EPA-600/4-76-029c
June 1976

Environmental Monitoring Series

EMPIRICAL TECHNIQUES FOR ANALYZING AIR QUALITY AND METEOROLOGICAL DATA

Part III. Short-Term Changes in Ground-Level Ozone Concentrations: An Empirical Analysis

Environmental Sciences Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service. Springfield, Virginia 22161.

EPA-600/4-76-029c
June 1976

EMPIRICAL TECHNIQUES FOR ANALYZING AIR QUALITY AND METEOROLOGICAL DATA
Part III: Short Term Changes in Ground-Level Ozone Concentrations: An Empirical Analysis

by

Leo Breiman and W. S. Meisel
Technology Service Corporation
2811 Wilshire Boulevard
Santa Monica, California 90403

Contract No. 68-02-1704

Project Officer

Kenneth L. Calder
Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
ENVIRONMENTAL SCIENCES RESEARCH LABORATORY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Environmental Sciences Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

PREFACE

This is the third of a trilogy reporting work performed under EPA contract no. 68-02-1704 examining the potential role of state-of-the-art empirical techniques in analyzing air quality and meteorological data. The three companion reports are entitled as follows:

I. The Role of Empirical Methods in Air Quality and Meteorological Analyses

II. A Feasibility Study of a Source-Oriented Empirical Air Quality Simulation Model

III. Short-Term Changes in Ground-Level Ozone Concentrations: An Empirical Analysis

ABSTRACT

This volume reports the results of a preliminary exploratory analysis of ozone and ozone precursor data. Parcels of air were tracked mathematically as they traveled over a region of the Los Angeles basin. The one- and two-hour changes in ozone levels in a parcel were related to previous hourly readings in that parcel of reactive hydrocarbons, methane, nitrous oxide, nitrogen dioxide, ozone, solar radiation, and temperature data. The data were formed by spatial interpolation of the hourly pollutant and wind field measurements taken during the summer of 1973 at seven stations scattered over the Los Angeles basin.

Through a nonlinear, nonparametric regression technique, we found that virtually all of the predictive capability was contained in three variables:

(1) the current ozone level,

(2) the current solar radiation reading (Langleys per hour), and

(3) the current nitrogen dioxide level.

These three variables explained 71% of the variance in the next two-hour change in ozone and 60% of the variance of the noisier one-hour ozone change data.

A continuous piecewise linear multivariate approximation to the two-hour change data was used to explore and model the relationship between the ozone change and the current ozone, nitrogen dioxide, and solar radiation values, explaining 61% of the variance. The qualitative conclusions are that there are basically two regimes:

(1) Below the average of one-hour ozone levels, the ozone change is determined largely by the solar radiation and nitrogen dioxide levels, with larger values of these latter two related to larger values of the ozone change. The largest positive changes in ozone occur in this regime.

(2) At above-average one-hour ozone levels, the ozone level has a strong negative association with ozone change, and moderate to high levels of nitrogen dioxide and solar radiation are associated with low to moderately above-average changes in ozone.

The specific equations of the empirical model provide a quantitative statement of the relationship which produces a correlation between predicted and actual values of 0.78 over 1800 samples. This approach can be extended to the derivation of a full set of empirical difference equations for the main chemical species, incorporating emissions.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

The suggestions of EPA reviewers, particularly the project monitor, Kenneth L. Calder, have led to significant improvements in the clarity of this report.

Without the able aid of Saul Miller, William Liles, and Mike Teener of Technology Service Corporation in programming and data base preparation, this report literally could not have been completed.

1. INTRODUCTION

The intent of this study was to examine the possibility of deriving empirical difference equations that would describe the production of ozone (O_3) in the urban environment. One- and two-hour changes in ozone concentration were related to the concentrations of certain precursor pollutants and meteorological variables. As part of a three-part study of the role of empirical approaches in air quality and meteorological applications, the present study was limited in scope to an exploratory analysis of this concept. Accordingly, a difference equation for ozone alone was examined. Further, limitations on the scope of the study and the data available prevented consideration of a number of meteorological variables that could potentially be important. Emissions measurements were not available in the data base constructed for this exploratory analysis and were not used. Hence, the objective of this study was to obtain an indication of the degree to which the observed data could be explained despite the limitations indicated.

In the remainder of this introduction, the ozone problem is reviewed and the study summarized.

1.1 MOTIVATION

Ozone and its precursors have been investigated, studied, and analyzed extensively in the past five years. Since A. J. Haagen-Smit revealed his findings on ozone formation in photochemical oxidation of organic substances some twenty years ago, the subject of oxidant pollution has evolved from a local phenomenon in the Los Angeles area to a problem of national

importance. The air pollution literature abounds with articles by chemists, meteorologists, mathematical modelers, statisticians, and research scientists that discuss the chemical formation, transport, and destruction of oxidants in the troposphere and the stratosphere on spatial scales ranging from micro to synoptic.

Currently, the motor vehicle is the principal source of oxidant precursors in urban atmospheres. The emissions of hydrocarbons and oxides of nitrogen from motor vehicles augmented by industrial sources and fossil fuel power plants in our metropolitan complexes appear to be the source of oxidant concentrations some 50 to 100 miles away.

In recent years, several field studies of ozone and/or its precursors have been conducted in nonurban areas. Although these areas are remote from large population centers, ozone concentration levels exceed the national ambient standard of 160 $\mu g/m^3$ (0.08 ppm). At such times, it appears that meteorological factors such as fronts, low-level stability, mixing height, and high-pressure systems are directly related to observed ozone concentrations.

Excessive photochemical air pollution, characterized by ozone, is the source of eye irritation, headaches, and chest discomfort. Consequently, well founded oxidant control strategies are necessary to minimize the impact of ozone and its precursors. Such strategies cannot be developed without a clear understanding of the causes of high ozone concentration in rural and urban atmospheres.

There has been a great deal of research on the oxidant problem. The bibliography of this volume was limited to articles published after 1969.

There are more than fifty of them. Our review does not include photochemical models since a very thorough discussion of such work was recently published by Roth [45].

Many of the characteristics of ozone formation and transport are controversial. Some conclusions and claims reported in the literature follow:

(1) Oxidant air pollution may be characterized as the result of a complex series of photochemical reactions stemming from reactive hydrocarbon (RHC) and nitrogen oxide (NO_x) emissions. The rate and extent of the photochemical reactions are affected by precursor concentrations, solar radiation (SR), residence time in the air and current oxidant concentration [9,10,11,18,22,40,42].

(2) Several meteorological factors have been related to the spatial and temporal changes in observed oxidant concentrations. The important meteorological parameters are (a) temperature at the top of the inversion, (b) surface wind speed, (c) the intensity of solar radiation, (d) maximum surface temperature, (e) depth of the mixing layer, and (f) the water vapor mixing ratio [1,2,10,19,27,31,35,48].

(3) During weather periods that are conducive to air quality stagnation and thus to the onset of episode conditions, the buildup of oxidant concentrations aloft in the inversion layer is suspected of contributing substantially to the observed increases in surface concentrations. Topographical features such as land-sea contrasts and mountain-valley altitude differences play a fundamental role in these situations [3,5,6,25,26].

(4) Urban oxidant pollution is transported by the wind field over distances as great as 100 miles and adds to the pollution of rural communities [7,8,12,13,14,16,17,33,43,46,50,52].

(5) Results of field studies across the country have shown that local ozone control strategies are unlikely to succeed. In general, broad-scale regional oxidant control measures are required to meet NAAQS levels [18,19,30,33,50,52]. There are indications that ozone and/or its precursors can be transported over 1000 km to give readings in excess of 0.1 ppm.

(6) The chemistry of oxidant formation in the atmosphere is more complex than that suggested in the smog chamber. Some researchers claim that ozone formation is significantly influenced by aerosol-radical interactions and thus that photochemical models that omit heterogeneous chemistry cannot accurately represent the chemistry of the urban atmosphere [18,20,33,29,41].

(7) Sulfur compounds, CO and CH_4, are unimportant species in urban photochemistry [18].

(8) Even though there is a distinct reduction in the emission of primary pollutants between the hours of 5 a.m. and 1 p.m. on weekends as compared to weekdays, there is no corresponding reduction in O_3 concentrations. This is known as the "weekend effect." It is supportive evidence that the proportional mix, and not only absolute amounts of primary pollutants, is of importance in oxidant formation [15,44,53].

(9) In the Los Angeles area, the reduction in hydrocarbon emissions, due to automotive controls, in combination with the increase in NO_x emissions has contributed to the decrease in oxidant concentration [24,34].

It is obvious that the oxidant formation and transport process is quite complex. Unraveling the sometimes contradictory conclusions of researchers to obtain a clear insight into the effect of controls will not be simple to do. The present study uses available data to perform an exploratory analysis of the amount of information required to explain observed data. In the following subsection, that research is summarized.

1.2 TECHNICAL APPROACH

The intent of the present study is not to obtain definitive results on the process of ozone formation but to test a technical approach that has the potential of clarifying some of the issues and providing a relatively simple empirical model for estimating the impact of control measures. There are two key components of the technical approach.

(1) The development of empirical difference equations for one- and two-hour changes in a pollutant as functions of the last measured values of the pollutant, meteorological factors, and emissions. This approach is used to model the chemistry of the process.

(2) An interpolation approach to estimating the effect of transport and dispersion. An approximate trajectory is used to estimate the precursor concentrations.

It is beyond the scope of the present study to formulate these approaches in a full theoretical framework. We will, however, attempt to provide an intuitive framework.

Empirical Difference Equations

Suppose a set of differential equations describing a physical process was given by Equation 1:

$$\frac{dC_1}{dt} = F_1[C_1, C_2, \ldots, C_R; u_1, u_2, \ldots, u_S]$$

$$\frac{dC_2}{dt} = F_2[C_1, C_2, \ldots, C_R; u_1, u_2, \ldots, u_S]$$

$$\vdots$$

$$\frac{dC_R}{dt} = F_R[C_1, C_2, \ldots, C_R; u_1, u_2, \ldots, u_S] \quad . \tag{1}$$

where

$C_1, C_2, \ldots, C_R$ are state variables and

$u_1, u_2, \ldots, u_S$ are input variables.

(In the present context, C_i could be regarded as a concentration of a particular pollutant and u_i as an external influence such as a meteorological variable or an emission.)

In implementing such equations on the computer, one usually uses a numerical integration technique. One such technique is Euler integration, where the derivative is in essence approximated as a difference over a small time interval Δt:

$$\frac{dC_i}{dt} \approx \frac{C_i(t + \Delta t) - C_i(t)}{\Delta t} \quad . \tag{2}$$

Replacing the derivatives in Eq. (1) with an approximation such as in Eq. (2) results in a set of difference equations, for example,

$$\Delta C_i(t) = G_i[C_1(t), \ldots, C_R(t); u_1(t), \ldots, u_S(t)] \quad , \tag{3}$$

where

$$\Delta C_i(t) = C_i(t + \Delta t) - C_i(t) \quad .$$

Since the values of the state variables and the external variables are calculated only at discrete intervals, the result of the difference equations is only an approximation to the result given by the differential equations. In this context, the smaller Δt, the more accurate the approximation to the differential equation.

However, if one begins with the point of view that the intent is to obtain an empirical difference equation of the form of Eq. (3), the focus of the problem shifts. In the empirical case, the difference ΔC as well as values of C_i and u_i are obtained from measurements. This encourages use of a considerably larger interval Δt for the following reasons:

(1) Data is measured at discrete time intervals, and most historical data is available for large time increments such as one hour.

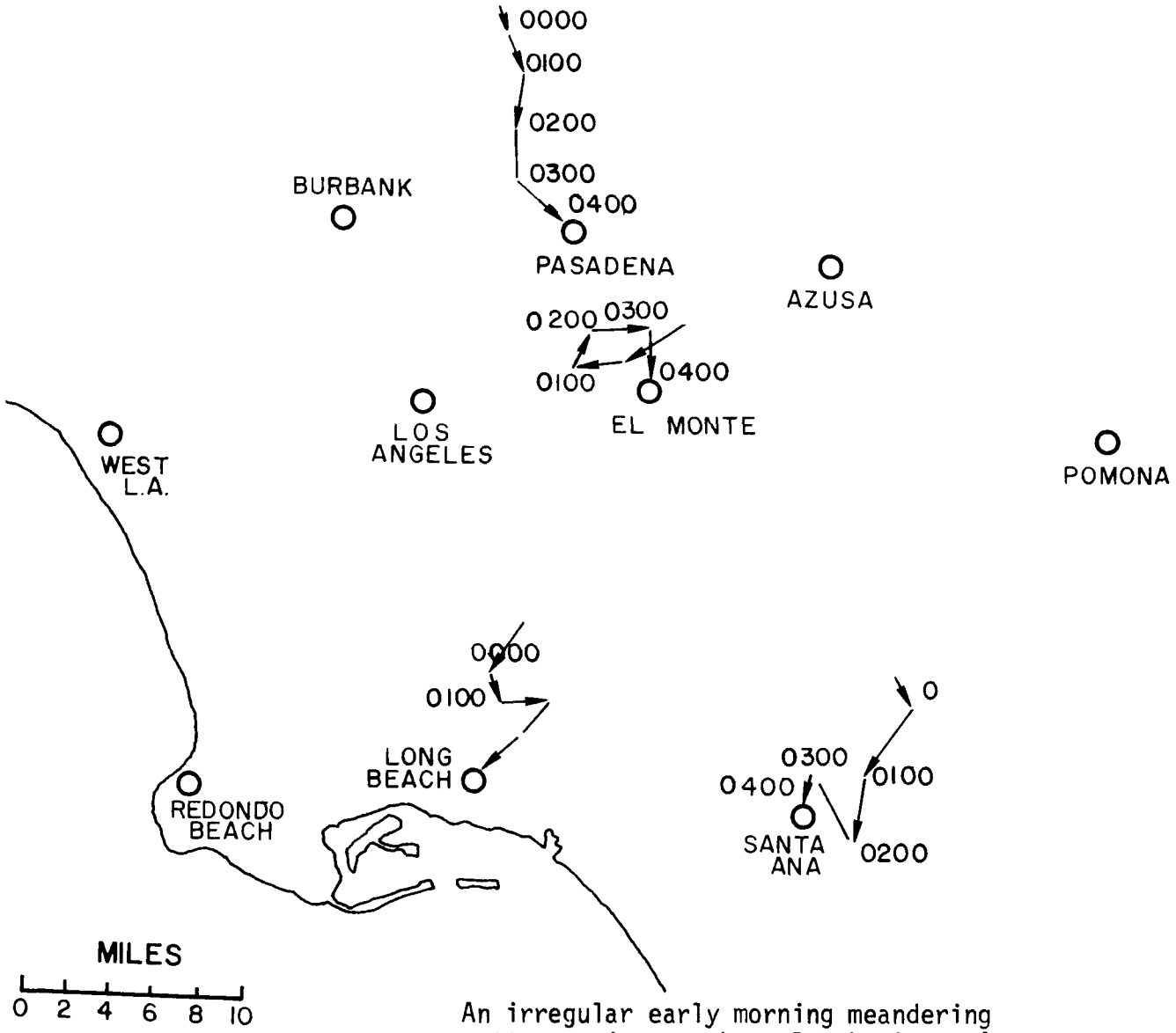
(2) Measuring short-term changes is similar to measuring derivatives of a continuous signal. The shorter the interval, the more susceptible the measurement is to "noise." (A brief discussion of this question is given in more formal terms in the Appendix.)

(3) It may require a much larger set of state variables and external variables to describe very short-term changes, even if the data were available. If the empirical model is to be used in a context where longer-term changes are of most interest, then attempting to explain those changes alone (rather than as a sequence of smaller changes) can potentially allow the use of fewer state variables at the expense of a less complete explanation of the process.

To implement this approach in full, difference equations for all the major pollutants and precursors on which measurements were available would be derived empirically. In the present exploratory analysis, we have simply chosen one of these equations (for ozone) and studied the degree to which we can relate the one-hour and two-hour changes in this species to precursor pollutants and to some meteorological variables. Emissions measurement were not easily available and could not be included within the scope of this study, nor could all potentially important meteorological variables be considered. The context in which the reader should then interpret the results is as the degree to which the change in ozone can be explained <u>despite</u> these limitations. Whatever degree of explanation of the variance in one- or two-hour changes in ozone we can achieve within these limitations can be improved when more of the omitted factors are taken into account. This analysis will thus provide a pessimistic estimate of the degree of success that can be expected in a full-scale implementation of the approach.

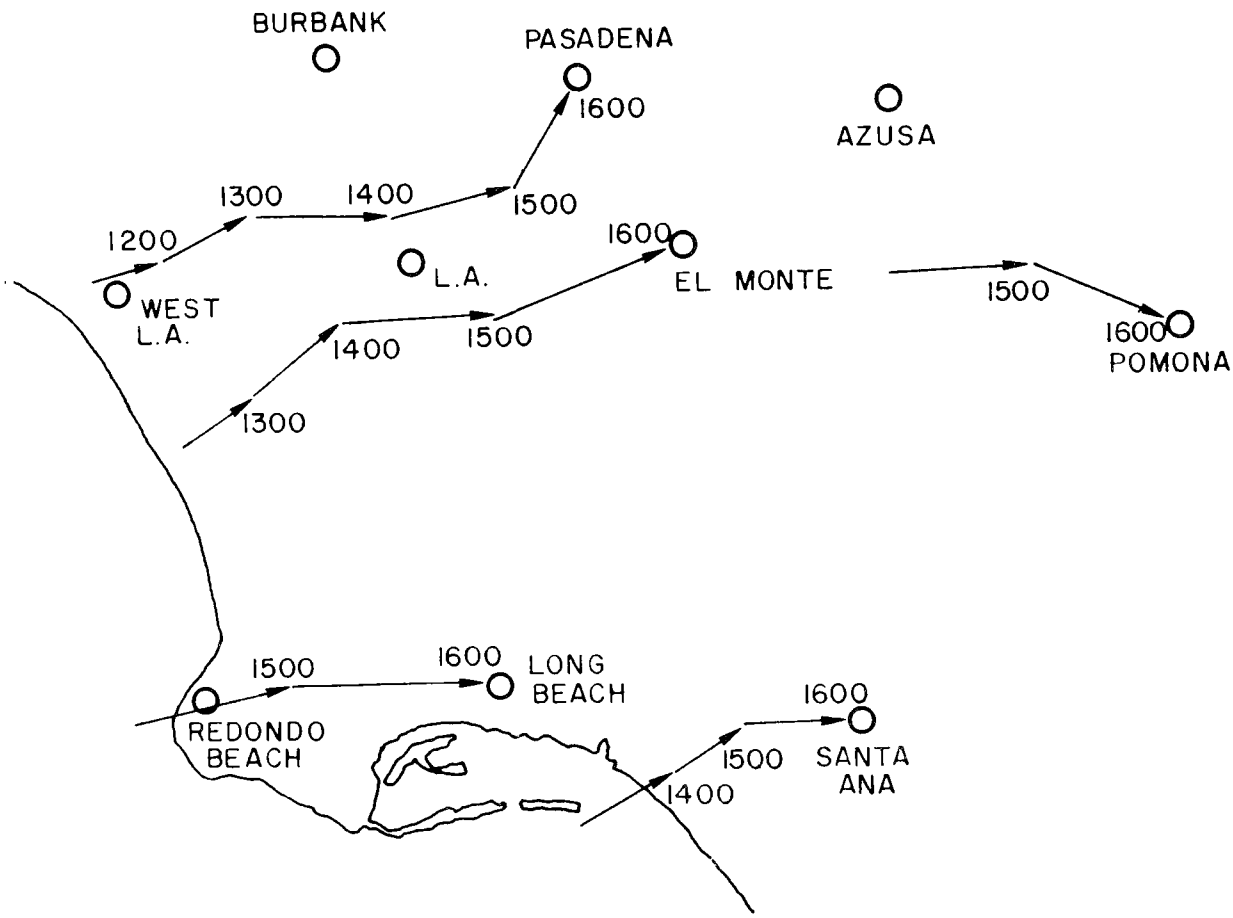
<u>Estimating Transport and Dispersion Effects</u>

Because of transport and dispersion, the concentration of ozone at a given location may result from concentrations of precursor pollutants occurring earlier in the day at other locations. We thus include in the anal ysis consideration of the transport and dispersion mechanism. For this exploratory analysis, we have adopted a rather simple model. The model estimates the trajectory of a "parcel" of air from ground-level measurements of the wind field. As illustrated in Figures 1 and 2, a parcel arriving at a given location at a given time (e.g., Pasadena at 1600 hours) is estimated, from the current wind direction, to have been at another location



An irregular early morning meandering pattern exists at Long Beach, Santa Ana, and El Monte. Pasadena, on the other hand, shows a northerly flow pattern due to nocturnal air drainage down the mountains combined with an offshore wind flow. The lengths of the arrows give an indication of how much the air has moved during an hour interval. None of the stations show more than 4 mph air movement for the early morning hours.

Figure 1. Estimated trajectory of air arriving at Pasadena, El Monte, Long Beach, and Santa Ana at 0400 September 29, 1969.



All of the stations show the dominance of onshore sea breezes with a tendency of higher velocities later in the afternoon. The more regular air trajectories of afternoon also show a greater air movement than early in the morning as shown by the greater lengths of the arrows. The deflecting influence of the Santa Monica mountains causes the air trajectory to curve northward as it approaches Pasadena.

Figure 2. Estimated trajectory of air arriving at Pasadena, El Monte, Long Beach, Santa Ana, and Pomona at 1600 September 29, 1969.

one hour earlier. The distance traveled from that direction is given by the current wind speed. The trajectory is tracked backwards to give a sequence of hourly locations. The values of pollutant levels at these points at the given times are obtained by an interpolation procedure described in the body of the report.

Figure 3 illustrates the basis for interpreting the trajectory estimation procedure statistically. It is important to emphasize that we are not assuming that this trajectory represents physical fact or that only ground-level wind measurements are important; we are simply taking an approximate approach to the dispersion/transport problem and attempting to determine how far this simple approach will carry us. To the degree that this approach is a reasonable average approximation to the effects of transport, it will suffice. This certainly does not imply that a more elaborate approach might not produce better results.

1.3 SUMMARY OF THE STUDY

We "tracked" parcels of air as they traveled over a region of the Los Angeles basin. We then related the one- and two-hour changes in O_3 levels in a parcel to previous hourly readings in that parcel of reactive hydrocarbons, methane, nitric oxide, nitrogen dioxide, ozone, solar radiation, and temperature data. The data were formed by interpolating the hourly pollutant and meteorological measurements taken during the summer of 1973 at seven stations scattered over the Los Angeles basin. However, the hourly solar radiation data were available only at Los Angeles Civic Center, and hourly temperature was available at only three locations in the region. Emission data were not used, so the emissions into a parcel of air during the change period produce part of the unexplained variance in the data.

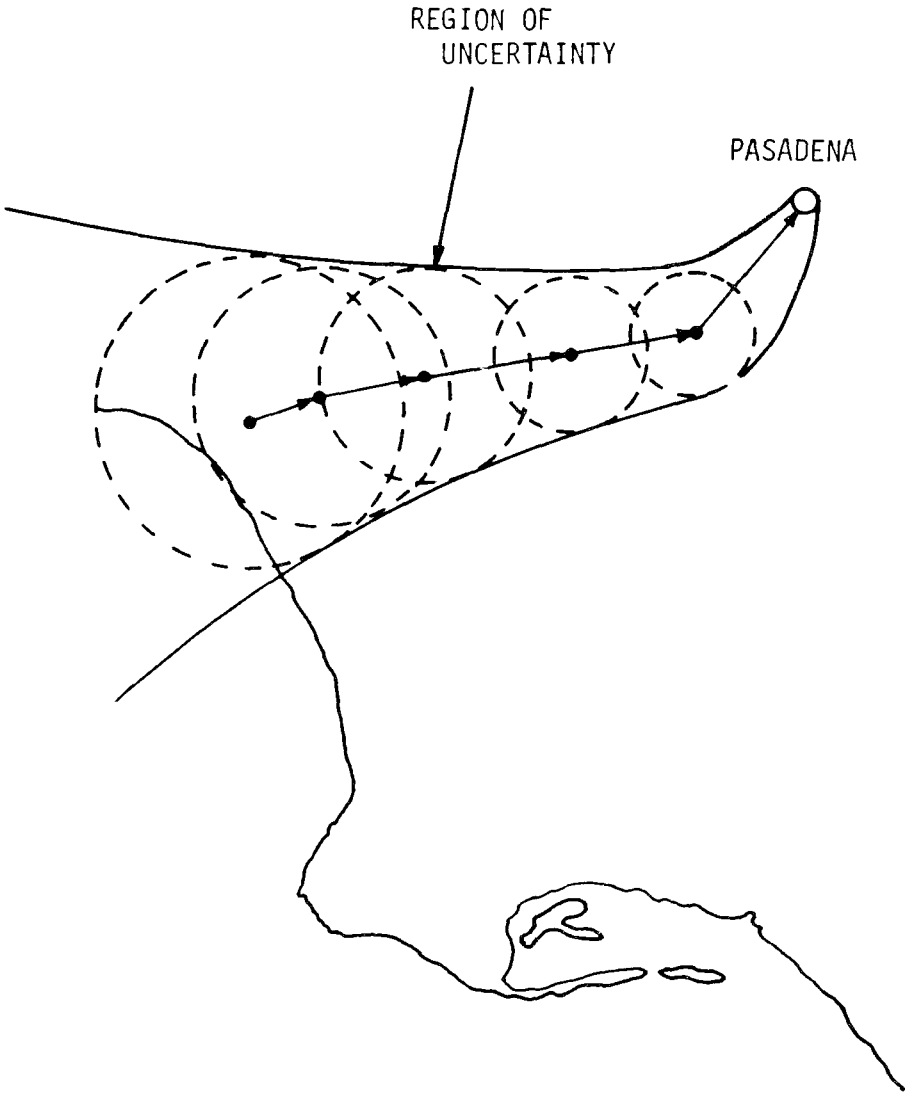


Figure 3. An illustration of the basis of a statistical interpretation of the trajectory estimation procedure.

The project was carried out in two phases: First, we attempted to discover which of the variables were most significantly related to the O_3 changes, and how strong the relationship was. Secondly, we proceeded to model and explore the functional relationship between the O_3 changes and the variables giving the best predictive capability.

Through extensive use of a nonlinear, nonparametric, exploratory regression technique developed by TSC [8], we found that virtually all of the predictive capability was contained in three variables,

- the current O_3 level
- the current solar radiation reading
- the current NO_2 level.

These three variables explained 71% of the variance in the next two-hour change in O_3 and 60% of the variance of the noisier one-hour O_3 change data. Adding other variables--such as current NO; current temperature; reactive hydrocarbons, current and three, four, and five hours earlier; and current methane--produced little additional increase in percent of variance explained.

We believe these results are quite gratifying; particularly since (1) we had to use the same current one-hour solar radiation data for all parcels of air over a region about 30 miles long by 10 miles wide and stretching from near the Pacific Ocean to hot, dry, and clear areas, and (2) for the interpolation over this region, data were available from only seven stations.

A TSC technique of fitting a continuous piecewise linear approximation to the data was used to both explore and model the relationship between the O_3 change and the current O_3, NO_2, and SR values. The qualitative conclusions are that there are basically two regimes:

- At below average O_3 levels, the O_3 change is determined largely by the SR and NO_2 levels, with larger values of these latter two related to larger values of the O_3 change. The largest positive changes in O_3 occur in this regime.
- At above average O_3 levels, the O_3 has a strong negative association with O_3 change, and moderate to high levels of NO_2 and SR are associated with low to moderately above-average changes in O_3.

The continuous function prediction $F(O_3, NO_2, SR)$ produced in this latter part of the study is fairly simple, consisting of eight hyperplane segments patched together in a continuous way. As a predictor of the two-hour O_3 change, it explains 61% of the variance. In view of its simplicity, we regard this as a promising approach to the data-analytic modeling of ozone production. A fuller account of the study follows.

2. PROJECT PURPOSE AND DESIGN

The purpose of this study was to perform a data analysis of the formation of ozone from its precursors, and to determine the effects of solar radiation (SR) and temperature (T) on the formation process.

To do this, we interpolated the wind field in a region of the Los Angeles basin so that we were able to track parcels of air as they moved through the basin. The pollutant readings at seven APCD stations were also interpolated so that we could keep hourly records of non-methane hydrocarbons (HC), methane (CH_4), ozone (O_3), and the nitrous oxides (NO, NO_2) along the trajectory of the parcel. We also had the hourly solar radiation readings at the Los Angeles Civic Center location of the APCD, hourly temperature readings at three representative locations in the basin, and the mixing height estimated for the Civic Center location. This data base was created largely from APCD data and reformatted for this study.

Our study was carried out over the five summer months, June through October 1973. About 7000 trajectories were formed and placed in the primary data base.

At each hour along each trajectory, we then extracted the following data:

O_3 for that hour (pphm)
NO " " " (pphm)
NO_2 " " " (pphm)
HC " " " (ppm)

CH_4 for that hour (PPM)

SR " " " (gm. cal./cm^2/hr)

T " " " (°F)

and ΔO_3, the change in O_3 over the next hour. Basically, we wanted to find a function F_1 such that the equation

$$\Delta O_3 = F_1(O_3, NO, NO_2, HC, CH_4, SR, T) \tag{4}$$

is a nearly "best" fit to the data.

Although mixing height was available in our data base, we did not use it as one of the independent variables since we did not have reliable estimates of its rate of change. Wind speed and direction were implicit variables since we were tracking the parcel of air. Furthermore, emissions data were not available.

Because we suspected that the hourly changes in O_3 might be subject to considerable noise, we also used a data base with ΔO_3 equal to the two-hour change in ozone and the other variables as above, except that SR is replaced by a two-hour SR total. (See Appendix for a discussion of this point.) We planned to use this two-hour data base to express the two-hour change in ozone level as

$$\Delta O_3 = F_2(O_3, NO, NO_2, HC, CH_4, SR, T) \tag{5}$$

Perhaps even more important, we wanted to determine which of the variables were most significantly related to the change in ozone. Therefore, we really had two objectives in this study:

(1) To find those subgroups of the variables most significantly related to the ozone change.

(2) To find the form of the function F providing the best fit to the data.

The remainder of the report will cover:

(1) A description of the interpolation program which produced the trajectories and pollutant values along the trajectories.

(2) The exploratory phase and the winnowing out of the significant variables.

(3) The relationship of ΔO_3 to the three variables O_3, NO_2 and SR which was found to be significant.

3. THE TRAJECTORY INTERPOLATION PROGRAM

In 1972, nine air quality monitoring stations in the greater Los Angeles Basin started to make hourly measurements of HC (hydrocarbons minus methane), CH_4 (methane), NO, and NO_2, along with the usual hourly pollutant measurements. Only seven of the nine stations are in the wind funnel of the basin. These are located at:

Pomona
Azusa
Pasadena
Burbank
Lennox
Whittier
Los Angeles Civic Center

These locations are shown in Figure 4. When connected by lines, they form a convex polygon with only the Los Angeles Civic Center location in the interior. All of these stations are also wind measurement stations, but the wind data at Lennox was not available. We assigned to Lennox wind data at Los Angeles County APCD sampling station #76, less than three-quarters of a mile away over flat terrain.

The area included in the polygon formed the area under study in this project. It was triangulated as shown in Figure 4 with an extra point added near Pomona that formed an additional triangle. The reason for adding this extra point will be explained below. This point was assigned pollutant levels and wind data identical to that at Pomona.

Given that a point was in any triangle, the pollutant levels were interpolated to that point from the measurements at the three vertices in a straightforward way: using the three values of the variable being interpolated at the vertices as heights, a plane was passed through the three

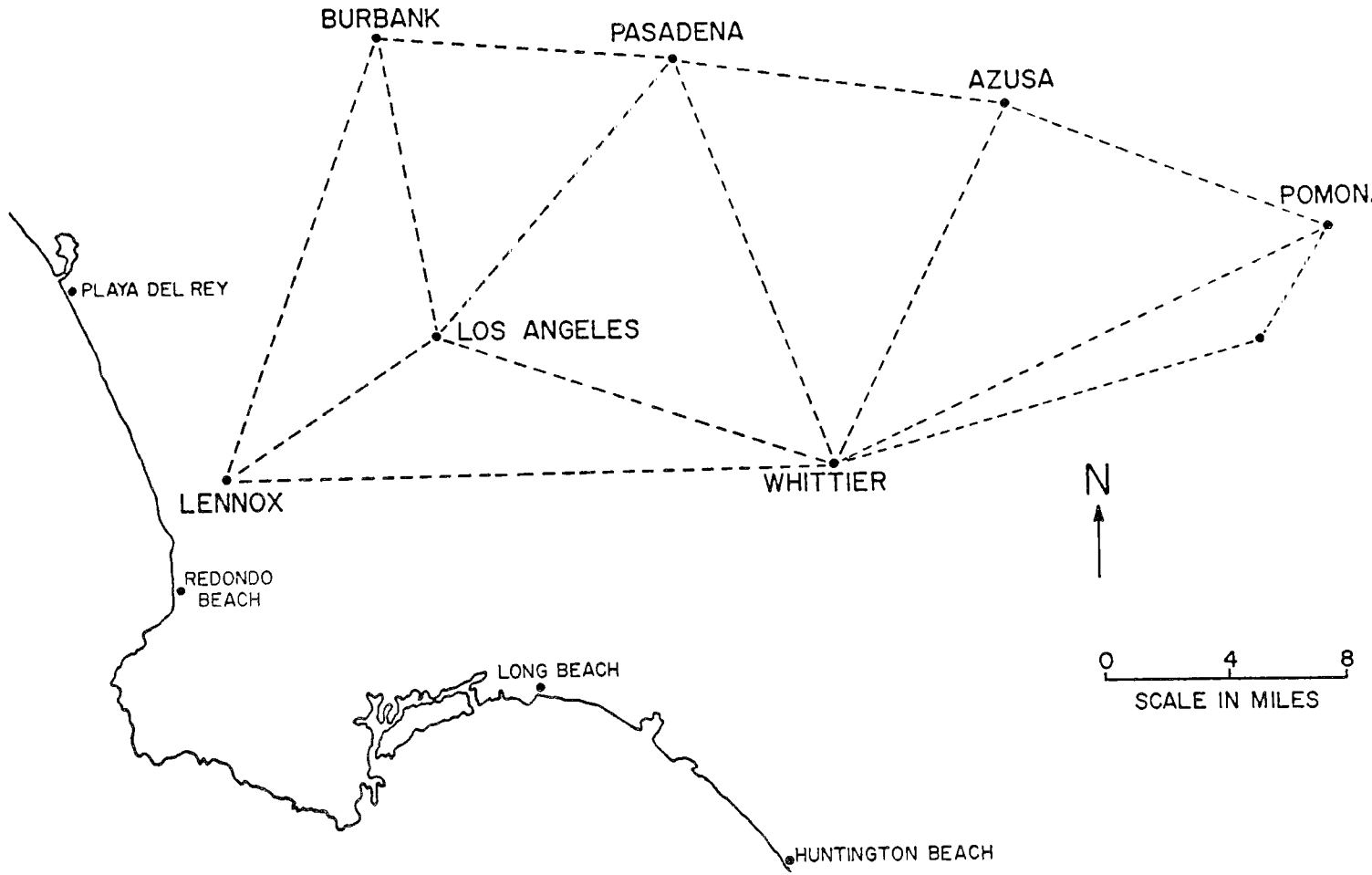


Figure 4. The study region.

points and the value of the variable at the interior point was the height of the plane above the point in question. (See Figure 5.)

At each station the wind speed and direction were combined to find the components of the wind W_N, W_E in the north direction and the east direction. Then W_N was interpolated over the region in exactly the same way as the pollutant interpolation. The same was true for W_E.

Using this interpolation scheme, we could start with any point $\underline{x}_1$ in the polygon, get pollutant levels and wind speeds at $\underline{x}_1$ for any given hour, and then use the wind speed at $\underline{x}_1$ to find the location of the air mass the hour before. Calling this prior location $\underline{x}_2$, we repeated the procedure, finding pollutant values and wind speeds at $\underline{x}_2$.

In this way we built up a trajectory

$$\underline{x}_1 \leftarrow \underline{x}_2 \leftarrow \underline{x}_3 \leftarrow \underline{x}_4 \leftarrow \cdots$$

that terminated when we found a prior point outside of the polygon.

The prevailing wind pattern for these five summer months is southerly near the coast, is southwesterly near downtown Los Angeles, and curves to westerly near Pomona.

We used trajectories ending at Burbank, Los Angeles, Pasadena, Azusa, and Pomona. The prevailing winds were such that the trajectories tended to stay in the polygon until they exited by crossing over the line between Lennox and Whittier. The extra triangle was added near Pomona so that the first point prior to Pomona would not be out of the polygon.

We had additional meteorological data:

- hourly temperature measured at Lockheed Airport--Burbank, L.A. Civic Center, and Ontario Airport
- hourly solar radiation measured at L.A. Civic Center
- daily mixing height computed for the L.A. Civic Center location

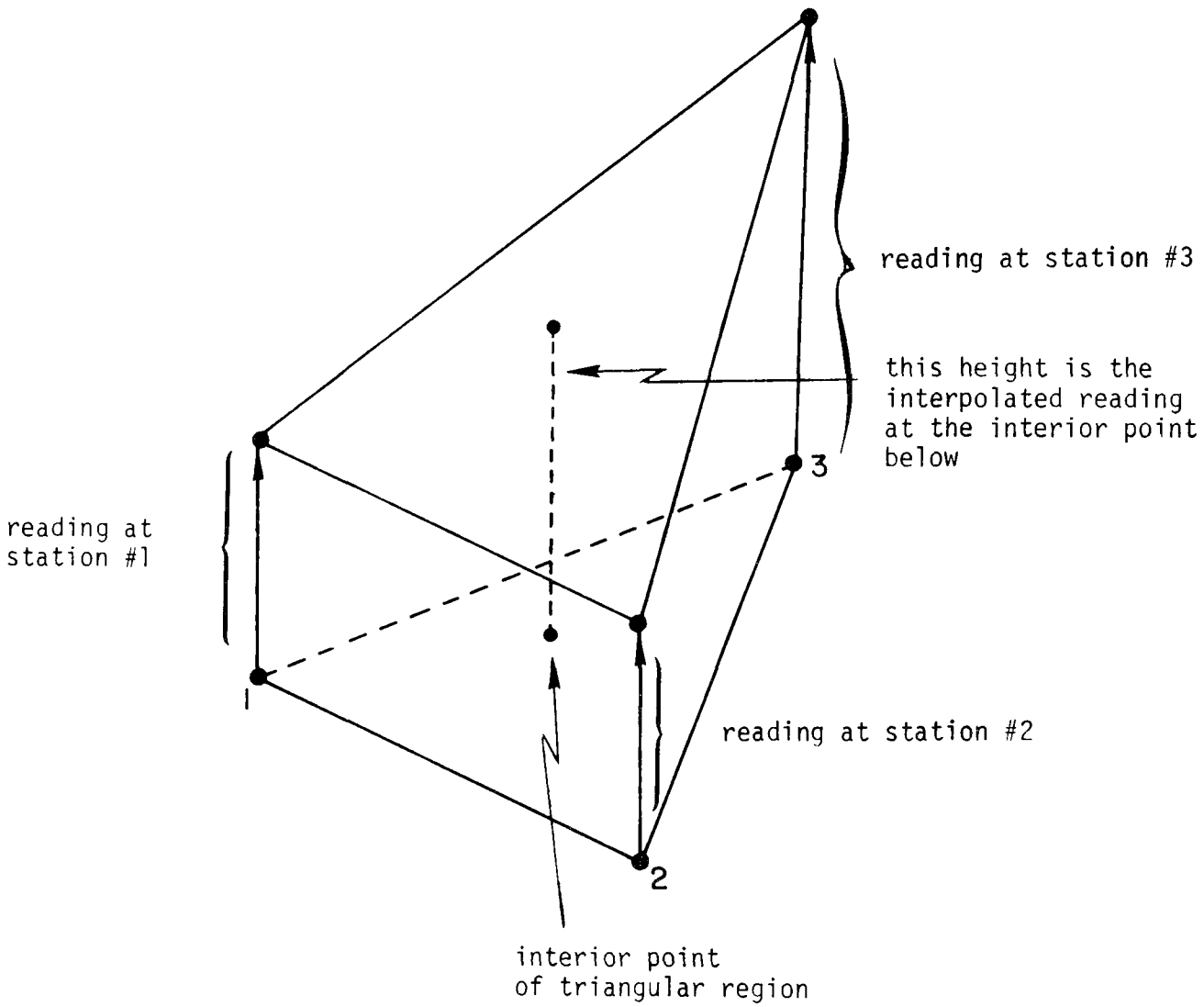


Figure 5. Interpolating pollutant levels over a triangular region.

The last two variables were used for all locations in the polygon, but the temperature assigned to a triangle was that measured at whichever one of the three temperature-measurement stations was closest to the triangle.

A master file of trajectories was constructed as follows:

- Starting with a given location $\underline{x}_1$ and time, find the corresponding values of the pollutants, wind speeds, and meteorological variables.
- Use the wind speeds to find the predecessor point $\underline{x}_2$ and repeat the first step.
- Keep going until the point passes out of the polygon.

We use the five stations listed above as starting points and traced trajectories backwards from 8 a.m. through 5 p.m. at hourly intervals. We did not compute trajectory values prior to 5 a.m. This gave us ten trajectories per day for each station.

4. THE EXPLORATORY PHASE

From the bank of 7000 trajectories, we extracted over 12,000 data vectors of the form

$$(\Delta O_3,\ O_3,\ NO,\ NO_2,\ HC,\ CH_4,\ SR,\ T)\ ,$$

where ΔO_3 was a one-hour change and over 8000 vectors where ΔO_3 was a two-hour change. However, a large majority of these vectors had small, usually positive ΔO_3 readings, as shown in Tables 1 and 2.

For our purposes, the data vectors with small values of ΔO_3 are largely uninformative, and keeping all of them in the working data base would swamp it with largely irrelevant information. For that reason, we sampled from the two histogram bins containing most of the ΔO_3 values to reduce the total number of data vectors in those bins to about 50% of the total. For instance, in the two-hour data, we selected the first 400 entries in the range 1.9 ± 2.8 and the first 500 in the range 7.5 ± 2.8 and rejected the remainder from our working data base.

For the variable selection and exploratory phase, we used INVAR, a general method for estimating efficiently how much of the variability in the dependent variable can be explained by a subgroup of the independent variables [8]. This technique estimates the limiting value of percent of variance explained (PVE) by a "smooth" nonlinear model.* We first tested

*Percent of variance explained equals

$$100 \times \left[1 - \frac{\text{variance of error in prediction}}{\text{variance of dependent variable}}\right]$$

Table 1. Frequency Breakdown of One-Hour Changes in ΔO_3

ΔRange (pphm)[a]	Frequency (original)		Frequency (after sampling)	
-21.4 ± 2.7	2		2	
-16.0 ± 2.7	5		5	
-10.7 ± 2.7	22		22	
-5.3 ± 2.7	222		222	
0.0 ± 2.7	8426	94%	400	52%
5.3 ± 2.7	3391		400	
11.7 ± 2.7	407		407	
16.0 ± 2.7	51		407	
21.3 ± 2.7	21		21	
26.7 ± 2.7	6		6	

[a]actual class interval length is 5.34

Table 2. Frequency Breakdown of Two-Hour Changes in ΔO_3

ΔRange (pphm)[a]	Frequency (original)		Frequency (after sampling)	
-20.7 ± 2.8	2		2	
-15.0 ± 2.8	7		7	
- 9.4 ± 2.8	22		22	
- 3.7 ± 2.8	181		181	
1.9 ± 2.8	5171	89%	400	50%
7.5 ± 2.8	1985		500	
13.2 ± 2.8	550		550	
18.8 ± 2.8	100		100	
24.4 ± 2.8	32		32	
30.1 ± 2.8	8		8	

[a]actual class interval length is 5.64

all independent variables as individual predictors, then pairs of variables, and then added variables to find the best three, etc. Some results for single variables are tabulated in Table 3. The most significant individual variables (in approximate order of importance) are SR, NO_2, T, and O_3.

Exploring pairs of variables, we found the results shown in Table 4. Other pairs were run that resulted in lower percent of variance explained than those in the table.

Triplets of variables were then explored with one really significant improvement showing. Some results are shown in Table 5.

The final significant increase occurred when we added temperatures to O_3, NO_2, SR. But, somewhat strangely, the increase was significant only for the data base of one hour ΔO_3. Here we obtained

Variables	One-Hour PVE
O_3, NO_2, SR, T	65.9

Our original thinking was that SR and T might be highly dependent. However, as the scatter plot (Figure 6) of 300 hourly readings of SR and T shows, the two are not closely related, particularly at high levels of SR. Still, the evidence is that temperature is much less strongly associated with ΔO_3 than is SR. In all of the INVAR runs using HC and CH_4, neither of them significantly increased the PVE. For instance, when HC and CH_4 were individually added to the variables NO_2, NO, O_3, and SR, the maximum increase in PVE was 2.1%.

Table 3. Percent of Variance Explained (Single Variables)

Variable	One hour ΔO_3	Two hour ΔO_3
O_3	15.8	24.1
NO	13.4	12.9
NO_2	25.7	41.7
HC	19.2	15.6
CH_4	16.6	19.3
SR	30.3	36.7
T	17.1	35.8

Table 4. Percent of Variance Explained (Pairs of Variables)

Variables	One hour ΔO_3	Two hour ΔO_3
NO_2, O_3	40.8	55.0
NO_2, SR	42.3	49.6
NO_2, T	38.7	52.9
NO_2, NO	34.8	50.1
O_3, SR	53.0	58.8
SR, T	43.6	52.1
O_3, T	33.2	46.7

Table 5. Percent of Variance Explained (Triplets of Variables)

Variables	One-hours ΔO_3	Two-hour ΔO_3
O_3, NO_2, SR	60.2	71.1
NO_2, NO, SR	46.7	53.8
O_3, NO_2, T	46.7	64.9

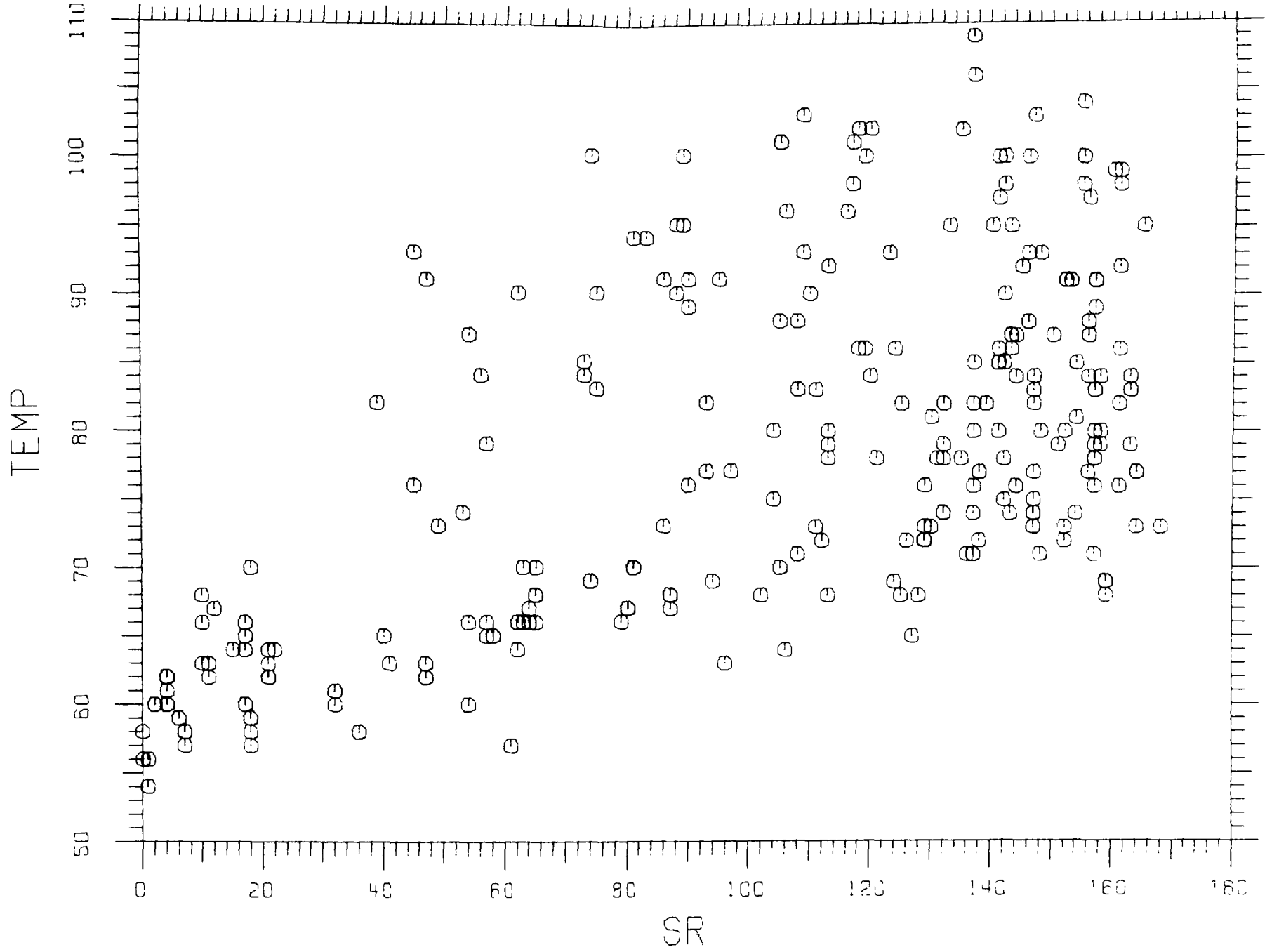


Figure 6. Temperature vs. solar radiation.

Similarly, the effect of NO was small. When NO was added to the three variables O_3, NO_2, and SR, there was no increase in PVE.

We considered these results to be highly promising in view of the approximations involved in the interpolation program used to determine the trajectories and pollutant values. In fact, we had not expected to be able to do this well.

Puzzles that remained were the small effect of the HC and the fairly large effect of the O_3. The O_3, by itself, is not particularly significant in relation to ΔO_3, but O_3 combined with SR has the highest PVE in the one-hour ΔO_3 case. On the other hand, HC, which contains the reactive hydrocarbons, is surprisingly insignificant in its effect on O_3 production. Examining the HC records, we found that the maximum HC levels occurred early in the morning. Table 6 classifies the HC readings by hour and by level for all seven stations over the five-month period. The peak HC levels occur mainly at the 6 a.m., 7 a.m., and 8 a.m. readings. This might indicate rapid removal of HC in the presence of solar radiation and other reactive pollutants.

Table 7 gives the frequency breakdown of maximum daily O_3 at Pomona versus maximum daily HC at the Los Angeles Civic Center. It does indicate that there is some association between HC levels and later O_3 levels.

The O_3 effect may be due partly to its being scavenged to produce a negative change in O_3 levels. We also considered it possible that the O_3 variable is a surrogate that reads the progress and strength of a slow reaction that started hours earlier with the take-up of the HC.

Table 6. Frequency of HC Levels (ppm) by Hour For All Stations

Hour	0	1	2	3	4	5
5 a.m.	599	359	41	12	2	0
6	519	407	71	14	1	1
7	470	412	92	13	6	0
8	496	423	79	12	2	1
9	538	416	47	11	1	0
10	567	406	37	3	0	0
11	593	388	26	5	1	0
12 noon	615	380	17	1	0	0
1 p.m.	661	343	8	0	0	0
2	676	330	8	0	0	0
3	680	329	2	2	0	0
4	690	318	3	2	0	0
5 p.m.	688	319	6	0	0	0

Table 7. Relation of O_3 Levels at Pomona to HC Levels at Los Angeles Civic Center

Max HC (Civic Center, ppm)	Max O_3 (Pomona, pphm) 0-10	11-20	21-30	31-40
0	9	16	4	2
1	17	48	17	1
2	5	11	10	0
3	1	1	1	0
4	0	0	1	0

To see if this was consistent with the data, we constructed new data bases in which the HC was replaced by:

1. The HC on the trajectory three hours earlier.
2. The average HC over three, four, and five hours earlier, if the trajectory stayed in the region that long, otherwise, the average over the period three and four hours earlier.

The data bases constructed this way still had a dominant number of ΔO_3 values close to zero and these were reduced by sampling until the new data bases contained 1000-1500 points each. Using both one- and two-hour changes yielded four new bases. The new bases were different than the old because data was taken from a point on a trajectory only if the trajectory had been in the region for three or four hours already.

The INVAR runs on this new data did not reveal a significant contribution from the earlier HC concentration, as can be seen in Table 8. There is, in the last two entries of the second column of Table 8, some possible evidence of increase in PVE with addition of HC. But the 64.0 PVE may be an unstable high value due to over-fitting. (Although the INVAR program contains a test to prevent over-fitting, the random character of the procedure infrequently produces too much splitting and a slight over-fitting.)

These results concluded, in the main, the exploratory phase of our study. The variables found most significantly related to ΔO_3 were O_3, NO_2, and SR. A puzzling result was the failure to find any significant relationship with HC or NO.

Table 8. PVEs Computed for the Four New Data Bases

Variables	One-Hour ΔO_3 Data Bases		Two-Hour ΔO_3 Data Bases	
	HC(-3)	HC(-3,-4,-5)	HC(-3)	HC(-3,-4,-5)
HC	6.6	5.7	9.5	7.5
HC_4	7.1	5.7	9.4	15.5
NO	4.9	4.4	7.0	4.0
NO_2	19.7	22.0	33.6	35.5
O_3	14.4	19.2	26.0	23.9
SR	23.6	24.4	27.1	34.2
T	19.7	23.8	23.1	25.9
O_3,SR	35.5	42.7	44.6	54.4
HC,SR	34.9	33.1	36.2	39.5
HC,NO_2,SR	40.1	42.4	48.1	54.5
HC,O_3,SR	43.1	47.8	52.1	55.4
O_3,NO_2,SR	49.8	57.7	62.2	62.4
HC,NO_2,O_3,SR	52.9	64.0(60.8)	61.4	62.7

Another exploratory task undertaken was to study the effects of solar radiation, temperature, and mixing heights alone--the key meteorological variables--on ozone production. We looked at trajectories that ended at one of the five "entry" stations at the time of maximum O_3 reading for that day at that station. We found the O_3 reading at the time the parcel of air entered the study region and formed the variable

$$\log(\Delta O_3) = \log[O_3(\text{end}) - O_3(\text{enter})]$$

We deleted all trajectories such that $O_3(\text{end}) - O_3(\text{enter}) \leq 0$. This dependent variable measures the change over the whole trajectory and essentially measures the final ozone level achieved at the end of the trajectory, since the ozone level at the beginning of the trajectory will generally be small. This variable is hence of a different character than the shorter changes.

Along each such trajectory we computed the sum of the solar radiation (SSR) for all the hours the trajectory took until it left the region, the average temperature (AVT) over the trajectory, and the mixing height (MH) for the day. We then did a linear stepwise regression with $\log(\Delta O_3)$ as the dependent variable and

log SSR, log AVT, log MH

as the independent variables. The results are shown in Table 9.

Table 9. Results of Linear Stepwise Regression of the Dependent Variable Log (ΔO_3)

Variable	PVE
log SSR	42.3
log SSR, log MH	45.7
log SSR, log MH, log AVT	45.8

That log SSR had a PVE as high as 42% by itself with the form of the equation restricted to be log linear is a bit surprising. Even more surprising is the conclusion that adding log MH and log AVT made very little difference in the PVE. This analysis supports the dominance of solar radiation observed in the short-term ozone changes. We did not run INVAR on this data as previous work with similar variables [10] had indicated a good fit with a log-linear model. However, we feel that this should be checked in future work.

5. THE RELATIONSHIP OF ΔO_3 TO O_3, NO_2, AND SR

In this part of the study, we focus on the nature of the functional relationship between the two-hour change in O_3 and the O_3, NO_2, SR variables. Throughout this analysis, we used the 1800 data points in the reduced data base except in the plots, where every sixth point was used. Although it is generally difficult to understand a multidimensional relationship from single variable regression, we plotted O_3, NO_2, and SR individually vs ΔO_3 (Figures 7, 8, and 9) to see what could be understood on this level. The scatter is immediately apparent, and the only apparent systematic relationship is positive correlation between NO_2 and ΔO_3.

The next step in the analysis was to stratify the ΔO_3 by the O_3, SR levels and take the average ΔO_3 in each bin. The results are shown in Table 10. Here, an underlying pattern emerges.

- High SR is associated with high ΔO_3
- High O_3 is associated with low ΔO_3

However, to get a continuous functional form for the relationship of ΔO_3 to O_3, NO_2, and SR, we used EFAP, a program developed by TSC [32, 37], that fits a piecewise linear continuous regression surface to the data. Since the surface generated by EFAP is smoother and less general than that used in INVAR estimates, EFAP generally does not achieve the level of PVE obtained by INVAR. Conversely, because INVAR usually splits the region into many subregions and fits a discontinuous surface, the resulting fit is difficult to interpret physically. The continuous piecewise linear

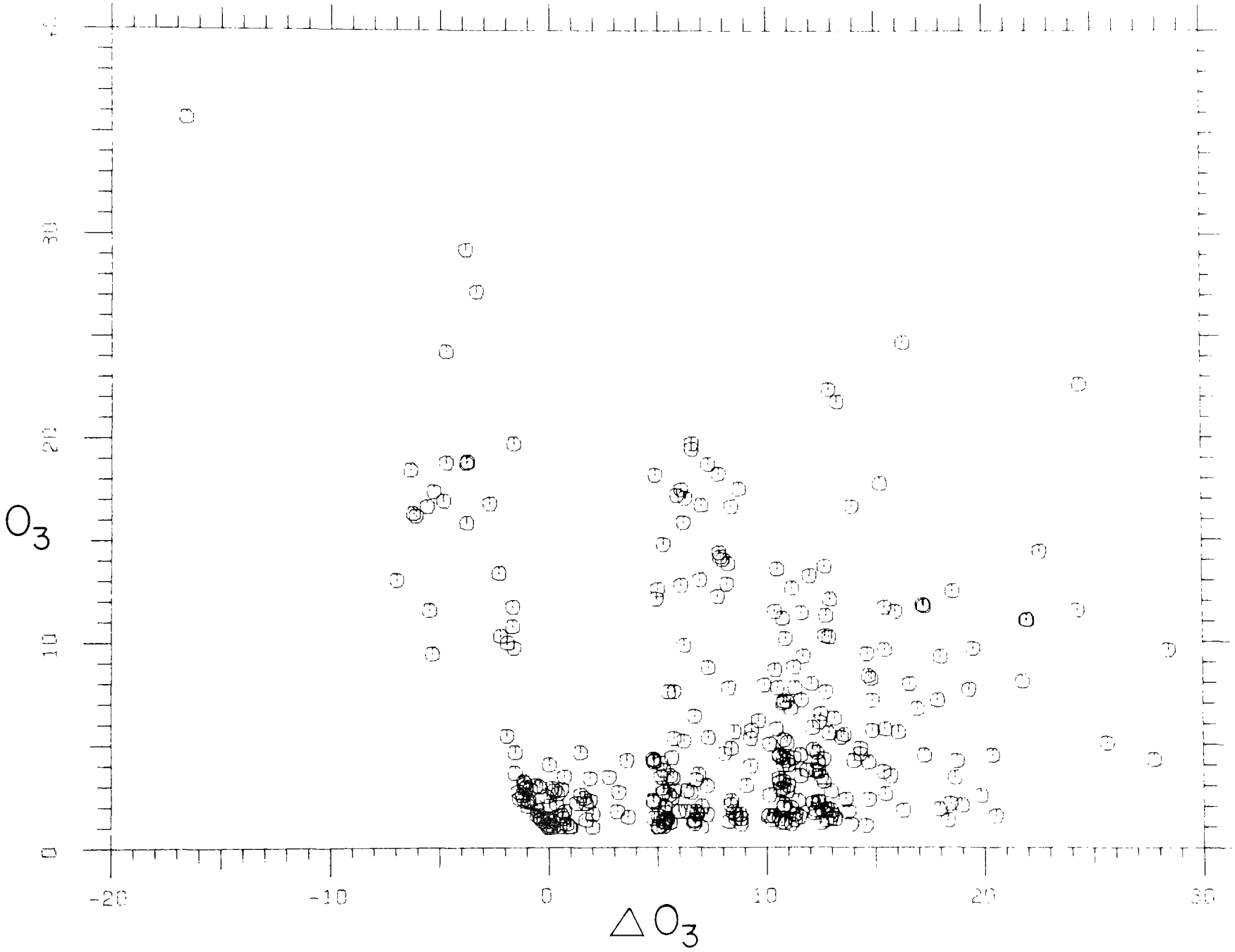


Figure 7. O_3 vs ΔO_3 (every 4th point sampled).

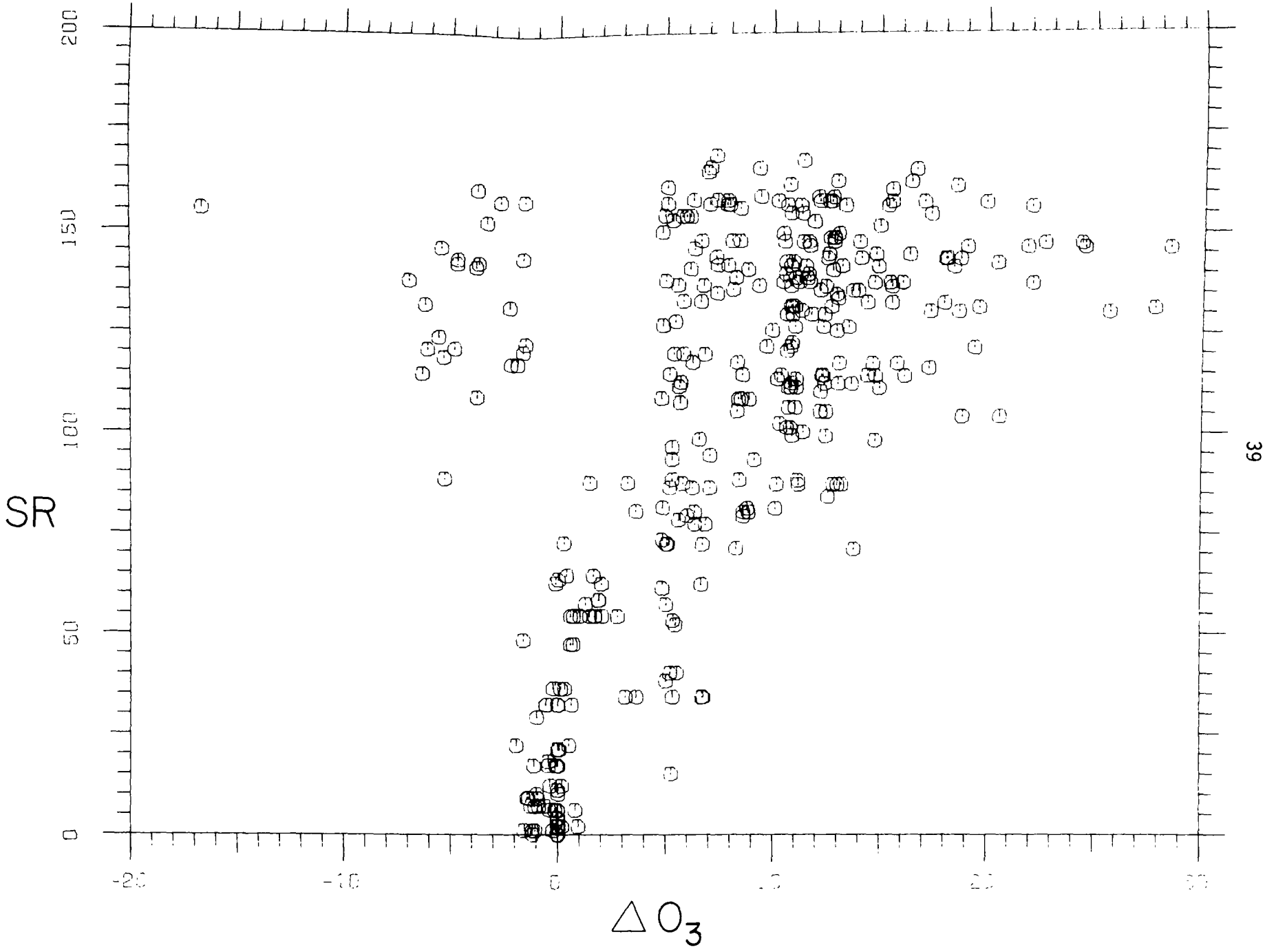


Figure 8. SR vs ΔO_3 (every 4th point sampled).

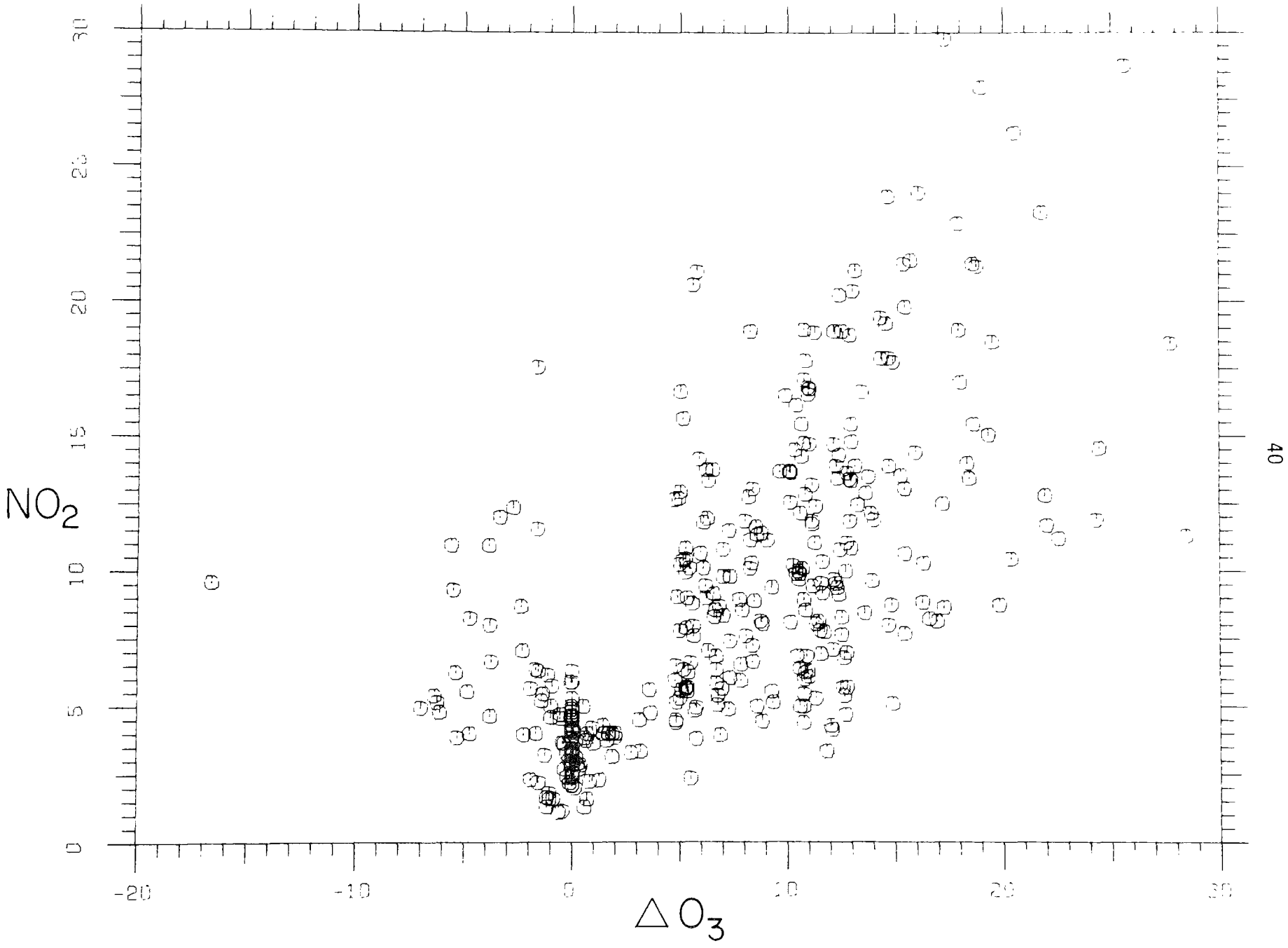


Figure 9. NO_2 vs ΔO_3 (every 4th point sampled).

Table 10. Two-Hour ΔO_3 Averages Stratified by SR and O_3. Entries in table are ΔO_3. (SR in Langleys averaged over two-hour periods; O_3, ΔO_3 in pphm.)

O_3	SR				
	0-40	40-80	80-120	120-160	160-200
0-5	.3	3.9	9.3	12.2	12.3
5-10	-2.1	.1	8.8	11.5	9.8
10-15	a	7.0	.0	10.1	8.3
15-20	a	a	-1.2	6.3	7.7
20-25	a	a	-4.8	.2	3.4[b]
25-30	a	a	-3.4[b]	-1.9	5.8[b]
30-35	a	a	-13.1[b]	2.0[b]	a
35-40	a	a	-10.0[b]	-12.2[b]	a
40-45	a	a	a	-11.6[b]	a

[a]No samples in bin

[b]Average is based on six or fewer samples

surface generated by EFAP gave a PVE of 60.7%. The final fitted surface is fairly simple, consisting of a continuous patching together of eight hyperplane segments. The equations for five of these are tabulated later. Three-dimensional slices of this surface are graphed in Figs. 10, 11, and 12. Of the eight regions, there are three together that contain only 1.0% of the total number of points. We will ignore these and restrict our analysis to the relevant information contained in the functional fit to ΔO_3 in the five other regions.

As a quick preliminary summary, in Table 11 we give the means of all variables corresponding to the points in each region.

Table 11. Means of Variables by Region

Percent of Points	ΔO_3	O_3	NO_2	SR
Overall (100)	7.1	6.1	9.0	100
Region 1 (46)	3.7	3.6	4.9	73
Region 2 (33)	11.0	4.6	11.9	118
Region 3 (8)	1.2	20.4	7.3	139
Region 4 (7)	14.7	5.4	20.3	119
Region 5 (5)	8.7	16.7	12.8	149

In Table 12 the mean values are characterized by region.

Table 12. Mean Value Characteristics

	ΔO_3	O_3	NO_2	SR
Region 1	very low	very low	very low	very low
Region 2	high	low	above average	above average
Region 3	very low	high	below average	high
Region 4	high	below average	high	above average
Region 5	above average	high	above average	high

This layout of mean values is itself interesting. Region 1, containing almost half of the sample points, is representative of low pollution levels, low O_3 production, and low solar radiation. Region 2, with 33% of the points, contains data with above average mean NO_2 and solar radiation levels, below average O_3 levels, and high positive changes in O_3. The other three regions, with a total of 20% of the sample points, represent more extreme conditions.

As can be seen in Table 13, the standard deviations of the variables in these regions give more information.

Table 13. Standard Deviations of Variables by Region

	ΔO_3	O_3	NO_2	SR
Overall	7.0	6.2	5.2	52.8
Region 1	5.2	3.1	1.9	59.9
Region 2	4.4	3.4	2.5	32.1
Region 3	8.8	6.1	2.1	18.5
Region 4	5.7	2.5	3.1	19.3
Region 5	7.9	4.2	2.1	11.0

Region 1, which contains the low mean values, has low standard deviations in all variables except for SR, which fluctuates over a wide range. Region 2 is similar. Regions 3, 4, and 5 have, generally, moderate standard deviations in the independent variables and larger deviations in the ΔO_3.

The equations of the EFAP hyperplanes in each region are given in Table 14.

Table 14. EFAP Equations for ΔO_3

Region				
1	-.14 (O_3)	+.87 (NO_2)	+.054 (SR)	-3.9
2	-.097 (O_3)	+.52 (NO_2)	+.056 (SR)	-1.4
3	-.87 (O_3)	+.86 (NO_2)	+.029 (SR)	+9.0
4	-.092 (O_3)	+.28 (NO_2)	+.059 (SR)	+2.3
5	-.82 (O_3)	+.26 (NO_2)	+.034 (SR)	+15.1

Before discussing these results, since the size of the above coefficients depend on the scaling of the variables, we introduce normalized variables by dividing the original variables by their overall standard deviations, i.e., denoting normalized variables by super asterisks,

$$O_3^* = O_3/6.2,\ NO_2^* = NO_2/5.2,\ SR^8 = SR/52.8$$

The equations are given in terms of the normalized variables, in Table 15.

Table 15. Normalized Equations for ΔO_3
(ΔO_3 not normalized)

Region					(PVE)[a]
1	-0.9 (O_3^*)	+4.5 (NO_2^*)	+2.8 (SR^*)	-3.9	58.6
2	-0.6 (O_3^*)	+2.6 (NO_2^*)	+2.9 (SR^*)	-1.4	28.8
3	-5.4 (O_3^*)	+4.4 (NO_2^*)	+1.5 (SR^*)	+9.0	26.1
4	-0.57(O_3^*)	+1.4 (NO_2^*)	+3.1 (SR^*)	+2.3	17.5
5	-5.1 (O_3^*)	+1.4 (NO_2^*)	+1.8 *SR^*)	+15.1	8.4

[a]The Percent Variance Explained (PVE) is for the samples of that subregion alone. This number is misleadingly low in that it does not reflect the continuity constraint but is useful in comparing significance of the linear equations in different regions.

From the table, then in Region 1 the predicted value of ΔO_3 would be $-0.9(O_3^*) + 4.5(NO_2^*) + 2.8(SR^*) - 3.9$.

The most drastic change between the five equations is in the coefficient of O_3^*. The implication is that when there are generally low pollutant and SR values, the dominant factors are the NO_2 concentration and SR intensity. But, in the region where the values are generally above average, the O_3 concentration has a strong inhibiting effect on the change in O_3. Further, the role of SR is substantially diminished.

The equations in Regions 2, 3, 4, and 5 should be cautiously evaluated since the PVEs are 28.8, 26.1, 17.5, and 8.4, respectively. This implies that in these regions the mean value of ΔO_3 is almost as good a predictor as the given equations. The table of mean values is equally informative in these regions. Looking at the mean values in Regions 2, 3, 4, and 5, it is clear that

- High ΔO_3 values are associated with low O_3 and high or above average values of NO_2 and SR
- Low to moderately above average ΔO_3 values are associated with high O_3, even in the presence of above average or high NO_2 and SR.

The negative values of ΔO_3 largely correspond to points in Region 2 with some in Region 5. These are also the two regions with above average to high O_3.

We can split the qualitative conclusions above into two regimes:

- At below average O_3 levels, O_3 has little to do with the resulting ΔO_3. The dominant variables are NO_2 and SR. The

relationship is monotonic in the sense that increasing NO_2 or SR leads, on the average, to higher ΔO_3 values.

- At above average or high O_3 levels, O_3 has a strong negative effect on ΔO_3. Increasing NO_2 and SR values are offset by increasing O_3 values, and the associated ΔO_3 levels are only moderately above average. This apparent self-limiting effect of oxidant has been noted by others [18,19].

This behavior can be seen graphically in Figures 10, 11, and 12. Figure 10 is a plot of the EFAP regression surface with SR fixed at its mean value of 100; Figure 11 is a plot of the surface with NO_2 fixed at its average of 9.0, and Figure 12 uses O_3 set at its average value of 6.1.

As the final output of this study, we have not only arrived at some understanding of the association between change in O_3 levels and its precursors; but we also have constructed a simple model which predicts changes in O_3 level as a function of current O_3, NO_2, and SR measurements with a correlation between predicted and actual values of .8. Equation (6) for this preliminary empirical model is given below; the form of the equation is a characteristic of the methodology [37].

For the continuous piecewise-linear model for two-hour ozone concentration changes, the change in O_3 over the next two hours is computed by calculating the linear equations A, B, C, D, E, F below with current values of the variables; taking the maximum of A, B, and C and multiplying that by -5.125; adding the result to the product of -1.167 and the maximum of D, E, and F; and finally, adding that result to 10.48. We have

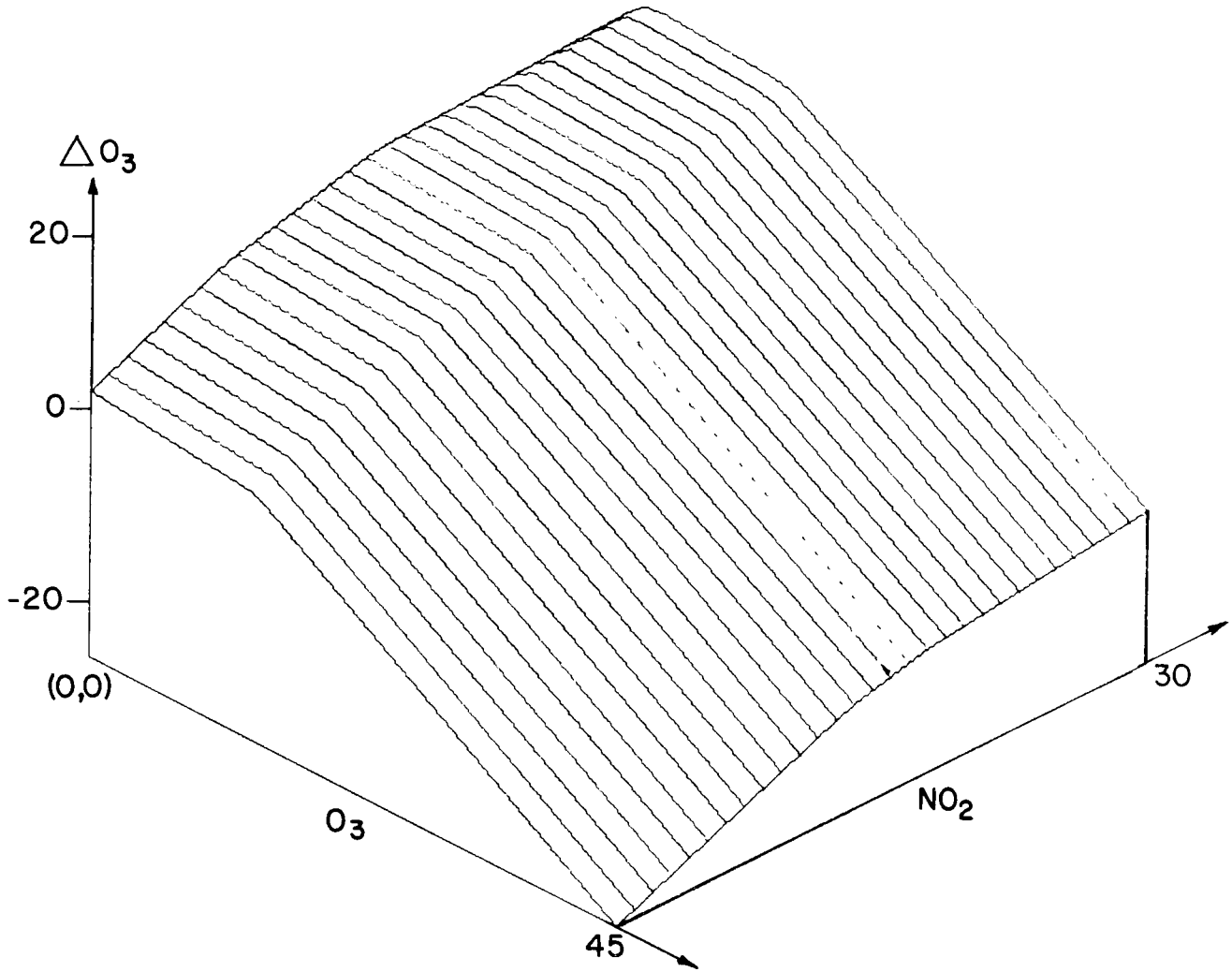


Figure 10. Graph of EFAP regression surface, with SR = 100.

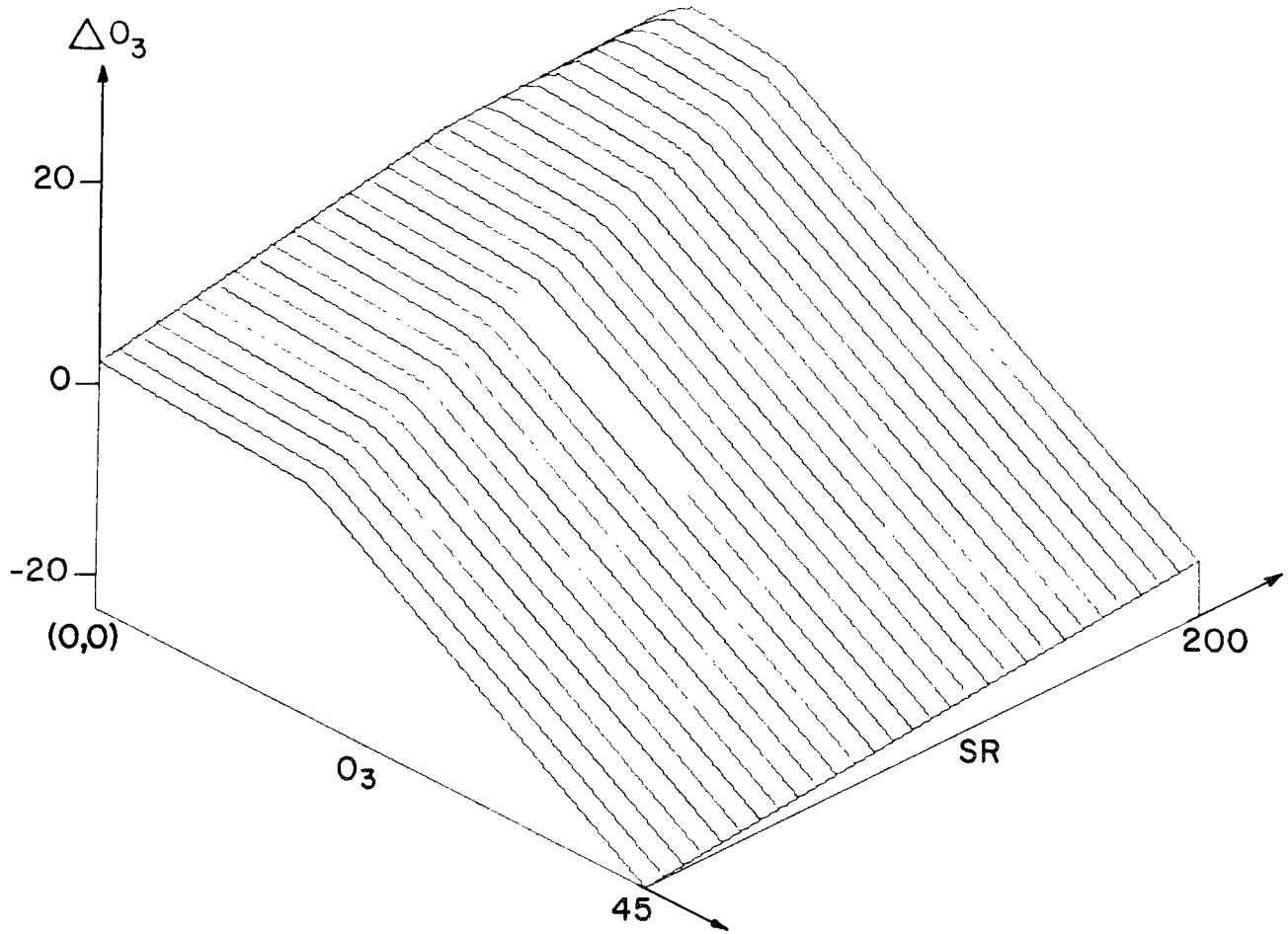


Figure 11. Graph of EFAP regression surfaces, NO_2 = 9.0.

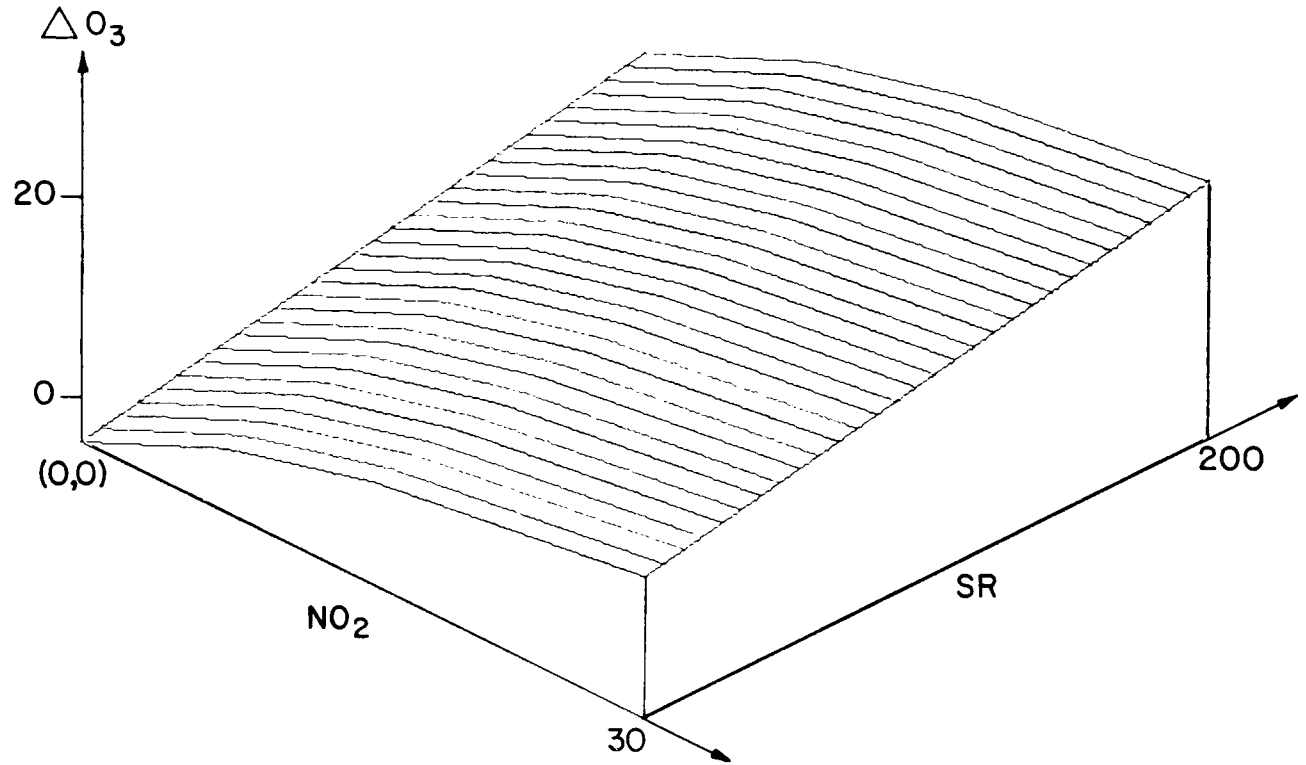


Figure 12. Graph of EFAP regression surface, O_3 = 6.1.

$$\Delta O_3 = -5.125 \cdot \max(A,B,C) - 1.167 \cdot \max(D,E,F) + 10.48 \qquad (6)$$

where:

$A = -0.2146 \cdot X_1 - .0701 \cdot X_2 - .002268 \cdot X_3 + .9376$

$B = .02114 \cdot X_1 - .1013 \cdot X_2 - .01075 \cdot X_3 + 2.275$

$C = .1638 \cdot X_1 - .09855 \cdot X_2 - .005938 \cdot X_3 - .2263$

$D = .02709 \cdot X_1 - .3015 \cdot X_2 + .001298 \cdot X_3 + 2.304$

$E = -.009565 \cdot X_1 + .0005252 \cdot X_2 - .001079 \cdot X_3 + .2306$

$F = -.0144 \cdot X_1 + .2066 \cdot X_2 - .003171 \cdot X_3 - 2.943$

and

$X_1 = O_3$ concentration (PPHM)

$X_2 = NO_2$ concentration (PPHM)

X_3 = solar radiation (gm. cal./cm^3/hr.)

6. CONCLUSIONS

The intent of the present study was to explore the possibility of developing empirical difference equations describing the production of ozone in the atmosphere. We examined the degree of predictability of one- and two-hour changes of ozone under very limiting conditions (no emissions data, a restricted set of meteorological variables, and only ground-level measurements). Only the change in ozone and not the other chemical species was modeled in the present analysis.

The results obtained were of two sorts:

(1) Three variables seem to produce the main effect in the change in ozone levels:

- the current O_3 level;
- the current solar radiation reading; and
- the current NO_2 level.

(2) A model relating the two-hour change in ozone to these variables was derived, and an attempt was made to extract the qualitative implications of the quantitative model. The predictions of this final empirical model were correlated with actual measurements with a correlation coefficient of .78 over 1800 samples. (This corresponds to explaining 61% of the variance.)

These results were quite encouraging. They suggest that a more complete implementation of the concepts suggested in the introduction of this report could result in a useful empirical model and some qualitative insights into aspects of the photochemical pollution problem.

REFERENCES

1. Air Pollution in California, Annual Report 1974. Air Resources Board State of California. March 1975.

2. Altshuller, A. P. Evaluation of Oxidant Results at Camp Sites in the United States. JAPCA. 25, January 1975.

3. Angell, J. K., et al. Three-Dimensional Air Trajectories Determined from Tetroon Flights in the Planetary Boundary Layer of the Los Angeles Basin, JAM. 11: 451-471, April 1972.

4. Baboolal, L. B., I. H. Tombach, and M. I. Smith. Mesoscale Flows and Ozone Levels in a Rural California Coastal Valley. First Conference on Regional and Mesoscale Modeling, Analysis, and Prediction of the AMS. Las Vegas. May 6-9, 1975.

5. Blumenthal, D. L., et al. Three Dimensional Pollutant Gradient Study--1972 Program. Meteorology Research, Inc., Altadena. May 18, 1973.

6. Blumenthal, D. L., et al. Determination of the Feasibility of Long-Range Transport of Ozone or Ozone Precursors, Final Report. MRI to RPA. Contract No. 68-03-1462. November 1974.

7. Blumenthal, D. L., and W. H. White. The Stability and Long Range Transport of Ozone or Ozone Precursors. Meteorology Research, Inc. (68th Annual APCA Meeting. Boston. June 1975.)

8. Breiman, L., and W. S. Meisel. General Estimates of the Intrinsic Variability of the Data in Nonlinear Regression Models. Research for the U.S. Air Force Office of Scientific Research under Contract #F44620-71-C-0093, March 19, 1975, submitted to JASA.

9. Bruntz, S. M. Ozone Concentrations in New Jersey and New York: Statistical Association with Related Variables. Bell Laboratories. May 1974.

10. Bruntz, S. M., et al. The Dependence of Ambient Ozone on Solar Radiation, Wind, Temperature, and Mixing Height. (Symposium on Atmospheric Diffusion and Air Pollution. Santa Barbara. September 9-13, 1974.) Published by Amer. Meteor. Society, Boston, Mass.

11. Buffington, P. D., and E. A. Bartczak. Ozone: Chemical Action and Reaction in the Lower-Level Transport Winds. Regional Air Pollution Control Agency. Dayton, Ohio. (68th Annual APCA meeting. Boston. June 1975.)

REFERENCES (Cont.)

12. California Air Resources Board Bulletin, February 1975.

13. Cavanaugh, L. A. Atmospheric Photochemical Smog Measurements over San Francisco Bay. Progress Report No. 4. Stanford Research Institute Project 2092, March 15, 1973.

14. Chock, David P., and Susanna B. Levitt. A Space-Time Correlation Study of Oxidant and Carbon Monoxide in the Los Angeles Basin, General Motors Research Laboratories. Warren, Michigan. (68th Annual APCA Meeting. Boston. June 1975.)

15. Cleveland, William, S. Sunday and Workday Behavior of Photochemical Air Pollutants in New Jersey and New York. No date.

16. Cleveland, William S., et al. Using Robust Statistical Methods in Analyzing Air Pollution Data with Applications to New York - New Jersey Photochemistry. (APCA Annual Meeting. Denver. June 1974.) APCA No. 74-76.

17. Cleveland, William S., and Bert Kleiner. The Transport of Photochemical Air Pollution from the Camden-Philadelphia Urban Complex. Bell Laboratories, 1974.

18. Cleveland, William S., et al. Chemical Kinetic and Data Analytic Studies of the Photochemistry of the Troposphere. Bell Laboratories. (EPA Symposium on the Atmospheric Chemistry of NO_x. Washington, D.C.. February 1975.)

19. Cleveland, William S., et al. The Analysis of Ground-Level Ozone Data from New Jersey, New York, Connecticut, and Massachusetts: Data Quality Assessment and Temporal and Geographical Properties. (68th National APCA Meeting. Boston. June 1975.)

20. Corn, Morton, et al. Photochemical Oxidants: Sources, Sinks, and Strategies. J. APCA. 25, January 1975.

21. Dabberdt, Walter F., et al. Studies of Air Quality On and Near Highways. First Interim Report. Prepared for Dept. of Transportation, FHWA, July 1974.

22. Demerjian, Kenneth L., et al. The Mechanism of Photochemical Smog Formation. In: Advances in Environmental Science and Technology, James N. Pitts, Jr., and Robert L. Metcalf, editors, Volume 4, 1974.

REFERENCES (Cont.)

23. Development of an Air Pollution Model for the San Francisco Bay Area, Second Semiannual Report to NSF by Lawrence Livermore Laboratory. February 1974.

24. Duckworth, Spencer, and Robert W. McMullen. Can We Ever Meet the Oxidant Standard? California Air Resources Board. (68th Annual APCA Meeting, Boston, Mass. June 1975.)

25. Edinger, James, G., et al. Penetration and Duration of Oxidant Air Pollution in the South Coast Air Basin of California. JAPCA, 22, November 1972.

26. Edinger, James G. Vertical Distribution of Photochemical Smog in Los Angeles Basin. Environmental Science & Technology. 7, March 1973.

27. Environmental Quality--1974, The Fifth Annual Report of the Council on Environmental Quality, U.S. Gov't. Printing Office. December 1974.

28. Episode Contingency Plan Development for the Metropolitan Los Angeles Air Quality Control Region. Final Report. Prepared by TRW for EPA, Region IX, December 1973.

29. Eschenroeder, Alan Q., et al. Evaluation of Transportation Plan Impacts on Photochemical Smog. Environmental Research & Technology, Inc.. (68th Annual APCA Meeting. Boston. June 1975.)

30. Gloria, H. R., et al. Airborne Survey of Major Air Basins in California. JAPCA. 24: 645-652, 1974.

31. Holmes, J. R., and F. Bonamassa. Application of the Results of Recent Chamber Studies to the Control of Photochemical Oxidant. Proceedings of the NERC-RTP Conference on Smog-Chamber Experimentation. October 24-25, 1974.

32. Horowitz, Alan, W. S. Meisel, and D. C. Collins. The Application of Repro-Modeling to the Analysis of a Photochemical Air Pollution Model. Prepared for EPA under contract #68-02-1207. December 31, 1973.

33. Investigation of Ozone and Ozone Precursor Concentrations at Nonurban Locations in the Eastern United States. Contract No. 68-02-1077. Prepared by Research Triangle Institute for EPA, May 1974.

34. Leonard, M. J., et al. Effects of the Motor Vehicle Control Program on Hydrocarbon Concentrations in the Central Los Angeles Atmosphere. Los Angeles County APCD. (68th Annual APCA Meeting. Boston. June 1975.)

REFERENCES (Cont.)

35. Martinez, E. L. Temporal-Spatial Variations of Nonurban Ozone Concentrations and Related Meteorological Factors. (Conference on Air Quality Measurements. Austin. March 10-11, 1975.)

36. McCollister, George M., and Kent R. Wilson. Linear Stochastic Models for Forecasting Daily Maxima and Hourly Concentrations of Air Pollutants. Atmospheric Environment. 9: 417-423, March 1975.

37. Meisel, W. S., and D. C. Collins. Repro-Modeling: An Approach to Efficient Model Utilization and Interpretation. IEEE Trans. on Systems, Man, and Cybernetics. SMC-3. (4): 349-359, July 1973.

38. Meisel, W. S., The Role of Empirical Methods in Air Quality and Meteorological Analyses. Interim Report to Office of Research and Development, EPA, Contract No. 68-02-1704, December 1974.

39. Meisel, W. S., and D. C. Collins. Continuous Piecewise Linear Regression. In preparation.

40. Paskind, Jack, and John R. Kinosian. Hydrocarbon, Oxides of Nitrogen and Oxidant Pollutant Relationships in the Atmosphere over California Cities. (Annual APCA Meeting. Denver. June 9-13, 1974.)

41. Perkins, William A., Jr. The Los Angeles Reactive Pollutant Program. (LARPP Symposium. Santa Barbara, California. November 12-14, 1974.)

42. Photochemical Smog and Ozone Reactions. In: Advances in Chemistry Series, Robert F. Gould, editor. American Chemical Society. 1972.

43. Price, J. H., et al. Estimation of Minimum Achievable Oxidant Levels by Trajectory Analysis: Implications for Oxidant Control. Texas Air Control Board. (68th Annual APCA Meeting. Boston. June 1975.)

44. Relationship of Oxidant Peak, High-Hour and Slope Values as a Guide in Forecasting Health-Effect Days. Final Report to State of California Air Resources Board, by Bay Area APCD Technical Services Division, February 1973.

45. Roth, P. M. Photochemical Air Pollution Simulation Models: An Overview and Appraisal. In: Proceedings of the Symposium on Chemical Aspects of Air Quality Modeling. Lawrence D. Kornreich, editor. Sponsored by TUCAP and EPA. April 17-19, 1974.

46. Rubino, R. A., et al. Ozone Transport. Connecticut Dept. of Environmental Protection. (68th Annual APCA Meeting. Boston. June 1975.)

REFERENCES (Cont.)

47. Severs, R. K. Simultaneous Total Oxidant and Chemiluminescent Ozone Measurements in Ambient Air. JAPCA. 25, April 1975.

48. Stephens, E. R. Chemistry and Meteorology in an Air Pollution Episode. JAPCA. 25, May 1975.

49. Sticksel, P. R. The Stratosphere as a Source of Ozone. Battelle-Columbus Laboratories. (68th Annual APCA Meeting. Boston. June 1975.)

50. Stoswik, Jr., William N., and Peter E. Coffey. Rival and Urban Ozone Relationships in New York State. JAPCA. 24: 564-568, 1974.

51. Trijonis, John C. Economic Air Pollution Control Model for Los Angeles County in 1975. Environmental Science and Technology. 8, September 1974.

52. Vaughan, Leland M., and Alexander R. Stankunas. Field Study of Air Pollution Transport in the South Coast Air Basin. Final Report to State of California Air Resources Board. July 1974.

53. White, Warren H., and Paul T. Roberts. The Nature and Origin of Visibility-Reducing Aerosols in Los Angeles. Meteorology Research, Inc. (68th Annual APCA Meeting. Boston. June 1975.)

CONCLUDING REMARKS OF PROJECT OFFICER

In discussing a first draft of this report, some insights were developed that appeared to provide an even stronger motivation than was originally in mind for the exploratory research that is described. Although some attempt has been made to reflect this in the final account, this is only done briefly. These concluding remarks represent the project officer's proposal of a context for the research; they will hopefully be useful in further development of the idea of an empirically-based, emissions-oriented photochemical pollution model.

For simplicity we consider a two-dimensional horizontal atmospheric mean motion $\underline{V}(\underline{r},t)$, and assume that turbulent mixing in the vertical is sufficient to produce a uniform distribution of all pollutants from the ground surface up to the mixing height $H(\underline{r},t)$. The effects of horizontal turbulent dispersion on the concentration field will be disregarded. This is the somewhat idealized photochemical model situation first proposed by Wayne, Kokin and Weisburd and extensively developed under earlier EPA contracts [See "Controlled Evaluation of Reactive Environmental Simulation Model (REM)," Vols. I & II, NTIS PB 220 456/8 and PB 220 457/6, Feb. 1973]. It simulates the photochemical reactions that occur in a parcel or vertical column of air of variable height moving along a wind trajectory. The time-rate of change of concentration of the i^{th} pollutant, following the motion of the parcel, is given by the conservation equation

$$\frac{D[HC_i]}{Dt} = (R_i + Q_i)H$$

where R_i = rate of generation of i^{th} species by chemical reaction

Q_i = rate of source emissions of species i,

and

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \underline{V} \cdot \underline{\text{grad}} \ ,$$

denoting a Lagrangian time-derivative. The quantity R_i may be a function of the concentrations $C_1, C_2, \ldots C_N$ of all the species that are involved in the process, and also of such meteorological variables as solar radiation intensity (SR) and temperature (T), so that the equation above may be written

$$\frac{DC_i}{Dt} + \frac{C_i}{H}\frac{DH}{Dt} = R_i(C_1, C_2, \ldots C_N, SR, T) + Q_i \qquad (i=1, 2, \ldots N)$$

where, in the Lagrangian framework, all quantities are to be regarded as functions of time that are dependent on the position $\underline{r} = \underline{r}(t)$ of the parcel. These basic conservation equations constitute a set of simultaneous ordinary differential equations that are identical with those for chemical reactions in a well-mixed batch reactor of variable volume. Not all of the functions R_i will involve all of the $C_i (i = 1, 2, \ldots N)$. Also, some of the Q_i will certainly be zero (all those for the secondary pollutant species). In the simple formulation under consideration the set of functions $R_i(C_1, C_2, \ldots C_N)$, $i = 1, 2, \ldots N$ contain the purely chemical aspects of the photochemical pollution problem and the (numerical) solution of the above set of simultaneous equations would provide the desired species concentrations as functions of the Lagrangian time-variable t. In the earlier modeling activities of Wayne, et al, referred to above, experimental data on the chemical kinetics that

constitute the set of functions R_i were based on ad hoc photochemical chamber experiments. However, the approach of the present report may, in principle, be regarded as an initial step in the direction of empirical determination of these functions from actual observations of the species concentrations in the ambient atmosphere. More specifically, the analysis attempts to develop an equation, by multivariate regression techniques, for the Lagrangian-trajectory time-rate of change of concentration of just the single pollutant ozone (O_3) in terms of the estimated concentrations of ozone, hydrocarbons not including methane (HC), methane (CH_4) and the nitrous oxides (NO, NO_2) together with the solar radiation intensity (SR) and air temperature (T). In this initial feasibility analysis this is done without regard to the variations in the atmospheric mixing depth, i.e., assuming that H = constant so that DH/Dt = 0; also since ozone is a secondary pollutant, the emissions strength $Q_i = 0$ (for $i = O_3$). In order to develop a complete prediction model along the lines suggested, it would, of course, be necessary to develop in some optimal fashion the full set of coupled equations for the chemical reaction functions $R_i (i = 1, 2, \ldots N)$. Possibly, however, some of these functions might be developed empirically from atmospheric air quality data and some from chamber tests. Since the set would have to include emissions of primary pollutants ($Q_i \neq 0$), they might then (like the conventional type photochemical model as suggested, for example, by Wayne, et al) hopefully be used to study various control strategies as applied to these primary species.

It should be emphasized that the air quality and wind data available for an analysis of the type being suggested, will be Eulerian data comprising the

(i) Species concentration field $C_i(\underline{r}, t)$, $i = 1, 2, \ldots N$

(ii) Wind velocity field $\underline{V}(\underline{r},t)$.

In terms of these (and with H = constant) the governing equation above would become

$$\frac{\partial C_i(\underline{r},t)}{\partial t} + \underline{V}(\underline{r},t) \cdot \underline{\text{grad}}\ C_i(\underline{r},t) = R_i(C_1, C_2, \ldots C_N) + Q_i \quad .$$

For i = 1, 2, ...N, this is now a set of partial differential equations, and analysis (or interpretation) of the relationship between the local time-rate of species concentration (as provided at the fixed observing stations) and the other variables would now be much more complicated, because of the dependence on the wind velocity field $\underline{V}$. By adopting a Lagrangian analysis following the air motion, the changes in concentration due to advection are effectively decoupled from the central problem. However, the Lagrangian trajectory must, of course, be estimated from the Eulerian wind field, as is done in the report.

The preceding suggestions only relate to the empirical determination of the chemical kinetics module of an overall emission-oriented photochemical pollution model. The meteorological module that is implied by the Lagrangian trajectory concept is very simplistic and much simpler, of course, than those under consideration by some air quality modelers at the present time, and which involve the use of sophisticated atmospheric transport and turbulent diffusion equations. However, in another volume of this trilogy on empirical techniques, an analysis is developed for the feasibility of determining atmospheric transport and dispersion functions by appropriate empirical

analysis of air quality data, and in this case for an inert pollutant, uncomplicated by any effects of chemical reactions. This suggests the possibility of eventually developing an emission-oriented approach that might combine, with realistic complexity, both the meteorological and chemical aspects of the problem in an empirical fashion.

Research Triangle Park, N.C.
October 1975

Kenneth L. Calder

APPENDIX

Statistical Noise in One-Hour and Two-Hour Changes

One-hour changes are usually more difficult to predict than two-hour changes because two-hour changes are usually less susceptible to measurement and interpolation error. This can be illustrated by the following simplified model: Suppose the error in estimating the change in O_3 due to all sources (e.g., wind field interpolation and pollutant interpolation) during the first hour is e_1 and during the second hour is e_2. Suppose there is a true change during both hours of amount δ. Defining the error ratio as the ratio of standard deviation of the error in measuring the change to the change itself, then the error ratio for the one-hour change is σ/δ where σ is the standard deviation of e_1. Assuming that e_2 has the same standard deviation as e_1 and that they are uncorrelated, the standard deviation of $e_1 + e_2$ is $\sigma\sqrt{2}$. Hence the error ratio for the two-hour change is

$$\frac{\sigma\sqrt{2}}{2\delta} = \frac{1}{\sqrt{2}} \frac{\sigma}{\delta}$$

about 2/3 as large as the one-hour error rate.

Many other models of the process will illustrate the same characteristic.

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

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-76-029c		

4. TITLE AND SUBTITLE	5. REPORT DATE
EMPIRICAL TECHNIQUES FOR ANALYZING AIR QUALITY AND METEOROLOGICAL DATA. Part III. Short-Term Changes in Ground-Level Ozone Concentrations: An Empirical Analysis	June 1976
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Leo Breiman, William S. Meisel	8. PERFORMING ORGANIZATION REPORT NO. TSC-PD-132-1
9. PERFORMING ORGANIZATION NAME AND ADDRESS Technology Service Corporation, 2811 Wilshire Boulevard, Santa Monica, California 90403	10. PROGRAM ELEMENT NO. 1AA009
	11. CONTRACT/GRANT NO. EPA 68-02-1704
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Sciences Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711	13. TYPE OF REPORT AND PERIOD COVERED Final May 74-Oct 75
	14. SPONSORING AGENCY CODE EPA-ORD

15. SUPPLEMENTARY NOTES
This is the last of three reports examining the potential role of state-of-the-art empirical techniques in analyzing air quality and meteorological data.

16. ABSTRACT

An empirical analysis of ambient air quality data for the Los Angeles Basin is used to relate the one- and two-hour changes in oxidant levels in the urban environment to the preceding levels of precursor pollutants and to meteorological variables. The intent was to demonstrate the feasibility of developing a set of empirical difference equations for the production of oxidant over time. The main variables determining one- and two-hour oxidant changes were identified using nonparametric regression techniques. A model for the oxidant changes was developed using nonlinear regression techniques. The implications of the model are discussed.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
* Air pollution		13B
* Ozone		07B
* Meteorological data		04B
* Regression analysis		12A
Empirical equations		
* Mathematical models		

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	73
	20. SECURITY CLASS *(This page)*	22. PRICE
	UNCLASSIFIED	

EPA Form 2220-1 (9-73)