290       FOR J = 1 TO 30
07300          IF HR8_GRAPH_SW$ = 'ON'
07310             PRINT PLOT$( I,J,2);
07320         ELSE
07330              PRINT ' ';
07340          ENDIF
07350          IF HR1_GRAPH_SW$ = 'ON'
07360              PRINT PLOT$ (I,J,1);
07370         ELSE
07380              PRINT ' ';
07390          ENDIF
07400          LINE_COUNT = LINE_COUNT + 1
07410       NEXT J
07420      PRINT
07430     NEXT I
07440
07450     PRINT TAB (10);
07460     FOR A = 1 TO 63
07470        PRINT HORIZ_AXIS_1$(A);
07480     NEXT A
07490     PRINT
07500     PRINT TAB (11);
07510     PRINT' 0         5         10         15         20         25         30'
07520     PRINT RPT$(' ',35); 'RANK'
07530    LINE_COUNT = LINE_COUNT + 1
07540
07550  REM  PRINT DOUBLE GRAPH HEADINGS
07560     IF MAXIMUM_1(1,1) > 15 THEN GO TO 7630
07570     REM CALCULATE SPACING
07580     X = 58 - (2 * MAXIMUM_1(1,1) + 27)
07590     FOR SPACE = 1 TO INT (X/2)
07600       PRINT
07610     NEXT SPACE
07620     GO TO 7690
07630     REM  NOT ENOUGH ROOM ON PAGE FOR BOTTOM GRAPHS
07640     I = FN_PAGE_HEADING_BLOCK (PAGE)
07650     PAGE = PAGE + 1
07660     PRINT IN$;'GRAPHICAL SUMMARY OF ';PERIOD$;' MAXIMUMS, CONT.'
07670     PRINT
07680     PRINT
07690     PRINT  '    OBSERVED MAXIMUMS (*) '
07700     PRINT'    (I= INTERPOLATED VALUE)'
```

```
07710       PRINT
07720       PRINT RPT$(' ',21);'1 HOUR';
07730     PRINT RPT$(' ',34);'8 HOUR'
07740
07750
07760     REM CREATE HORIZONTAL AXIS
07770       HORIZ_AXIS$(1) = '0'
07780       HORIZ_AXIS$(2) = ' '
07790       HORIZ_AXIS$(3) = '+'
07800       FOR A = 4 TO 27
07810       IF FP((A-3)/6)=0 THEN HORIZ_AXIS$(A)='+' ELSE HORIZ_AXIS$(A)='-'
07820       NEXT A
07830    MAT TEMP_HORIZ$ = HORIZ_AXIS$
07840    MAT PLOT$ = (' ')
07850
07860   REM FILL PLOTTING MATRIX FOR MAXIMUM GRAPHS
07870   REM ONE HOUR
07880   TS1$= 'OFF'
07890   IF HR1_GRAPH_SW$ = 'ON'
07900      G=1
07910      NUM_STARS=1
07920      DAY=POSITION_1(1,Z)
07930      STAR_HOUR=MAX_TIME_1(DAY,Z)-1
07940      GOSUB 15480 ! FILL PLOT
07950      MAT HORIZ_AXIS_2$=HORIZ_AXIS$
07960      IF TWELVE=1 THEN TS1$= 'ON'
07970   ENDIF
07980
07990   REM 8 HOUR
08000   TS8$='OFF'
08010   MAT HORIZ_AXIS$ = TEMP_HORIZ$
08020   IF HR8_GRAPH_SW$ = 'ON'
08030      G=2
08040      NUM_STARS=8
08050      DAY=POSITION_8(1,Z)
08060      STAR_HOUR=MAX_TIME_8(DAY,Z)-1
08070      GOSUB 15480 ! FILL PLOT
08080      MAT HORIZ_AXIS_3$=HORIZ_AXIS$
08090      IF TWELVE=1 THEN TS8$='ON'
08100   ENDIF
08110
08120    REM PRINT PLOT MATRIX
08130      IF LONG_AXIS_SW$= 'ON' THEN TOP = MAXIMUM_1(1,1) +1 &
08140       & ELSE TOP = MAXIMUM_1(1,1)
08150      FOR I = TOP TO 1 STEP -1
08160       PRINT USING PLOT_FORM$, LABEL$ (I), AXIS$ (I);
08170        FOR J = 1 TO 24
08180          PRINT PLOT$ (I,J,1);
08190       NEXT J
08200       PRINT '    ';
08210       PRINT USING PLOT_FORM$, LABEL$ (I), AXIS$ (I);
08220       FOR J = 1 TO 24
08230         PRINT PLOT$ (I,J,2);
08240       NEXT J
08250        PRINT
```

```
08260    NEXT I
08270
08280      PRINT TAB (10);
08290      FOR A = 1 TO 27
08300        PRINT HORIZ_AXIS_2$(A);
08310      NEXT A
08320      PRINT '              ';
08330       FOR A = 1 TO 27
08340         PRINT HORIZ_AXIS_3$(A);
08350       NEXT A
08360       PRINT
08370
08380      PRINT TAB (9);
08390   IF TS1$='ON'    ! 1 HR GRAPH REQUIRES 12-12 SCALE
08400     PRINT '  1200  1800  0000  0600  1200';
08410     IF TS8$ = 'ON'  ! 8 HR GRAPH REQUIRES 12-12 SCALE
08420       PRINT RPT$(' ',10);' 1200  1800  0000  0600  1200'
08430     ELSE     ! 8 HR GRAPH REQUIRES 0-24 SCALE
08440       PRINT RPT$(' ',10);' 0000  0600  1200  1800  0000'
08450     ENDIF
08460   ELSE   ! 1 HR REQUIRES 0-24 SCALE
08470     PRINT '  0000  0600  1200  1800  0000';
08480     IF TS8$ = 'ON'    !8 HR GRAPH REQUIRES 12-12 SCALE
08490       PRINT RPT$(' ',10);' 1200  1800  0000  0600  1200'
08500     ELSE   ! 8 HR GRAPH REQUIRES 0-24 SCALE
08510       PRINT RPT$(' ',10);' 0000  0600  1200  1800  0000'
08520     ENDIF
08530   ENDIF
08540     REM TIME
08550      PRINT RPT$ (' ',19);'START HOUR';RPT$(' ',29);'START HOUR'
08560   GO TO 8570
08570  NEXT Z
08580  GO TO 16500
08590
08600  !*****************************************************************
08610
08620   REM  THIS IS THE FUNCTION THAT FILLS THE PRINT LINE CONSTANTS
08630     DEF FN_FILL_PRINT_CONSTANTS
08640     IF RVALUE <= 0
08650        STAR$=' '  %  MONTH = 0  %  DAY = 0
08660        YEAR = 0  %     START_HOUR = 0   % END_HOUR = 0
08670        GOTO 9010
08680     ENDIF
08690     IF HOUR = 8 THEN END_HOUR =START_HOUR+8 ELSE END_HOUR = START_HOUR+1
08700     IF END_HOUR >= 24 THEN END_HOUR = END_HOUR - 24
08710     IF HOUR = 8 THEN S = DAYS_8 ELSE S = DAYS_1
08720     IF START_HOUR = 24 THEN START_HOUR = 0
08730
08740     REM LOOK FOR I FLAGS
08750     INTP$=''
08760     IF HOUR = 1
08770       S$=FLAGS$(PRINT_DAY,START_HOUR+3)
08780       IF S$='Q' OR S$='I' OR S$='C' THEN INTP$=' (I)'
08790     ELSE
08800       STARTHR=START_HOUR
```

```
08810       IF PP=5 AND STARTHR<5 THEN STARTHR=STARTHR+24
08820        FOR J = STARTHR+3 TO STARTHR+10
08830          IF J>26
08840            S$=FLAGS$(PRINT_DAY+1,J-24)
08850            IF S$='Q' OR S$='I' OR S$='C' THEN INTP$=' (I)'
08860          ELSE
08870            S$=FLAGS$(PRINT_DAY,J)
08880            IF S$='Q' OR S$='I' OR S$='C' THEN INTP$=' (I)'
08890          ENDIF
08900        NEXT J
08910      ENDIF
08920      IF INTP$=' (I)' THEN INTP_FLAG$ = 'ON'
08930      IF NOCTURNAL_SWITCH$ = 'ON'
08940        IF START_HOUR < 9
08950          PRINT_DAY = PRINT_DAY + 1  ! NOCTURNAL START HOUR IN NEXT DAY
08960        ENDIF
08970      ENDIF
08980      MONTH = HOLD(PRINT_DAY,1)
08990      DAY = HOLD ( PRINT_DAY,2)
09000      YEAR = HOLD ( PRINT_DAY,3)
09010      FNEND
09020
09030 !**********************************************************************
09040
09050  DEF FN_PRINT_LINE
09060    IF RVALUE <=0
09070      PRINT USING EMPTY$, PERIOD$, HOUR, 'INSUFFICIENT DATA'
09080    ELSE
09090     PRINT USING FORM1$, PERIOD$,HOUR,S,SUFF$,RVALUE,INTP$,MONTH,DAY&
09100        &,YEAR,START_HOUR, END_HOUR
09110    ENDIF
09120  FNEND
09130
09140 !**********************************************************************
09150
09160  REM THIS IS THE HELP MESSAGE
09170      OPEN #5, 'C08 WHELP', INPUT
09180       ON EOF GO TO 9220
09190      LINPUT #5, A$
09200      PRINT A$
09210      GO TO 9190
09220      CLOSE #5
09230      LINPUT CR$
09240      GO TO 1850
09250
09260 !**********************************************************************
09270
09280   REM THIS IS THE IO ERROR MESSAGE
09290    PRINT ' THAT IS NOT AN ACCESSIBLE FILE.  TRY AGAIN.'
09300    PRINT
09310   GO TO 1850
09320
09330 !***********************************************************************
09340
09350  DEF FN_PAGE_HEADING_BLOCK (PAGE)
```

```
09360     PRINT CHR$(12); RPT$(CHR$(0),65)
09370     PRINT HED10$;
09380     PRINT HED11$;PAGE
09390     PRINT HED12$
09400     PRINT HED13$
09410     PRINT
09420     PRINT HED14$; PROJECT_DESCRIPTION$
09430     PRINT HED15$; SITE_CODE$
09440     PRINT HED16$; DATE$;
09450     PRINT HED17$; FIL$
09460     PRINT % PRINT
09470  FNEND
09480
09490 !************************************************************
09500
09510  DEF FN_OUTLIER_HEADING_BLOCK
09520     PRINT IN$; 'OUTLIER SUMMARY:'
09530     PRINT % PRINT
09540     PRINT IN$;'VALUE'
09550     PRINT IN$;'(PPM)';RPT$(' ',10);'DATE';RPT$(' ',12);'TIME'
09560     PRINT IN$; RPT$('-',38)
09570  FNEND
09580
09590  !***********************************************************
09600
09610  DEF FN_OUTLIER (D,H)
09620    FLAGS$(D,H) = 'O'
09630    WORK (D,H) = -1
09640    OUTLIER_COUNT = OUTLIER_COUNT + 1
09650  FNEND
09660
09670 !*************************************************************
09680
09690  REM   THIS SUBROUTINE INTERPOLATES MISSING VALUES
09700  IF FLAGS$(D,H) <> 'T'
09710    IF WORK(D,H+1) <0
09720      IF WORK(D,H+2) < 0
09730        FOR P = 0 TO 2      ! FLAG THREE MISSING IN A ROW
09740         IF FLAGS$(D,H+P) = ''
09750          FLAGS$(D,H+P) = 'T'
09760         ENDIF
09770        NEXT P
09780      ELSE  ! 2 MISSING IN A ROW.  INTERP BY AVG AND INCREMENT
09790        IF WORK(D,H-1)>=0 AND WORK(D,H+2) >= 0
09800        INCREMENT = ABS(WORK(D,H-1) - WORK(D,H+2)) / 3
09810        IF WORK (D,H-1) > WORK (D,H+2)
09820          WORK(D,H) = IP(WORK(D,H-1) - INCREMENT + 0.5)
09830          WORK(D,H+1) = IP(WORK(D,H+2) + INCREMENT + 0.5)
09840          FLAGS$(D,H) = 'I'   % FLAGS$(D,H+1) = 'I'
09850        ELSE
09860          WORK(D,H) = IP(WORK(D,H-1) + INCREMENT + 0.5)
09870          WORK(D,H+1) = IP( WORK(D,H+2) - INCREMENT + 0.5)
09880          FLAGS$(D,H) = 'I'   % FLAGS$(D,H+1) = 'I'
09890        ENDIF
09900         ELSE
```

```
09910         FLAGS$(D,H)='T'
09920          FLAGS$(D,H+1)='T'
09930       ENDIF
09940     ENDIF
09950   ELSE   ! INTERPOLATE BY AVERAGING
09960     IF WORK (D,H-1)>=0 AND WORK(D,H+1)>=0
09970     WORK(D,H) = IP((WORK(D,H-1) + WORK(D,H+1))/2)
09980     IF FLAGS$(D,H)='O' THEN FLAGS$(D,H)='Q' ELSE FLAGS$(D,H)='I'
09990      ELSE
10000        FLAGS$(D,H)='T'
10010    ENDIF
10020   ENDIF
10030 ENDIF
10040   RETURN
10050
10060 !**********************************************************************
10070
10080  REM THIS SUBROUTINE PRINTS OUTLIER CHARTS
10090     IF CHART_COUNT >6
10100       I = FN_PAGE_HEADING_BLOCK (PAGE)
10110       PAGE = PAGE + 1
10120       CHART_COUNT = 1
10130       PRINT IN$;'SUMMARY OF OUTLIERS:'
10140     ENDIF
10150
10160 REM FILL VARIABLES
10170  REM DATE
10180  IF FP(CHART_COUNT/2) = 0
10190  DATE2$ = VALUE$(HOLD(D,1))+'/'+VALUE$(HOLD(D,2))+'/'+VALUE$(HOLD(D,3))
10200  ELSE
10210  DATE1$ = VALUE$(HOLD(D,1))+'/'+VALUE$(HOLD(D,2))+'/'+VALUE$(HOLD(D,3))
10220  ENDIF
10230     START_TIME = H-8
10240     IF START_TIME <= 0
10250      IF D>1
10260         IF CONSEQ_DAYS$(D) = 'YES'
10270              DAY = D - 1
10280              START_TIME = START_TIME + 24
10290         ELSE
10300              DAY = D
10310              START_TIME = 1
10320         ENDIF
10330       ELSE
10340         DAY = D
10350         START_TIME = START_TIME + 24
10360      ENDIF
10370     ELSE
10380       DAY = D
10390     ENDIF
10400     IF FP(CHART_COUNT/2)=0
10410        W=4   %  X=6   %  Y=5    ! CHART ON RIGHT SIDE
10420     ELSE
10430        W=1   %  X=3   %  Y=2    ! CHART ON LEFT SIDE
10440     ENDIF
10450
```

```
10460      REM FILL TIME
10470      HOUR = START_TIME
10480      FOR P = 0 TO 10
10490        IF HOUR > 23
10500           HOUR = HOUR - 24
10510           DAY = DAY + 1
10520        ENDIF
10530       IF HOUR < 10
10540          OUTLIER$(P+1,W) = '0'+ VALUE$( HOUR)
10550       ELSE
10560          OUTLIER$(P+1,W) = VALUE$(HOUR)
10570       ENDIF
10580       REM FILL OUTLIER AND INTERP AND ENVIRONMENT
10590       IF HOUR + 3 = H          ! TIME OF OUTLIER
10600         IF FLAGS$(D,H) = 'Q'        ! INTERPOLATED OUTLIER
10610           OUTLIER$(P+1,X) = VALUE$(WORK(DAY,H))
10620           OUTLIER$(P+1,Y) = '*' + VALUE$(HOLD(DAY,H+1))
10630         ELSE
10640           OUTLIER$(P+1,X) = 'NI'
10650           OUTLIER$(P+1,Y) = '*'+ VALUE$(HOLD(DAY,H+1))
10660           NI_FLAG$ = 'ON'
10670         ENDIF
10680       ELSE        ! NOT OUTLIER
10690         OUTLIER$(P+1,X) = ' '
10700         OUTLIER$(P+1,Y) = VALUE$(WORK(DAY,HOUR + 3))
10710       ENDIF
10720
10730        HOUR = HOUR + 1
10740     NEXT P
10750
10760       CHART_TOTAL = CHART_TOTAL + 1
10770
10780       IF FP(CHART_COUNT/2) <>0
10790         IF CHART_TOTAL <> OUTLIER_COUNT
10800         GO TO 11020
10810         ELSE
10820           PRINT RPT$(' ',15);DATE1$
10830           PRINT OUTLIER_COLUMN_1$
10840           PRINT OUTLIER_COLUMN_2$
10850           PRINT RPT$('-',37)
10860         ENDIF
10870       ELSE                    !BOTH CHARTS FULL
10880          PRINT RPT$(' ',17);DATE1$;RPT$(' ',33);DATE2$
10890          PRINT OUTLIER_COLUMN_1$;'     ';OUTLIER_COLUMN_1$
10900           PRINT OUTLIER_COLUMN_2$;'          ';OUTLIER_COLUMN_2$
10910         PRINT RPT$(' ',6); RPT$('-',28); RPT$(' ',11); RPT$('-',28)
10920    ENDIF
10930      FOR LL = 1 TO 11
10940       IF FP(CHART_COUNT /2) = 0   ! BOTH CHARTS
10950         PRINT USING OUTLIER_FORM_2$, OUTLIER$(LL,1),OUTLIER$(LL,2) ,&
10960          & OUTLIER$(LL,3),OUTLIER$(LL,4),OUTLIER$(LL,5),OUTLIER$(LL,6)
10970       ELSE    ! ONE CHART
10980   PRINT USING OUTLIER_FORM_1$, OUTLIER$(LL,1),OUTLIER$(LL,2),OUTLIER$(LL,3)
10990       ENDIF
11000      NEXT LL
```

```
11010     PRINT
11020     CHART_COUNT = CHART_COUNT + 1
11030 .   RETURN
11040
11050 !*********************************************************************
11060
11070  REM THIS IS THE CONVERSION ERROR MESSAGE
11080
11090  PRINT' THE SYSTEM HAS FOUND A NON-NUMERIC DATA ITEM IN THE '
11100  PRINT'WRONG PLACE IN YOUR DATA FILE.  PLEASE CHECK THE USER '
11110  PRINT'DOCUMENTATION OR THE HELP MESSAGE FOR INFORMATION ABOUT'
11120  PRINT'STRUCTURING DATA FILES.  CONSIDER THE POSSIBILITY OR AN *'
11130  PRINT'IN THE WRONG COLUMN.  WOULD YOU LIKE TO HAVE THE HELP'
11140  PRINT'MESSAGE DISPLAYED?  TYPE YES OR NO AND RETURN.'
11150
11160  INPUT CONVERSION_ERROR$
11170  IF CONVERSION_ERROR$ = 'YES' THEN GO TO 9160  ELSE GO TO 16500
11180
11190 !**********************************************************************
11200
11210 !  STATISTICAL ANALYSIS OF OUTLIER DATA
11220
11230 DIM ARRAY(1000),POS(1000)
11240 DIM CHI_TEST_05 (90)
11250 DIM GRUBBS_CHECK(90)
11260 DIM CNT (100)
11270
11280 ! CHI SQUARE VALUES AT 5%
11290 MAT READ CHI_TEST_05
11300 DATA 3.841, 5.991, 7.815, 9.488, 11.070
11310 DATA 12.592, 14.067, 15.507, 16.919, 18.307
11320 DATA 19.675, 21.026, 22.362, 23.685, 24.996
11330 DATA 26.296, 27.587, 28.869, 30.144, 31.410
11340 DATA 32.671, 33.924, 35.172, 36.415, 37.652
11350 DATA 38.885, 40.113, 41.337, 42.557, 43.773
11360 DATA 44.97, 46.17, 47.37, 48.57, 49.77
11370 DATA 50.97, 52.17, 53.37, 54.57, 55.76
11380 DATA 56.93, 58.10, 59.27, 60.44, 61.61
11390 DATA 62.78, 63.95, 65.12, 66.29, 67.50
11400 DATA 68.66, 69.82, 70.98, 72.14, 73.30
11410 DATA 74.46, 75.62, 76.78, 77.94, 79.08
11420 DATA 80.23, 81.38, 82.53, 83.68, 84.83
11430 DATA 85.98, 87.13, 88.28, 89.43, 90.53
11440 DATA 91.67, 92.81, 93.95, 95.09, 96.23
11450 DATA  97.37, 98.51, 99.65,100.79,101.88
11460 DATA 103.01,104.14,105.27,106.40,107.53
11470 DATA 108.66,109.79,110.92,112.05,113.14
11480
11490  ! CHI SQUARE VALUE AT 1%
11500    CHI_TEST_01 = 6.635
11510
11520  ! GRUBBS TEST VALUES AT 1%
11530  MAT READ GRUBBS_CHECK
11540 DATA 0, 0, 1.155, 1.492, 1.749, 1.944
11550 DATA 2.097, 2.221, 2.323, 2.410, 2.485
```

```
11560 DATA 2.550, 2.607, 2.659, 2.705, 2.747
11570 DATA 2.785, 2.821, 2.854, 2.884, 2.912
11580 DATA 2.939, 2.963, 2.987, 3.009, 3.029
11590 DATA 3.049, 3.068, 3.085, 3.103
11600 DATA 3.119, 3.135, 3.150, 3.164, 3.178
11610 DATA 3.191, 3.204, 3.216, 3.228, 3.240
11620 DATA 3.251, 3.261, 3.271, 3.282, 3.292
11630 DATA 3.302, 3.310, 3.319, 3.329, 3.336
11640 DATA 3.345, 3.353, 3.361, 3.368, 3.376
11650 DATA 3.383, 3.391, 3.397, 3.405, 3.411
11660 DATA 3.418, 3.424, 3.430, 3.437, 3.442
11670 DATA 3.449, 3.454, 3.460, 3.466, 3.471
11680 DATA 3.476, 3.482, 3.487, 3.492, 3.496
11690 DATA 3.502, 3.507, 3.511, 3.516, 3.521
11700 DATA 3.525, 3.529, 3.534, 3.539, 3.543
11710 DATA 3.547, 3.551, 3.555, 3.559, 3.563
11720
11730  ! E FUNCTION
11740  DEF FN_E(X) = A * EXP (B*X)
11750
11760
11770
11780 ! FILL WORKING ARRAY
11790   ! REPEAT 4 TIMES; ONCE FOR EACH PERIOD
11800   FOR PERIOD = 1 TO 4
11810    MAT ARRAY = (-99)
11820     ON PERIOD GO TO  11830,11830,11830,11870
11830    ! REM MORNING, MIDDAY, AND EVENING
11840      BEGIN_HOUR = PS(PERIOD+1)+2  % END_HOUR = PE(PERIOD+1)+2
11850      HOURS_IN_PERIOD = END_HOUR - BEGIN_HOUR + 1
11860      GO TO  12080! FILL ARRAY
11870    ! REM NOCTURNAL
11880      HOURS_IN_PERIOD = PE(5) - PS(5) + 1
11890      ! FILL ARRAY
11900      X=0
11910      NUMBER_OF_VALUES=0
11920      FOR D = 1 TO NUMBER_OF_RECORDS
11930        FOR H = PS(PERIOD+1)+2 TO 26
11940          X=X+1
11950          IF WORK(D,H)>0 THEN NUMBER_OF_VALUES = NUMBER_OF_VALUES + 1
11960          ARRAY(X) = INT(WORK(D,H)+.5)
11970        NEXT H
11980        IF D+1 < NUMBER_OF_RECORDS        ! SKIP LAST RECORD
11990          FOR H = 3 TO PE(PERIOD+1)-22
12000            X=X+1
12010          IF WORK(D+1,H)>0 THEN NUMBER_OF_VALUES=NUMBER_OF_VALUES + 1
12020            ARRAY(X) = INT(WORK(D+1,H)+.5)
12030          NEXT H
12040        ENDIF
12050      NEXT D
12060      GOTO  12200
12070
12080 ! FILL ARRAY FOR MORNING, MIDDAY, AND EVENING PERIODS
12090    NUMBER_OF_VALUES = 0
12100   X=0
```

```
12110   MAT ARRAY = (-99)
12120   FOR D = 1 TO NUMBER_OF_RECORDS
12130     FOR H = BEGIN_HOUR TO END_HOUR
12140       X= X+ 1
12150        IF WORK(D,H)=>0 THEN NUMBER_OF_VALUES = NUMBER_OF_VALUES + 1
12160        ARRAY(X) = INT(WORK(D,H) +.5)
12170     NEXT H
12180   NEXT D
12190
12200
12210  IF NUMBER_OF_VALUES<50 THEN GOTO 13220
12220
12230   ! SORT ARRAY
12240  MAT POS=DIDX(ARRAY)
12250  MAT ARRAY=DSORT(ARRAY)
12260
12270  ! INITIALIZATIONS
12280    UPPER_1 = INT( NUMBER_OF_VALUES / 100 +.5)
12290    UPPER_20 = INT( NUMBER_OF_VALUES / 5 + .5)
12300        LOOP
12310          IF ARRAY (UPPER_20) = ARRAY ( UPPER_20 + 1)
12320              UPPER_20 = UPPER_20 +1
12330           ELSE
12340              GO TO  12370
12350          ENDIF
12360        ENDLOOP
12370    IF UPPER_1 = 0 THEN UPPER_1 = 1
12380    UPPER_20_END =NUMBER_OF_VALUES-UPPER_20
12390    UPPER_1_END = NUMBER_OF_VALUES - UPPER_1
12400
12410
12420  ! COUNT THE NUMBER OF READINGS OF EACH VALUE IN UPPER 20%
12430  P=0  %  MAT CNT = (0)
12440  X=1
12450  LOOP
12460      IF ARRAY(X)>=0 THEN CNT(ARRAY(X)+1)=CNT(ARRAY(X)+1) + 1
12470      X=X+1
12480  UNTIL X > UPPER_20 - 1
12490  ! INSPECT UPPER 1% OF DATA FOR GAP
12500   FOR N = 1 TO UPPER_1
12510    IF ARRAY(N) - ARRAY(N+1) > 1
12520       GOSUB  12610! GAP TEST
12530       IF NO_OUTLIERS$ = 'ON' THEN GO TO 12560  !TOP VALUE NOT OUTLIER
12540    ENDIF
12550  NEXT N
12560 NEXT PERIOD
12570 GOSUB 13290      ! GRUBBS TEST
12580 GOSUB 13760      ! TIME SEQUENCE TEST
12590 RETURN
12600
12610 !**************************** GAP TEST *******************************
12620
12630  !  THIS SUBROUTINE CHECKS FOR THE LIKELIHOOD OF A GAP
12640  SUM_OF_VALUES = 0   % SUM_OF_SQ_VALUES = 0
12650  OXI_LOG = 0
```

```
12660  SUM_OF_LOGS = 0   %  SUM_2_PAR_1 = 0
12670  NUM_CLASSES = 0
12680  NO_OUTLIER$ = 'OFF'
12690
12700  ON ZDIV IGNORE
12710  ON OFLOW IGNORE
12720  ON UFLOW IGNORE
12730
12740   FOR X = N+1 TO UPPER_20
12750    ! DETERMINE VALUES REQUIRED TO COMPUTE 2 PARAMETER EXPONENTIAL FIT
12760    IF ARRAY(X) > 0
12770      IF ARRAY(X+1) <> ARRAY (X)
12780         SUM_OF_VALUES = SUM_OF_VALUES + ARRAY(X)
12790         SUM_OF_SQ_VALUES = SUM_OF_SQ_VALUES + ARRAY(X)**2
12800         NUM_CLASSES = NUM_CLASSES + 1
12810         SUM_OF_LOGS = SUM_OF_LOGS + LOG(CNT(ARRAY(X)+1))
12820         SUM_2_PAR_1 = SUM_2_PAR_1 + ARRAY(X) * LOG(CNT(ARRAY(X)+1))
12830      ENDIF
12840    ENDIF
12850  NEXT X
12860
12870  ! COMPUTE 2 PARAMETER EXPONENTIAL FIT
12880   B= (SUM_2_PAR_1-((1 /(NUM_CLASSES))*SUM_OF_VALUES*SUM_OF_LOGS))/ &
12890      & (SUM_OF_SQ_VALUES - ((1 /(NUM_CLASSES)) * SUM_OF_VALUES**2))
12900  IF B > 1E+70
12910    NO_OUTLIER$='ON'
12920     RETURN
12930  ENDIF
12940
12950  ! HYPOTHETICAL UPPER TAIL DISTRIBUTION
12960  A= EXP((SUM_OF_LOGS/(NUM_CLASSES))-(B*( SUM_OF_VALUES/(NUM_CLASSES))))
12970     IF A = 0 THEN A = 1E-15
12980
12990  CHI_SUM = 0
13000 FOR Z = N+1 TO UPPER_20
13010      IF ARRAY(Z+1) <> ARRAY(Z)
13020     CHI_SUM = CHI_SUM + ((CNT(ARRAY(Z)+1) - FN_E(ARRAY(Z))) **2)  &
13030     & / FN_E(ARRAY(Z))
13040      ENDIF.
13050 NEXT Z
13060  IF CHI_SUM > CHI_TEST_05 (NUM_CLASSES -1)
13070
13080     RETURN
13090 ENDIF
13100 ! COMPUTE CHI SQUARE VALUE FOR GAP
13110 CHI_SQUARE = FN_E(ARRAY(N+1)+1) - FN_E(ARRAY(N)) &
13120     &+((CNT(ARRAY(N)+1) - FN_E(ARRAY(N+1)+1)) **2 / FN_E(ARRAY(N+1)+1))
13130
13140
13150 IF CHI_SQUARE > CHI_TEST_01      ! OUTLIER
13160      FOR X=1 TO N
13170         D=1+IP((POS(X)-1)/HOURS_IN_PERIOD)
13180         H=BEGIN_HOUR+2+FP((POS(X)-1)/HOURS_IN_PERIOD)*HOURS_IN_PERIOD
13190        I=FN_OUTLIER(D,H)
13200      NEXT X
```

```
13210 ELSE
13220    NO_OUTLIER$ = 'ON'
13230 ENDIF
13240  ON ZDIV SYSTEM
13250  ON OFLOW SYSTEM
13260  ON UFLOW SYSTEM
13270 RETURN
13280
13290  !****************** GRUBBS TEST ******************************
13300
13310 ! THIS TESTS FOR THE ACCEPTIBILITY OF THE T STAT OF A VALUE
13320  FOR H = 3 TO 26
13330    ! INITIALIZATIONS
13340    SUM_OF_LOGS=0 % SUM_OF_SQ_LOGS=0 % NUMBER=0
13350    VMAX = -100
13360
13370    ! CALCULATE VALUES NEEDED FOR XBAR AND STANDARD DEVIATION
13380    FOR D = 1 TO NUMBER_OF_RECORDS
13390      IF WORK(D,H) > 0
13400         SUM_OF_LOGS = SUM_OF_LOGS + LOG(WORK(D,H))
13410         SUM_OF_SQ_LOGS = SUM_OF_SQ_LOGS + LOG(WORK(D,H))**2
13420         IF WORK(D,H) > VMAX THEN VMAX = WORK(D,H)
13430         NUMBER = NUMBER + 1
13440      ENDIF
13450    NEXT D
13460
13470    IF NUMBER < 3 THEN GOTO 13730
13480    ! CALCULATE MEAN AND STANDARD DEVIATION
13490    SUM_OF_LOGS_SQ = SUM_OF_LOGS **2
13500    XBAR = SUM_OF_LOGS / NUMBER
13510    IF SUM_OF_SQ_LOGS-(SUM_OF_LOGS_SQ/NUMBER)<0
13520      XXXX=0
13530    ELSE
13540      XXXX=SUM_OF_SQ_LOGS-(SUM_OF_LOGS_SQ/NUMBER)
13550    ENDIF
13560    STD_DEV=(XXXX/(NUMBER-1))**0.5
13570
13580    ! CALCULATE OUTER LIMIT VALUE
13590    DIFF = STD_DEV * GRUBBS_CHECK(NUMBER)
13600
13610    !CHECK VALUES AGAINST OUTER LIMIT
13620    IF LOG(VMAX)-XBAR <=DIFF  THEN GOTO 13730
13630    FOR D = 1 TO NUMBER_OF_RECORDS
13640    IF WORK(D,H) > 0
13650      HHLOG = LOG(WORK(D,H))
13660    ELSE
13670      GOTO  13720
13680    ENDIF
13690      IF HHLOG-XBAR > DIFF+0.0000000001
13700     I= FN_OUTLIER(D,H)
13710      ENDIF
13720    NEXT D
13730  NEXT H
13740 RETURN
13750
```

```
13760  !****************** TIME SEQUENCE TEST ***************************
13770
13780   ON ZDIV IGNORE
13790   ON OFLOW IGNORE
13800   ON UFLOW IGNORE
13810   FOR D = 1 TO NUMBER_OF_RECORDS
13820    FOR H = 3 TO 26
13830       FOR I= -2 TO 2
13840         IF WORK(D,H+I)<0 THEN GO TO 14030
13850       NEXT I
13860       DELTM=ABS(WORK(D,H-1)-WORK(D,H-2))
13870       DELTP=ABS(WORK(D,H+1)-WORK(D,H+2))
13880       IF DELTM<1 THEN DELTM=1
13890       IF DELTP<1 THEN DELTP=1
13900       XMAX=MAX(WORK(D,H-1)+DELTM,WORK(D,H+1)+DELTP)
13910       XMIN=MIN(WORK(D,H-1)-DELTM,WORK(D,H+1)-DELTP)
13920       IF WORK(D,H)<=XMAX AND WORK(D,H)>=XMIN THEN 14030
13930      IF WORK(D,H)>XMAX
13940         R=(WORK(D,H)-XMAX)/(WORK(D,H)-XMIN)
13950       ELSE
13960         R=(XMIN-WORK(D,H))/(XMAX-WORK(D,H))
13970       ENDIF
13980
13990      ! COMPARE R VALUE TO CRITICAL VALUE
14000      IF R > .5
14010      I= FN_OUTLIER (D,H)
14020      ENDIF
14030     NEXT H
14040   NEXT D
14050   ON ZDIV SYSTEM
14060 ON  OFLOW SYSTEM
14070 ON UFLOW SYSTEM
14080 RETURN
14090
14100 !*****************************************************************
14110
14120 ! SUBROUTINE TO DETERMINE IF NUMBER OF DAYS IS SUFFICIENT
14130 TOTAL = 1
14140 FOR D = 1 TO NUMBER_OF_RECORDS
14150   TOTAL=TOTAL*(1-PROB(D,PP,TT))
14160 NEXT D
14170 IF TOTAL>0.075 THEN SUFF$='$' ELSE SUFF$=' '
14180 IF SUFF$='$' THEN SUFF_FLAG$='ON'
14190 RETURN
14200
14210 !*********************************************************************
14220
14230 ! SUBROUTINE  TO CALCULATE DAY OF WEEK (DOW)
14240 DD=HOLD(D,2)
14250 IF HOLD(D,1)<=2
14260   YY=HOLD(D,3)+1899
14270   MM=HOLD(D,1)+13
14280 ELSE
14290   YY=HOLD(D,3)+1900
14300   MM=HOLD(D,1)+1
```

```
14310 ENDIF
14320 N=IP(365.25*YY+0.0001)+IP(30.6*MM+0.0001)+DD-621049
14330 DOW=FP(N/7)*7+1
14340 RETURN
14350
14360 !************************************************************
14370
14380 ! SHIP CO DATA TO AQDHS
14390     PRINT 'ENTER 4 DIGIT AREA CODE.     ';
14400     INPUT AREA_CODE$
14410     IF LEN (AREA_CODE$) = 4 THEN GO TO 14440
14420     PRINT'AREA CODE MUST BE 4 DIGITS. PLEASE RIGHT JUSTIFY AND ZERO FILL.'
14430     GO TO  14390
14440     PRINT 'ENTER 2 DIGIT PROJECT CODE.     ';  % INPUT PROJECT_CODE$
14450     IF LEN(PROJECT_CODE$)=2 THEN GO TO 14480
14460     PRINT' CODE MUST BE 2 DIGITS. PLEASE RIGHT JUSTIFY AND ZERO FILL.'
14470     GO TO 14440
14480  UF$ = '         ########'
14490 X = CMS('USERID')
14500 INPUT USING UF$,U$
14510 DIST$ = STR$(U$,2,1)
14520 !
14530 OK = CMS('STATE JOBCARD MVS A')
14540 IF OK <> 0 THEN 14600  !  NO FILE,  GO BUILD ONE
14550   OPEN #5,'JOBCARD MVS A',INPUT
14560   LINPUT #5,JOBCARD$
14570   CLOSE #5
14580   GOTO 14900
14590 !
14600 PRINT "      CAN'T FIND YOUR JOBCARD FILE."
14610 PRINT "      PLEASE ENTER THE FOLLOWING INFORMATION:"
14620 PRINT % PRINT "      TMS CODE ?";
14630 INPUT TMS$
14640 PRINT "      MSP CODE ?";
14650 INPUT MSP$
14660 PRINT "      LABOR COST CODE ?";
14670 INPUT LCC$
14680 PRINT "      YOUR NAME ?";
14690 INPUT NAME$
14700 IF IDX(NAME$,' ') = 0 THEN 14720
14710  NAME$="'"+NAME$+"'"
14720  IF LEN(NAME$) < 21 THEN 14750
14730  PRINT ' NAME MUST BE LESS THAN 20 CHARACTERS. RE-ENTER.'
14740  GOTO 14690
14750 JOBCARD$ = '//'+U$+' JOB (T'+DIST$+'100P,'+TMS$        &
14760   &         +','+MSP$+','+LCC$+'),'+NAME$
14770 PRINT % PRINT TAB(6);JOBCARD$
14780 PRINT "      OK (Y/N) ?";
14790 LINPUT AN$ % AN$ =STR$(UPRC$(AN$),1,1)
14800 IF AN$ = 'Y'
14810     OPEN #5,'JOBCARD MVS A',OUTPUT
14820     PRINT #5,JOBCARD$
14830     CLOSE #5
14840     PRINT "      'JOBCARD MVS' HAS BEEN ADDED TO YOUR ACCOUNT"
14850   ELSEIF AN$ = 'N'
```

```
14860       GOTO 14620
14870    ELSE
14880       GOTO 14770
14890 ENDIF
14900 !  WRITE JCL FOR SENDING DATA TO AQDHS
14910    OPEN #5,'COSHIP JCL O',OUTPUT
14920    PRINT #5,JOBCARD$
14930    PRINT #5,'//* MAIN  CLASS=D,REGION=190K'
14940    PRINT #5,'//* FORMAT PU,DDNAME=,DEST=TS1.'+U$
14950    PRINT #5,'//STEPADD EXEC TMENV00A'
14960    PRINT #5,'$2'    !    $=ACTION,  2=ADD
14970         !INPUT CARD FORMAT
14980         !COL          FIELD           DESCRIPTION
14990         !--------------------------------------
15000         ! 1           FORM CODE       ALWAYS 1
15010         ! 2-3         STATE CODE      ALWAYS 05
15020         ! 4-7         AREA CODE
15030         ! 8-10        SITE CODE
15040         ! 11          AGENCY CODE     ALWAYS F
15050         ! 12-13       PROJECT CODE
15060         ! 14          TIME CODE       1 FOR HOURLY DATA
15070         ! 15-16       YEAR CODE
15080         ! 17-18       MONTH CODE      1-12
15090         ! 19-20       DAY CODE        1-31
15100         ! 21-22       START HOUR      MUST BE 00 OR 12
15110         ! 23-27       PARAMETER CODE  42101 FOR CO
15120         ! 28-29       METHODS CODE    USUALLY 11 FOR CO
15130         ! 30-31       UNITS CODE      07 FOR PPM
15140         ! 32          DECIMAL CODE    0-4 FOR NUMBER OF DIGITS RIGHT
15150         !                             OF DECIMAL POINT
15160         ! 33-36       DATA
15170         ! 37-40         .             12 4 DIGIT DATA FIELDS
15180         !  .            .             RIGHT JUSTIFIED, 0 FILLED
15190         !  .            .
15200         ! 77-80         .
15210         !
15220    ! WRITE DATA TO FILE FOR AQDHS
15230       BEGIN_HOUR$ = '00'
15240       FOR D = 1 TO NUMBER_OF_RECORDS
15250         FOR K = 1 TO 2
15260           PRINT #5, '105'; AREA_CODE$; SITE_CODE$;
15270           PRINT #5, 'F'; PROJECT_CODE$; '1';
15280           PRINT #5 USING '##', HOLD(D,3);
15290           PRINT #5 USING '@##@@##@', HOLD(D,1), HOLD(D,2);
15300           PRINT #5 USING '##',BEGIN_HOUR$;
15310           PRINT #5, '4210111071';
15320           FOR H = VALUE(BEGIN_HOUR$)+4 TO VALUE(BEGIN_HOUR$)+15
15330             IF HOLD(D,H) = -1
15340                PRINT #5, '9999';
15350             ELSE
15360                PRINT #5 USING '@####@', HOLD(D,H) *10;
15370             ENDIF
15380           NEXT H
15390           PRINT #5
15400           IF BEGIN_HOUR$='00' THEN BEGIN_HOUR$='12' ELSE BEGIN_HOUR$='00'
```

```
15410        NEXT K
15420      NEXT D
15430 CLOSE #5
15440 X = CMS(' EXEC VNET SUBMIT COSHIP JCL O')
15450 RETURN
15460
15470 !*************************************************************
15480
15490 REM SUBROUTINE TO FILL PLOTTING MATRIX
15500 !  TWELVE: 1 = 12-12 SCALE, 0 = 0-0 SCALE
15510 !  FIRST : 1 = PRINT LEFT HALF OF GRAPH, 0 = NOPRINT
15520 !  SECOND: 1 = PRINT RIGHT HALF OF GRAPH, 0 = NOPRINT
15530
15540 IF STAR_HOUR<5
15550   TWELVE=1
15560   IF DAY<>1
15570     IF CONSEQ_DAYS$(DAY-1)='YES'
15580        DAY=DAY-1
15590        FIRST=1
15600        SECOND=1
15610        GOTO 15810
15620     ENDIF
15630   ENDIF
15640   FIRST=0
15650   SECOND=1
15660   GOTO 15810
15670 ELSEIF STAR_HOUR >15
15680   FIRST = 1
15690   TWELVE=1
15700   IF CONSEQ_DAYS$(DAY)='YES'
15710     SECOND=1
15720   ELSE
15730       SECOND=0
15740   ENDIF
15750 ELSE
15760   TWELVE=0
15770   FIRST=1
15780   SECOND=1
15790 ENDIF
15800
15810 STARS=0
15820 IF FIRST=1
15830   IF TWELVE=0 THEN START=4 ELSE START=16
15840   POST=0
15850   FOR H=START TO START+10
15860     POST=POST+1
15870     VV=WORK(DAY,H)
15880     IF STAR_HOUR=H-3 THEN STARS=NUM_STARS
15890     IF VV<0
15900       IF STARS>0 THEN STARS=STARS-1
15910       GO TO 16130
15920     ELSEIF VV=0
15930       IF STARS>0
15940         MARK$='*'
15950         STARS=STARS-1
```

```
15960       ELSE
15970         MARK$='.'
15980       ENDIF
15990       HORIZ_AXIS$(POST)=MARK$
16000       GOTO 16130
16010     ELSE
16020       FLAG$=FLAGS$(DAY,H)
16030       IF STARS>0
16040         MARK$='*'
16050         STARS=STARS-1
16060       ELSEIF FLAG$='I' OR FLAG$='Q' OR FLAG$='C'
16070         MARK$='I'
16080       ELSE
16090         MARK$='.'
16100       ENDIF
16110       PLOT$(VV,POST,G)=MARK$
16120      ENDIF
16130   NEXT H
16140 ENDIF
16150 IF SECOND=1
16160   IF TWELVE=0 THEN START=15 ELSE START=3
16170   IF TWELVE=1 AND FIRST=1 THEN DAY=DAY+1
16180   POST=11
16190   FOR H=START TO START+12
16200     POST=POST+1
16210     VV=WORK(DAY,H)
16220     IF STAR_HOUR=H-3 OR STAR_HOUR=H+21 THEN STARS=NUM_STARS
16230     IF VV<0
16240       IF STARS>0 THEN STARS=STARS-1
16250       GO TO 16470
16260     ELSEIF VV=0
16270       IF STARS>0
16280         MARK$='*'
16290         STARS=STARS-1
16300       ELSE
16310         MARK$='.'
16320       ENDIF
16330       HORIZ_AXIS$(POST)=MARK$
16340       GOTO 16470
16350     ELSE
16360       FLAG$=FLAGS$(DAY,H)
16370       IF STARS>0
16380         MARK$='*'
16390         STARS=STARS-1
16400       ELSEIF FLAG$='I' OR FLAG$='Q' OR FLAG$='C'
16410         MARK$='I'
16420       ELSE
16430         MARK$='.'
16440       ENDIF
16450       PLOT$(VV,POST,G)=MARK$
16460      ENDIF
16470   NEXT H
16480  ENDIF
16490  RETURN
16500  END
```

APPENDIX C - Derivation of Equations 5 and 6

APPENDIX C

Derivation of Equations 5 and 6

From a set of six ordered, independent values, $[x_1,...,x_i,...,x_6]$ with $x_i > x_{i+1}$, assume that a subset of r members is randomly chosen. The number of ways r members can be drawn is

$$_6C_r = \frac{6!}{r!(6-r)!} \quad .$$

Let x_m be any one of the six members of the original ordered set. Consider a second subset of values containing $\ell = 6-m$ members such that $x_i < x_m$ for all x_i contained in the subset. The number of ways of drawing r members from this subset is

$$_\ell C_r = \begin{cases} \frac{\ell!}{r!(\ell-r)!} \ , & \ell \geq r \\ 0 \ , & \ell < r \quad . \end{cases}$$

Similarly, the number of ways of drawing r members from the subset including x_m such that $x_i \leq x_m$ is $_{(\ell+1)}C_r$. The total number of ways of drawing a subset of r members containing x_m such that $x \leq x_m$ is $_{(\ell+1)}C_r - {}_\ell C_r$. This is the number of ways subsets of r members having x_m as a maximum value can be chosen from the original six member set. The probability of x_m occurring as the maximum value of a subset of r members is the ratio of the number of ways of drawing r members containing x_m as a maximum to the total number of ways of drawing r members. This can be written as

$$P(m|r) = \frac{_{(\ell+1)}C_r - {}_\ell C_r}{_6C_r} \quad .$$

Assume that two additional values, y_1 and y_2, are added to the original six member ordered set such that

$$\sum_{j=1}^{2} \sum_{i=1}^{6} D_{ij} = 2,$$

where D_{ij} is a dummy variable defined as

$$D_{ij} = \begin{cases} 1 , & x_i = y_j \\ 0 , & x_i \neq y_j \end{cases} .$$

The values y_1 and y_2 may be distributed 36 ways among the six member set. These arrangements can be subdivided with respect to x_m into the following six modes:

	Mode	Number of Possible Arrangements
1.	$y_1 > x_m$, $y_2 > x_m$	$(m-1)^2$
2.	$y_1 > x_m$, $y_2 = x_m$ or $y_1 = x_m$, $y_2 > x_m$	$2(m-1)$
3.	$y_1 > x_m$, $y_2 < x_m$ or $y_1 < x_m$, $y_2 > x_m$	$2(m-1)(6-m)$
4.	$y_1 = x_m$, $y_2 = x_m$	1
5.	$y_1 = x_m$, $y_2 < x_m$ or $y_1 < x_m$, $y_2 = x_m$	$2(6-m)$
6.	$y_1 < x_m$, $y_2 < x_m$	$(6-m)^2$

Letting $g=m-1$ and $\ell=6-m$, the total arrangements for the six modes can be shown to equal 36 as follows:

$$\begin{aligned}\text{Total Ways} &= g^2+2g+2g\ell+1+2\ell+\ell^2 \\ &= (g+\ell)^2+2(g+\ell)+1 \\ &= 36.\end{aligned}$$

The probability of occurrence for each mode is the ratio of the number of ways y_1 and y_2 can be arranged in accordance with the mode to the total number of ways y_1 and y_2 can be arranged for all modes. A specific probability for each mode, k, can also be assigned to the chances of drawing a subset of r members containing x_m as a maximum. This can be expressed in the following general form,

$$P(m|r)_k = \frac{1}{_8C_r}\left[(\ell_1+1)^{C_r} - (\ell_2)^{C_r}\right] .$$

where

$$\ell_1 = \ell + \sum_{j=1}^{2}({}_1D_j) \quad \text{and} \quad \ell_2 = \ell + \sum_{j=1}^{2}({}_2D_j)$$

with

$${}_1D_j = \begin{cases} 1, & x_m \geq y_j \\ 0, & x_m < y_j \end{cases}$$

and

$${}_2D_j = \begin{cases} 1, & x_m > y_j \\ 0, & x_m \leq y_j \end{cases} .$$

The probability of occurrence of the kth mode, P(k), and the probability (multiplied by ${}_8C_r$) of x_m being the maximum value of an r member subset for the kth mode, $P(m|r)_k$, are given below for each of the six modes:

Mode (k)	P(k)	$P(m\|r)_k \cdot {}_8C_r$
1	$g^2/36$	${}_{(\ell+1)}C_r - {}_{\ell}C_r$
2	$g/18$	${}_{(\ell+2)}C_r - {}_{\ell}C_r$
3	$g\ell/18$	${}_{(\ell+2)}C_r - {}_{(\ell+1)}C_r$
4	$1/36$	${}_{(\ell+3)}C_r - {}_{\ell}C_r$
5	$\ell/18$	${}_{(\ell+3)}C_r - {}_{(\ell+1)}C_r$
6	$\ell^2/36$	${}_{(\ell+3)}C_r - {}_{(\ell+2)}C_r$

The combined probability of a specific mode occurring and of x_m being the maximum value of an r member subset given that mode is the product $P(k) \cdot P(m|r)_k$. The total probability, $P(m|r)$, is a summation of the individual modal probabilities:

$$P(m|r) = \sum_{k=1}^{6} P(k) \cdot P(m|r)_k \quad .$$

Upon substitution, this yields Equation 6.