Long-term Modeling of Regional Ozone Concentrations and Control Strategies

Thesis by

Darrell Alan Winner

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1998

(Submitted June 10, 1998)

Acknowledgments

Many people have helped me along the journey of this research. Glen Cass, my advisor, provided insight into the world of air pollution research. I appreciate the time he spent mentoring me and guiding this research. Rob Harley introduced me to the CIT photochemical airshed model and helped me learn how to use it. Greg McRae and Cliff Davidson introduced me to environmental engineering at Carnegie Mellon.

Joe Cassmassi, Kevin Durkee, Xinqiu Zhang, and Satoru Mitsutami at the South Coast Air Quality Management District provided key data to make this research possible.

Several people helped me to tame the myriad of computers and operating systems used in this research. Ken McCue, in addition to keeping the EQL VAX cluster humming, helped decipher several odd computer tape formats. Edith Huang's time and expertise made running the airshed model on the parallel computer an enjoyable task. John Salmon shared his insights on the use of parallel computers.

This research was supported by the Caltech Center for Air Quality Analysis and the Palace Knight program of the U.S. Air Force. This research was performed in part using the Intel Gamma System operated by Caltech on behalf of the Concurrent Supercomputing Consortium. This work was supported in part by a grant of HPC time from the DoD HPC Center, the Arctic Region Supercomputing Center.

This thesis would never have been finished without the loving support of my wife, Rebecca. It is dedicated to her and my two little girls who fill my life with joy, Emily and Megan.

Abstract

The capability to accurately model the long-term frequency of occurrence of high and low ozone concentrations in the urban and regional atmosphere is developed. A full year of hourly ozone concentration predictions is generated throughout Southern California using a photochemical airshed model driven by automated assimilation of routine meteorological observations. Statistical measures of airshed model performance for the full year are comparable to results achieved by episodic models for 2-3 days of simulation that are driven by data from expensive special field measurement programs.

The effect of alternative emission control strategies on the long-term frequency distribution of daily peak ozone concentrations in Southern California is determined. Ozone isopleth diagrams are constructed in terms of the number of days per year that the current U.S. Federal 8 h average and former 1 h average ozone standards will be exceeded at all feasible combinations of air basin-wide reactive organic gases and oxides of nitrogen emissions. It is found that the frequency of violation of the former 1 h average ozone standard set at 0.12 ppm ozone could be reduced to approximately 25 days per year through stringent emissions controls under 1987 meteorological conditions. No practical way is evident to even come close to meeting the new 8 h average standard set at 0.08 ppm ozone. It is shown that the days with the highest historically observed ozone concentrations are not necessarily the hardest days to bring below the air quality standard.

Use of synthetic meteorological data is explored as an efficient means to meet the input data requirements of airshed models that must operate over long periods of tine. A

semi-Markov process is used to generate a time series of synoptic weather conditions that statistically resembles the occurrence and persistence of historically observed weather. Local weather variables then are drawn for each day that are representative of the meteorological potential for ozone formation. These values are used as the initial conditions for a prognostic mesoscale meteorological model that generates the meteorological fields needed by a photochemical airshed model. The procedure produces daily peak 8 h average ozone concentration predictions that compare well with historical observations.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Introduction to the urban ozone problem

Ozone is a molecule containing three oxygen atoms which is formed by chemical reactions in the atmosphere involving reactive organic gases (ROG), nitrogen oxides (NO_x), and sunlight (Haagen-Smit, 1952). The main source of NO_x emissions to the atmosphere is fossil fuel combustion while ROG are emitted from a wide variety of anthropogenic and biogenic sources involving motor vehicle exhaust, solvent and gasoline evaporation, industrial processes and vegetation (National Research Council, 1991). The primary route of ozone formation in urban atmospheres is a chain reaction involving the photolysis of nitrogen dioxide which produces singlet oxygen, the reaction of singlet oxygen with molecular oxygen to form ozone, and the reaction of nitric oxide with ozone to form nitrogen dioxide which completes the chain.

$$NO_2 + h\upsilon \xrightarrow{k_1} NO + O \tag{1.1}$$

$$O + O_2 + M \xrightarrow{k_2} O_3 + M \tag{1.2}$$

$$NO + O_3 \xrightarrow{k_3} NO_2 + O_2 \tag{1.3}$$

If ozone concentrations are placed in approximate steady state, then the above photolysis cycle yields ozone concentrations that depend on solar radiation intensity and on the ratio of NO_2 to NO.

$$[O_3] \approx \frac{k_1[NO_2]}{k_3[NO]} \tag{1.4}$$

Additional reactions involving organic gases (RH) and radical species act to further convert NO to NO_2 in the atmosphere without consuming ozone. Thereby the NO_2 to NO ratio increases which results in increasing ozone production.

$$RH + OH\cdot + O_2 \rightarrow RO_2\cdot + H_2O \quad (1.5)$$

$$RO_2\cdot + NO \rightarrow NO_2 + RO\cdot \quad (1.6)$$

$$RO\cdot + O_2 \rightarrow R'CHO + HO_2\cdot \quad (1.7)$$

$$HO_2\cdot + NO \rightarrow NO_2 + OH\cdot \quad (1.8)$$

$$RO\cdot \xrightarrow{\Delta} \text{Products} \quad (1.9)$$

$$RO\cdot \xrightarrow{\text{Isomerization}} \text{Products} \quad (1.10)$$

$$R'CHO + OH\cdot \rightarrow RO_2\cdot + CO + H_2O \quad (1.11)$$

The ultimate sinks for ozone are destruction by chemical reaction, deposition to the earth's surface, or photolysis.

Ozone is a trace species in the atmosphere. If the entire atmospheric ozone volume were collapsed to a pressure of one atmosphere, it would only form a 3 mm thick layer (National Research Council, 1991). The downward mixing of stratospheric air was originally thought to be the dominant source of ozone in the remote troposphere (Junge, 1962). If that were the case, maximum ozone concentrations in the remote troposphere should occur in the spring since exchange between the stratosphere and troposphere is most effective at this time of year (Danielsen, 1968; Mahlman and Moxim, 1978). However, rural areas near anthropogenic sources of ozone precursors show a broad maximum in the seasonal variation of ozone concentrations which continues into the

summer. Average surface-level ozone concentrations have increased by 1 to 2% annually over the past 30 years in Europe (Logan, 1985; Janach, 1989) and transport from anthropogenic sources in North America appears to play a major role (Parrish *et al.*, 1993). Satellite data show a general increase in ozone concentrations over the temperate latitudes of the Northern Hemisphere and in large areas of the Southern Hemisphere downwind of widespread biomass burning (Watson *et al.*, 1990; Fishman *et al.*, 1991). Increases in background ozone concentrations would alter the chemical composition of the troposphere by increasing the oxidizing capacity of the atmosphere.

In urban areas such as Los Angeles and the Northeastern United States, ozone concentrations produced by reaction between anthropogenic ROG and NO_X can exceed 200 ppb over short time periods, roughly a factor of five above global background levels. Since the formation of high ozone concentrations requires energy from sunlight, and inversion layers that trap ozone precursors and ozone, meteorology plays a large role in the ultimate concentrations of ozone formed. In particular, the peak hourly values of ozone measured in a given year depend on the meteorological conditions of that year. This makes measuring the effectiveness of a control program difficult because the potential for ozone formation varies depending on the year. Several recent studies have attempted to adjust for the effect of changing meteorology from year to year in order to determine that part of the trend in ozone concentrations that is due to changes in precursor emissions (Chock *et al.*, 1982; Kumar and Chock, 1984; Davidson, 1986, 1993; Kuntasal and Chang, 1987). The general consensus is that the basin-wide ozone peak in the South Coast Air Basin (SoCAB) that surrounds Los Angeles has moved eastward over

time and that the number of Stage I episode days (maximum hourly ozone concentration greater than or equal to 200 ppb) has decreased over time.

1.1.1 Detrimental effects of ozone

Exposure to ozone for 6 to 7 h at relatively low concentrations (i.e., 80–120 ppb) has been found to significantly reduce lung function in normal, healthy people during periods of moderate exercise. Symptoms may include chest pain, coughing, nausea, and pulmonary congestion (Lippmann, 1989; Beckett, 1991). Repeated exposure to ozone over a period of years can cause permanent lung damage and accelerated aging of the lungs. The annual value of the health benefits that would accrue if the Federal ambient air quality standard for ozone was attained in the Los Angeles area alone has been estimated to be $2.7 billion per year (Hall *et al.*, 1992). Ozone also damages trees, crops, and materials (Skarby and Sellden, 1984). Ozone is estimated to cause several billion dollars of agricultural crop yield losses each year in the United States alone (Heck *et al.*, 1982).

1.1.2 Air quality standards

The Federal government has established standards for atmospheric pollutants. In the Clean Air Act Amendments of 1970, Congress specified that National Ambient Air Quality Standards (NAAQS) would be set. Over the subsequent years, such standards have been established for six criteria air pollutants: ozone, nitrogen oxides, sulfur dioxide, carbon monoxide, lead, and particulate matter. Two types of NAAQS exist. Primary standards are designed to protect public health. Secondary standards protect

public welfare, including protection against damage to vegetation, materials, and visibility. Until 1997, the primary and secondary NAAQS for ozone were set at a limit of 120 ppb for a maximum daily 1 h average, not to be exceeded more than once per year over a three year averaging period. Federal legislation also required states to submit a specific plan that described the methods that would be used to reach and maintain concentrations below the NAAQS for each criteria pollutant. These plans are known as State Implementation Plans (SIP) and were initially required to be submitted by the States in 1972. Under those plans, compliance with NAAQS was to be attained by 1977. In 1977 Congress amended the attainment schedule to extend it until 31 December 1987. However, by 1990 the air in approximately 100 locations in the U.S. with a combined population of over 100 million people still exceeded the NAAQS for ozone.

In 1990 Congress again extended the Clean Air Act, and again pushed back the date by which air quality standards for ozone must be met. The South Coast Air Basin (SoCAB) of California that surrounds the Los Angeles area is the only area classified as an extreme ozone standard non-attainment area (1987–1989 design value of 330 ppb; design value is the fourth highest 1 h average ozone concentration observed over 3 years). Areas that have ozone concentrations significantly above the ozone air quality standard must reduce emissions rapidly enough to achieve a 15% reduction in ROG emissions by 1996 and an average reduction of 3% ROG per year thereafter until attainment of the 1 h ozone standard of 120 ppb. Major NO_x sources must reduce emissions at the same rate required for major ROG sources unless the EPA finds that there is no benefit from further NO_x control.

The national ambient air quality standard for ozone in the United States was revised in 1997 to require that the 3 year moving average of the fourth highest 8 h average O_3 concentration of the year not exceed 0.08 ppm O_3 (Federal Register, 40 CFR Part 50, Vol. 62, No. 138). In those areas where O_3 concentrations historically have exceeded standards, the former requirement that the 1 h average O_3 concentration at each air monitoring site not exceed 0.12 ppm O_3 more than once per year when averaged over 3 years still applies as well. The U.S. Environmental Protection Agency also proposed but did not actually adopt a long-term O_3 exposure standard averaged over a growing season designed to protect against damage to crops. The present national ambient air quality standard for NO_2 in the United States is stated as 0.053 ppm not to be exceeded over an annual averaging period. Given the progression toward setting air quality standards for photochemically reactive air pollutants that apply over long averaging times or that are based on events that occur over a greater number of days per year, the ability to effectively design emission control strategies requires procedures for modeling the response of O_3 and NO_2 concentrations defined over averaging times longer than the 1 h peaks that occur within a two or three day smog episode.

Development of photochemical airshed models that can calculate annual average concentrations would also provide one step toward chemically reactive models for secondary aerosol formation that can operate over annual time periods. The annual average standard for fine particulate matter concentrations, rather than the daily standard, is the binding constraint on airborne particle control requirements in most air basins.

1.1.3 Photochemical airshed models

Photochemical airshed models are large computer codes that solve the atmospheric diffusion equation in the presence of the chemical reactions that lead to ozone formation. These air quality models provide the preferred method for studying the effects of proposed emissions control strategies directed at ozone abatement in advance of their adoption.

Eulerian photochemical model development at Caltech has a long history. The first model of this type was developed in the early 1970's (Reynolds *et al.*, 1973, 1974; Roth *et al.*, 1974) and used to examine the effect of the 1977 EPA control plan for the Los Angeles Basin based on the meteorological conditions of 29 September 1969 (Reynolds and Seinfeld, 1975). By the early 1980's a new comprehensive modeling system (McRae *et al.*, 1982b; McRae and Seinfeld, 1982) with an updated chemical mechanism (Falls and Seinfeld, 1978; Falls *et al.*, 1979) and improved input data generation (Goodin *et al.*, 1979, 1980) was used to model air quality over the period 26–28 June 1974. The chemistry of nitrogen-containing air pollutants was extended to include more gaseous species, additional nighttime reactions, and a thermodynamic equilibrium treatment of ammonium nitrate aerosol production (Russell *et al.*, 1988a). This version of the model and its associated 30 August 1982 – 1 September 1982 model performance evaluation data set were used to study the control of nitrogen-containing pollutants (Russell *et al.*, 1988b), the spatial patterns in pollutant responses to emission reductions (Milford *et al.*, 1989), the effect of alternate fuel use in motor vehicles

(Russell *et al.*, 1990; McNair *et al.*, 1992b), and modeling and control of the deposition of nitrogen-containing air pollutants (Russell *et al.*, 1993).

A more current version of the CIT airshed model (Harley *et al.*, 1992b, 1993) uses a revised dry deposition module (Russell *et al.*, 1993) and an extended version of the LCC chemical mechanism (Lurmann, Carter, and Coyner, 1987). The LCC chemical mechanism includes 8 lumped organic species classes: C4+ alkanes, ethene, C3+ alkenes, monoalkyl benzenes, di- and trialkyl benzenes, formaldehyde, C2+ aldehydes, and ketones. The mechanism has been extended to explicitly include the chemistry of methane, methanol, ethanol, methyl tert-butyl ether (MTBE), isoprene, hydrogen peroxide, and sulfur dioxide. This extended chemical mechanism contains 35 differential species, 10 steady-state species, and 107 chemical reactions.

1.1.4 Control strategy modeling

Careful selection of appropriate emission control strategies for ozone abatement benefits from the integration of air quality models with economic models. The problem of minimizing the cost of air pollution control can be formulated as a mathematical programming problem (Trijonis, 1974):

Select $\vec{x}$ that minimizes $C[\vec{x}]$ subject to

$$\vec{Q}[\vec{E}(\vec{r},t,\vec{x}),\vec{M}(\vec{r},t),\vec{P}] \leq \vec{s} \tag{1.11}$$

where the emissions for n primary pollutants, E_i, i = 1, n, are functions of space, $\vec{r}$, time, t, and control measures, $\vec{x}$. The meteorology, $\vec{M}$, is also a function of space and time. The air pollutant concentrations for k primary and secondary pollutants, Q_i, i = 1, k, are a

function of the emissions, meteorology, and chemical reaction parameters, $\vec{P}$. The emissions must be controlled such that the pollutant concentrations, $\vec{Q}$, are held below the applicable air quality standards, $\vec{s}$. Since the number of space and time points typically involved in air quality models is quite large, a single number is frequently sought that can more compactly characterize the air quality. Some examples include the yearly second highest 1 h average ozone concentration, the mean of the 30 highest 1 h average ozone concentrations, the yearly mean, and the numbers of hours per year that ozone concentrations exceed the NAAQS at a specific location.

Air quality descriptors are stochastic variables because the meteorology contains random fluctuations. Thus, Q is in fact a probability distribution calculated from E, P, and the probability distribution associated with M. Within the range of alternative emissions levels where the constraint that the air quality meets the standards has been satisfied, the goal is to minimize the total cost, C, of the control measures, x_i, i = 1, j.

One method for displaying the relationship between NO_X emissions, ROG emissions, and atmospheric O_3 concentrations is through an isopleth diagram (Haagen-Smitt and Fox, 1954; Milford *et al.*, 1989) in which the peak O_3 concentrations for a single day are plotted along lines of constant O_3 concentration that are a function of the combinations of ROG and NO_x emissions into the air basin of interest. Trijonis (1974) formulated a method for identifying the least cost approach to control of photochemical smog in Los Angeles that combined O_3 isopleth diagrams with a control cost-emission level relationship developed using a linear programming model. Trijonis (1974) used a highly empirical emission level–air quality relationship that specified the number of days

per year that the standards for O_3 and NO_2 are exceeded in Los Angeles County at any given level of basin-wide ROG and NO_X emissions based on organization of historically observed data on O_3, ROG, and NO_x concentrations in the atmosphere. This set of observed empirical relationships between historical ROG and NO_X emissions and the resulting O_3 air quality was organized in the form of an O_3 isopleth diagram labeled according to the number of days per year that air quality standards would be exceeded at various levels of ROG and NO_x emissions. The ROG and NO_X emission levels that would provide a feasible solution to the O_3 control problem then can be determined by inspection from this diagram. Curves showing the least-cost combination of emission controls needed to reach any specified level of ROG and NO_x emissions then can be overlaid on that O_3 isopleth diagram. The least cost solution to the O_3 control problem is the point where the control cost curves first touch the feasible air quality region where the O_3 and NO_2 air quality standards are both satisfied.

Trijonis' O_3 control strategy optimization approach provides a very interesting formulation of the problem that has never been pursued using chemically explicit ozone models because it would require a model for predicting the long-term frequency of occurrence of the daily ozone peaks at all levels of ROG and NO_x control. Such a model has heretofore been considered to be beyond present technical capabilities due to the great many days that would have to be analyzed at many different levels of ROG and NO_x emissions.

1.2 Motivation and Objectives

The objectives of this research are threefold: (1) to develop the capability to accurately model the long-term frequency of occurrence of high and low ozone concentrations in the urban and regional atmosphere, (2) to link that capability to control strategy performance analyses like those proposed by Trijonis (1974) that require calculation of the effect that emissions controls would have on the long-term frequency distribution of atmospheric ozone concentrations and (3) to explore the possibility that synthetic meteorological data could be used to speed the generation of the massive quantities of airshed model input data required by the long-term modeling process.

Traditionally, photochemical air quality models have been applied to analyze intense smog episodes of two to three days duration. The motivation for seeking to examine longer modeling periods is twofold. First, health effects research now shows that ozone affects public health over long periods of time, not just during a few 1 h peak events (Rombout et al., 1986; Lioy and Dyba, 1989). As a result of these findings, the form of the health-based ozone air quality standard in the United States was changed in 1997 to address a longer averaging time as well as a greater number of days per year. The new standard is based on the three year moving average of the fourth highest daily 8 h average ozone concentration that occurs each year. Models are needed that can be used to develop information on the frequency of occurrence of ozone concentrations above various target levels and for long averaging times in order to predict the frequency of exposure of the public to O_3 concentrations above the level of the new air quality standards.

Second, long-term modeling methods are needed to study how emission controls can be deliberately designed to reduce ozone concentrations defined over averaging times longer than the 1 h peaks that occur within a two or three day smog episode. The new O_3 air quality standard is directed at controlling a point on the long-term frequency distribution of O_3 concentration events that is far removed from the extreme represented by the highest 1 h average O_3 concentration observed each year. As a result, a potentially large number of O_3 events in each three year period now need to be analyzed when calculating the effect of proposed future emissions controls on compliance with the new O_3 air quality standards. It is highly likely that the spatial distribution of non-compliance areas and the response to controls will be very different for ozone standards defined over longer averaging times than for control programs directed at a few extreme events as has been the case in the past.

The ability to perform massive numbers of airshed model calculations would permit more thorough exploration of the likely effect of candidate emission control programs in advance of their adoption, and assist in the identification of the most cost-effective control programs. Despite many laws and cumulative abatement expenses exceeding $300 billion ($20 billion annually), numerous areas of the United States have not yet attained compliance with Federal air quality standards for ozone. A careful study of available control strategies is necessary not only to determine how to meet air quality standards, but also how to do so while spending the least amount of money. Prior studies have estimated that identification of the least costly sequence of control equipment can lead to attainment of an air quality standard at less than half the cost of applying the best

available control equipment to the largest sources first (Cass and McRae, 1981). As explained earlier, methods exist for the cost optimization of O_3 control strategies (Trijonis, 1974) that have never been used in conjunction with state-of-the-art photochemical airshed models because the computations would require an ability to predict how the frequency distribution of long-term O_3 concentrations would shift in response to emission controls. The purpose of the second objective stated above is to close this gap by determining how to best generate O_3 isopleth diagrams stated in terms of the number of days per year above specified levels as a function of all possible levels of ROG and NO_x control while using state-of-the-art photochemical airshed models to perform the analyses.

Traditionally, deterministic modeling of long-term ozone concentrations has involved not only large amounts of computer resources, but also huge amounts of time and effort to gather and prepare the input data needed for such calculations. For instance, while the research described in this thesis has reduced the amount of computer time needed for photochemical airshed model calculations to the equivalent of 3.5 min per day of simulation (see chapter 3 for details), the calculations that describe the 3 day August 1987 episode of the Southern California Air Quality Study (SCAQS) required over 6 months of effort just to set up the input data (Harley et al., 1993). Clearly, even if sufficient computer resources are available that would support calculations over long periods of time, those calculations are blocked until the input data can be generated more efficiently. Thus, the third objective of this research is to determine if the effort required to generate the meteorological fields required by photochemical airshed models based on

deterministic methods might be reduced greatly by turning to probabilistic methods based on the use of synthetic meteorology. The concept of synthetic meteorology is to draw inputs to the airshed model from the probability distribution of meteorological variables rather than directly from the records of specific historical days. Suppose that the atmospheric system can be defined in terms of the joint probability distribution of the system variable fields–temperature, wind, mixing height, solar radiation, humidity, and emissions. Then the input data for the large numbers of events that need to be modeled in order to reproduce the expected frequency distribution of high and low ozone concentrations at various emissions levels over long periods of time might be drawn quickly from the joint distribution of meteorological characteristics.

In the course of this thesis research, the long-term distribution of high and low ozone concentrations will be used to gain insights into the air pollution control situation in Southern California. The effect of emission controls on the frequency distribution of days exceeding the Federal 1 h ozone concentration standard and the new Federal 8 h O_3 standard in Southern California will be determined. This development provides key information that would permit the O_3 control strategy optimization problem of Trijonis (1974) to be worked for the first time in Southern California using models that feature an explicit description of atmospheric chemistry.

1.3 Approach

The input data sets needed to drive photochemical airshed models over long periods of time can be developed by two different methods. The first is the automated generation of meteorological fields directly from the historical observations for each day examined. The second is the generation of synthetic meteorology representative of the long-term distribution of events in an airshed. Each of these approaches will be investigated in turn.

In chapter 2, a model evaluation study is conducted to assess the accuracy and speed with which the distribution of O_3 and NO_2 concentrations throughout the year can be computed using a deterministic photochemical airshed model for O_3 formation driven by objective analysis of measured meteorological parameters. This study sets a baseline against which the performance of models driven by synthetic meteorological data sets can be compared. Previous research involving photochemical airshed models (e.g., Russell *et al.*, 1988ab; Harley *et al.*, 1993ab, Chico *et al.*, 1993) has generally utilized meteorological fields that have been painstakingly created for a few days with great care paid to human evaluation of the data. It is simply not feasible to develop hand-crafted meteorological inputs for use in a photochemical airshed model over the thousands of hours of simulation in a year. It is necessary to develop procedures that automate the generation of the required meteorological input fields, including hourly fields for winds, mixing depth, cloud cover, solar radiation, relative humidity, and temperature. The performance of a photochemical airshed model applied to the Southern California area

over a full year when driven by a fully automated procedure for generating meteorological input data is examined in detail in chapter 2.

Photochemical airshed models are the preferred tools for use in evaluation of the effect of proposed emissions control strategy changes in advance of their adoption. Thus, important measures of model performance include not just how well the model can reproduce historically observed O_3 concentrations, but also how faithfully it can be expected to respond to changed emissions inputs. In chapter 3, an analysis is conducted to examine the sensitivity of the predicted O_3 concentration response to changes in ROG and NO_x emissions as a function of the way that the inflows of ROG and NO_x across the boundaries of the modeling region are represented. It is found that O_3 control strategy performance predictions are so sensitive to the way in which boundary values are specified that it would be wise to greatly expand the boundaries of the modeling region traditionally used in Southern California in order to place the boundaries far over the ocean and the desert at locations where relatively clean air can be represented as entering the modeling region. As a result of this analysis, an expanded Southern California modeling region is used in chapters 2, 4, and 5 of this thesis, with calculations performed over an area stretching nearly from Mexico on the south to north of Point Conception, and from the Pacific Ocean west of Point Conception to the desert east of the San Jacinto Mountains.

In chapter 4, the emission control strategy modeling approach of chapter 3 is combined with the deterministic long-term O_3 modeling procedures developed in chapter 2. The objective is to develop methods for representing the way that the long-term

frequency distribution of high and low O_3 concentrations can be expected to shift in response to emissions controls. From this work, O_3 isopleth diagrams are generated that depict the number of days per year with O_3 concentrations above specified levels as a function of all possible combinations of ROG and NO_x control. Example calculations are worked for the case of O_3 control in Southern California. This analysis in turn paves the way for the O_3 control strategy optimization techniques developed by Trijonis (1974) to be applied using state-of-the-art mechanistic airshed models to generate the necessary data that describe the emissions / air quality relationship for O_3. This analysis also yields important insights into the extent of the ozone air quality improvement that is possible in the Los Angeles area given historically observed meteorological conditions. The effect that the switch from 1 h to 8 h averaging times will have on the magnitude of the O_3 air quality standard compliance problem in Southern California will be analyzed.

In chapter 5, generation of synthetic meteorological data is explored as an alternative method for creation of the huge meteorological input data sets that are required by models that calculate the effect of pollutant emissions on the long-term frequency distribution of regional O_3 concentrations. Stochastic models have been used extensively in the past to generate synthetic data for use in hydrological models (Thomas and Fiering, 1963; Fiering, 1967; Jackson, 1975). Generally the spatial nature of the problem is simplified so that time series analysis methods can be used (Box and Jenkins, 1976). A good example of this kind of analysis is provided by Jensen (1976) where synthetic streamflows were used to evaluate the operation of dams on the Colorado River.

In the case of air pollution modeling, three-dimensional meteorological data fields must be generated that describe winds, temperatures, inversion base heights, relative humidity, and solar radiation intensity. In this work, that process is begun by classifying the frequency of occurrence of the major synoptic weather patterns in the location of interest as a function of the time of year. Synoptic climatology seeks to explain the local climate in terms of large-scale atmospheric circulation (Barry and Perry, 1973) and provides a basis for statistically classifying the frequency of occurrence of major weather patterns as a function of time of year. Once the weather patterns for a geographic area have been classified in this manner, the process of generating synthetic meteorological data for use in an air quality model is begun by using a random number generator to draw a synoptic weather class appropriate to the intended place and time of year. Then progressive draws of key weather variables from the probability distributions of such parameters within that synoptic weather class proceeds until enough meteorological parameters have been drawn to initialize an advanced prognostic meteorological model. The prognostic meteorological model driven by the stochastically selected inputs and by local terrain data then proceeds to create the detailed meteorological fields required by the photochemical air quality model. The distributions of O_3 concentration predictions provided by the air quality model when driven by synthetic meteorological data then are compared to the historically observed distributions of O_3 concentrations. Comparisons are made between the model performance characteristics obtained when the modeling process is conducted using historically observed versus synthetically generated meteorological data sets.

Finally, the results of this research are summarized in chapter 6 and opportunities for further research are identified.

Chapter 2 Modeling the long-term frequency distribution of regional ozone concentrations

2.1 Introduction

The newly established national ambient air quality standard for ozone in the United States requires that the fourth highest 8 h average O_3 concentration each year, averaged over three years, not exceed 0.08 ppm O_3 (Federal Register, 40 CFR Part 50, Vol. 62, No. 138). In those areas where O_3 concentrations historically have exceeded standards, the former requirement that the fourth highest 1 h average O_3 concentration in three years at each air monitoring site not exceed 0.12 ppm O_3 still applies. In regions of the U.S. where the standards are exceeded, standard violations often occur on many days per year. While air quality modeling efforts directed at evaluating the likely effect of proposed emission controls usually are focused on understanding the single day with the highest observed O_3 concentration, the actual need is to evaluate and adopt emissions controls that will bring all days of the year below the intended O_3 concentration limit. Since it is not at all obvious *a priori* that the day with the highest observed O_3 is the hardest event to control, methods are needed that can model the shift in the entire frequency distribution of high and low O_3 concentration events throughout the year that will occur in response to emission controls. Indeed, the new O_3 standard is stated as a

---

Reference: Winner D.A. and Cass G.R. (1998) Modeling the long-term frequency distribution of regional ozone concentrations. *Atmospheric Environment*, in press.

statistical expectation over a 3 yr period with a fairly large number of days allowed above the 0.08 ppm limit before the standard has been exceeded. This again suggests that an ability to model the long-term frequency distribution of hourly O_3 concentrations would be desirable.

Recently, serious consideration also has been given to alternate forms of the U.S. national ambient air quality standard for O_3 such that compliance would be judged over even longer averaging times. The U.S. Environmental Protection Agency proposed but did not actually adopt a long-term O_3 exposure standard designed to protect against damage to crops stated in terms of the summation of all hourly O_3 concentrations that are greater than or equal to 0.06 ppm O_3 over a growing season (SUM06). The present national ambient air quality standard for NO_2 in the United States is already stated as 0.053 ppm not to be exceeded over an annual averaging period. Procedures for modeling very long periods of time are needed to study how emission controls can be deliberately designed to reduce O_3 and NO_2 concentrations defined over averaging times longer than the 1 h peaks that occur within a 2–3 day smog episode.

Further, annual average standards exist that limit particulate matter concentrations. In many locations, including Southern California as well as the northeastern U.S., approximately half of the fine particle concentrations (particle diameter < 2.5 μm) arise from gas-to-particle conversion products, including sulfates, nitrates, and secondary organic aerosol (e.g., Gray *et al.*, 1986). Photochemical airshed models for secondary aerosol formation are needed to study such problems (e.g., Russell *et al.*, 1988ab; Binkowski and Shankar, 1995). In order to study compliance with annual

average particulate air quality standards, methods are needed to model very long periods of record.

The present chapter is the first of a series that describes methods for efficiently modeling emissions–air quality relationships that govern O_3, NO_2, and particulate matter concentrations over very long periods of time (e.g., years and longer). Previous work by Tesche and McNally (1991a) sets the stage for this effort by their demonstration of a photochemical airshed model simulation of 30 days duration in which performance comparable to short episodic modeling was achieved. In this chapter, a baseline model evaluation study is conducted to assess the accuracy and speed with which the distribution of O_3 and NO_2 concentrations throughout the year can be computed using a deterministic photochemical airshed model for O_3 formation driven by objective analysis of measured meteorological parameters. A key aspect of this work involves the use of procedures that automate the generation of the meteorological fields needed by such models, including hourly fields for winds, mixing depth, cloud cover, solar radiation, relative humidity, and temperature. In most cases in the past (e.g., Russell *et al.*, 1988; Harley *et al.*, 1993ab, Chico *et al.*, 1993), meteorological fields have been painstakingly created for a few days at a time with great care paid to human evaluation of data from vertical temperature soundings, for example. Hand-crafted meteorological inputs are simply not feasible for use over the thousands of hours of simulation in a year. Several general questions arise about the feasibility of simulating the long-term distribution of high and low O_3 concentrations. How well will a photochemical airshed model perform over a full year when driven by a fully automated procedure for generating meteorological

input data? How well can such a long-term model reproduce proposed compliance measures such as the fourth highest 8 h average O_3 concentration of the year? The present analysis based on objective interpolation of meteorological observations further serves as a baseline against which other approaches such as the long-term use of prognostic meteorological models (Seaman *et al.*, 1995; Lyons *et al.*, 1995; Lu and Turco, 1995) or the use of stochastic meteorological simulations can be compared. Methods developed here will be illustrated by application to the air quality situation that exists in Southern California.

2.2 Description of model

2.2.1 CIT photochemical airshed model

Eulerian photochemical airshed models have been developed to study the formation, fate, and control of urban and regional O_3, NO_2, and particulate matter. These models are based on a mathematical description of pollutant emissions, transport, chemistry, and dry deposition (for surveys see Seinfeld, 1988; Seigneur, 1995). With a model that successfully simulates atmospheric transport and chemical reaction, the effect of reduced emissions of oxides of nitrogen (NO_X) and/or volatile organic compounds (VOC) can be examined.

The CIT photochemical airshed model (Harley *et al.*, 1993ab) is used for execution of the O_3 formation calculations in the present study. This Eulerian photochemical model numerically solves the atmospheric diffusion equation (Lamb and Seinfeld, 1973; McRae *et al.*, 1982)

$$\frac{\partial c_i}{\partial t} + \nabla \bullet (\vec{u} c_i) = \nabla \bullet (K \nabla c_i) + R_i + Q_i \tag{2.1}$$

where c_i is the ensemble mean concentration of species i, $\bar{u}$ is the mean wind velocity at location $\bar{x}$ at time t, K is the eddy diffusivity tensor (assumed to be diagonal), R_i is the chemical reaction rate of species i which depends on the pollutant species $c_1 \ldots c_n$ and on the temperature T, and Q_i is the elevated source emission rate of species i at location $\bar{x}$ at time t.

A no-flux boundary condition is applied at the top of the modeling region. The boundary condition at the earth's surface requires that the upward flux of each chemical species equals the ground-level emissions minus the dry deposition flux:

$$-K_{zz} \frac{\partial c_i}{\partial z} = E_i - v_g^i c_i \tag{2.2}$$

where K_{zz} is the vertical eddy diffusivity, E_i is the ground-level emission flux of species i, and v_g^i is the deposition velocity for species i.

The version of the CIT airshed model used here employs a revised dry deposition module based on surface resistance values (Russell *et al.*, 1993) and an extended version of the LCC chemical mechanism (Lurmann *et al.*, 1987). The LCC chemical mechanism represents organic gas emissions through 8 lumped organic species classes: C4+ alkanes, ethene, C3+ alkenes, monoalkyl benzenes, di- and trialkyl benzenes, formaldehyde, C2+ aldehydes, and ketones. The LCC mechanism has been extended by Harley *et al.* (1993ab) to explicitly include the chemistry of methane, methanol, ethanol, methyl tert-butyl ether (MTBE), isoprene, hydrogen peroxide, and sulfur dioxide. This extended

mechanism contains 35 differential species, 10 steady-state species, and 107 chemical reactions.

2.2.2 Model performance evaluation using SCAQS data

The CIT model has been tested extensively using experimental data from the August 1987 episode of the Southern California Air Quality Study (SCAQS; Harley *et al.*, 1993ab). When supplied with the best estimates of pollutant emissions in the Los Angeles area for the August SCAQS episode, the model predicts O_3 concentrations that display +1% normalized bias and an average station peak prediction accuracy of -7%. While the O_3 peak predicted at the time and location of the maximum observed O_3 concentration on 28 August 1987 is 33% below the observed value, the peak O_3 concentration unpaired in space is only 13% below the observed peak. The predictions for O_3 precursors show a normalized bias of +22% for total NO_2 and -12% for reactive hydrocarbons during the August SCAQS episode.

The CIT model also has been used to simulate the 24–25 June 1987 SCAQS episode (McNair *et al.*, 1996). The statistics for model performance for O_3 are similar to the August episode with the exception of an improved prediction accuracy for the basin-wide maximum (-3%). The predictions for total NO_2 show a normalized bias of +30%.

2.3 Modeling the long-term frequency distribution of ozone concentrations

2.3.1 Deterministic modeling of the inputs to long-term ozone calculations

The input data sets needed to drive photochemical airshed models over long periods of time can be developed by several different methods. The first approach, which is used in this chapter, employs the automated generation of meteorological fields directly from the historical observations for each day examined. A second approach, which will be examined in chapter 5, involves the generation of synthetic meteorological data that statistically resembles the long-term distribution of events in an airshed without seeking to reproduce the exact meteorological conditions of specific historical days. To set a baseline for evaluation of the characteristics of synthetic meteorological data sets, a long-term deterministic model simulation will be constructed first. The necessary input data will be assembled to execute the CIT model for each day over the period of a year using Southern California as the test area. The major issues that arise when organizing such a calculation involve the meteorological and emissions inventory inputs to the model. The year 1987 is chosen for examination because extensive research already has been done to improve the emissions inventory for that period based on respeciation of key source emission profiles plus assimilation of data from special studies of vehicle emissions within highway tunnels (Ingalls *et al.*, 1989; Harley *et al.*, 1992). The year 1987 also is of continuing importance because it contains the base year meteorological conditions that are used for O_3 air quality modeling and control strategy evaluation under the newly proposed 1997 implementation plan for photochemical smog control in the South Coast

Air Basin (SoCAB) that surrounds Los Angeles (South Coast Air Quality Management District, 1996).

2.3.2 Meteorology

Photochemical airshed modeling studies require input data that describe surface meteorology as well as winds aloft, solar radiation levels, and mixing depths. A description of the methods used to automate the construction of these meteorological fields follows.

2.3.2.1 Temperature and humidity

The approach used to generate the temperature and water vapor concentration fields required by the model is direct interpolation of observations. The weighted interpolation scheme used is a second-degree polynomial fitting procedure with an r^{-2} distance weighting scheme (Goodin *et al.*, 1979). Investigations by Kumar and Russell (1996) conclude that in areas with a dense meteorological measurement network, diagnostic meteorological fields generated by the method of Goodin et al. are superior to meteorological fields generated by current prognostic models. Interpolation barriers prevent interpolation through mountain ranges. The large number of surface meteorological stations in or near California's SoCAB are shown in Figure 2.1a. All of these stations report temperature, humidity, wind speed, and wind direction. Data at all of these sites were obtained from the National Climatic Data Center (NCDC), South Coast Air Quality Management District (SCAQMD), California Irrigation Management

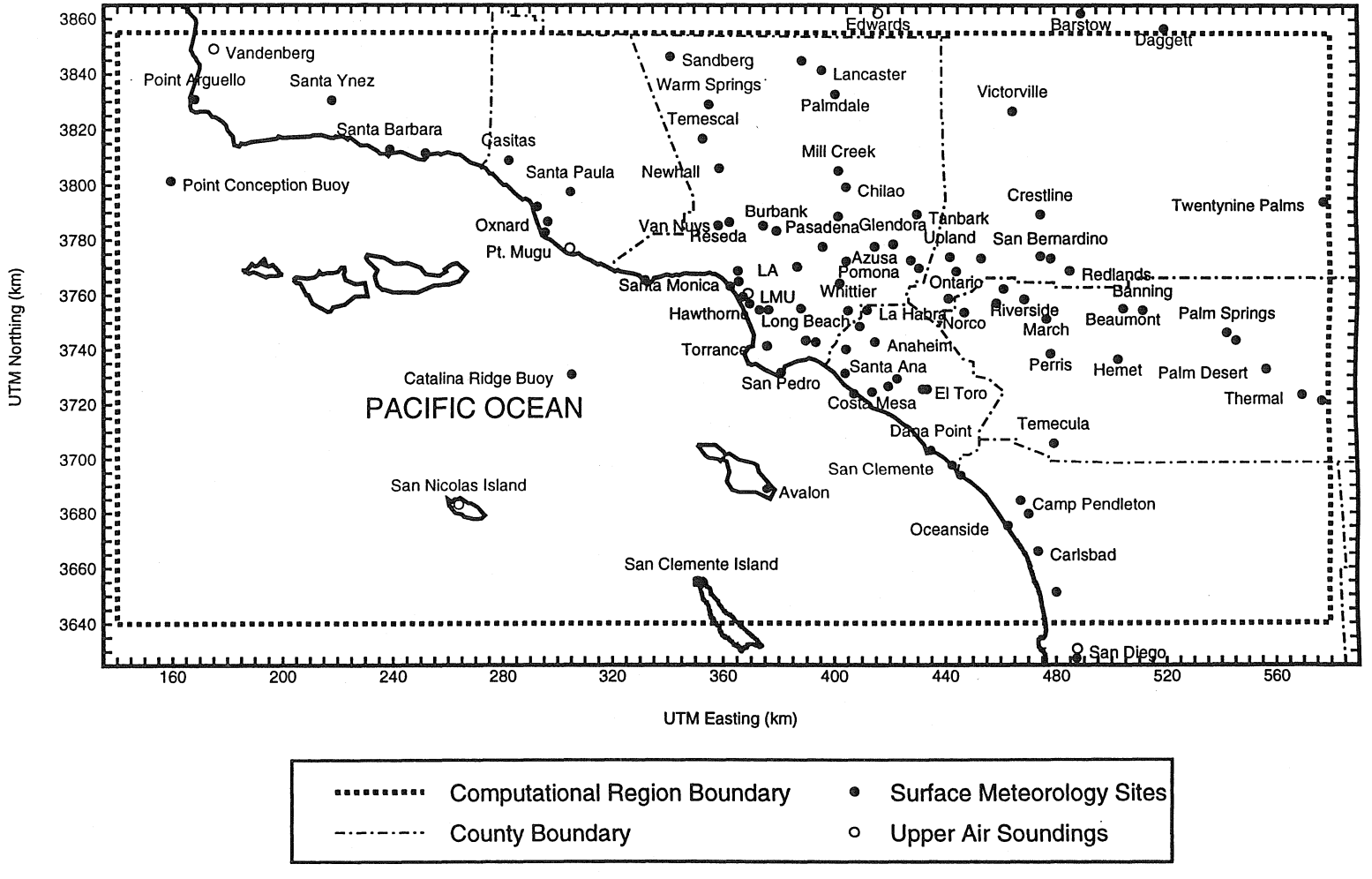

Figure 2.1a Southern California, showing the region over which air quality and emissions data were developed; the computational region for the air quality model; the county boundaries; and surface and upper air meteorology sites.

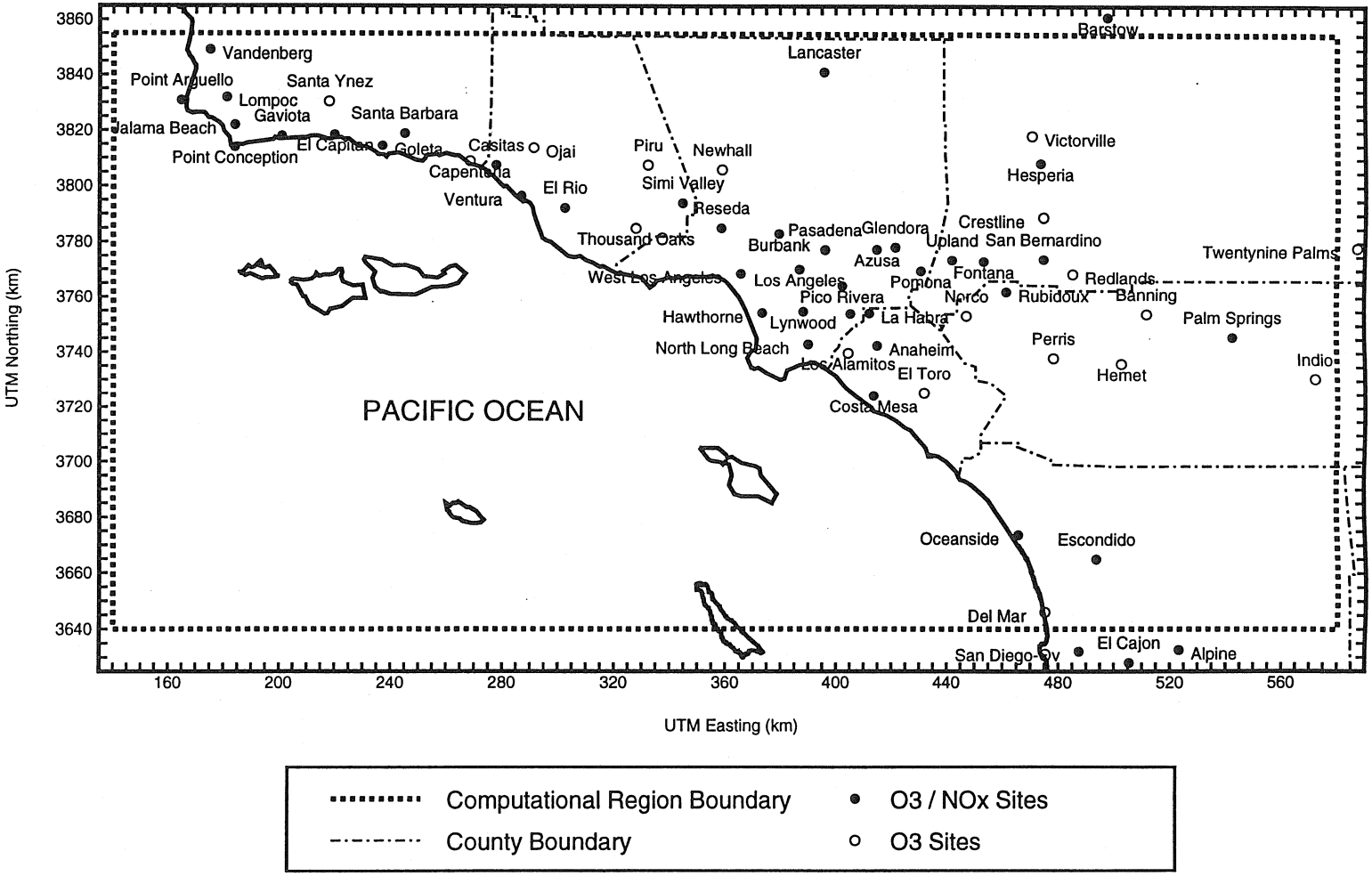

Figure 2.1b Southern California, showing the region over which air quality and emissions data were developed; the computational region for the air quality model; the county boundaries; and ozone and nitrogen oxides measurement sites.

Information System (CIMIS), Remote Automatic Weather Stations (RAWS) system, and California Air Resources Board (CARB) for the entire year 1987. The temperature and humidity fields were constructed by automating the interpolation procedure just described over a grid system of 5 km by 5 km cells spanning the modeling region shown in Figure 2.1.

2.3.2.2 Wind fields

The horizontal wind fields at each of 5 vertical levels are required as inputs to the airshed model. The 5 vertical layers in the model extend from 0–38.5 m, 38.5–154 m, 154–308 m, 308–671 m, and 671–1100 m above ground-level. Previous analysis of predicted peak concentrations of CO, O_3, and NO_2 as a function of the number of layers in the model shows that the incremental change in concentration predictions is large when vertical resolution is increased from 3 to 5 layers, but small when increased from 5 layers to 10 or 25 layers (McRae et al., 1992). Use of 5 layers achieves accuracy while conserving computing time. Objective analysis is used to generate the surface-level horizontal wind fields. Inverse distance squared weighting is applied on the x- and y-components of the observations, with barriers used to prevent interpolation through mountain ranges. The dense network of surface measurement sites used is shown in Figure 2.1a. There are six sites within the domain of Figure 2.1a at which vertical soundings of the winds aloft are taken several times daily: LMU (0500 and 1100 h PST); San Diego (0400, 1600); Vandenberg AFB (0400, 1600, occasionally 1000); Pt. Mugu (0400, 1000, 1600; soundings not taken at each time every day); San Nicolas Island

(0400, 1000, 1600; soundings not taken at each time every day); and Edwards AFB (various times, soundings not taken every day). These sites are spaced widely throughout the modeling region: four locations along the coast from the northern to the southern edge of the modeling region, one site on an offshore island in the middle of the Pacific Ocean portion of the modeling region, and one site in the desert at the north edge of the modeling region. An objective analysis program (Goodin *et al.*, 1980) is used to generate a 3-D wind field for each hour based on wind measurements up to 1500 m above ground level combined with the surface-level horizontal winds and mixing depths. A smoothing procedure is applied to reduce the divergence of the wind field. The vertical winds are computed from continuity considerations by the objective analysis program.

2.3.2.3 Solar radiation

Solar radiation data affect the model in several ways. Total solar radiation is an important parameter in determining the stability parameters (Pasquill-Gifford stability class, Monin-Obukhov length) used to calculate turbulent diffusion coefficients. The incident ultraviolet radiation is used to scale the photolysis rates of photochemical reactions. It is especially important to account for the effect of clouds on the photolysis rate of NO_2.

In the model, the total solar radiation flux is first computed as a function of the solar zenith angle while assuming clear sky conditions (McRae *et al.*, 1982). These clear sky values next are scaled as a function of location in the airshed according to actual measurements of solar radiation and cloud cover. Direct measurements of solar radiation,

available at CELA and CIMIS sites, are combined with cloud cover data that are available at most airports and military bases. The cloud cover measurements are first correlated with direct solar radiation measurements according to the procedure of Holtslag and Ulden (1983) to produce estimated solar radiation values at the cloud cover data sites. These values then are interpolated spatially using an r^{-2} weighting procedure to produce a scaling factor field that is used to correct the clear sky solar radiation values as a function of position throughout the modeling region.

Routine measurements of ultraviolet radiation are taken only at one site (Central Los Angeles), and there are large amounts of missing data in the first half of 1987. Special measurements of ultraviolet radiation were taken in the second half of 1987 at four sites (Central Los Angeles, Long Beach, Claremont, and Rubidoux) as a part of the SCAQS experiments that are thought to be of much higher quality. Using the SCAQS special study data and routine total solar radiation measurements, a regression line was determined that relates ultraviolet radiation to the total solar radiation measurements. This expression is used to estimate ultraviolet radiation levels from the routine total solar radiation measurements at sites with total solar radiation monitors. These ultraviolet radiation estimates which include the effects of cloud cover are used to produce an ultraviolet scaling factor field via an r^{-2} weighted spatial interpolation scheme. The scaling factor field is then applied to correct the hourly clear sky ultraviolet radiation fields produced by the model.

2.3.2.4 Mixing depths

Hourly mixing depths are calculated from the interpolated surface-level air temperature field plus the air temperature versus altitude data taken from upper air soundings at the 6 locations cited previously. Typical summer conditions in Southern California during the day result in well-defined inversion layers atop neutral and unstable layers near the surface. Typical summer conditions during the night produce slightly stable to neutral conditions near ground-level in urban areas with ground-based inversions in rural areas. Normal monitoring activities in Southern California provide upper air soundings typically 2–3 times per day at six sites. Since none of these sites are in the inland portion of the SoCAB near the eastern edge of the modeling region, a method must be adopted for estimating changes in mixing depths in the inland portions of the air basin during the course of the day. The method used in the present study for predicting mixing depth fields as a function of time and space at inland portions of the SoCAB is to determine the height at which the adiabatic lapse rates calculated from the local surface temperature field intersect the vertical temperature profile measured that morning near the coast according to Holzworth's (1967) method. This method tends to slightly overpredict the mixing depths in the inland portions of the SoCAB but has been found to be more accurate than the use of a prognostic meteorological model for this purpose (Cassmassi and Durkee, 1991). Figure 2.2 shows the inversion base heights predicted by this method compared to special measurements made as a part of the SCAQS study that are not a part of the routine set of soundings used to generate the mixing depth predictions Comparison of predicted versus observed mixing depths shows an average bias of 39.0 m.

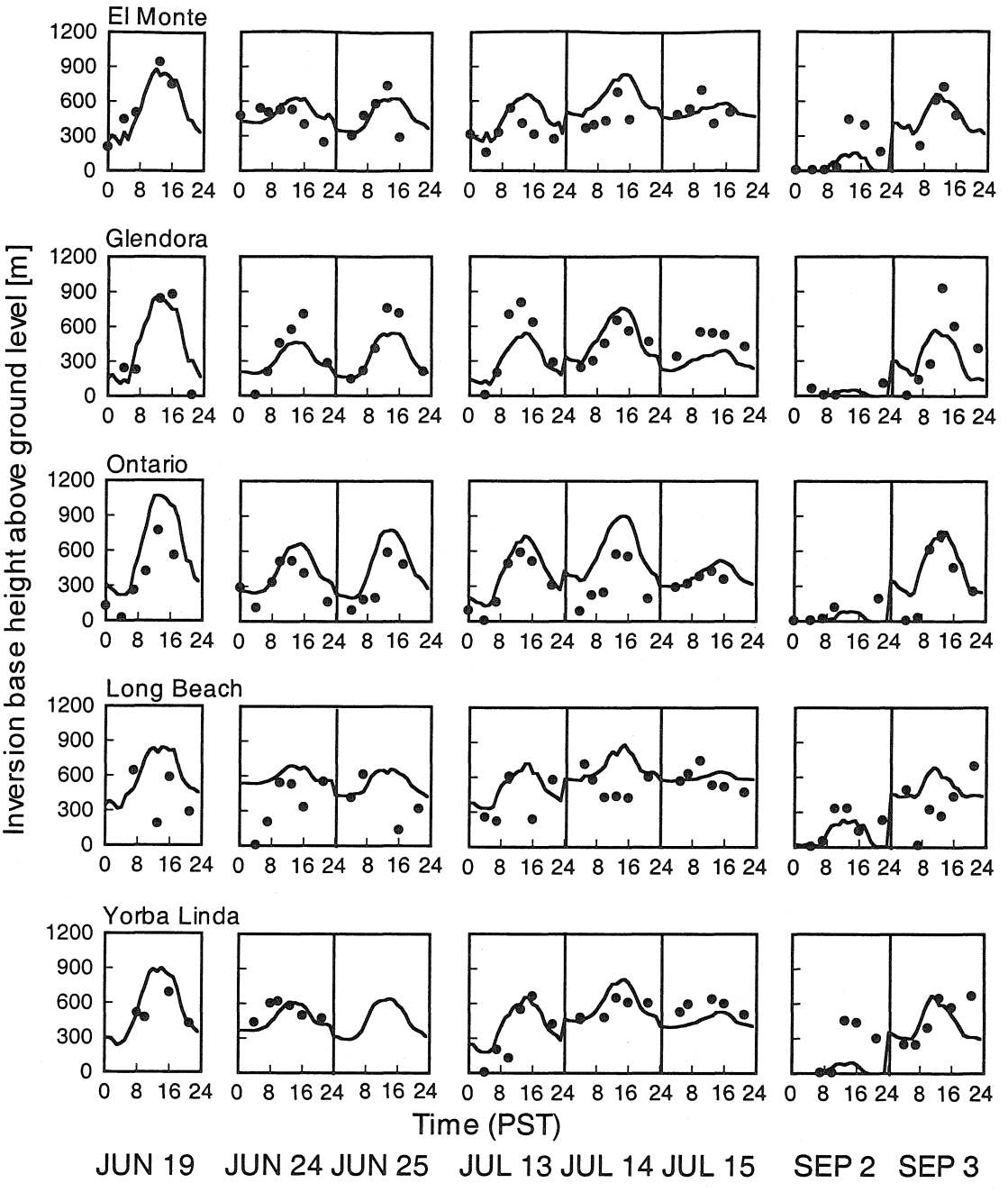

Figure 2.2 Inversion base height above ground level [m]. Comparison of predicted values from routine meteorological data via the Holzworth method to observed values generated from special soundings taken during the Southern California Air Quality Study.

2.3.3 Boundary conditions

Historically measured pollutant concentrations are used to specify the hourly boundary conditions that are supplied to the model. Figure 2.1b shows the network of stations which routinely measure O_3, NO, and NO_2. Additionally, on a typical day there are 18 sites routinely reporting total hydrocarbons (THC) or non-methane hydrocarbons (NMHC), 40 sites reporting SO_2, and 30 sites reporting CO data. Surface-level boundary conditions are specified by hourly interpolation of measurements made near the boundaries of the modeling region as described by Winner *et al.* (1995; this thesis chapter 3). In order to best estimate NMHC concentrations in remote areas near the boundaries of the modeling region, monthly average data that were collected in other years at four additional sites also are used along with the 1987 measurements (Anacapa Island, Barstow, Lancaster, and Palm Springs). Twelve of the hydrocarbon monitoring stations measure only THC. For those sites, NMHC is estimated from the THC data according to the procedure of Harley *et al.* (1993a).

The upwind boundaries of the air quality modeling region customarily used for photochemical airshed modeling in the Los Angeles area are located about 30–40 km west of West Los Angeles over the Catalina Channel and the Santa Barbara Channel areas of the near-coastal Pacific Ocean (Russell *et al.*, 1988; Chico *et al.*, 1993; Harley *et al.*, 1993ab; McNair *et al.*, 1996). Compelling evidence exists that the air over the ocean to the west of Los Angeles (generally thought of as "upwind") is contaminated with anthropogenic air pollutants from continental sources at concentrations much higher than the cleaner air found over the mid-Pacific Ocean (Winner *et al.*, 1995). In order to

minimize the effect of boundary conditions on O_3 concentration predictions in the SoCAB, the modeling region has been greatly increased in size as recommended by Winner *et al.* (1995) to cover the area shown in Figure 2.1. The western boundary of the modeling region now extends beyond the western boundary of the continent at Point Conception, CA. This places the western, inflow boundary of the modeling region in an area of lower pollutant concentrations than was used in previous modeling studies of the SoCAB. This larger modeling region contains a larger fraction of the emission sources that produce regional O_3 concentrations than in the past.

Measurements of NMHC concentrations are available at two sites on the western edge of the continent near Point Conception (Jalama Beach, Point Arguello). These values are used to set the inflow NMHC boundary conditions for the modeling region. Measurements at these sites are not taken every day; on average data are available at one or both of these sites every 3 days during the time period 6 am to 9 am local time. Hourly values are established by interpolation between the available measurements. The lower NMHC value measured at these two sites also is utilized at the southwest corner of the modeling region. This value has a mean of 27.4 ppbC with a standard deviation of 20.9 ppbC. The surface-level NMHC boundary values along the entire western edge of the modeling region are established by interpolation between the Point Conception area data and the value at the southwest corner of the modeling region. The surface-level NMHC values along the southern edge are established by interpolation between the southwest corner and data in the San Diego area. Surface-level boundary conditions for pollutants

other than NMHC are as described by Winner *et al.* (1995), except that CO concentrations along the western and southern boundaries also are lowered to 120 ppb.

Boundary conditions for pollutant concentrations in model levels above the ground level are set using historically measured values as described in Table 2 of Winner *et al.* (1995) except that the inflow NMHC and CO concentrations along the western and southern boundaries of the modeling region are reduced to reflect the fact that these boundaries have been pushed far out over the ocean. Measured pollutant concentrations over the Pacific Ocean along the western boundary of the SoCAB have been reviewed by Main *et al.* (1990, 1991). For inflow of clean air under prevailing westerly wind flow far from the Los Angeles coastline, Main *et al.* (1990) recommend inflow boundary conditions of 10–30 ppbC of NMHC and 30–40 ppb of O_3. New upper-level inflow boundary conditions to the model along the western and southern edges of the modeling region are set to the higher end of this range at 30 ppbC NMHC, 40 ppb O_3, and 120 ppb CO. Each ppbC of NMHC is speciated as 0.095 ppbv ALKA, 0.017 ppbv ETHE, 0.018 ppbv ALKE, 0.015 ppbv TOLU, and 0.016 ppbv AROM as specified by Harley *et al.* (1993a; Table IV) based on detailed analysis of organic gas data collected upwind of Los Angeles during the 1987 SCAQS experiments.

2.3.4 Initial conditions

Initial conditions used to begin the model simulation are based on interpolation of pollutant concentration measurements made at the air monitoring sites shown in Figure 2.1b. The model is initiated using data from 27 December 1986 which includes 69 sites

reporting O_3, 46 sites reporting NO and NO_2, 18 sites reporting NMHC or THC, 40 sites reporting SO_2, and 29 sites reporting CO. NMHC concentrations are calculated and supplemented as was described above when discussing the boundary conditions.

Five days (27–31 December 1986) are simulated as "spin up" days in order to completely flush the initial conditions out of the modeling region. After the first day of simulation, every successive day uses the final, predicted concentrations from the previous day to initialize the simulation.

2.3.5 Emissions inventory

Preparation of the emissions inventory for this study begins with the 1987 emission inventory developed for the SoCAB and adjacent areas (Los Angeles, Orange, Riverside, San Bernardino, and Ventura counties) by CARB. Measurements made in the Van Nuys highway tunnel during the 1987 SCAQS experiments indicate that the official CARB emission inventory for 1987 understates the hot stabilized exhaust hydrocarbon and CO emissions from light-duty motor vehicles by about a factor of two to three (Ingalls *et al.*, 1989; Harley *et al.*, 1993ab, Harley and Cass, 1995). In the present study, emission data for motor vehicles are employed that both are corrected for the effect of hourly changes in ambient temperature on emissions and scaled upward to match the emission rates measured during the Van Nuys tunnel study. Day-specific motor vehicle emissions for the SoCAB as a function of time and location throughout 1987 are computed using the Caltrans model DTIM2 (California Department of Transportation, 1994) in conjunction with motor vehicle emission factors from the CARB motor vehicle

emission factor program (EMFAC7F) and the hourly, day-specific surface temperature fields. Then the hot stabilized exhaust portion of the emissions are scaled upward by a factor of 2.25, which is the ratio of the tunnel study to EMFAC7F predictions for hot stabilized emissions from the Los Angeles vehicle fleet. Ideally, the motor vehicle emissions for weekend days would be based on weekend-specific traffic patterns. Unfortunately, no weekend traffic pattern information exists for Southern California, and thus the presently available weekday traffic patterns are used for all days of the week. Model performance statistics computed for weekends versus weekdays show that model performance was slightly better on weekends, thereby indicating that weekend emissions were adequately represented in the model.

In order to minimize the effect of upwind boundary conditions on O_3 concentration predictions in the Los Angeles area, the modeling region used for the present study is greatly extended when compared to previous studies. San Diego and Santa Barbara counties are now appended to the modeling region along with large areas of the Pacific Ocean. Gridded emission inventories for these additional areas are necessary. While 1987 county-wide emissions totals are available, detailed gridded emissions inventories have not been prepared previously for these outlying counties for 1987. For Santa Barbara County, the necessary gridded emissions data are prepared by scaling the spatially and temporally resolved 1984 emission inventory prepared for the SCCCAMP study (Tesche and McNally, 1991b) to match the county-wide emissions totals for 1987 on a major source category by source category basis (California Air Resources Board, 1990a) and subsequently adjusted to account for the increased motor

vehicle exhaust emissions described previously. For San Diego County, the spatially and temporally resolved emissions inventory for 1989 (Jackson, 1995) is scaled on a major source category by source category basis to match the county-wide emissions totals for 1987 (California Air Resources Board, 1990a). Again, the hot exhaust portion of the CARB motor vehicle emissions estimates are scaled upward according to the results of the Van Nuys tunnel study. At the conclusion of this process, the resulting emissions over the entire 450 km by 225 km grid of 5 km by 5 km cells that covers the modeling region of Figure 2.1 totals 2924 metric tons d^{-1} NMHC compared to 2304 metric tons d^{-1} NMHC for the smaller modeling region used previously by Harley *et al.* (1993a; the above comparison is for the temperatures of 28 August 1987). Speciation of the NMHC emissions within the model follows the revised source composition profiles previously compiled by Harley *et al.* (1992).

While not significant compared to 1987 emissions of anthropogenic hydrocarbons, the seasonal variation of biogenic hydrocarbons is important for simulations with greatly reduced anthropogenic emissions. Thus, scale factors adapted from the seasonally-resolved biogenic emissions inventory for the South Coast Air Basin reported by Benjamin *et al.* (1997) are used to modify the SCAQS 1987 August biogenic hydrocarbon inventory which totaled 125 t d^{-1}. The scale factors are 0.30, 0.63, 1.00, 0.40 respectively for winter (Jan.–Mar.), spring (Apr.–Jun.), summer (Jul.–Oct. 10), and fall (Oct. 11–Dec.). Note that 1–10 Oct. 1987 was a period of high O_3 concentrations and temperatures. The summer scale factor is used for this period.

At the end of this process of emission inventory generation, analysis showed that day-specific emissions vary considerably over the course of the year, depending largely on day-to-day temperature fluctuations. The daily NMHC emissions from motor vehicles in the entire modeling region range from 1335 metric tons d^{-1} to 2227 metric tons d^{-1} and the combined NMHC emissions from all sources in that region vary from 2580 metric tons d^{-1} to 3527 metric tons d^{-1} depending on the day of interest.

2.4 Evaluation of model performance

The photochemical airshed model was applied to the 1987 input data set defined over the modeling region of Figure 2.1. The year-long simulation required 311 h of elapsed time on a Digital AlphaStation 600 5/333 scientific workstation for the case employing the motor vehicle emissions inventory that was scaled upwards to match the emissions rates seen in the Van Nuys tunnel experiment. A second simulation employing the motor vehicle emissions calculated according to EMFAC7F without rescaling also was conducted for purposes of comparison. Faster execution is possible on a supercomputer (Winner *et al.*, 1995). However, it is important to note that it is practical to run such long-term simulations on relatively inexpensive desktop machines as was done here.

A formal evaluation of model performance conducted over the course of the one year simulation is necessary to establish a baseline for the speed and accuracy of such calculations. Since a vast amount of information is generated by simulation of one year, measures of model performance will be sought that reveal the degree to which the

frequency of occurrence of high and low O_3 and total NO_2 concentrations can be reproduced by the model. When judging model performance, it is important to keep in mind the likely accuracy of the component parts of photochemical airshed models in general. In a recent review of the subject, Russell and Dennis (1998) find that the uncertainty in O_3 predictions due to the accuracy of the rate coefficients and formulation of the chemical mechanism is about ±30%, uncertainty in the emissions inputs also are important and vary depending on location (the inventory used here has been corrected to remove many of the typical biases), and uncertainties due to meteorological inputs are typically in the range 5-15%. The numerical scheme used to solve the horizontal transport equations should introduce uncertainties of no more than ±10%. Other lesser uncertainties come from the numerical routines used to integrate the chemical rate equations (less than the horizontal advection uncertainties), from the organic speciation profiles (±10% or less), from the use of only 5 vertical layers in the model (±5% or less), from the dry deposition parameterization, and from boundary conditions. When these uncertainties are compounded, it is expected that overall normalized gross errors of about ±40 to ±50% are the best that can be expected when a large pool of airshed model predictions are compared against observations. Nevertheless, if the model is unbiased, it may represent the overall occurrence of high and low O_3 events quite well.

Time series plots of the sequence of predicted and observed 1 h average O_3 concentrations are shown in Figures 2.3–2.5 for a site in the western Los Angeles basin (central Los Angeles), a downwind site in the eastern zone of the SoCAB (Fontana), and a site at high elevation in the San Gabriel Mountains (Crestline). The central Los Angeles

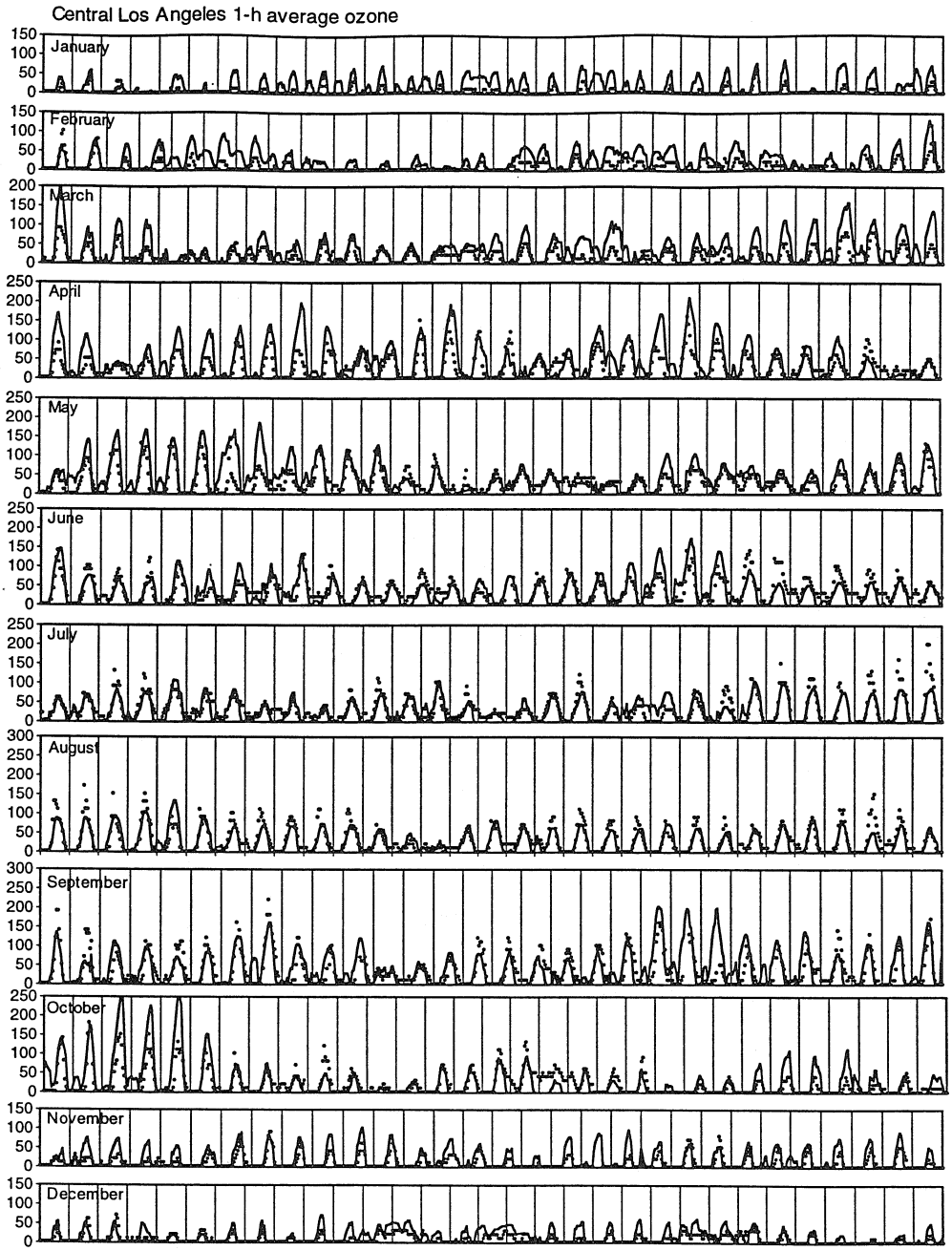

Figure 2.3 Time series of 1 h average O_3 concentrations [ppb] in the western South Coast Air Basin at Central Los Angeles throughout calendar year 1987: model predictions (solid line) versus observed values (circles).

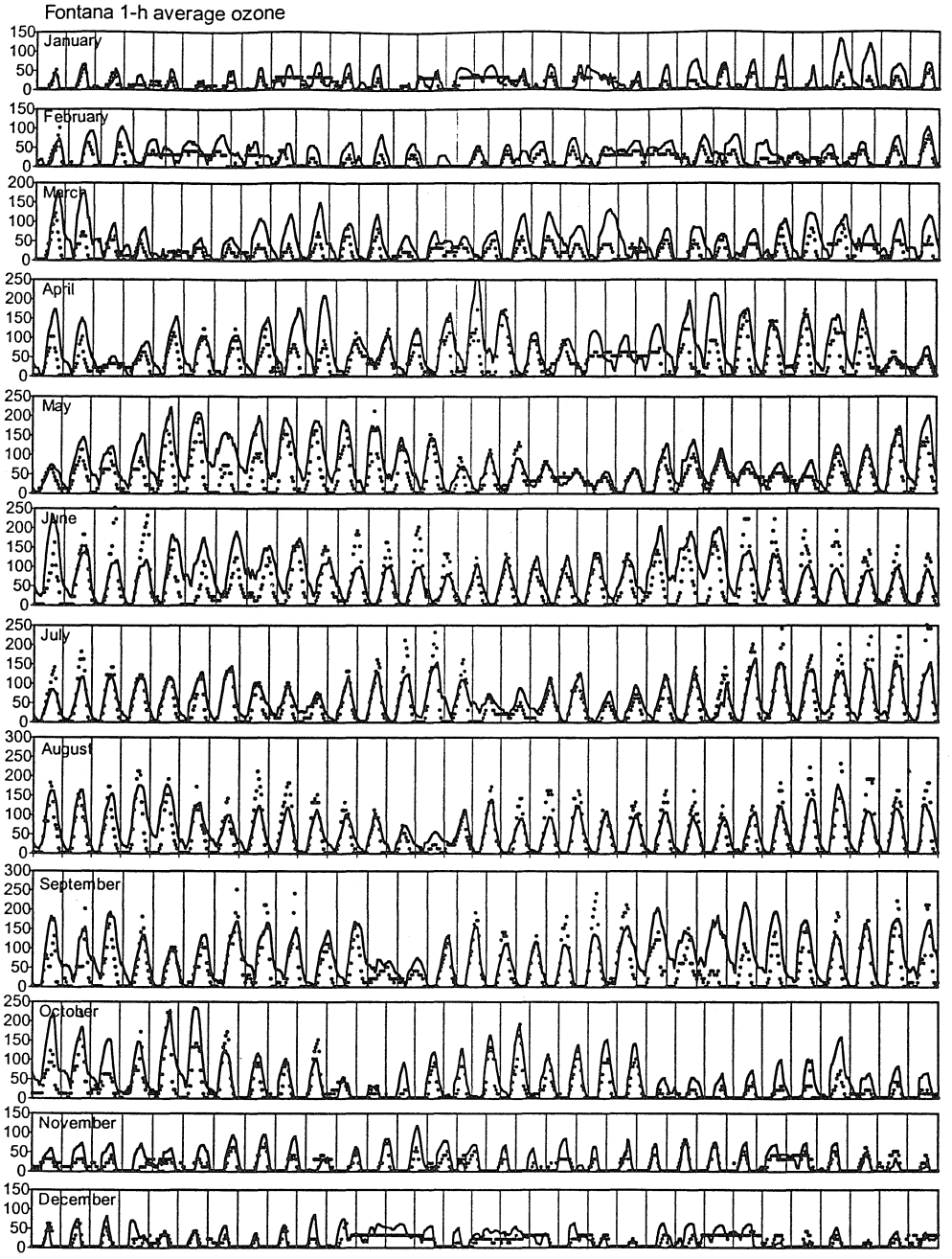

Figure 2.4 Time series of 1 h average O_3 concentrations [ppb] in the eastern South Coast Air Basin at Fontana throughout calendar year 1987: model predictions (solid line) versus observed values (circles).

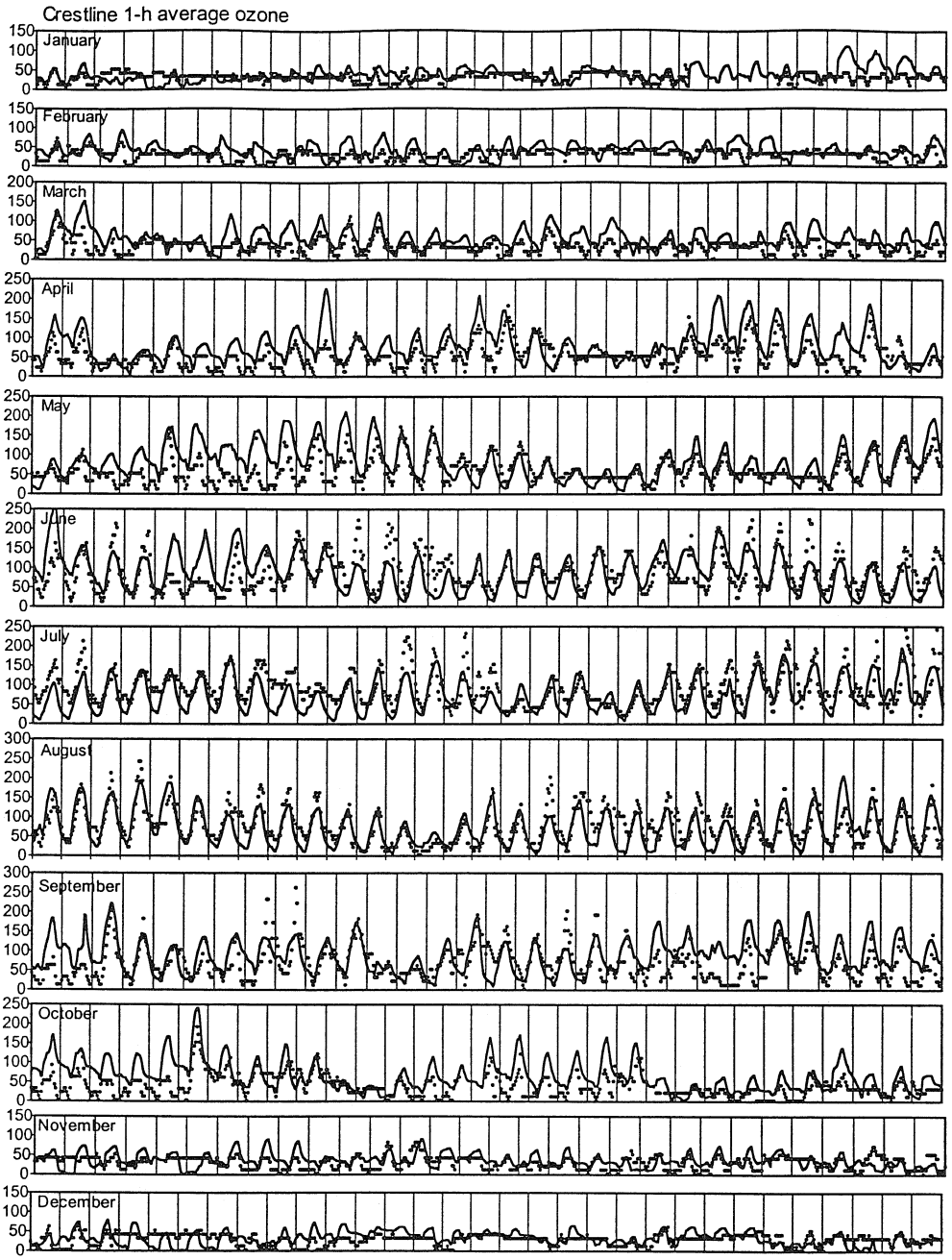

Figure 2.5 Time series of 1 h average O_3 concentrations [ppb] in the San Bernardino mountains at Crestline throughout calendar year 1987: model predictions (solid line) versus observed values (circles).

site is located in the area of highest traffic and population density. The Fontana site is typical of the heavily polluted photochemical smog receptor sites downwind of Los Angeles. The Crestline site is of interest because it has the highest annual average O_3 concentrations in the SoCAB (0.051 ppm; California Air Resources Board, 1987). Crestline has such a high annual average largely because sparse vehicle traffic in the mountains leads to less scavenging of O_3 at night by NO. Results shown are for the case with motor vehicle emissions scaled upwards to match the measurements made during the Van Nuys tunnel study. We believe that this is the best estimate for on-road vehicle emissions for the time period studied. Predicted O_3 values show reasonable agreement with observations over the course of the year, with occasional overpredictions and underpredictions as is the case for any set of photochemical airshed model results.

To quantify the model assessment, a statistical comparison of model predictions and observations above the cutoff concentrations for 1 h average O_3 concentrations is given in Table 2.1, based on data from all 54 O_3 monitoring sites in the modeling region shown in Figure 2.1b. There are 71,558 and 198,769 pairs of hourly predictions and observations for O_3 and NO_2, respectively, represented in the statistical summary of Table 2.1. As shown by the model bias statistics in Table 2.1, the model on average predicts O_3 concentrations to within -3 to +5 ppb during the smog season and 0 to +8 ppb over the entire year, depending on the emission inventory used. While the change in model predictions summarized in Table 2.1 as motor vehicle emissions are changed may look small (change of 8 ppb bias over our annual average), it is important to note that the measured annual average O_3 concentration over all monitoring sites in the modeling

Table 2.1 Statistical analysis of model performance for O_3 and total NO_2 (1 h avg.)

statistical measure[a]	model with EMFAC7F emissions	model with 2.25 x hot exhaust emissions[b]	typical performance [c]
Smog season (May–Oct)			
O_3 (1 h avg.)			
bias (pphm)	-0.3	0.5	
gross error (pphm)	2.7	2.7	
σ of residuals (pphm)	3.1	3.1	
normalized bias (%)	2	11	±15
normalized gross error (%)	31	33	±35
peak prediction accuracy[d]			
average station (%)	-2	7	
region-wide peak (%)	-22	-9	±20
total NO_2[e] (1 h avg.)			
bias (pphm)	0.0	0.1	
gross error (pphm)	1.9	2.0	
σ of residuals (pphm)	2.5	2.5	
normalized bias (%)	11	13	
normalized gross error (%)	54	55	
Entire year			
O_3 (1 h avg.)			
bias (pphm)	0.0	0.8	
gross error (pphm)	2.7	2.8	
σ of residuals (pphm)	3.0	3.1	
normalized bias (%)	6	15	±15
normalized gross error (%)	32	35	±35
peak prediction accuracy[d]			
average station (%)	5	14	
region-wide peak (%)	-5	7	±20
total NO_2[e] (1 h avg.)			
bias (pphm)	0.4	0.5	
gross error (pphm)	2.2	2.3	
σ of residuals (pphm)	2.7	2.9	
normalized bias (%)	21	23	
normalized gross error (%)	60	62	

[a]Bias and gross error statistics are calculated according to the recommendations of Tesche *et al.* (1990). The statistics use all pairs of hourly predicted and observed concentrations ≥ 6 pphm for O_3 and 2 pphm for NO_2.

[b]From the CIT airshed model using on-road vehicle hot-exhaust emissions scaled upward to match measurements made in the Van Nuys tunnel.

[c]California Air Resources Board acceptance criteria based on technical guidance for "typical" model performance (CARB, 1990b).

[d]The average station peak prediction accuracy statistic compares the measured peak values at all air monitoring locations within the modeling region to the highest 1 h average prediction for each day at each site that occurs within 3 h of the time of the measured peak. The values are computed as prediction minus observation divided by the observed values for each site every day and then averaged over all sites and days. The region-wide peak prediction accuracy statistic compares the highest 1 h average O_3 concentration predicted within a radius of 25 km and within ±3 h of the place and time of the single highest 1 h average O_3 value observed within the modeling region on each day. Results for each day are computed as prediction minus observation divided by the observed value and then averaged over all days.

[e]Total NO_2 includes NO_2 plus all other reactive nitrogen compounds that are measured as if they were NO_2 by chemiluminescent NO_x monitors (e.g., HNO_3 and PAN).

region is only 30 ppb, due in part to scavenging of O_3 at night by fresh NO emissions. The region-wide peak 1 h average O_3 concentration at the highest concentration site on each day is underpredicted on average by only 9% during the smog season and is overpredicted by 7% during the entire year when the emission inventory is used that was adjusted to conform to the Van Nuys tunnel experiment. The daily peak 1 h O_3 concentration averaged over all sites is overpredicted by 7% during the smog season and 14% over the entire year. Model performance statistics are within the range typically achieved in past studies of two or three days duration where great attention has been paid to preparation of hand-crafted meteorological fields based on expensive special field measurement programs (e.g., SCAQS). In Table 2.2, an analysis of model performance is presented that is stratified by O_3 concentration range, location, and time of day. Model gross error statistics are very similar over all subsets of the data examined. Table 2.2 confirms that there is a tendency to underpredict peak O_3 concentrations by 10-12% on days that exceed the 0.12 ppm 1 h Federal O_3 standard, as can be seen in Figures 2.3-2.5.

The long-term frequency distributions of the predicted and observed daily region-wide maximum 1 h average O_3 concentration are shown in Figure 2.6a and b. The upper tails of these distributions are fairly similar above the 0.12 ppm 1 h average O_3 standard. Observations somewhat exceed the predicted values at the upper tail of the concentration distribution, with three days observed to have O_3 values greater than 0.30 ppm and two days observed at 0.30 ppm. The highest predicted O_3 value is in the range 0.28–0.30 ppm. The largest differences between predicted and observed daily maximum region-wide 1 h average O_3 values occur at low concentrations where the measured values show a sharp

Table 2.2 Statistical analysis of model performance (2.25 x hot exhaust emissions) for O_3 (1 h avg.) stratified by ozone concentration, location, and time of day.

	Normalized bias[a] (%)	normalized gross error[a] (%)	peak prediction accuracy[b] (%) average station	peak prediction accuracy[b] (%) region-wide peak
Concentration level				
all values ≥ 0.06 ppm	15	35	14	7
all values ≥ 0.08 ppm	3	29	2	-1
all values ≥ 0.12 ppm	-15	28	-12	-10
Location				
all sites in modeling region	15	35	14	7
all SoCAB sites	17	39	16	12
Los Angeles and Orange County SoCAB sites	15	40	17	13
San Bernardino and Riverside County SoCAB sites	19	38	15	19
Time of day				
all hours	15	35		
morning (0600 – 0900 PST)	3	30		
Later in day (1100 – 1500 PST)	18	35		

[a]Bias and gross error statistics are calculated according to the recommendations of Tesche *et al.* (1990). The statistics use all pairs of predicted and observed 1 h average concentrations ≥ 6 pphm.

[b]The average station peak prediction accuracy statistic compares the measured peak values at all air monitoring locations within the modeling region to the highest 1 h average prediction for each day at each site that occurs within 3 h of the time of the measured peak. The values are computed as prediction minus observation divided by the observed values for each site every day and then averaged over all sites and days. The region-wide peak prediction accuracy statistic compares the highest 1 h average O_3 concentration predicted within a radius of 25 km and within ±3 h of the place and time of the single highest 1 h average O_3 value observed within the modeling region on each day. Results for each day are computed as prediction minus observation divided by the observed value and then averaged over all days.

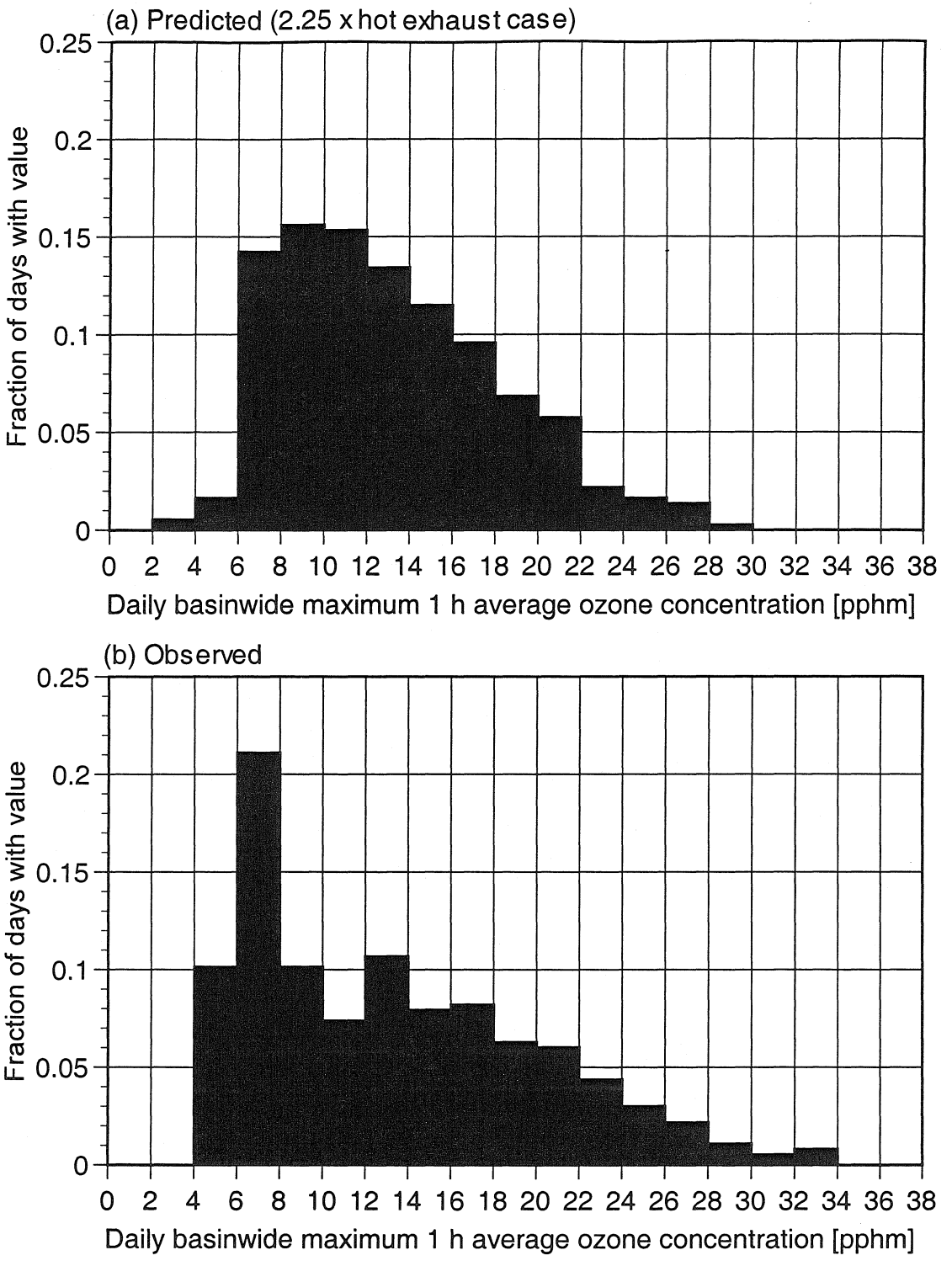

Figure 2.6 Frequency of occurrence of daily region-wide maximum 1 h average ozone concentrations (a) air quality model predictions (b) observed values.

peak in the range 0.06–0.08 ppm while the model predictions are more smoothly distributed over the range 0.06–0.12 ppm. Differences between predicted and observed daily peak O_3 at such low concentrations are especially sensitive to an accurate knowledge of the background O_3 values at the boundaries of the modeling region. Those background O_3 values are not known with high accuracy because daily O_3 measurements over the ocean upwind of the airshed are not available; only typical average values along the western boundary of modeling region are known. In spite of this situation, the number of days per year predicted to have a region-wide peak O_3 concentration below 0.12 ppm totals 171; very close to the 178 days observed to fall below 0.12 ppm O_3.

In addition to examination of model performance for O_3, it is desirable to confirm that the model is capable of simultaneously reproducing the concentrations of the ROG and NO_x that act as O_3 precursors. While detailed data on organic air pollutant concentrations are not available on a daily basis, the model used here has previously been evaluated for its ability to predict both the concentrations of lumped organic species used by the model as well as the concentrations of 53 specific gas-phase organic compounds during an August 1987 period in the middle of the year studied here (Harley et al., 1993b; Harley and Cass, 1995). It was found that total reactive hydrocarbons could be modeled with a normalized bias of -26% and that most individual organic species concentrations could be predicted within a normalized bias of ±50%.

The model performance evaluation for O_3 is accompanied by a comparison of model predictions and observations for NO_2. Figure 2.7 shows the time series of predicted and observed 1 h average NO_2 concentrations at the Pomona–Walnut Valley air

Figure 2.7 Time series of 1 h average total NO_2* concentrations [ppb] in Los Angeles County at the Pomona–Walnut monitoring site throughout calendar year 1987: model predictions (solid line) versus observed values (circles).

monitoring station of the South Coast Air Quality Management District. That site reported the highest annual average NO_2 concentration in Southern California in 1987, making it one of the highest, if not the highest, NO_2 sites in the United States. The measured annual mean NO_2 concentration reached 0.055 ppm at Pomona–Walnut that year, exceeding the national ambient air quality standard for NO_2. Figure 2.7 shows good agreement between measured and predicted NO_2 concentrations. A statistical analysis of model results for NO_2 based on comparison of 1 h average measurements and predictions at all available air monitoring sites is given in Table 2.1. Chemiluminescent NO_x monitors measure reactive nitrogen species such as HNO_3 and PAN as if they were NO_2. For this reason model predictions for NO_2 and other reactive nitrogen compounds that are measured as if they were NO_2 are summed to give predictions for "total NO_2" before predicted values are compared to observations.

Now that model predictions for NO_2 and other reactive nitrogen species can be made on an annual average basis, it is interesting to examine the effect of the upward measurement bias of the chemiluminescent NO / NO_2 monitors on measured compliance with the national ambient air quality standard for NO_2. In the case of the Pomona–Walnut station shown in Figure 2.7, the model predicts the annual average of reactive nitrogen species that would be measured as if they were NO_2 by chemiluminescent NO_x monitors to be 0.048 ppm compared to a measured annual mean of 0.055 ppm. Only 75% of the predicted annual average value is actually NO_2 according to the model; the remainder is HNO_3, PAN, and other minor nitrogenous air pollutants. The Pomona–Walnut site would fall below the 0.053 ppm annual average NO_2 standard if NO_2 were measured accurately.

The most recent changes to the U.S. national ambient air quality standard for O_3 emphasize the need to be able to predict the occurrence of 8 h average concentrations that exceed 0.08 ppm O_3. The time series of 8 h average O_3 concentrations thus was computed at each monitoring site; an example of this time series is shown in Figure 2.8. The San Bernardino site is typical of the heavily polluted photochemical smog receptor sites downwind of Los Angeles. A statistical comparison of 8 h average O_3 predictions and observations is given in Table 2.3 based on the time series of 8 h average values at all of the O_3 monitoring sites within the modeling region of Figure 2.1.

In Figure 2.9, the measured cumulative frequency distribution of the region-wide daily maximum 8 h average O_3 concentration is graphed based on the highest value at any single station on each day. The model results shown are from the case where the motor vehicle emissions inventory was adjusted to correspond to the results of the Van Nuys tunnel study. Two comparisons are given. The case with paired predictions and observations shows the distribution of 8 h average O_3 predictions at exactly the same site and times as the measured peaks. The second case with "unpaired" predictions and observations shows the distribution of the highest 8 h average O_3 value predicted within a 25 km radius and 3 h of the time of the measured peak, thus allowing a search for higher O_3 values predicted in the neighborhood of the peak monitoring site each day. The observed frequency distribution above 0.08 ppm O_3 is seen to fall between the paired and unpaired model predictions, approaching closely the distribution of paired model predictions at the highest concentrations.

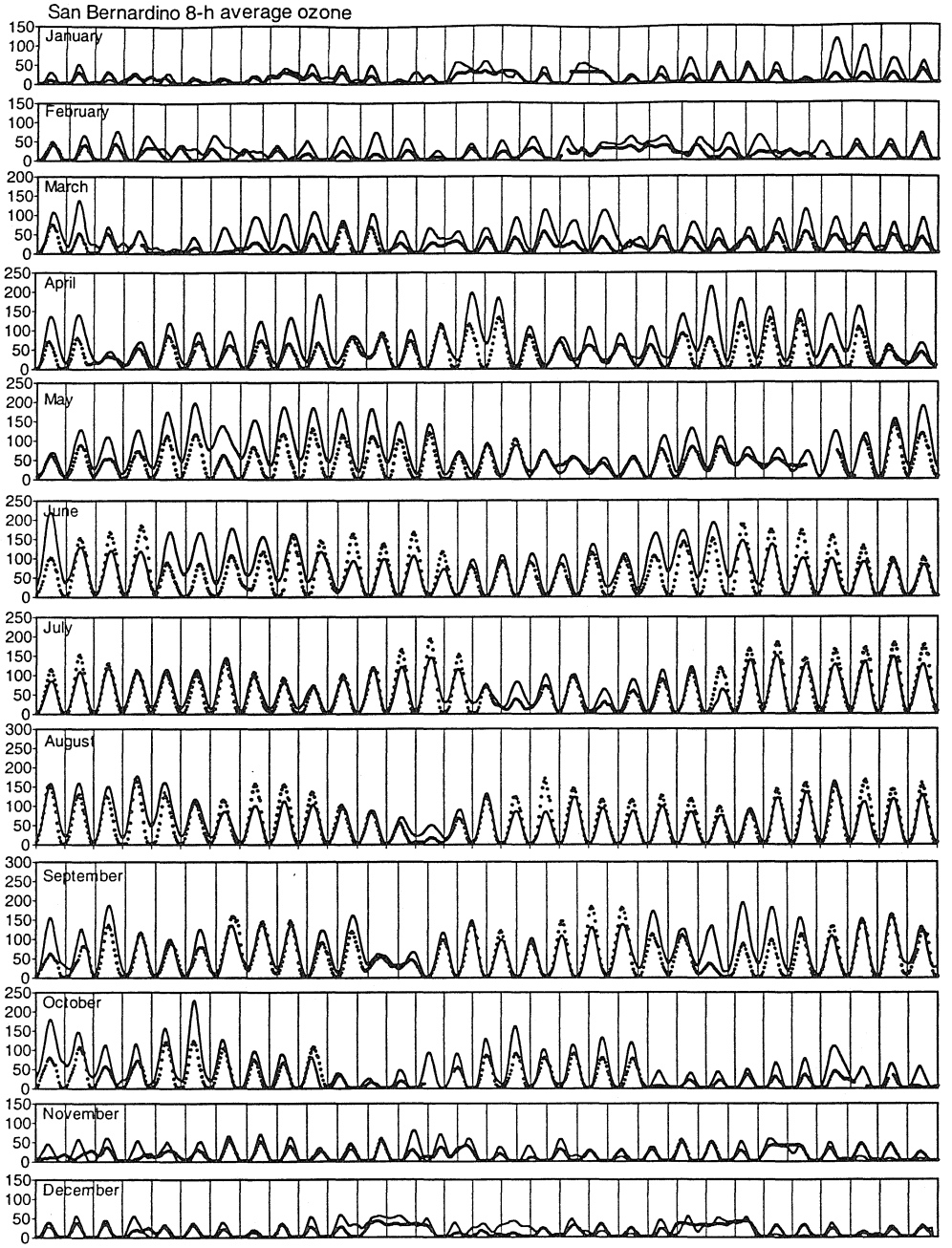

Figure 2.8 Time series of 8 h average O_3 concentrations [ppb] in the eastern South Coast Air Basin at San Bernardino throughout calendar year 1987: model predictions (solid line) versus observed values (circles).

Table 2.3 Statistical analysis of model performance for O_3 (8 h avg.).

statistical measure[a]	model with EMFAC7F emissions	model with 2.25 x hot exhaust emissions[b]
Smog season (May–Oct)		
O_3 (8 h avg.)		
bias (pphm)	0.5	1.1
gross error (pphm)	2.1	2.3
σ of residuals (pphm)	2.3	2.4
normalized bias (%)	14	23
normalized gross error (%)	35	38
peak prediction accuracy[c]		
average station (%)	2	11
region-wide peak (%)	-10	3
Entire year		
O_3 (8 h avg.)		
bias (pphm)	0.8	1.3
gross error (pphm)	2.1	2.3
σ of residuals (pphm)	2.3	2.4
normalized bias (%)	18	26
normalized gross error (%)	37	40
peak prediction accuracy[c]		
average station (%)	5	14
region-wide peak (%)	3	15

[(a)]Bias and gross error statistics are calculated according to the recommendations of Tesche *et al.* (1990). The statistics use all pairs of predicted and observed 8 h average concentrations ≥ 4 pphm.

[(b)]From the CIT airshed model using on-road vehicle hot-exhaust emissions scaled upward to match measurements made in the Van Nuys tunnel.

[(c)]The average station peak prediction accuracy statistic compares the measured peak values at all air monitoring locations within the modeling region to the highest 8 h average prediction for each day at each site that occurs within 3 h of the time of the measured peak. The values are computed as prediction minus observation divided by the observed values for each site every day and then averaged over all sites and days. The region-wide peak prediction accuracy statistic compares the highest 8 h average O_3 concentration predicted within a radius of 25 km and within ±3 h of the place and time of the single highest 8 h average O_3 value observed within the modeling region on each day. Results for each day are computed as prediction minus observation divided by the observed value and then averaged over all days.

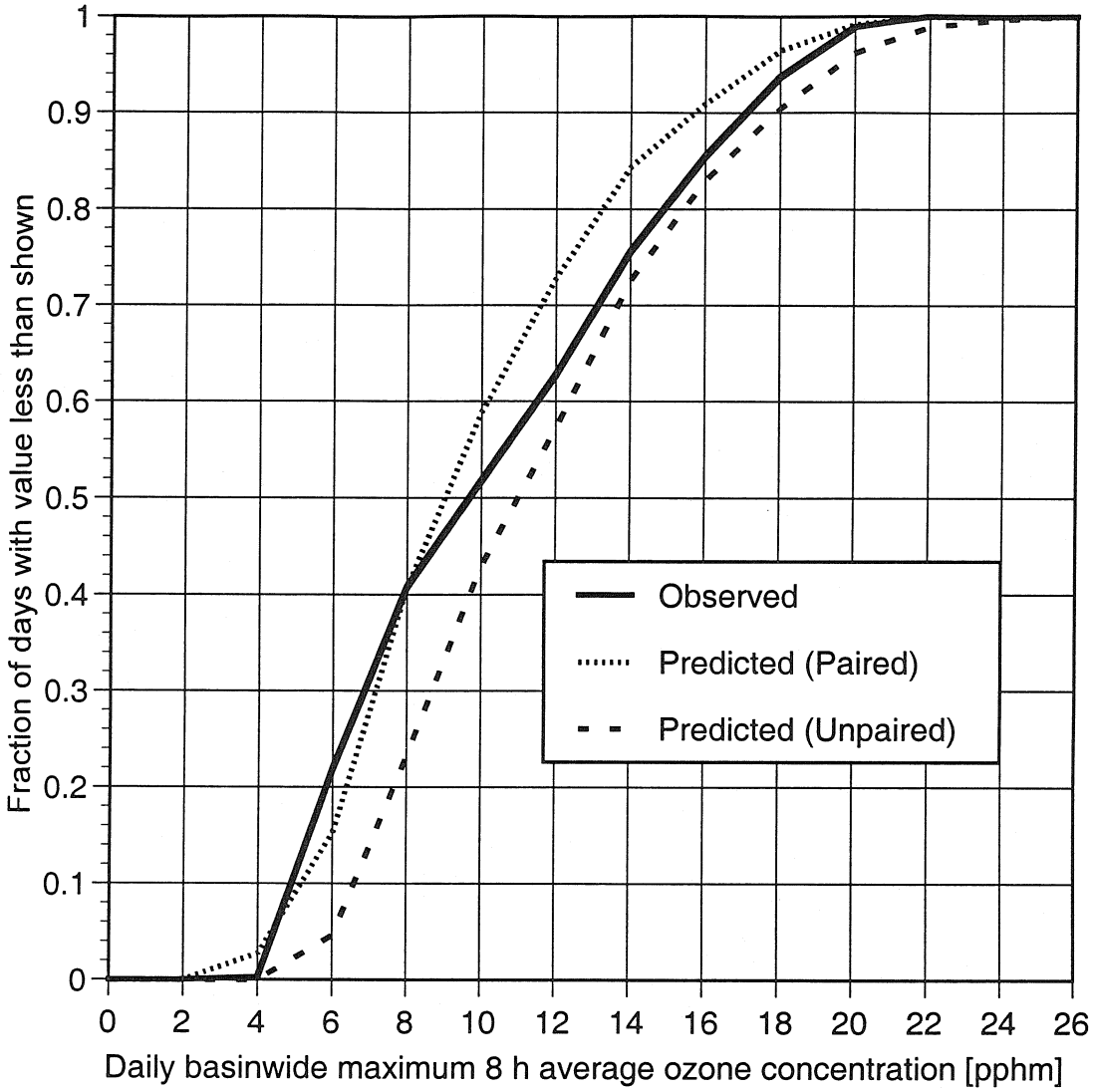

Figure 2.9 Cumulative frequency of occurrence of the daily regionwide maximum 8 h average ozone concentrations. Paired predictions are those at the time and place of the observed 8 h daily maximum O_3 concentration. Unpaired predictions are the highest 8 h average O_3 concentrations predicted within ±3 h and ±25 km of the time and location of the daily regionwide maximum 8 h average O_3 concentration.

2.5 Conclusions

Procedures have been developed that permit the relationship between air pollutant emissions and O_3 and NO_2 concentrations to be predicted over very long periods of time. While it is not possible to prove that the modeling procedures are completely without faults, the model has been challenged by more than 100 times the number of events normally encountered in the simulation of a 3 day episode, over a much wider than usual range of meteorological conditions, and over NMHC emissions that change based on daily temperature fields from a high of 3527 metric tons d^{-1} to a low of 2580 metric tons d^{-1}. Model performance statistics achieved when running for long periods of computation based on meteorological fields developed through automated assimilation of routine observations in Southern California are comparable to results usually achieved by episodic models driven by data from expensive special field studies such as the SCAQS experiments. Model results obtained by the present method can be used in the future to examine the effect of proposed emissions controls on the frequency of occurrence of high O_3 concentration events as well as to study the control of pollutants such as NO_2 whose concentration is regulated in the U.S. over annual averaging times. Analysis of annual average predictions for NO_2 at one of the highest concentration sites in the U.S. indicates that violation of the NO_2 standard at that site occurs because species in addition to NO_2 (e.g., HNO_3 and PAN) are measured as if they were NO_2 by the chemiluminescent NO_x monitors that are in routine use at governmental air monitoring sites.

Chapter 3 Effect of boundary conditions

3.1 Introduction

Eulerian photochemical airshed models have been developed to study the formation and fate of ozone, based on a mathematical description of chemistry, physics, meteorology, and emissions (for a survey see Seinfeld, 1988). One justification for the effort that goes into airshed modeling is that these models will be useful for evaluating emission control strategies. However, little thought has gone into seriously studying how best to perform such control strategy calculations. This is not a trivial issue because the computing resources consumed by such models can be so large as to limit the ability to systematically explore emissions control options (Cass and McRae, 1981). Typically, even if an air pollution control agency is able to gain access to a photochemical airshed model, the agency is only able to test one or two proposed control programs before they have exhausted their computational and personnel resources. Seldom are the computational resources available in order to fully explore seemingly subtle issues, such as how the boundary conditions supplied to the model will be affected by emission controls in adjacent areas.

Reference: Winner D.A., Cass G.R., Harley R.A. (1995) The Effect of Alternative Boundary Conditions on Predicted Ozone Control Strategy Performance: A Case Study in the Los Angeles Area. *Atmospheric Environment*, **29**, 3451–3464.

As we will show in the present study, the predicted effect of emission controls depends critically on how the estimate of future inflow boundary conditions responds to the control program. Initial conditions are of some importance, but their effect can be minimized by running the model for a period of time before using the results, in effect running away from the initial conditions. However, the effects of uncertainties in boundary conditions are not as easy to avoid. In particular, as the atmospheric concentrations of pollutants decrease due to the effect of emission reductions, it is reasonable to expect that the pollutant concentrations at the boundaries that are influenced by recirculation of aged pollutants from the airshed and by simultaneous emission control efforts in adjacent airsheds should also decrease. Inflow boundary conditions for modeling of historically observed air quality usually are set based on pollutant concentration measurements made at the edges of the modeling region. When performing calculations to represent future years, such measurements are not available. While the airshed model can dynamically supply its own outflow boundary conditions in future years, the airshed model itself cannot be used to predict the future year inflow boundary conditions. Yet the inflow boundary conditions affect the atmospheric concentrations computed within the modeling region. Thus, the atmospheric concentrations near the boundaries of the modeling region cannot be determined without knowing the boundary conditions in advance. Therefore, future year inflow boundary conditions must be supplied by assumption.

Due to the large amount of computational resources necessary to simulate the effect of different emission control strategies using a photochemical airshed model, most

previous studies have not been able to examine the effect of alternate future year boundary condition formulations over a wide range of reactive organic gases (ROG) and oxides of nitrogen (NO_X) emission rates. However, in those cases where the sensitivities of control programs to the effects of altered assumptions about boundary conditions have been explored, the effects of these assumptions have been found to be significant.

One of the first studies to use an airshed model to examine the effect of emission control strategies on ozone levels was conducted by Reynolds and Seinfeld (1975) to forecast the effect of the 1977 EPA Transportation Control Plan for Los Angeles. Two forecast emission inventories for 1977 were produced–one with the controls in effect and one without. These two emission cases were then used in an airshed model simulation based on historical meteorological data from 29 September 1969. This study found a difference of a factor of four in peak ozone levels at the Los Angeles County coast when comparing two different methods to set the future year initial and boundary conditions for NO_X and ozone. In this study the boundary and initial concentrations for CO, RHC, and NO_X for a future year simulation were set in proportion to the corresponding reductions in emissions of CO, RHC, and NO_X *within* the modeling region. The ozone initial and boundary concentrations were set based on the ozone steady-state relationship. Thus, specifying the ratio of NO_2 to NO fixes the ozone concentration. The NO_2 to NO ratio was calculated by two methods. The first method assumed that NO and NO_2 were both reduced in direct proportion to the reduction in total NO_X emissions in the region. Ozone boundary and initial conditions were then set based on the steady-state relationship between ozone, NO, and NO_2. The second method used a rollback relationship in which

total NO_X initial and boundary concentrations were taken to change in proportion to changes in NO_X emissions within the air basin. The ratio of NO_2 / NO (and hence ozone boundary and initial conditions) was assumed to change in proportion to hydrocarbon emissions within the air basin. The first method produced four times as much NO_2 as the second method with a corresponding noontime ozone concentration along the coastline that was four times higher than with the second method. Note that this is an extreme outcome; in that study the initial conditions were critical because the model simulation was only for 1 day. Also, the modeling region was smaller than will be used in the present study so the western boundary of the modeling region was closer to the coast.

In another study (Roth *et al.*, 1983; Tesche *et al.*, 1984), the emissions control requirements for the South Coast Air Basin necessary to attain the Federal ozone air quality standard by 1987 were examined. Airshed model simulations based on historical meteorological data from 26–27 June 1974 and 7–8 November 1978 were used in conjunction with several alternative emission inventory forecasts and control strategies for 1987. For each episode studied the model was executed for a 2 day period so that the results from the second day would be governed by emissions within the modeling region and not by initial conditions. The boundary conditions were not lowered as the emissions declined; historically measured boundary conditions were used.

A study of the effect of emission controls on ozone and nitrogen-containing pollutants in the Los Angeles area conducted by Russell *et al.* (1988) used the 1982 emission inventory for the South Coast Air Basin. Specific ROG and NO_X control measures were applied to the emission inventory in order to study future emission control

possibilities. The resulting effects on air quality were simulated using historical meteorological data from 30–31 August 1982. Historically observed boundary and initial conditions were maintained without alteration. A perturbation analysis of the effect of altered upper-level boundary conditions in the model was conducted. Reducing upper-level ozone boundary concentrations to 40 ppb resulted in only a 10 ppb reduction in predicted base case peak ozone concentrations on 31 August 1982.

In the most detailed study of ozone control to date, Milford *et al.* (1989) generated basin-wide ozone isopleth diagrams for the Los Angles area using a photochemical airshed model applied to 30–31 August 1982 historical conditions. Initial conditions were scaled downward in proportion to the changes in emissions studied while boundary conditions were maintained at historically observed levels. Cases with emission reductions of greater than 80% of the base case emissions were not presented because predicted peak ozone levels in cases nearing 100% control occurred on the far eastern boundary of the air basin where boundary conditions are uncertain.

The purpose of the present chapter is to explore the effect of assumptions made regarding boundary and initial conditions on the predicted performance of regional ozone control strategies. First, we will describe a computationally efficient approach to depicting the response of an air basin to emission controls. Then, we will examine the effect on predicted ozone concentrations of alternative assumptions about how boundary conditions at the edge of the air basin will change as a result of decreases in emissions upwind plus recirculation of aged pollutants from the air basin studied. Data from the South Coast Air Basin that surrounds Los Angeles will be used for example calculations.

We will show that accurate predictions of the effect of future controls in adjacent air basins plus the effect of recirculation of aged pollutants on the boundary conditions at the edge of the modeling domain are absolutely critical to selection of a successful ozone control strategy for Los Angeles.

3.2 Conceptual model of ozone control strategies

The effect of emission controls on ozone formation can be represented conceptually using an isopleth (lines of constant concentration) diagram such as Figure 3.1 (Haagen-Smit and Fox, 1954; Trijonis, 1974; Milford *et al.*, 1989). The horizontal axis represents the air basin-wide percentage control of the anthropogenic emissions of reactive organic gases (ROG), and the vertical axis represents the air basin-wide percentage control of the emissions of nitrogen oxides (NO_X). A line is drawn in Figure 3.1 along the diagonal ridge from lower left to the upper right. The area to the lower right of the ridge line has relatively high ROG / NO_X ratios. Ozone formation in this area is limited by the availability of NO_X since there is an ample supply of hydroperoxyl (HO_2) and organic peroxy (RO_2) radicals to convert nitric oxide (NO) to nitrogen dioxide (NO_2). Within this domain lowering both ROG and NO_X emissions or lowering NO_X emissions alone will result in lower peak ozone concentrations. To the upper left of the ridge line, the ROG / NO_X ratio is lower than along the ridge. Ozone formation in this area is ROG-limited since radicals are scavenged by the high amounts of NO_X present. When NO_X concentrations are lowered more of the radical pool can react with ROG

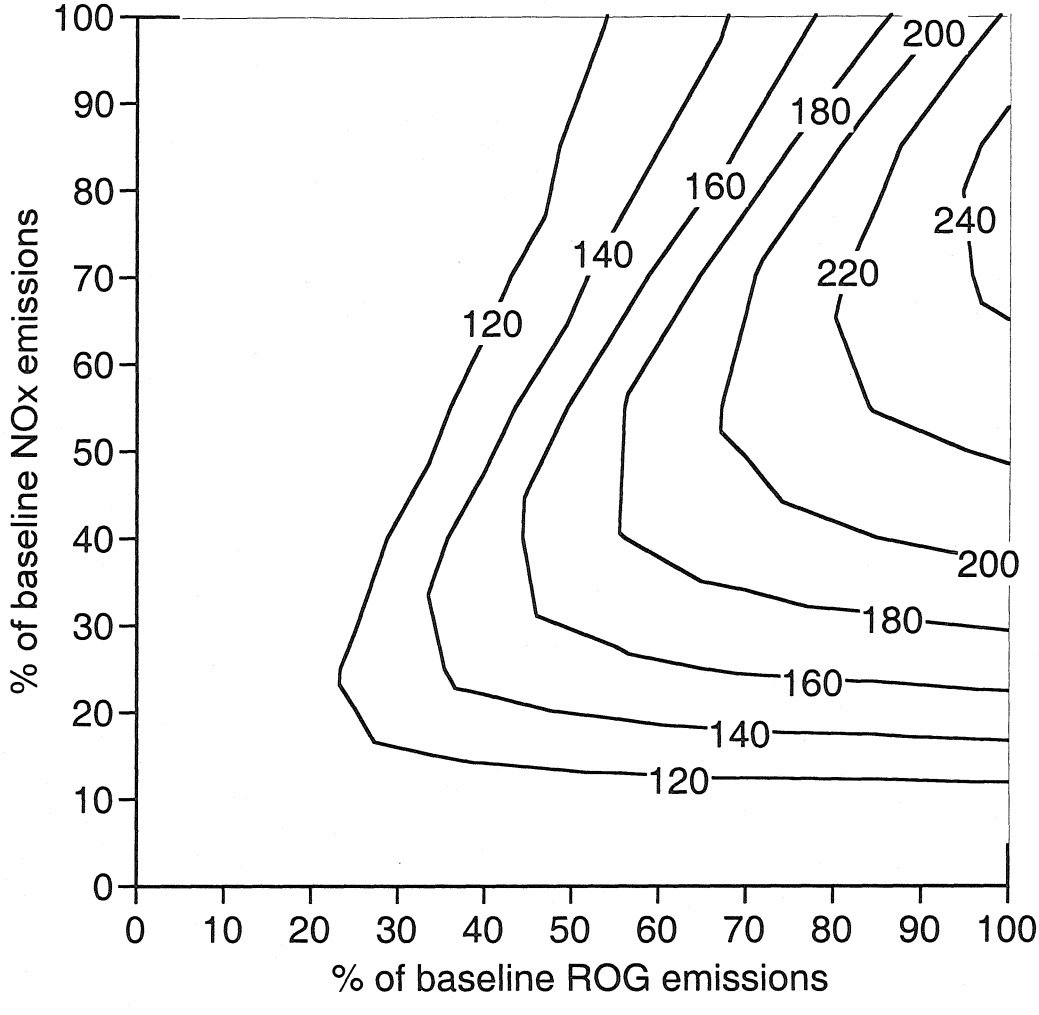

Figure 3.1 Ozone isopleth diagram showing the hypothetical response of peak 1 h average ozone concentrations within an air basin to changed levels of anthropogenic ROG and NOx emissions. Contour lines are lines of constant ozone concentration [ppb].

thereby promoting additional radical formation resulting in higher ozone concentrations. Within this upper left region of Figure 3.1, lowering ROG emissions alone or lowering both ROG and NO_X emissions can result in a decrease of ozone concentrations. Historically, ozone control efforts have emphasized reducing ROG emissions since the technology was initially less expensive and more readily available.

Isopleth diagrams are a widely understood method for communicating the effects of emission controls and have been constructed in the past using a variety of different air quality models. Haagen-Smit and Fox (1954) used physical smog chamber experiments to map the effect of altered inputs of ROG and NO_X on ozone formation. Trijonis (1974) used an empirical air quality model specific to the Los Angeles area to develop ozone isopleth diagrams that depicted the ozone response to altered ROG and NO_X emissions. The effect of changes in emissions was judged from ambient air quality data by examining afternoon ozone levels downwind of the central city on days with different early morning ROG and NO_X concentrations in the city center. The EKMA box models promoted by the U.S. Environmental Protection Agency (Dodge, 1977) have been used to estimate the response of ozone to control efforts in many locations throughout the United States. In the most detailed study of ozone controls conducted to date, Milford *et al.* (1989) generated isopleth diagrams using the much more complex results of a full photochemical airshed model. In that study, 46 basin-wide pairs of different reductions in ROG and NO_X emissions were used to study the predicted changes in the spatial distribution of ozone over the Los Angeles area. The predicted basin-wide peak ozone concentrations as well as daily peak ozone values at selected monitoring sites were

matched with the alternative ROG and NO_X emissions levels studied to form ozone isopleth diagrams.

Since photochemical airshed models are designed to simulate the actual physical and chemical processes occurring in the atmosphere, the use of an airshed model to generate isopleth diagrams should provide the most accurate results. A two-step process leading to control strategy selection is envisioned. First, ozone isopleth diagrams based on basin-wide percentage reductions in ROG and NO_X emissions from all sources are generated in order to locate the vicinity of the ROG and NO_X emissions levels at which compliance with the ozone air quality standard will occur. Next, model simulations that employ control measures applied to specific sources that will achieve the desired level of control can be used to refine control strategy performance. A major barrier to this approach is that the time required to generate a full isopleth diagram from a photochemical airshed model using a sequential computer is quite long.

Parallel computers offer an attractive increase in computing power that potentially can be used to speed the process of ozone isopleth diagram generation based on photochemical models. Since typically 90% of the computational time for an airshed model execution on a single processor is consumed by the chemical integration and vertical transport subroutine, the initial impulse has been to rewrite the actual airshed model transport and chemistry code in order to break the chemical calculations at any single time step in the model into many chemical calculations conducted in parallel (Carmichael *et al.*, 1989; Shin and Carmichael, 1992; Dabdub and Seinfeld, 1994). Thus, each computational node within a parallel computer has been used to solve the chemistry

and transport algorithms over a subset of the computational region. After modifying STEM-II, a photochemical airshed model, Saylor and Fernandes (1993) report a parallelized fraction, p, of 0.98 for their code. According to Amdahl's law (Amdahl, 1967) the maximum speed-up, s, possible on a parallel processing machine with N processors is:

$$s = \frac{1}{(1-p)+\frac{p}{N}} \tag{3.1}$$

Figure 3.2 shows the projected speed-up as a function of the number of processors, N, assuming p = 0.98. Note that as the number of processors increases, the corresponding maximum speed-up is less than linear–using 10 processors yields a speed-up by a factor of 8.5; using 100 processors yields a speed-up by a factor of 33.6; using 1000 processors yields a speed-up by a factor of 47.7. Another problem that arises when parallelizing the transport and chemistry code is that the airshed model computer code must be extensively rewritten in order to divide the chemistry and transport calculations into geographically distinct sub-regions over the airshed.

An alternative way to view the isopleth diagram generation problem is to view the problem as a matrix in ROG / NO_X emission space. Then a mesh of grid cells can be defined over that space such as in Figure 3.3, where each node in that grid system represents a different combination of across-the-board percentage reductions in anthropogenic ROG and NO_X emissions. The mesh of node points should be designed to completely span the range of possible emission reductions at a high enough density that accurate contour lines can be drawn. Higher resolution (tighter node spacing) can be used

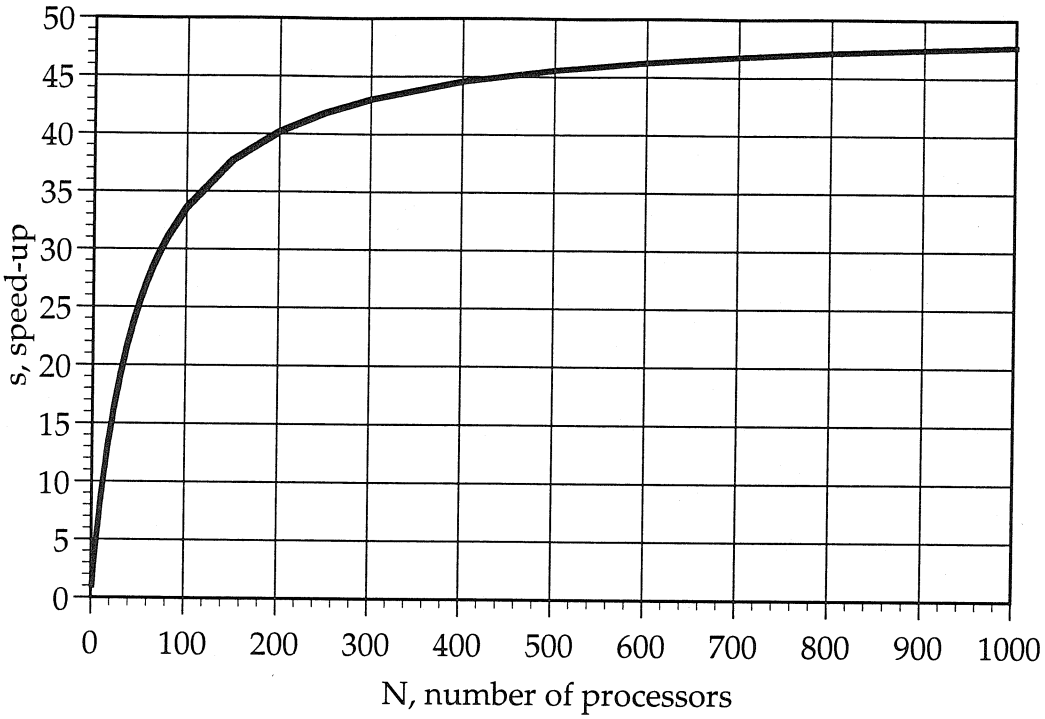

Figure 3.2 Amdahl's law showing the increase in computational speed as a function of increasing number of processors for a computer code with a parallelized fraction of 0.98 (i.e., 98% of operations occur in parallel).

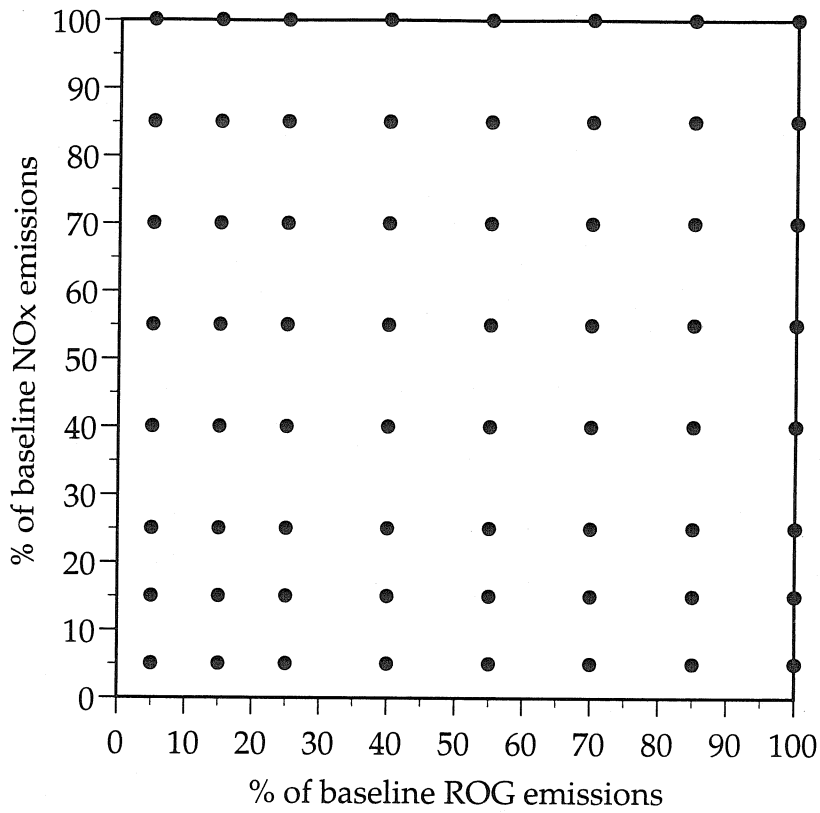

Figure 3.3 Decomposition of the ozone isopleth generation problem into a matrix of 64 combinations of emissions levels in ROG and NOx emission space.

in those areas of the isopleth diagram that are expected to yield peak ozone concentration values that are close to achieving compliance with the ozone air quality standard in the air basin studied. Each point of the mesh in ROG / NO_X emission space can be assigned to one node of a parallel computer with one complete copy of the airshed model assigned to each node. Thus, each node completes an airshed model simulation of the entire computational region using the same meteorological conditions. Only the emission inputs, and in some cases possibly the boundary and initial conditions, to each node change. The output is collected and contoured to form an isopleth diagram showing the air basin-wide 1 h average peak ozone concentration as a function of basin-wide ROG and NO_X emission levels. That 1 h peak ozone concentration is tracked as it moves in time and space within the air basin as the level of ROG and NO_X emissions is changed. As long as input and output bottlenecks are avoided, this method exhibits a *linear* speed-up. For a 64-node parallel computer, defining the problem as an array in ROG and NO_X emissions space produces a 64-fold speed-up of the calculation. This is not the same as the time required if a parallelized version of the airshed model is distributed over 64 nodes and then is used to successively process the 64 combinations of ROG and NO_X emissions needed to produce an isopleth diagram. Amdahl's law holds that a parallelized code with a parallelized factor of 0.98 run successively over the necessary ROG and NO_X emissions combinations would exhibit a speed-up of only 28.3-fold. In the present study, air basin-wide ozone isopleth diagrams will be generated using a 64-node parallel computer with the ROG and NO_X control problem defined as an array in ROG and NO_X space in order to capture the potential linear increase in computational speed.

3.3 The predicted response of South Coast Air Basin to emission controls

As an example of the ROG / NO_X emission space decomposition of the ozone control modeling problem, the response of the South Coast Air Basin that surrounds Los Angeles to changes in ROG and NO_X emissions will be examined. Figure 3.4 shows the South Coast Air Basin and surrounding areas as well as the computational region used for this study. The 27–28 August 1987 photochemical smog episode studied as part of the extensive Southern California Air Quality Study (SCAQS; see Lawson, 1990 for a description of the SCAQS experiment) will be used as the basis for test calculations.

3.3.1 CIT photochemical airshed model

Harley *et al.* (1993a) describe the application of the CIT photochemical airshed model to simulate the base case 27–28 August 1987 conditions in Southern California. The CIT airshed model (Harley *et al.*, 1992b, 1993a) solves the atmospheric diffusion equation with chemical reaction subject to boundary and initial conditions, data on the meteorological conditions, and the emissions of ozone precursors. The model uses a revised dry deposition module (Russell *et al.*, 1993) and an extended version of the LCC chemical mechanism (Lurmann *et al.*, 1987). The LCC chemical mechanism represents organic gas emissions through 8 lumped organic species classes: C4+ alkanes, ethene, C3+ alkenes, monoalkyl benzenes, di- and trialkyl benzenes, formaldehyde, C2+ aldehydes, and ketones. The LCC mechanism has been extended by Harley *et al.* (1993a) to explicitly include the chemistry of methane, methanol, ethanol, methyl tert-butyl ether (MTBE), isoprene, hydrogen peroxide, and sulfur dioxide. This extended mechanism

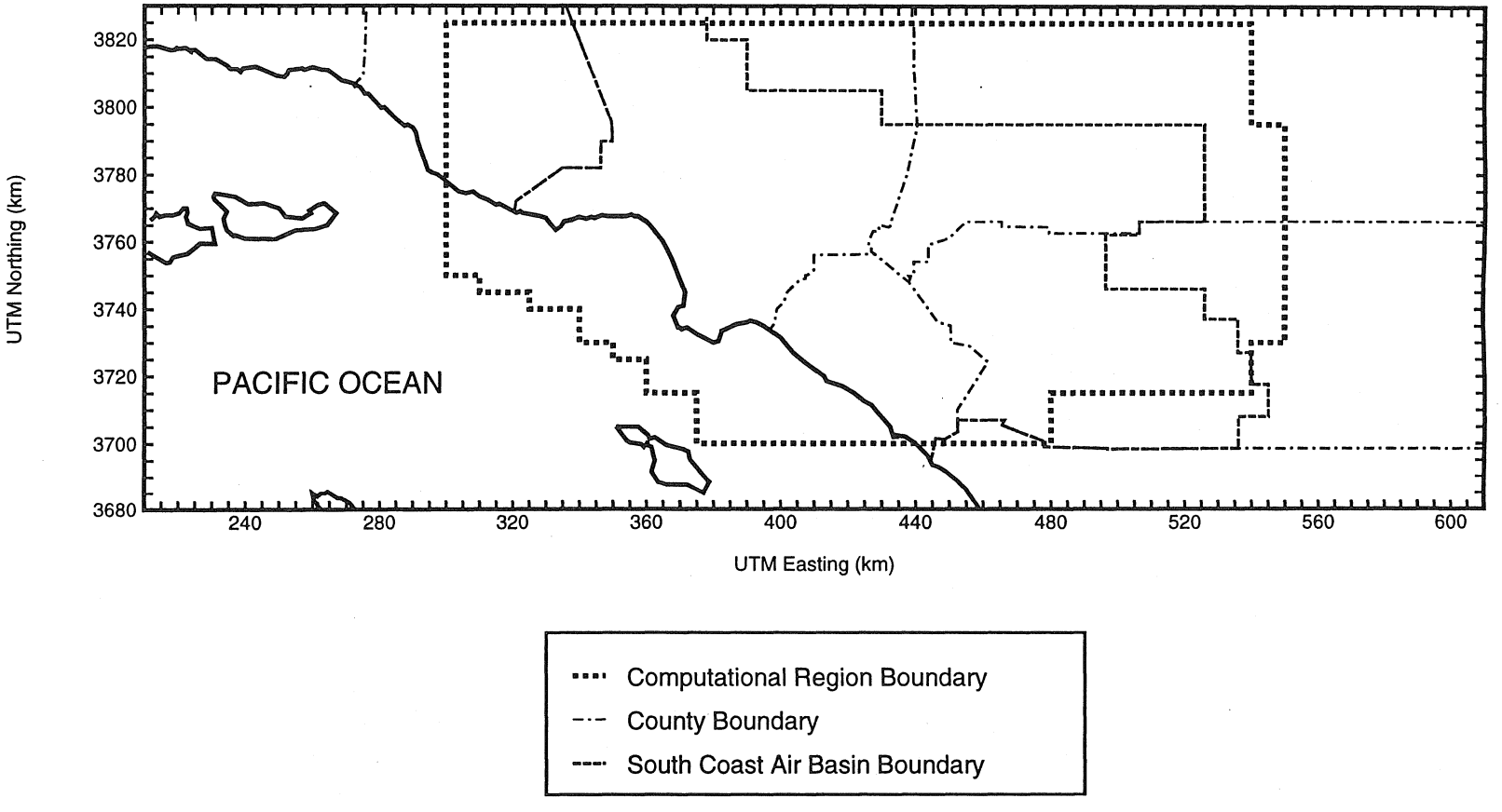

Figure 3.4 Southern California, showing the entire region over which air quality and emissions data were developed; the computational region for the air quality model; the legal boundaries for the South Coast Air Basin, and county boundaries.

contains 35 differential species, 10 steady-state species, and 107 chemical reactions.

3.3.2 Emission inventory

The emissions inventories for 27 and 28 August 1987 are similar but not identical, each having been derived by applying day-specific hourly temperature corrections and day-specific major point source emissions corrections to the official State of California emissions inventory. Researchers have suspected that the official State of California emission inventory understates hydrocarbon and carbon monoxide emissions from mobile sources. Recent field measurements in a highway tunnel (Ingalls *et al.*, 1989; Pierson *et al.*, 1990), remote sensing measurements of car exhaust (Bishop *et al.*, 1989; Bishop and Stedman, 1990; Lawson *et al.*, 1990), comparison of emission inventory and ambient ratios of CO, ROG, and NO_X (Fujita *et al.*, 1992), and an organic chemical tracer study (Harley *et al.*, 1992a) all support the finding that actual hydrocarbon emissions from mobile sources in Los Angeles are 2–3 times higher than in the official State of California emission inventory for 1987. The organic chemical tracer study (Harley *et al.*, 1992a) also concluded that there is a large excess of unburned gasoline in the atmosphere compared to what would be expected from the speciation of the official State of California emission inventory.

In the present study, emission factors for motor vehicles have been employed that reflect the emission rates measured within the Van Nuys highway tunnel during SCAQS. The emission inventory is identified as the 3X hot exhaust inventory described in detail by Harley *et al.* (1993a). The total emissions of ROG on 27 August 1987 over the entire

400 km by 150 km grid of 5 km by 5 km cells that covers the area shown in Figure 3.4 are 2,337 metric t d^{-1}, and the total emissions of NO_X are 1138 metric t d^{-1}. The computational domain for the airshed model simulation is shown in Figure 3.4, and it contains nearly all of the emissions within the larger area mapped.

3.3.3 Boundary conditions

Actual historically measured pollutant concentrations were used to specify the initial conditions and boundary conditions shown in Tables 3.1 and 3.2, respectively. The upwind boundaries of the air quality modeling domain customarily used for photochemical airshed modeling in the Los Angeles area are located about 30–40 km offshore over the Catalina Channel and the Santa Barbara Channel areas of the near-coastal Pacific Ocean, as seen in Figure 3.4. Compelling evidence exists that the air over the ocean to the west of Los Angeles (generally thought of as "upwind") is contaminated with anthropogenic air pollutants from continental sources at concentrations much higher than the cleaner air found over the mid-Pacific Ocean.

Recent work by Benkovitz *et al.* (1994) hints strongly that this zone of contaminated offshore air could be quite large in size. They applied a large-scale transport and chemical transformation model for aerosol sulfate formation over most of the northern hemispheric portions of North America, Europe, north Africa, and western Asia. They show a large cloud of elevated sulfate concentrations extending over 10° longitude to the west of the California, Baja California, and western Mexico coastlines over the

Table 3.1 Initial conditions [ppb] used for the simulation of air quality for 27–28 August 1987 in California's South Coast Air Basin.

	Historical Case					**Clean Case**				
	Vertical layer in model[a]					Vertical layer in model				
Species	Surface	Level 2	Level 3	Level 4	Level 5	Surface	Level 2	Level 3	Level 4	Level 5
NO_2	aq[b]	aq	10	10	10	0.5	0.5	0.5	0.5	0.5
NO	aq	aq	30	30	30	0.3	0.3	0.3	0.3	0.3
O_3	aq	0.5pd[c]	0.8pd	0.8pd	0.5pd	40	40	40	40	40
RHC[d]	aq	aq	aq	100	100	10	10	10	10	10
HCHO	aq	aq	aq	3	3	3	3	3	3	3
ALD2	aq	aq	aq	5	5	5	5	5	5	5
MEK	aq	aq	aq	4	4	4	4	4	4	4
CO	aq	aq	300	300	200	120	120	120	120	120

[a] The model has 5 vertical layers with elevations above ground level of: (1) 0–38 m; (2) 39–154 m; (3) 155–308 m; (4) 309–671 m; (5) 672–1100 m.

[b] An entry of aq means that the value is from an interpolated surface-level air quality field based on actual measurements made on 27–28 August 1987.

[c] An entry of pd means that the value is from an interpolated air quality field based on actual measurements from the previous day's mid-afternoon surface-level values.

[d] RHC expressed in units of ppbC.

Table 3.2 Inflow boundary conditions [ppb] used for the simulation of air quality for 27–28 August 1987 in California's South Coast Air Basin.

		Historical Case					Clean Case				
		Vertical layer in model[a]					Vertical layer in model				
Species	**Boundary**	Surface	Level 2	Level 3	Level 4	Level 5	Surface	Level 2	Level 3	Level 4	Level 5
NO_2	N, S, W	1	1	1	1	1	0.5	0.5	0.5	0.5	0.5
NO_2	East	aq[b]	aq	1	1	1	0.5	0.5	0.5	0.5	0.5
NO	N, S, W	1	1	1	1	1	0.3	0.3	0.3	0.3	0.3
NO	East	aq	aq	1	1	1	0.3	0.3	0.3	0.3	0.3
O_3	North	aq	70	70	70	60	40	40	40	40	40
O_3	East	aq	aq	60	70	70	40	40	40	40	40
O_3	S, W	40	40	40	40	40	40	40	40	40	40
RHC[c]	North	aq	100	100	100	100	10	10	10	10	10
RHC	East	aq	aq	100	100	100	10	10	10	10	10
RHC	S, W	100	100	100	100	100	10	10	10	10	10
HCHO	N, E	aq	aq	3	3	3	3	3	3	3	3
HCHO	S, W	3	3	3	3	3	3	3	3	3	3
ALD2	N, E	aq	aq	5	5	5	5	5	5	5	5
ALD2	S, W	5	5	5	5	5	5	5	5	5	5
MEK	N, E	aq	aq	4	4	4	4	4	4	4	4
MEK	S, W	4	4	4	4	4	4	4	4	4	4
CO	N, E	aq	200	200	200	200	120	120	120	120	120
CO	S, W	200	200	200	200	200	120	120	120	120	120

[a] The model has 5 vertical layers with elevations above ground level of: (1) 0–38 m; (2) 39–154 m; (3) 155–308 m; (4) 309–671 m; (5) 672–1100 m. Mixing depths vary in space and time and were interpolated from 8 sites at which upper air soundings were taken six times per day. The maximum mixing depth measured in the air basin during this episode was less than 1100 m.

[b] An entry of aq means that the value is from an interpolated surface-level air quality field based on actual measurements made on 27–28 August 1987. The north and east boundaries of the modeling region are over land where detailed surface air quality measurements are available on these days that permit estimation of spatial variations along those boundaries.

[c] RHC expressed in units of ppbC.

Pacific Ocean that results from westward transport of continental emissions (see their Plates 3b, 4b, and 5). Presumably hydrocarbons and NO_X emissions from transportation and industrial sources accompany these anthropogenic sulfur oxides emissions.

Air quality for total non-methane hydrocarbons (TNMHC) over the mid-Pacific Ocean near Hawaii has been studied by Moore *et al.* (1991). They report an average TNMHC concentration of 3.4 ppbC at that remote area. Moore *et al.* (1991) also have characterized hydrocarbon air quality over the Santa Barbara Channel portion of the near coastal Pacific Ocean close to where the western boundary of the modeling region of Figure 3.4 of the present chapter approaches the shore. Data from their Table 3.3 show two conditions. Roughly one-third to one-half of the samples analyzed show "clean" Pacific Ocean air with an average TNMHC concentration of 12–16 ppbC extending from the ocean at least up to 1600 m elevation. But on average, the hydrocarbon concentrations along what would be a part of the inflow boundary of our region exist in a thick nearly uniformly mixed layer with TNMHC concentrations in the range 40–59 ppbC from the ocean up to 1000 m elevation, with cleaner air having an average of 15 ppbC TNMHC above 1000 m elevation. This average of all samples contains the "clean air" cases. Therefore, at those times when "clean air" is not observed in the mixed layer, it can be calculated that the average TNMHC concentrations were in the range 94–103 ppbC from 0–400 m elevation and 53 ppbC from 400 m to 1000 m elevation.

Measured pollutant concentrations over the Pacific Ocean along the western boundary of the South Coast Air Basin have been reviewed by Main *et al.* (1990, 1991). For future year conditions, assuming inflow of clean air under prevailing westerly wind

flow, Main *et al.* (1990) recommend inflow boundary conditions of 10 to 30 ppbC of ROG, 30 to 40 ppb of ozone, and 0.5 to 1.0 ppb of total NO_X. Harley *et al.* (1992a) report that the sum of all carbonyls in fresh ROG emissions from catalyst-equipped gasoline engine exhaust represents only 3% by weight of total ROG, with the balance consisting of hydrocarbons. In contrast, Main *et al.* (1990) report that air over the Pacific Ocean contains a much higher fraction of carbonyls – in the range of 12–33% on a carbon atom basis. This suggests that the ROG concentrations entering the modeling region consist of aged emissions, with the excess carbonyls formed as atmospheric oxidation products of other precursor hydrocarbon emissions. Main *et al.* (1990) also note that the remaining hydrocarbons are predominately low-reactivity light alkanes. Also, the available measurements are consistent with either a decreasing ROG concentration profile above ground level, or with a constant inflow ROG concentration at all heights. The data of Moore *et al.* (1991) show a nearly constant TNMHC concentration profile with height up to 1000 m under clean air conditions. These observations and recommendations were used to specify the inflow boundary conditions shown in Table 3.2 for the clean air conditions.

As part of 1987 SCAQS experiment, pollutant concentrations were measured at San Nicolas Island, which is located 140 km offshore (WSW of Long Beach), and above the ocean in aircraft flights at two airway intersections located 10 miles offshore of Long Beach and Palos Verdes respectively. These data have been reviewed by Main *et al.* (1990). The inner quartile range (i.e., central half) of measured non-methane hydrocarbon (NMHC) concentrations was 46–101 ppbC at San Nicolas Island, and 73–176 ppbC in the

offshore aircraft flights. Clearly these measured NMHC concentrations are well above the values of clean marine air cited previously. Main *et al.* (1990) agree that NMHC concentrations measured offshore during SCAQS were elevated above clean marine air levels due to the influence of onshore emissions. As onshore emissions decline, inflow boundary conditions would be expected to decline, eventually approaching the cleaner mid-Pacific Ocean values at zero anthropogenic emissions over the continent. Ozone concentrations measured during SCAQS at San Nicholas Island ranged from 28 to 40 ppb. Measured NO_X concentrations averaged less than 10 ppb, the detection limit of the analyzer. Actual measured offshore pollutant concentrations from the SCAQS experiment were used to specify boundary conditions for the historical conditions shown in Table 3.2, as these data are specific to the actual episode studied here.

3.3.4 Model performance evaluation using SCAQS data

The CIT model has been extensively tested using experimental data from the August 1987 Southern California Air Quality Study (Harley *et al.*, 1993a). When supplied with the best current estimates of pollutant emissions in the Los Angeles area, the model predicts ozone concentrations that display +1% normalized bias, an average station peak prediction accuracy of -7%; while the ozone peak predicted near the location of the maximum observed ozone concentration is -33%, the peak ozone concentration unpaired in space is only 13% below the observed peak. This underprediction of the basin-wide maximum ozone peak may be due in part to the statistical phenomenon of regression towards the mean (Fairly, 1993) because the model predictions will always have less

variance than the observations due to averaging and filtering of model inputs. Also, the model behaves in a deterministic fashion, while the actual atmosphere behaves stochastically. Thus, one would expect to find differences between model predictions and atmospheric measurements (Rao *et al.*, 1985). The predictions for ozone precursors have a normalized bias of +22% for total NO_2 and -12% for reactive hydrocarbons (Harley *et al.*, 1993a).

Additional model verification studies have been completed using the June and December SCAQS episodes (McNair *et al.*, 1992).

3.3.5 Model implementation on a parallel computer

The CIT airshed model was configured to generate ozone isopleth diagrams using the parallel computation scheme described earlier in this chapter. An Intel Gamma Hypercube System was used for this purpose. The Intel Gamma, located at the Caltech Concurrent Supercomputing Facilities (CCSF), contains 64 40MHz i860 computational nodes and 6 i/o nodes connected to 5.2 Gbytes of total disk space. As a first step in demonstrating the possibility of doing massive numbers of airshed model simulations quickly, two isopleth diagrams describing the ozone response to emission controls for 28 August 1987 were generated. Each of these diagrams depicts the effect of emissions controls on ozone air quality in the Los Angeles area in the presence of alternative assumptions about the boundary conditions in future years. The emission controls are applied to both the 27 and 28 August data sets, and multiple day simulations are used in all cases to compute the predicted ozone peaks for 28 August. The problem of ozone

isopleth generation was decomposed into a grid system of 64 different control combination points each representing a different degree of basin-wide reductions in anthropogenic ROG and NO_X emissions (see Figure 3.3). As basin-wide anthropogenic ROG emissions are lowered in our calculations, the biogenic organic vapor emissions present in the emission inventory (Harley *et al.*, 1993a) are not lowered. The entire problem was dispatched as a single job to the Intel Gamma. This method produces a *linear* speed-up in the computations as opposed to the declining speed-up with using more nodes typically encountered in parallelized airshed model runs because the emission control problem has been placed essentially 100% in parallel (p=1.0).

The generation of one isopleth diagram requires 128 days of simulation (64 2-day calculations for the period 27–28 August 1987). The two isopleth diagrams generated require a total of 853 min (14.2 h) on the Intel Gamma to calculate the 256 days of simulation. Importantly, only two separate jobs had to be submitted, one for each isopleth diagram, so that the time spent waiting between hundreds of sequential simulations is eliminated. Each set of 128 simulated days requires 500 Mbytes of disk storage space for the standard output (ground level concentrations for all 35 species tracked by the model for all hours, last hour concentrations for all species in all grid cells for restart, and diagnostics files).

The magnitude of this calculation can be compared to prior studies. The CIT airshed model takes approximately 6 h to complete one simulated day on a MicroVax 4000/60 computer and 5 h to compute one day of simulation on a DECstation 5000 computer. The "longest" calculations that have been reported previously on such models

include 200 days of simulation (120 h of CRAY X-MP/48 time; McRae *et al.*, 1988) needed to define air basin-wide contours of ozone concentrations as a function of alternative levels of ROG and NO_X emissions for a single day, 28 August 1982 (Milford *et al.*, 1989), and 30 days of simulation (one week SUN-4 CPU time) in time series for June, 1985 (Tesche and McNally, 1991). The speed of the present calculations also can be compared to the speed of a parallelized version of the CIT photochemical airshed model. Dabdub and Seinfeld (1994) report that such a version of the present model requires 2000 s (33.3 min) of elapsed time when using 64 nodes on an Intel parallel computer with the same node speed to complete 1 day of simulation for the same episode in Los Angeles. At that speed, 142 h of elapsed time would be required to perform the 256 days of simulation reported in the present chapter. The calculations reported in the present study are both faster and larger than in those prior studies.

3.3.6 Effect of alternative boundary and initial conditions

The effect of boundary and initial conditions on control program performance is examined using two different sets of simulations in this study. In both cases, emission levels of NO_X and ROG are reduced to obtain the 64 different combinations shown in Figure 3.3. These percentage reductions are applied "across-the-board" to all anthropogenic sources in the entire modeling region. The relative spatial distribution of anthropogenic emissions is not changed. First, historically measured pollutant concentrations are used to set boundary and initial conditions at all levels of ROG and NO_X control examined. This simulation produces the highest predicted ozone

concentrations. Second, boundary and initial conditions are set to very low concentrations based on clean air values measured over the Pacific Ocean far upwind of the air basin (Moore *et al.*, 1991; Main *et al.*, 1990, 1991). These are similar to the remote background values cited in the 1991 Air Quality Management Plan adopted by the South Coast Air Quality Management District (South Coast Air Quality Management District, 1990).

For both cases, the model is executed for 27 and 28 August 1987. Table 3.1 shows the initial condition values used for the historical and clean air simulations. The historical case initial conditions amount to 1788 metric t RHC, 589 t NO_2 and 973 t NO, which are roughly equal to one day's emissions within the modeling region. The clean air case initial conditions are lower; 149 metric t RHC, 28 t NO_2, and 11 t NO. These values are used only at the start of the first day of simulation (27 August 1987). At the end of the first day of simulation, the concentrations of all species are saved and these values are used to start the second day of simulation (28 August 1987). Only results from the second day, 28 August are used for isopleth construction. The first day is used to minimize the effect of initial conditions on model predictions for August 28. Table 3.2 shows the boundary condition values for the historical and clean air cases. Note that some values used in setting the historically observed boundary and initial conditions are taken from an interpolated air quality field. This means that air quality measurements made near the boundaries of the entire region shown in Figure 3.4 were used along with air quality measurements made within the South Coast Air Basin to create pollutant concentration fields over the entire region via spatial interpolation. The modeling domain shown in Figure 3.4 was then cut from those larger concentration fields, with the concentrations

falling at the edges of the modeling domain used as the historically observed boundary conditions. The distribution of the inflow boundary concentrations and initial condition concentrations used during the historical case model simulation is shown in Figure 3.5.

In Figure 3.6a, the ozone control isopleth diagram is constructed based on the 1 h peak ozone concentration that occurs anywhere within the modeling region of Figure 3.4 for the case where the boundary and initial conditions are based on the clean Pacific Ocean values given in Tables 3.1 and 3.2. In that case, pollutant inflows across the boundaries of the modeling region are 688 metric t d^{-1} ROG and 35 metric t d^{-1} NO_X, compared to emissions within the area mapped in Figure 3.4 of 2025 metric t d^{-1} ROG and 1128 metric t d^{-1} NO_X on 28 August 1987. Substantial amounts of reactive organics flow across the boundaries of the modeling region even at the low ROG concentrations present over the mid-Pacific Ocean, in part because the modeling domain is so large. Nevertheless, these organics inflows are at low concentrations compared to the concentrations observed near emissions sources within the urban area. Note that the ozone isopleth diagram for the clean boundary and initial condition case of Figure 3.6a exhibits ozone inhibition at high NO_X emission rates (i.e., in the upper left half of that diagram, increasing NO_X emissions leads to reduced ozone concentrations). Ozone concentrations are lower than the U.S. Federal 1 h average ozone air quality standard of 120 ppb over a variety of combinations of ROG and NO_X control.

Figure 3.6b shows the outcome of the same control calculations when historically observed boundary and initial conditions given in Tables 3.1 and 3.2 remain unaltered as pollutant emissions are changed within the modeling domain. In that case, inflows across

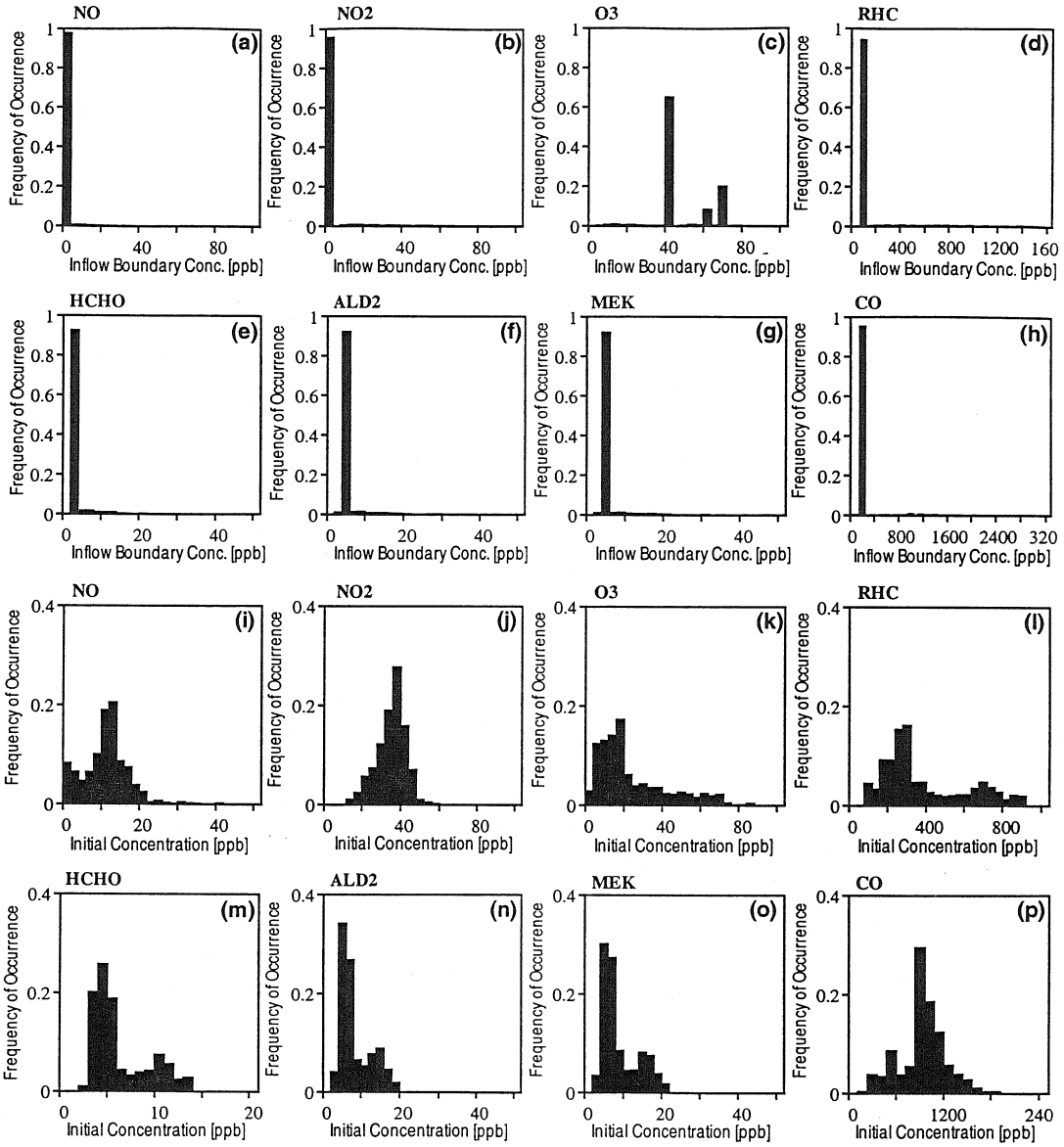

Figure 3.5 The frequency of occurrence of initial concentrations and inflow boundary concentrations at the edges of the modeling region. (a)-(h) fraction of the grid cells having inflow boundary values in the ranges shown for the historical case. (i)-(p) fraction of ground-level grid cells having initial condition values in the ranges shown at 00:00 hours August 27, 1987, for the historical case simulation.

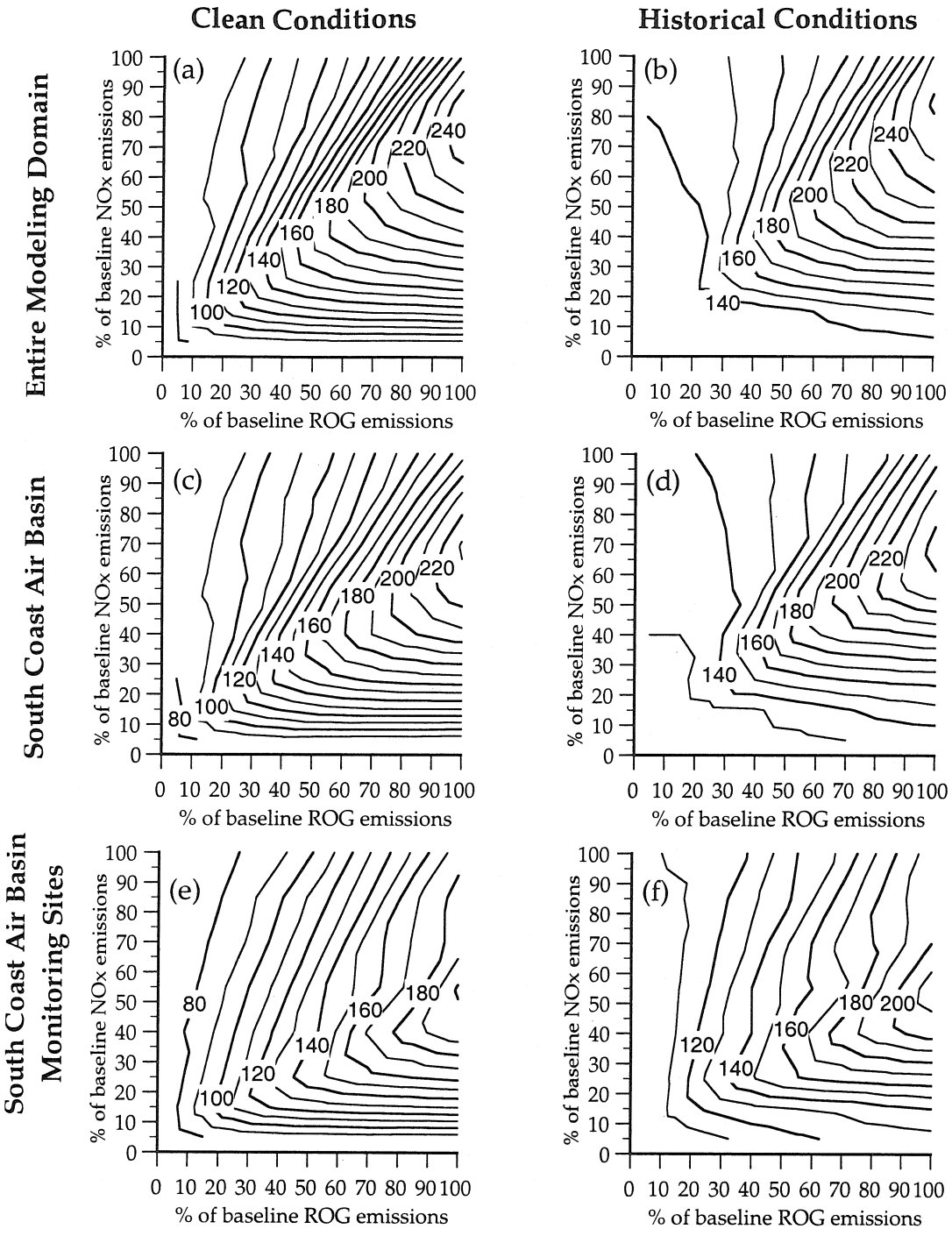

Figure 3.6 Isopleth diagrams showing the response of the predicted peak 1 h average ozone concentration on 28 August 1987, to alternative levels of basin-wide anthropogenic ROG and NOx emission reductions in the Los Angeles area. Fig. 6a and b is based on peak concentrations within the entire modeling domain for (a) clean, (b) historical boundary and initial conditions; Fig. 6c and d is based on peak concentrations within the SoCAB for (c) clean, (d) historical boundary and initial conditions; Fig. 6e and f are based on values predicted at SoCAB monitoring sites for (e) clean, (f) historical boundary and initial conditions.

the boundaries of the modeling domain are 1066 metric t d^{-1} of low reactivity alkanes, ethene, and MEK; 690 metric t d^{-1} of more reactive organic vapors, and 96 metric t d^{-1} of NO_X stated at the molecular weight of NO_2. These organic vapor inflows are due to flows of a widespread dilute mixture of relatively low reactivity organics (by comparison to fresh urban emissions) having the character of significantly aged air pollution (Moore *et al.*, 1991). Much of these inflows occur at elevations high above ground level and do not interact directly with the mixed layer adjacent to the ground at most times. The quantities of ROG involved are 1068 t d^{-1} greater than the 688 t d^{-1} clean mid-Pacific Ocean air inflow case and could be due in part to recirculation of aged emissions from the Los Angeles area. An increased inflow of approximately 1000 t d^{-1} is equivalent to less than one-half of a day's organics emissions within the Los Angeles area. Also, emissions in adjacent air basins influence the boundary conditions. The South Coast Air Basin that contains Los Angeles accounts for only about half of the ROG emissions in the greater Southern California area (California Air Resources Board, 1990). While the ROG inflows seem large, the important point is that detailed field measurements discussed earlier show that the pollutant concentrations leading to these inflows are present historically for whatever reason, and detailed modeling of the emissions and air quality relationships for organics in this air basin employing the historical emissions and boundary conditions is in good agreement with the detailed composition of the organics measured within the air basin on 28 August 1987 (Harley and Cass, 1994, 1995; Harley *et al.*, 1993b).

Note from Figure 3.6b that the ozone isopleths in the case of historical initial and boundary conditions change shape relative to Figure 3.6a. The peak 1 h average ozone

concentrations within the modeling domain are not inhibited by increasing NO_X emissions if hydrocarbon emissions already have been severely restricted, and the U.S. Federal ozone standard of 120 ppb is never attained. Thus, changed assumptions about the status of boundary conditions in future years results in major changes in the predicted outcome of control programs designed to approach compliance with air quality standards.

When ozone control isopleth diagrams are constructed based on the highest 1 h peak ozone concentration observed anywhere in the modeling domain of Figure 3.4, then predicted peak ozone levels often occur in the desert far downwind (north or east) of the urban area (all of the peak values are still within the computational modeling domain; none occur on the edges). This is particularly true for those high NO_X / ROG ratios that cause the production of the ozone peak to be delayed. If high ozone concentrations that occur outside the air basin are disregarded, then the ozone control isopleth diagrams can be redrawn by considering only the highest 1 h average ozone concentrations that occur within the South Coast Air Basin boundaries shown in Figure 3.4. Those results are shown in Figure 3.6c and d for the clean and historical boundary condition cases, respectively. By definition, the peak ozone concentrations within the air basin must be less than or equal to those within the larger modeling domain. Therefore, Figures 3.6c and d for the air basin response to control shows lower ozone concentrations at the same emissions levels than is the case for the larger modeling domain diagrams of Figures 3.6a and b. Qualitatively, however, the results are similar to the previous cases. When the clean boundary conditions are used in the model, NO_X inhibition results and the U.S. Federal ozone standard could be attained within the air basin for the meteorological

conditions studied here with no further NO_X control and only 43% ROG control. In the presence of historically observed boundary and initial conditions, compliance with the U.S. Federal ozone standard is virtually impossible.

A third interesting case can be examined in Figure 3.6e and f in which only the peak ozone concentrations that would be observed at the locations of air quality management district air monitoring sites within the South Coast Air Basin are used to specify the values in the ozone control isopleth diagrams. The 1 h peak ozone concentrations likely to be measured at those air monitoring sites existing in the air basin in 1987 are noticeably lower than at the predicted basin-wide peak. This is expected because the basin-wide peak probably will not occur exactly at an air monitoring site. A number of combinations of ROG and NO_X emissions exist that would produce *measured* compliance with the U.S. Federal ozone standard in this case.

3.4 Conclusions

When constructing airshed model predictions of the effect of emission control programs on future air quality, future pollutant fluxes across the boundaries of the air quality modeling domain must be supplied by assumption. In the present study, ozone control isopleth diagrams based on photochemical airshed model predictions of the peak 1 h average ozone concentration generated within the air basin or modeling domain over nearly all possible combinations of basin-wide ROG and NO_X control have been used to study in detail the effects that alternative assumptions regarding future boundary and initial conditions can have on predicted peak ozone levels. Such control isopleth diagrams

based on basin-wide percentage emissions reductions can be used to locate the vicinity of the emissions levels for the air basin at which compliance with ozone air quality standards will occur, and such calculations can be refined by modeling the effect of specific combinations of pollutant abatement technologies that achieve the desired level of control.

It was found for the case of the Los Angeles area that when very clean future boundary and initial conditions are assumed, predicted peak ozone concentrations exhibit inhibition by NO_X emissions at high NO_X / ROG levels. In that case, attainment of the Federal ozone standard of 120 ppb under the meteorological conditions studied here would be comparatively easy. However, if boundary conditions are set based on historically observed pollutant concentrations near the edges of the Los Angeles modeling region, then NO_X inhibition is greatly reduced. In this case, it may be virtually impossible to predict attainment of the U.S. Federal ozone standard in the Los Angeles area when starting from the hydrocarbon emission levels in the 3X hot motor vehicle exhaust emission inventory that is inferred from the 1987 Van Nuys tunnel experiments.

Clearly as anthropogenic emissions within an airshed such as the Los Angeles basin begin to approach zero, assumptions made about the future magnitude of the pollutant fluxes across the boundaries of the airshed become more important. Tremendous energy has been expended to define pollutant emission control opportunities within the Los Angeles area, while comparatively little is known about likely future pollutant inflows. Clearly, greater insight into the latter issue should be sought. A program of experiments that better characterize air quality far offshore could be combined

with the use of multi-scale nested grid models in order to extend the boundaries of the Southern California modeling domain so far offshore that the effect of onshore emissions on boundary conditions over the ocean would no longer raise important questions. Moore *et al.* (1991) show that a reservoir of air that approximates our clean air case boundary condition and that does not vary from day to day exists over the mid-Pacific Ocean. This far offshore background condition could be relied upon as a stable boundary condition for use in future emission control strategy calculations that would not have to be tied to the circumstances of a particular onshore pollution episode. Moore *et al.* (1991) show that air of nearly that quality appears occasionally in the near-coastal Santa Barbara, CA, area. The question remains, what is the minimum distance offshore needed to reliably place the airshed model boundary within this clean air during a particular season of the year. The answer to this question needs to be determined experimentally by mapping offshore pollutant concentrations over a suitably long period of time, probably by aircraft flights.

Emissions control questions like the boundary value problem studied in the present chapter can be attacked rapidly if photochemical airshed models can be used to quickly generate basin-wide ozone control isopleth diagrams. The procedure developed here involves decomposition of the isopleth generation process into a matrix of control combinations that are then dispatched simultaneously to the nodes of a parallel computer. This method is easy to implement and produces a linear speed-up in the computations as opposed to the declining speed-up with more nodes typically encountered when airshed models are adapted to run on highly parallel computer systems. The equivalent of 256 d of airshed model simulation were executed on a parallel computer at the equivalent of 3

min 20 s elapsed time per day of simulation. The parallel computer used here, the Intel Gamma Hypercube, is only a 64-node machine; machines with greater than 500 nodes are presently available. It is now possible through control isopleth diagram generation on parallel computer systems to quickly attack air quality control problems that in the past would have seemed to be intractable.

Chapter 4 Modeling the control of the long-term frequency distribution of regional ozone concentrations

4.1 Introduction

Despite considerable efforts over the past 30 years, numerous areas of the United States have not attained compliance with Federal air quality standards for ozone. As of 09 October 1997, 59 of the 151 air quality control regions (AQCR's) in the U.S. still experienced ozone concentrations that exceed the former 1 h Federal standard of 0.12 ppm O_3 not to be exceeded for more than one hour per year averaged over 3 years (Title 40, Code of Federal Regulations, part 81). The population living in these areas exceeds 102 million people. It is estimated that an even larger number of AQCR's are out of compliance with the new U.S. Federal air quality standard set in 1997 at 0.08 ppm O_3 over an 8 h averaging time not to be exceeded more than 3 days per year averaged over 3 years.

Each of the air quality control regions that exceeds the former 1 h standard has had in place since the 1970's an approved plan that was accepted by the government as adequate to attain compliance with that ozone standard long ago. In pursuit of these plans, annualized compliance costs of approximately $26.1 billion per year are being incurred as of 1990 (EPA, 1997). Clearly careful study of available control strategies is necessary not only to determine how to meet air quality standards, but also how to do so while spending the least amount of money. Economic studies have estimated that careful selection of the least costly set of control equipment for a region as a whole can lead to a particular level

of air quality improvement at less than half the cost of the common practice of applying the best available control equipment to the largest sources first (Cass and McRae, 1981).

The cost associated with an air pollution control strategy can be viewed in light of cost-benefit analysis. Using a purely economic approach, an air pollution control agency would seek to minimize the combined total of the costs associated with achieving the air quality standards plus the cost of the damages caused at that level of air quality. The control costs include direct costs such as the costs of capital and maintenance as well as indirect costs such as job losses and enforcement costs. The damage costs arise from adverse health effects, crop losses, material damage, and impaired visibility. The total damage cost at a given level of air quality is extremely difficult to estimate especially in a large, diverse urban area such as Los Angeles. Thus, cost-benefit analysis has not actually been used in many cases. Instead, the economic issue typically is reduced to the question of how to achieve the mandated air quality objective with the lowest total expenditure on emission controls.

One key prerequisite to finding the lowest total cost of emission controls to meet a given air quality objective is that the relationship between emission changes and air quality changes must be understood. For an air quality standard like the new Federal 8 h average ozone standard that is stated in probabilistic terms (not more than 3 days above the standard per year on average over 3 years), understanding how one or a few days per year change in response to emission controls is inadequate. For a secondary pollutant like ozone that is formed in the atmosphere by non-linear chemical reactions which depend as much on pollutant precursor ratios as on the amount emitted, it is not at all obvious that

the days with the highest observed concentrations historically are the hardest ones to bring below the air quality standard. Yet the almost universal practice has been to model the effect of emission controls on atmospheric ozone concentrations for one or a few historically observed high ozone events when evaluating proposed emission control programs.

Trijonis (1974) formulated a method for identifying the least costly approach to control of photochemical smog in Los Angeles. The control cost–emission level relationship was determined using a simple, graphical non-linear optimization technique. Contour lines were constructed that show the trade-off between reactive organic gases (ROG) emissions and oxides of nitrogen emissions at fixed levels of total control cost (i.e., isocost lines in 2-D ROG and NO_x space). The emission level–air quality relationship used by Trijonis to demonstrate this technique in 1974 was founded on a highly empirical model that determined the number of days per year that the standards for ozone and nitrogen dioxide are exceeded in Los Angeles County at any given level of air basin-wide ROG and NO_X emissions based on relationships between ROG and NO_x concentrations observed at central Los Angeles in the morning and observed ozone concentrations downwind later in the day. The effect of NO_x controls on NO_2 concentrations daily was estimated by a simple linear rollback model applied to each day's historically observed NO_2 concentrations. This set of observed empirical relationships between historical ROG and NO_X emissions and the resulting O_3 and NO_2 air quality was organized in the form of an ozone isopleth diagram that was constructed to show the number of days per year that the ozone standard will be exceeded at all possible

combinations of air basin-wide ROG and NO_x emissions. The diagram was accompanied by a scale showing the number of NO_2 standard violations per year at any level of NO_x control. All ROG and NO_X emission levels that would provide a feasible solution to the O_3 and NO_2 control problems can be determined by inspection from this diagram. The cost of control curves can be overlaid on the same diagram. The least-cost solution is the point where the minimum cost curve first touches the feasible air quality region. Trijonis, incidentally, concluded that the Los Angeles area could not be brought into literal compliance with the ozone standard, but that very substantial improvements in air quality could be made. Reduction of ozone air quality standard violations to a frequency of 20 per year seemed possible.

The mathematical programming formulation of Trijonis (1974) provides a very interesting method for attacking the cost optimization of emission control strategies for photochemical smog abatement. It has never been pursued using chemically explicit mechanistic Eulerian airshed models for ozone formation because it would require a model for the response of the frequency distribution of all daily ozone peaks for the year to all possible combinations of emission controls.

Recently methods have been demonstrated for modeling the long-term frequency distribution of daily ozone concentrations using a chemically-explicit Eulerian photochemical airshed model (Winner and Cass, 1998; this thesis chapter 2). The procedure involves automated generation of the day-specific emissions inputs and meteorological fields needed to drive the CIT photochemical airshed model (Harley *et al.*, 1993ab) for a period of a year or longer based on data obtained routinely in Southern

California. The model can be executed in time series for the year on an engineering workstation. Model outputs are summarized in terms of the frequency distribution of 1 h and 8 h daily peak ozone concentrations (and NO_2 if desired). Time series plots showing predicted and observed ozone concentrations for the year at all monitoring sites can be generated. Model performance statistics for the entire year are comparable to results typically obtained when using laboriously hand-crafted model input data fields for a 3 day smog episode obtained from expensive special field studies.

In the present work, we show that this modeling procedure can be employed to examine how the frequency distribution of days above both the former 1 h average and current 8 h average U.S. Federal ambient air quality standards for O_3 respond to changes in ROG and NO_x emissions. These air quality model results are cast into the format of Trijonis' emission level - air quality relationship, showing ozone isopleth diagrams in ROG and NO_x emissions space where contours of constant air quality are labeled according to the number of days per year with ozone above a specified goal. Methods are illustrated by application to the greater metropolitan Los Angeles area, and the implications resulting from this analysis for ozone control in Southern California are discussed.

4.2 Background

4.2.1 Mathematical programming formulation

The problem of minimizing the cost of air pollution control can be formulated as a mathematical programming problem (Trijonis, 1974):

Select $\vec{x}$ that minimizes $C[\vec{x}]$ subject to

$$Q[E(\vec{r},t,\vec{x}),M(\vec{r},t),P] \leq \vec{s} \tag{4.1}$$

The emissions for n primary pollutants, E_i, i = 1, n, are functions of space, $\vec{r}$, time, t, and control measures, $\vec{x}$. The meteorology, M, is also a function of space and time. The air quality for k primary and secondary pollutants, Q_m, m = 1, k, is a non-linear function of the emissions, meteorology, and chemical reaction parameters, P. Vector $\vec{s}$ represents the limits on allowable air quality for each of the m pollutants. Since the number of points in space and time typically involved in air quality models is quite large, a single number is frequently used to characterize the air quality. Some examples include the yearly second highest 1 h average ozone concentration, the yearly mean, and the number of days or hours per year that ozone concentrations exceed the Federal ozone standard at a specific location.

The air quality parameters are stochastic variables because the meteorology contains random fluctuations. Thus, Q is in fact a probability distribution calculated from E, P, and the probability distribution associated with M. The main objective in the present work is to extract Q from a very large number of airshed model runs, and to display Q graphically at various levels of E and $\vec{s}$ taking M and P as given. Once the region where the constraint that the air quality meets various possible goals, $\vec{s}$, has been determined, the door in principle is opened to the opportunity to conduct further study of how to minimize the total cost, C, of the control measures, $\vec{x}_j$, j = 1, p by a method like the one used by Trijonis (1974).

A two-step process leading to control strategy selection is envisioned. First, ozone isopleth diagrams based on the many possible combinations of basin-wide percentage reductions in ROG and NO_X emissions from all sources are generated in order to locate the vicinity of the ROG and NO_X emissions levels at which compliance with the ozone air quality standard or goal will occur. Next, model simulations that employ control measures applied to specific sources that will achieve the desired level of control can be used to refine control strategy performance. In the present chapter, we concentrate on the first step in this process.

4.2.2 Isopleth diagrams

The effect of emission controls on ozone formation can be represented conceptually using ozone isopleth (lines of constant concentration) diagrams (Haagen-Smit and Fox, 1954; Trijonis, 1974; Milford *et al.*, 1989). The horizontal axis represents the air basin-wide percentage control of the anthropogenic emissions of reactive organic gases (ROG), and the vertical axis represents the air basin-wide percentage control of the emissions of nitrogen oxides (NO_X). Lines of constant ozone concentrations are drawn at all possible levels of NO_x and ROG emissions. The area to the lower right of a line along a diagonal ridge running from lower left to the upper right of such diagrams has relatively high ROG / NO_X ratios. Ozone formation in this area is limited by the availability of NO_X since there is an ample supply of hydroperoxyl (HO_2) and organic peroxy (RO_2) radicals to convert nitric oxide (NO) to nitrogen dioxide (NO_2); the ratio of NO_2 to NO and solar intensity determine the resulting ozone concentrations. Within this domain lowering NO_X

emissions will result in lower peak ozone concentrations. To the upper left of the ridge line, the ROG / NO_X ratio is lower than along the ridge. Ozone formation in this area is ROG-limited since radicals are scavenged by the high amounts of NO_X present. When NO_X concentrations are lowered, more of the radical pool can react with ROG thereby promoting additional radical formation resulting in higher ozone concentrations. Within this upper left region of an ozone isopleth diagram, lowering ROG emissions alone or possibly even increasing NO_X emissions can result in a decrease of ozone concentrations. Historically, ozone control efforts in Southern California have emphasized reducing ROG emissions since the technology was initially less expensive and more readily available.

A variety of different air quality models have been used to construct ozone isopleth diagrams in the past. Haagen-Smit and Fox (1954) used physical smog chamber experiments to map the effect of altered inputs of ROG and NO_X on ozone formation. Trijonis (1974) used an empirical air quality model specific to the Los Angeles area to develop ozone isopleth diagrams that depicted the ozone response to altered ROG and NO_X emissions. The effect of changes in emissions was judged from ambient air quality data by examining afternoon ozone levels downwind of the central city on days with different early morning ROG and NO_X concentrations in the city center. The EKMA box models promoted by the U.S. Environmental Protection Agency (Dodge, 1977) have been used to estimate the response of ozone to control efforts in many locations throughout the United States.

Since photochemical airshed models are designed to simulate the actual physical and chemical processes occurring in the atmosphere, the use of an airshed model to

generate emission control predictions should provide the most accurate results. One of the first studies to use an airshed model to examine the effect of emission control strategies on ozone levels was conducted by Reynolds and Seinfeld (1975) to forecast the effect of the 1977 EPA Transportation Control Plan for Los Angeles. Two forecast emission inventories for 1977 were produced – one with the controls in effect and one without. These two emission cases were then used in an airshed model simulation based on historical meteorological data from 29 September 1969. In another study (Roth *et al.*, 1983; Tesche *et al.*, 1984), the emissions control requirements for the South Coast Air Basin (SoCAB) that surrounds Los Angeles necessary to attain the Federal 1 h average ozone air quality standard by 1987 were examined. Airshed model simulations based on historical meteorological data from 26–27 June 1974 and 7–8 November 1978 were used in conjunction with several alternative emission inventory forecasts and control strategies for 1987. A study of the effect of emission controls on ozone and nitrogen-containing air pollutants in the Los Angeles area conducted by Russell *et al.* (1988ab) used the 1982 emission inventory for the South Coast Air Basin. A matrix of specific ROG and NO_X control measures was applied to the emission inventory in order to study future emission control possibilities. The resulting effects on air quality were simulated using historical meteorological data from 30–31 August 1982. The above studies, and others like them, produce predictions of the effect of emission controls on one or a few days, do not cast their results into the form of an ozone isopleth diagram, and do not examine all possible levels of control (although the study by Russell *et al.* (1988b) does map 10 feasible combinations of ROG and NO_x control).

The first study to generated basin-wide ozone isopleth diagrams for the Los Angles area using a photochemical airshed model was conducted by Milford *et al.* (1989). They examined conditions over the 30–31 August 1982 historical episode studied by Russell *et al.* (1988), calculated peak ozone concentrations at 46 different levels of basin-wide ROG and NO_x control, and then contoured the results to form an ozone isopleth diagram for 31 August 1982.

Winner *et al.* (1995; this thesis chapter 3) used a computationally efficient approach to generate ozone isopleth diagrams on a massively parallel computer. The CIT photochemical airshed model (Harley *et al.*, 1993ab) was used to compute ozone concentrations in the Southern California area for the conditions of the 27–28 August SCAQS episode at 64 combinations of ROG and NO_x control and with varying assumptions about boundary conditions. This study demonstrated that either the boundaries of the modeling region need to be expanded into areas with cleaner air or that alternatively an accurate forecast of the effect of future emission control programs on pollutant inflows across the boundaries of the Los Angeles modeling region will be absolutely critical to the selection of a successful ozone control strategy.

4.3 Approach

4.3.1 Model description

In the present work the approaches of Winner *et al.* (1995; this thesis chapter 3) and Winner and Cass (1998; this thesis chapter 2) are combined. The model for the frequency distribution of long-term average ozone concentrations is used to generate an

isopleth diagram showing the frequency of exceedence of ozone standards as a function of the basin-wide level of ROG and NO_x control. The technical vehicle for performing these calculations is the CIT photochemical airshed model (Harley *et al.*, 1993ab). This Eulerian photochemical model numerically solves the atmospheric diffusion equation (Lamb and Seinfeld, 1973; McRae *et al.*, 1982)

$$\frac{\partial c_i}{\partial t} + \nabla \bullet (\vec{u} c_i) = \nabla \bullet (K \nabla c_i) + R_i + S_i \tag{4.2}$$

where c_i is the ensemble mean concentration of species i, $\vec{u}$ is the mean wind velocity at location $\vec{x}$ at time t, K is the eddy diffusivity tensor (assumed to be diagonal), R_i is the chemical reaction rate of species i which depends on the pollutant species $c_1 \ldots c_n$ and on the temperature T, and S_i is the elevated source emission rate of species i at location $\vec{x}$ at time t.

A no-flux boundary condition is applied at the top of the modeling region. The boundary condition at the earth's surface requires that the upward flux of each chemical species equals the ground-level emissions minus the dry deposition flux:

$$-K_{zz} \frac{\partial c_i}{\partial z} = E_i - v_g^i c_i \tag{4.3}$$

where K_{zz} is the vertical eddy diffusivity, here E_i is the ground-level emission flux of species i, and v_g^i is the deposition velocity for species i.

The version of the CIT airshed model used here employs a revised dry deposition module based on surface resistance values (Russell *et al.*, 1993) and an extended version of the LCC chemical mechanism (Lurmann *et al.*, 1987). The LCC chemical mechanism represents organic gas emissions through 8 lumped organic species classes: C4+ alkanes,

ethene, C3+ alkenes, monoalkyl benzenes, di- and trialkyl benzenes, formaldehyde, C2+ aldehydes, and ketones. The LCC mechanism has been extended by Harley *et al.* (1993ab) to explicitly include the chemistry of methane, methanol, ethanol, methyl tert-butyl ether (MTBE), isoprene, hydrogen peroxide, and sulfur dioxide. This extended mechanism contains 35 differential species, 10 steady-state species, and 107 chemical reactions.

4.3.2 CIT airshed model performance evaluation using SCAQS data

Extensive testing of the CIT model has been performed using experimental data from the August 1987 episode of the Southern California Air Quality Study (SCAQS; Harley *et al.*, 1993ab). Based on the best estimates of pollutant emissions in the Los Angeles area for the August SCAQS episode, the model predicts O_3 concentrations that display +1% normalized bias and an average station peak prediction accuracy of -7%. Although the O_3 peak predicted at the time and location of the maximum observed O_3 concentration on 28 August 1987 is 33% below the observed value, the peak O_3 concentration unpaired in space is only 13% below the observed peak. The predictions for O_3 precursors show a normalized bias of +22% for total NO_2 and -12% for reactive hydrocarbons during the August SCAQS episode.

The 24–25 June 1987 SCAQS episode also has been simulated using the CIT model (McNair *et al.*, 1996). With the exception of an improved prediction accuracy for the basin-wide maximum (-3%), the statistics for model performance for O_3 are similar to the August episode. The predictions for total NO_2 show a normalized bias of +30%.

A baseline model evaluation study for the entire calendar year 1987 (Winner and Cass, 1998) demonstrated that it is possible to successfully model ozone concentrations using routinely measured air quality and meteorological data in the SoCAB. When supplied with the best estimates of pollutant emissions in the Southern California modeling region shown in Figure 4.1 with motor vehicle emissions scaled upward to match the 1987 Van Nuys Tunnel study (Ingalls *et al.*, 1989; Harley *et al.,* 1993a) and calculated according to daily temperature variations, the model predicts 1 h average O_3 concentrations for the entire year 1987 that display +15% normalized bias and an average station peak prediction accuracy of +14%. The region-wide peak O_3 concentration for each day is on average 7% above the observed value over the entire year. The predictions for O_3 precursors show a normalized bias of +23% for total NO_2; daily ROG data are unavailable but when such data are available for a few days during the SCAQS episodes that year, model performance for lumped (and even individual) ROG is favorable (Harley *et al.*, 1993b, Harley and Cass, 1995).

4.3.3 Emission inventory

The baseline 1987 Southern California emissions inventory for the present study is the same as for the baseline model evaluation study of Winner and Cass (1998). The emissions inventory is based on the 1987 emission inventory developed for the SoCAB and adjacent areas (Los Angeles, Orange, Riverside, San Bernardino, and Ventura counties) by the California Air Resources Board (CARB) plus separate emissions inventories for San Diego and Santa Barbara Counties that have been adjusted to 1987

conditions. Measurements made in the Van Nuys highway tunnel during the 1987 SCAQS experiments indicate that the official CARB emission inventory for 1987 understates the hot stabilized exhaust hydrocarbon and CO emissions from light-duty motor vehicles by about a factor of two to three (Ingalls *et al.*, 1989; Harley *et al.*, 1993ab, Harley and Cass, 1995). In the present study, emission data for motor vehicles are employed that reflect both the effect of hourly changes in ambient temperature on emissions and are scaled upward to match the emission rates measured during the Van Nuys tunnel study. Motor vehicle emissions for the SoCAB as a function of time and location throughout 1987 are computed using the Caltrans model DTIM2 (California Department of Transportation, 1994) in conjunction with motor vehicle emission factors from the CARB motor vehicle emission factor program (EMFAC7F) and the hourly, day-specific surface temperature fields. Then the hot stabilized exhaust portion of the emissions are scaled upward by a factor of 2.25, which is the ratio of the tunnel study to EMFAC7F predictions for hot stabilized emissions from the Los Angeles vehicle fleet. Ideally, the motor vehicle emissions for weekend days would be based on weekend-specific traffic patterns. Unfortunately, no weekend traffic pattern information exists for Southern California, and thus the presently available weekday traffic patterns are used for all days of the week. Model performance evaluation statistics show that model bias is slightly smaller on weekend days than on the other days of the week, and hence the motor vehicle emissions data used to represent the weekends are judged to be adequate.

In order to minimize the effect of upwind boundary conditions on O_3 concentration predictions in the Los Angeles area, the modeling region used for the

present study is greatly extended when compared to previous studies, as shown in Figure 1. San Diego and Santa Barbara counties are now appended to the modeling region along with large areas of the Pacific Ocean. Gridded emission inventories for these additional areas are necessary. While 1987 county-wide emissions totals are available, detailed gridded emissions inventories have not been prepared previously for these outlying counties for 1987. For Santa Barbara County, the necessary gridded emissions data are prepared by scaling the spatially and temporally resolved 1984 emission inventory prepared for the SCCCAMP study (Tesche and McNally, 1991a) to match the county-wide emissions totals for 1987 on a major source category by source category basis (California Air Resources Board, 1990a) and subsequently adjusted to account for the increased motor vehicle exhaust emissions described previously. For San Diego County, the spatially and temporally resolved emission inventory for 1989 (Jackson, 1995) is scaled on a major source category by source category basis to match the county-wide emissions totals for 1987 (California Air Resources Board, 1990a). Again, the hot exhaust portion of the CARB motor vehicle emissions estimates are scaled upward according to the results of the Van Nuys tunnel study. At the conclusion of this process, the emissions over the entire 450 km by 225 km grid of 5 km by 5 km cells totals 2924 metric t d^{-1} non-methane organic gases (NMOG) compared to 2304 metric t d^{-1} NMOG for the smaller modeling region used previously by Harley *et al.* (1993a; the above comparison is for the temperatures of August 28, 1987). Speciation of the NMOG emissions within the model follows the revised source composition profiles previously compiled by Harley *et al.* (1992).

The emissions inventory includes seasonally adjusted biogenic emissions. Scale factors adapted from the seasonally-resolved biogenic emissions inventory for the South Coast Air Basin reported by Benjamin *et al.* (1997) are used to modify the SCAQS August 1987 biogenic hydrocarbon inventory which totaled 125 t d^{-1}. The scale factors are 0.30, 0.63, 1.00, 0.40 respectively for winter (Jan.–Mar.), spring (Apr.–Jun.), summer (Jul.–Oct. 10), and fall (Oct. 11–Dec.). Note that 1–10 Oct. 1987 was a period of high ozone concentrations and temperatures. The summer scale factor is used for this period.

Table 4.1 shows the total base case NMOG emissions within the modeling domain of Figure 4.1 for the highest emission days of the year. Also calculated are the NMOG emissions at two high levels of control, 5% and 15% of base case anthropogenic NMOG emissions remaining.

4.3.4 Meteorological input fields

Input data fields describing temperature, humidity, solar radiation, mixing depth, and wind speed and direction are the same as utilized in the base case model evaluation study (Winner and Cass, 1998). These fields are generated from historical observations using a second-degree polynomial fitting procedure with an r^{-2} weighting scheme (Goodin *et al.*, 1979). The 3-D wind fields are generated using an objective analysis program (Goodin *et al.*, 1980) that combines measurements of horizontal winds aloft measured daily at 6 sites up to 1500 m above ground level with surface-level horizontal winds measured hourly at 70 sites. The vertical winds are computed from continuity considerations by the objective analysis program. The mixing depth fields are calculated hourly according to Holzworth's

Table 4.1 NMOG emissions within the entire modeling domain of Figure 4.1 on peak emissions days in 1987 for the base case and for two cases with high levels of anthropogenic emission control.

Date	Base case NMOG emissions [metric tons]	5% anthropogenic NMOG emissions remaining [metric tons d^{-1}]	15% anthropogenic NMOG emissions remaining [metric tons d^{-1}]
04-Oct	3527	295	635
03-Oct	3509	294	633
01-Sep	3430	290	621
01-Oct	3310	284	603
02-Sep	3219	280	589
27-Aug	3201	279	586
01-Aug	3199	279	586
22-Apr	3144	232	539
31-Jul	3142	276	578
04-Jun	3137	232	538
21-Apr	3132	232	537
02-Oct	3072	272	567
15-Apr	3057	272	565
08-Sep	3052	271	564
31-Aug	3046	271	563
05-Oct	3034	270	561
30-Sep	3029	270	561
22-Sep	3017	270	559
20-Apr	3012	226	519
30-Jul	3002	269	557
04-May	2991	225	516
05-May	2986	224	515

(1967) method from the interpolated surface-level air temperature field plus the air temperature versus altitude data from upper air soundings taken at 6 sites daily.

4.3.5 Boundary and initial conditions

Previous research (Winner *et al.*, 1995) shows that the assumptions made about the way that boundary conditions will change in the presence of an emissions control strategy can have a large effect on the predicted outcome of an emission control strategy for O_3 abatement. In order to minimize the effect of boundary conditions, the modeling region has been greatly enlarged to cover the area shown in Figure 4.1. As emission controls are applied both within the South Coast Air Basin and in neighboring air basins, a transition will probably occur from boundary conditions like those measured historically toward boundary conditions that approach mid-Pacific Ocean clean air background values at zero on-shore anthropogenic emissions. In spite of having pushed the upwind boundary of the modeling region into the cleanest air areas for which air quality data are still routinely available, boundary conditions along the western and southern edges of the modeling region shown in Figure 4.1 are still somewhat above mid-Pacific Ocean concentrations. There is no way to definitively capture the possibly non-linear effects of altered emissions on photochemistry that occurs outside the modeling region and that might affect future pollutant inflows along the boundaries of the modeling region. Instead, some assumptions must be made that approximate what might happen.

In order to capture the possible effect of emission controls both inside and outside the modeling region on the different ROG and NO_x control strategies, an interpolation

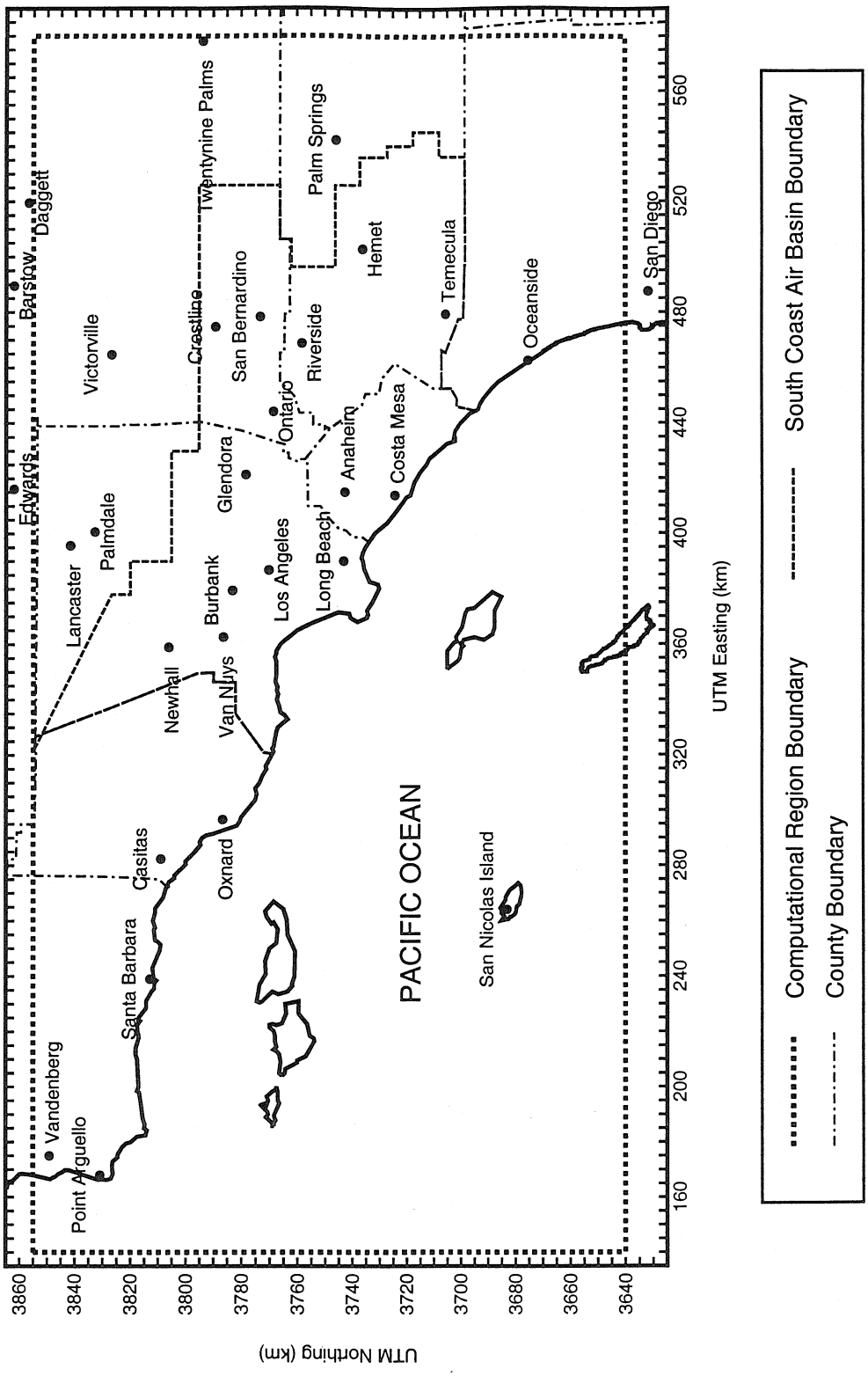

Figure 4.1 Southern California; showing the computational region for the air quality model, the legal boundaries of the South Coast Air Basin that forms the air quality control region for Los Angeles, and the county boundaries.

scheme was developed that sets initial conditions as well as boundary conditions along all edges of the model and at all elevations by interpolation between the historically measured case (Winner and Cass, 1998) and clean air case (Winner *et al.*, 1995) in proportion to the amount of emission control that is applied within the modeling domain. In effect, this simulation assumes that adjacent and upwind jurisdictions reduce their emissions by approximately the same amount as the reductions that occur within the modeling region. NO_x species (NO, NO_2) boundary and initial conditions are interpolated in proportion to the percentage change in NO_x emissions. ROG species boundary and initial conditions are interpolated in proportion to the percentage change in anthropogenic ROG emissions. Ozone boundary and initial conditions are interpolated in proportion to the average of the percentage NO_x and anthropogenic ROG control. Thus, at 100% of 1987 actual emission levels for NO_x and ROG (no control), the boundary and initial conditions are set at the historically observed levels while at 5% of the historical anthropogenic emission levels for NO_x and ROG (maximum control studied) the boundary and initial conditions approach, but do not quite reach, the mid-Pacific Ocean clean air values.

These interpolated boundary conditions may roughly approximate the combined effect of reduced emissions in the adjacent air basins (e.g., the Ventura and Santa Barbara areas to the west of Los Angeles) as emission controls are applied in those locations plus the effect of reduced emissions within the Los Angles area on pollutants recirculated across the air basin boundaries by reversals in wind direction. In reality, while the motor vehicle control program in California should reduce emissions in adjacent air basins at

the same rate as within the Los Angeles area, the stationary source control programs in adjacent air basins will be vigorous but less severe than within the Los Angeles area. Events in Mexico and the practices of ship operations in international waters likewise will affect pollutant fluxes across the boundaries. For that reason, boundary and initial conditions may not decline as rapidly as in the case with boundary and initial conditions interpolated in proportion to emission changes within the Los Angeles area.

The effect of initial conditions on the results for the simulation of 1 January 1987 is minimized by simulating five "spin-up" days (27–31 December 1986) in order to flush the initial conditions out of the modeling region. After the first day of simulation, every successive day carries the final predicted concentrations from the previous day over into the start of the next day's simulation; in other words, the model in effect runs continuously without stopping throughout the year 1987.

4.3.6 Isopleth diagram generation technique

Isopleth diagrams based on ozone concentrations predicted by the CIT photochemical airshed model are generated using the computationally efficient approach described by Winner *et al.*, 1995. A massively parallel supercomputer is used to simulate the effect of 64 different combinations of NO_x and ROG emission levels in Southern California over the calendar year 1987. Two weeks of CPU time using 64 nodes of a Cray T3E-600 computer are required to generate these 64 year-long simulations. While this currently represents a significant amount of computing resources, the continued rapid

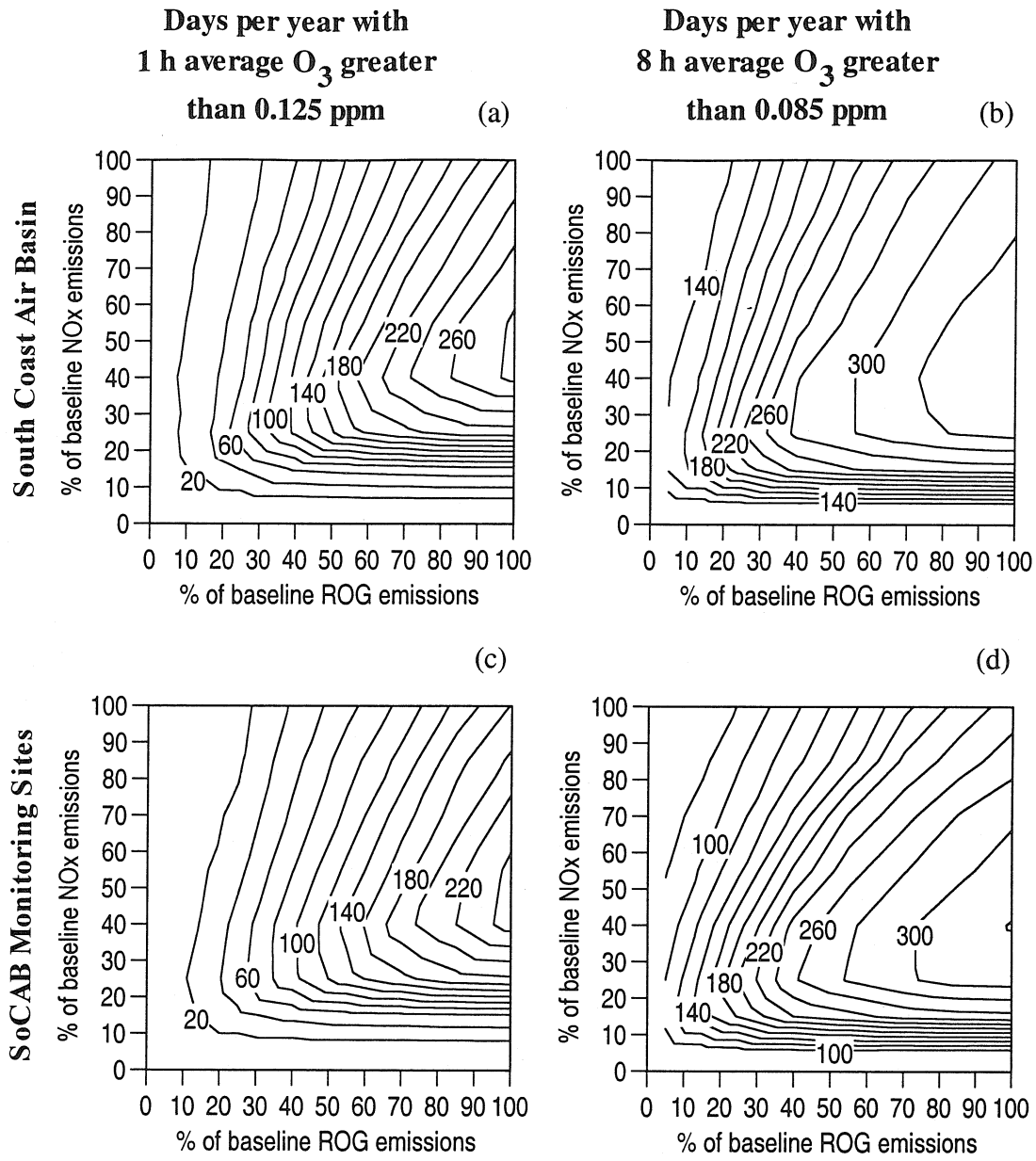

Figure 4.2 Isopleth diagrams showing the response of the predicted number of days of violation of the 1 h (a and c) and 8 h (b and d) ozone standard to alternative levels of basin-wide anthropogenic ROG and NO_x emission reductions in the Los Angeles area. Fig. 2a and b are based on predicted peak concentrations anywhere within the SoCAB; Fig. 2c and d are based on values predicted at SoCAB air monitoring sites.

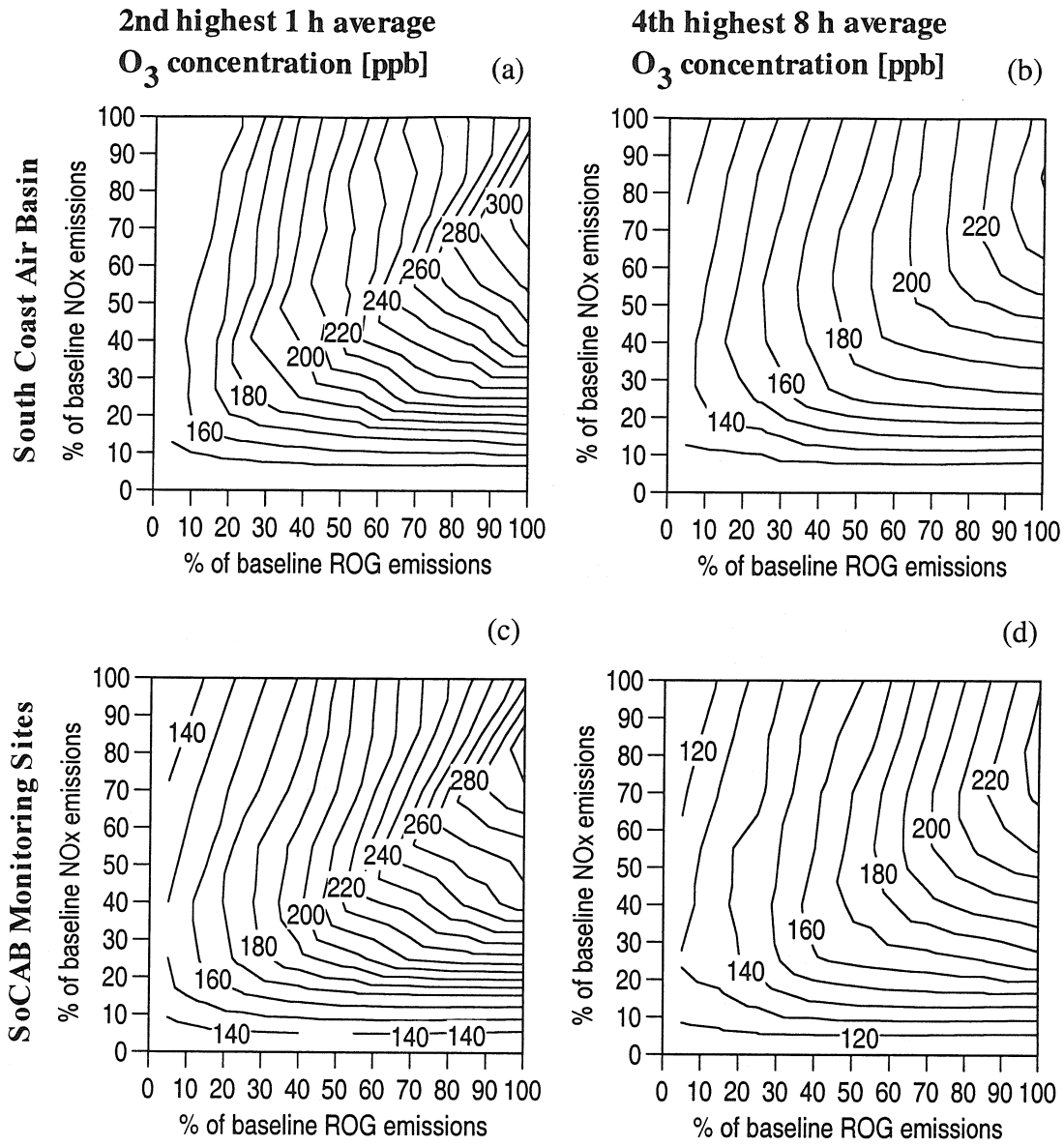

Figure 4.3 Isopleth diagrams showing the response of the predicted 2nd highest 1 h (a and c) and 4th highest 8 h (b and d) average ozone concentrations for 1987 to alternative levels of basin-wide anthropogenic ROG and NOx emission reductions in the Los Angeles area. Fig. 3a and b are based on concentrations anywhere within the SoCAB; Fig. 3c and d are based on values predicted at SoCAB air monitoring sites.

monitors in the SoCAB in 1987 (see Figure 2.2 for monitoring site locations). The upper right corner of the plot represents historical 1987 emissions. The number of days in violation of the 1 h O_3 standard in 1987 is predicted to be 181 compared to 172 such days actually observed. For the 8 h O_3 standard, the number of days during which that would have been exceeded at SoCAB air monitoring sites in 1987 is predicted to be 249 days compared to 216 days actually observed.

As seen in Figures 4.2cd and 4.3cd the 1 h and 8 h peak ozone concentrations likely to be measured at those air monitoring sites existing in the SoCAB in 1987 are lower than at the location of the daily predicted basin-wide peaks. This is expected because the basin-wide peak probably will not occur exactly at an air monitoring site. Each of the isopleth diagrams shows varying degrees of ozone inhibition at high NO_x emission rates (i.e., in the upper left portion of the diagram, increasing NO_x emissions leads to reduced ozone concentrations).

Compliance with the former 1 h Federal standard for O_3 is not predicted to be attained at any combination of anthropogenic emissions reductions studied (maximum level of control studied is 95% for both anthropogenic ROG and NO_x). However, much improvement can be made when measured against the former Federal 1 h standard. For example, 85% ROG control and 45% NO_x control would reduce the measured frequency of violation of the former Federal 1 h average O_3 standard within the SoCAB to about 25 days per year under 1987 meteorological conditions.

In contrast, a very large number of days per year are predicted to exceed the new Federal 8 h ozone standard even at high levels of ROG and NO_x control. This occurs

because even at very high levels of control, enough anthropogenic plus biogenic emissions remain to boost the 8 h average O_3 concentration within the SoCAB to above 0.085 ppm on 106 days per year when combined with the inflows pollutants in the background air even at low concentrations within the background air.

Another key result of this analysis is that the days that are hardest to control are not necessarily the days that produced the highest peak ozone concentrations under 1987 historical conditions. Table 4.2 lists the days in 1987 that displayed the highest measured 1 h and 8 h average O_3 concentrations under 1987 actual meteorological conditions and emission levels (100% ROG, 100% NO_x). Also shown in Tables 4.2 and 4.3 are the dates and magnitudes of the 10 highest 1 h average O_3 concentration peaks that are predicted at several combinations of emission control that fall as close as we can come to the 20 days per year of violation contour for the case of the 1 h O_3 standard evaluated only at SoCAB air monitoring sites given the 64 combinations of ROG and NO_x control for which we have detailed results. These combinations are 25% ROG remaining at 100% of 1987 NO_x emissions (13 predicted exceedences of the 1 h standard); 15% ROG, 55% NO_x remaining (17 exceedences); 15% ROG, 15% NO_x (22 exceedences); and 100% ROG, 15% NO_x (58 exceedences). At high levels of both ROG and NO_x control, the hardest days to control often are those with the highest ROG emissions as listed in Table 4.1. The peak ROG emission days are caused by high levels of ROG emissions from motor vehicles. This effect is due to high temperatures; not necessarily the maximum temperature inland, but instead this can be due to higher than average temperatures over the western part of SoCAB where the highest density of emissions sources are located

Table 4.2 Dates and magnitude [ppb] of 10 highest 1 h and 8 h average O_3 concentrations predicted at SoCAB monitoring sites under 1987 meteorological conditions and at various levels of emission control. Percentage values show the fraction of base case 1987 emissions remaining after control.

	Observed		100% ROG 100% NO_x		25% ROG 100% NO_x		15% ROG 55% NO_x		15% ROG 15% NO_x		100% ROG 15% NO_x	
1 h avg.												
1st	08 Sep	330	03 Oct	319	21 Apr	153	21 Apr	164	21 Apr	155	02 Sep	174
2nd	01 Sep	320	05 Oct	276	22 Apr	153	22 Apr	157	22 Apr	153	22 Apr	168
3rd	02 Oct	320	15 Apr	270	03 Oct	152	20 Apr	153	03 Oct	151	21 Apr	167
4th	31 Jul	300	01 Jun	267	22 Jun	148	03 Oct	149	20 Apr	141	05 Oct	166
5th	30 Sep	300	22 Sep	265	20 Apr	141	22 Jun	138	02 Oct	140	03 Oct	163
6th	28 Aug	290	06 Oct	263	04 May	133	23 Apr	132	05 Oct	137	02 Oct	158
7th	28 Sep	290	14 Apr	261	03 May	132	01 Jun	131	19 Apr	132	04 Oct	158
8th	20 Sep	280	23 Sep	255	05 May	132	01 Oct	130	03 May	131	30 Sep	150
9th	03 Oct	280	22 Apr	254	21 Jun	132	02 Oct	130	21 Sep	131	20 Apr	147
10th	02 Sep	270	05 May	246	30 Sep	132	03 May	128	04 Oct	131	01 Sep	147
8 h avg.												
1st	31 Jul	210	03 Oct	245	22 Apr	144	21 Apr	150	22 Apr	145	22 Apr	157
2nd	13 Jul	199	01 Jun	233	03 Oct	142	22 Apr	150	21 Apr	142	21 Apr	153
3rd	08 Sep	196	15 Apr	232	21 Apr	139	03 Oct	138	03 Oct	138	03 Oct	150
4th	29 Aug	194	06 Oct	229	22 Jun	133	20 Apr	138	20 Apr	128	05 Oct	145
5th	14 Jul	191	23 Sep	227	20 Apr	124	22 Jun	128	19 Apr	125	02 Sep	139
6th	30 Jul	190	22 Sep	225	23 Apr	121	23 Apr	124	23 Apr	125	04 Oct	138
7th	04 Aug	190	05 Oct	225	05 May	120	21 Jun	120	05 Oct	122	21 Sep	137
8th	20 Sep	189	22 Apr	222	04 May	119	02 Oct	119	21 Sep	122	23 Apr	136
9th	24 Jun	186	23 Apr	215	03 May	119	03 May	119	04 Oct	121	02 Oct	133
10th	28 Sep	186	06 May	209	21 Jun	117	05 May	119	02 Oct	121	20 Apr	133

Table 4.3 Dates and magnitude [ppb] of 10 highest 1 h and 8 h average O_3 concentrations predicted anywhere in the SoCAB under 1987 meteorological conditions and at various levels of emission control. Percentage values show the fraction of base case 1987 emissions remaining after control.

	100% ROG 100% NO_x		25% ROG 100% NO_x		15% ROG 55% NO_x		15% ROG 15% NO_x		100% ROG 15% NO_x	
1 h avg.										
1st	03 Oct	319	21 Apr	167	21 Apr	180	21 Apr	162	02 Sep	179
2nd	22 Sep	284	03 Oct	161	22 Apr	164	22 Apr	160	21 Apr	175
3rd	05 Oct	297	22 Apr	160	20 Apr	159	03 Oct	153	22 Apr	172
4th	14 Apr	277	20 Apr	159	03 Oct	157	20 Apr	146	05 Oct	172
5th	01 Jun	277	22 Jun	154	04 Oct	153	02 Oct	145	03 Oct	166
6th	08 May	275	04 Oct	151	22 Jun	141	05 Oct	142	02 Oct	165
7th	04 Oct	273	03 May	144	03 May	138	19 Apr	137	04 Oct	164
8th	15 Apr	270	05 Oct	140	30 Sep	138	21 Sep	137	01 Sep	161
9th	06 Oct	269	04 May	137	05 Oct	138	05 May	134	30 Sep	159
10th	23 Sep	261	05 May	137	04 May	137	23 Apr	132	20 Apr	153
8 h avg.										
1st	03 Oct	253	21 Apr	153	21 Apr	166	21 Apr	150	22 Apr	162
2nd	04 Oct	252	22 Apr	149	22 Apr	154	22 Apr	149	21 Apr	160
3rd	01 Jun	240	03 Oct	146	03 Oct	147	03 Oct	141	03 Oct	156
4th	15 Apr	234	20 Apr	146	20 Apr	145	20 Apr	135	05 Oct	149
5th	08 May	231	04 Oct	139	04 Oct	143	19 Apr	129	02 Sep	143
6th	23 Sep	230	22 Jun	138	22 Jun	131	04 Oct	127	20 Apr	140
7th	06 Oct	244	03 May	130	03 May	128	23 Apr	126	04 Oct	139
8th	21 Apr	230	21 Jun	129	23 Apr	126	05 Oct	124	21 Sep	138
9th	22 Sep	230	23 Apr	126	21 Jun	123	21 Sep	122	02 Oct	136
10th	05 Oct	229	05 Oct	123	05 May	122	05 May	121	23 Apr	136

(i.e., highest traffic density). Clearly, effort should be focused on the 02-06 October period and possibly also on the 20-23 April period in order to examine the effectiveness of emission control strategies.

Figure 4.4 shows the shifts in the frequency of peak O_3 concentrations as emission controls are applied. The emission control combinations are the same as those described when constructing Tables 4.2 and 4.3. Figure 4.4 clearly illustrates that the frequency distribution of peak 8 h average O_3 concentrations is altered as emissions controls are applied. The 8 h average O_3 concentrations experienced on the highest concentration days of the year fall much more rapidly than the mid to low values. This is further demonstrated in Figure 4.5 which shows the time series of the daily peak 8 h average O_3 concentrations before and after a control program that leaves 15% of 1987 anthropogenic ROG emissions and 55% of 1987 NO_x emissions remaining after control.

4.5 Conclusions

A method has been developed and demonstrated in which a photochemical airshed model driven by automated assimilation of routinely measured meteorological data can be used to predict how the frequency of occurrence of daily maximum 1 h and 8 h average O_3 concentrations will shift in response to widely varying levels of emissions control. Methods developed were illustrated by application to the air pollution control situation in Southern California. It was found that a program of stringent ROG and NO_x controls could reduce the frequency of occurrence of measured violations of the former U.S. Federal 1 h average O_3 standard to perhaps 25 days per year. The same calculations

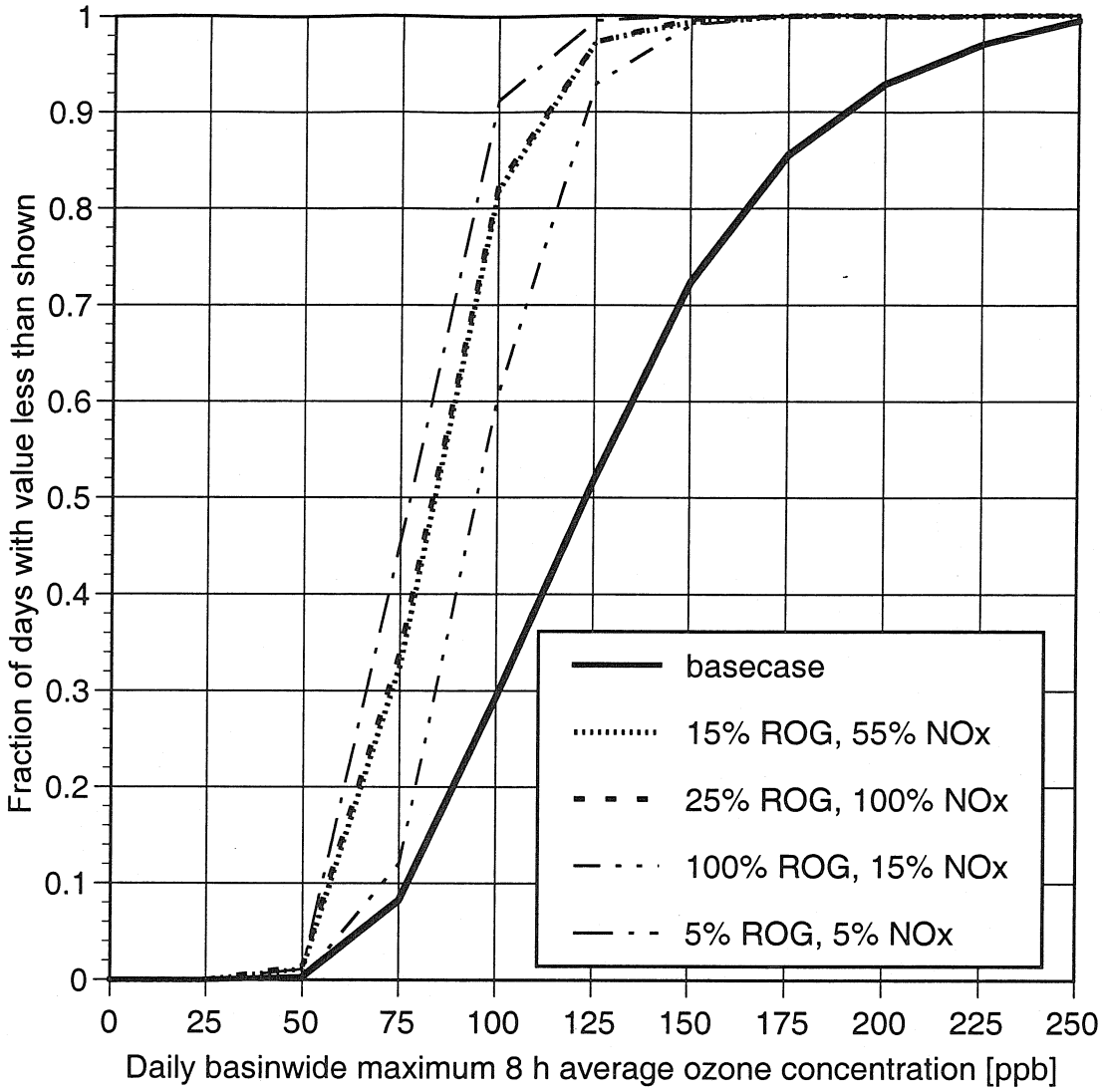

Figure 4.4 Cumulative frequency of occurrence of the daily regionwide maximum 8 h average ozone concentrations for August 1987 predicted anywhere within California's South Coast Air Basin for various levels of anthropogenic emissions remaining after control.

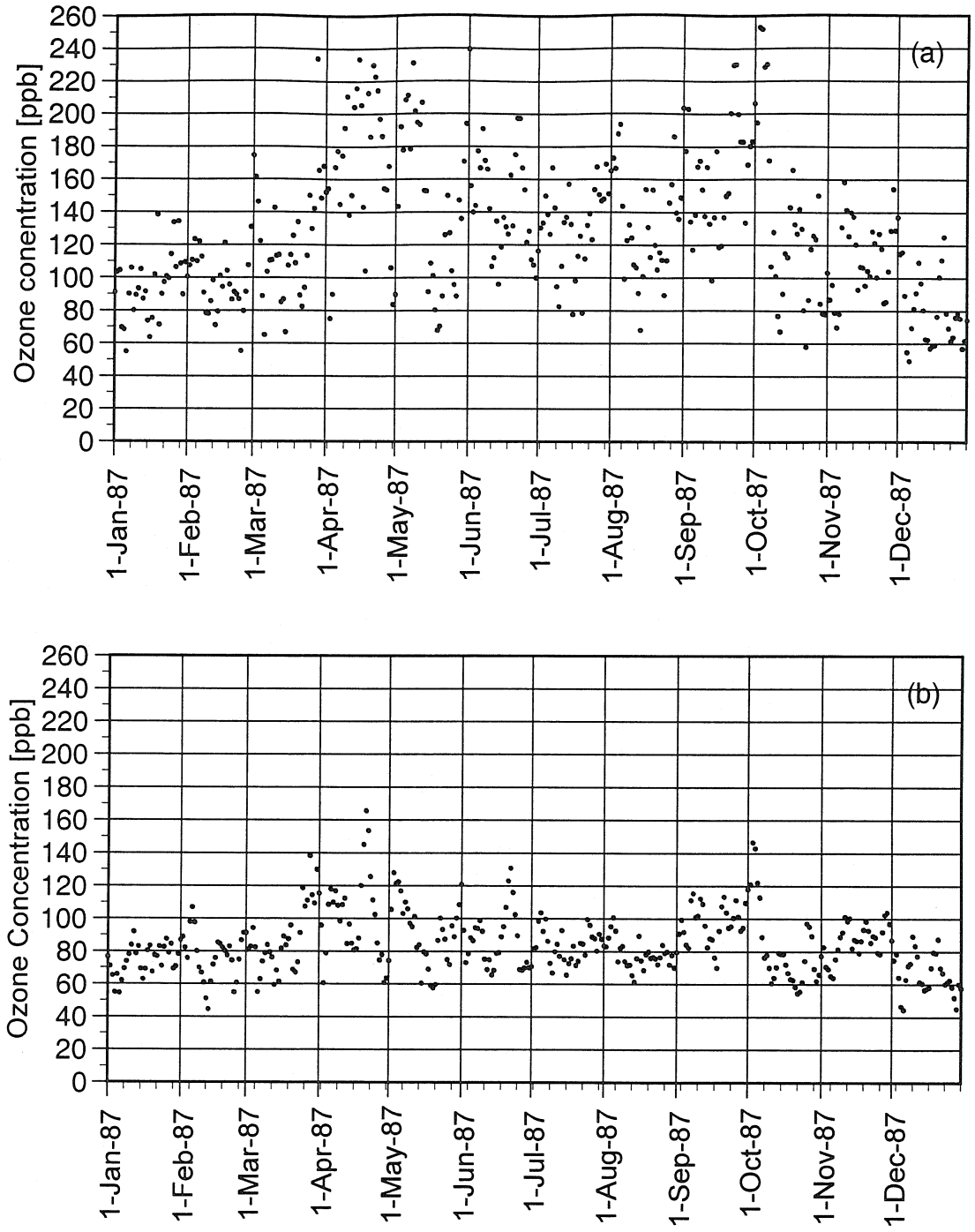

Figure 4.5 Time series of daily maximum 8 h average ozone concentration [ppb] predicted anywhere in the SoCAB: (a) historical 1987 emissions (b) 15% ROG, 55% NOx anthropogenic emission remaining control case.

show that the newly established 8 h average U.S. Federal O_3 standard will be exceeded nearly 100 days per year under 1987 meteorological conditions even at the most stringent levels of emission control considered here.

It was found that the days with the highest historically measured O_3 concentrations in the Los Angeles area are not necessarily the days which are most difficult to bring below the national ambient air quality standards for O_3. This argues further that the process leading to the selection of an appropriate level of emission control for an air basin needs to consider conditions observed on more than a few historical days with the highest measured concentrations. Methods demonstrated in the present study show that it is practical to examine essentially all days of the year when testing emission control strategy performance.

Chapter 5 Modeling the long-term frequency distribution of regional ozone concentrations using synthetic meteorology

5.1 Motivation

Deterministic modeling of long-term ozone concentrations not only involves large amounts of computer resources, but also huge amounts of time and effort to gather and prepare the input data. For instance, while the amount of computer time required for airshed model use has been reduced to the equivalent of 3.5 min per day of simulation (see chapter 3 for details), the 3 day August episode of the Southern California Air Study (SCAQS) required over 6 months of effort just to set up the input data. Clearly, full use of the computer resources is not available until the input data can be generated more efficiently. In particular, the effort to generate the meteorological data base required for use with airshed models might be reduced greatly by turning to probabilistic methods. The atmospheric system can be defined by the probability distribution of the system variable fields–temperature, wind, mixing height, solar radiation, and humidity. The concept of synthetic meteorology is to draw model inputs from these probability distributions of the meteorological variables.

5.2 Introduction

By incorporating information about the stochastic structure of weather sequences, a model that simulates alternative weather conditions via Monte Carlo simulation is developed. An important assumption required for this model is that the probability

distributions that describe the long-term, large-scale weather are stationary. This allows the estimation of model parameters from historical weather records.

Stochastic models have been used extensively to generate synthetic data for hydrological models (Thomas and Fiering, 1963; Fiering, 1967; Jackson, 1975). Generally the spatial nature of the problem is simplified so that time series analysis methods can be used (Box and Jenkins, 1976). A good example of this kind of analysis is provided by Jensen (1976) who evaluated alternative policies for the operation of dams on the Colorado River by using synthetic stream flow data to challenge various reservoir operation policies.

The synthetic weather model developed in the present research is most similar to those used in stochastic precipitation models (Bardossy and Plate, 1991; Hay *et al.*, 1991; Wilson *et al.*, 1992). These models simulate the sequence of precipitation events based on large-scale meteorological information.

5.3 Methods

One of the major problems encountered in generating synthetic meteorology is to determine the appropriate key variables. In a sense, the process can be thought of as drawing meteorological fields from a series of linked probability distributions. A random number can be used to draw a particular case from the probability distribution of the first, key variable. The following variables would then be drawn from their respective distributions based on probability distributions that are jointly coupled to the initial key variables.

5.3.1 Synoptic climatological classification

A likely candidate for choice of the first key variable is a synoptic climatology indicator which describes the general, large-scale weather conditions. Synoptic climatology seeks to explain the local climate in terms of large-scale atmospheric circulation (Barry and Perry, 1973). Thus, weather elements are combined into groups that represent the synoptic climate at a particular time. Early research in air pollution accumulation (Niemeyer, 1960; Holzworth, 1962) used synoptic climatological techniques to determine that high pollution episodes were associated with low elevation temperature inversion conditions–high pressure, calm winds, and clear skies.

Recent research has developed automated methods for the classification of synoptic conditions (Davis and Kalkstein, 1990; Davis and Walker, 1992). These methods use a group of key meteorological variables such as temperature, geopotential height, wind components, and dew point temperature to distinguish between characteristically different weather patterns. Through the use of an objective classification technique, Davis and Walker (1992) identified 13 synoptic types of circulation patterns over the Western United States based on rawinsonde data from 21 weather stations for the years 1979–1988. The 13 synoptic types resemble features observed in the general circulation of the atmosphere. The method of Davis and Walker (1992) will be used here with minor modifications to organize the historical weather data for the Western U.S. into days that fall into specific synoptic weather classes.

5.3.1.1 Principal component analysis

Principal component analysis (PCA) is used here to linearly transform the historical daily data on weather parameters from 21 weather monitoring sites in the western United States and northern Mexico over the period 1979 to 1992 onto orthogonal axes which are uncorrected with each other (Dillon and Goldstein, 1984). PCA determines the smallest number of linear combinations of a set of variables that retain as much of the information content of the original variables as possible. As such, it reduces the size of the data set by grouping together variables that basically convey the same information because they vary together in a systematic way. The components used in PCA are based on the eigenvectors computed from a matrix of data. PCA orders the components based on the magnitude of their eigenvalues, and then presents the components in decreasing order of significance in representing the original data. Thus, the first component explains the highest proportion of the total variance of the original data set and the second component accounts for as much of the residual variance as possible while remaining orthogonal to the first component. For the cases treated in this study, PCA reduces the description for each single day of weather observations from a total of 798 observed values to a new total of 6 component scores that describe that day.

5.3.1.2 Cluster analysis

Once the matrix of available historical weather variables has been developed into component scores for each day via PCA, a two-stage cluster analysis (Davis and Kalkstein, 1990; Davis and Walker, 1992) is used to group days from the historical

weather records that possess similar component scores from the PCA procedure. First, average-linkage clustering is used to select the number of clusters and to develop preliminary clusters. The final clusters are produced by k-means clustering which iterates the assignment of days to the clusters and cluster definitions to optimize the final classification.

Average-linkage clustering is an agglomerative hierarchical procedure (Dillon and Goldstein, 1984). Each day of weather component scores starts as a cluster by itself. The two clusters with the most similar weather conditions as described by the component scores are merged to form a new cluster that replaces the two old clusters. This merging of close clusters is repeated, reducing the number of clusters by one through each iteration. The two clusters which are merged have the smallest average squared distance between all component scores within the two clusters weighted by the number of members of each cluster. This average linkage statistic (L_{12}) is

$$L_{12} = \frac{1}{N_1 N_2} \sum_{m=1}^{N_1} \sum_{n=1}^{N_2} \left\| \vec{x}_m - \vec{y}_n \right\|^2 \qquad (5.1)$$

where N_1 is the number of days in cluster 1, N_2 is the number of days in cluster 2, $\vec{x}_m$ is the vector of component scores that describe day of observation m within cluster 1, and $\vec{y}_n$ is the vector of component scores that describes day of observation n within cluster 2.

The average linkage procedure produces compact spherical clusters. These clusters tend to have small variances between weather conditions for days within clusters and large separations between clusters. In fact, the average linkage statistic is composed of two parts–the variances of the clusters and the squared distances between the centroids

of clusters. The variance of a cluster k is the within-cluster sum of squares divided by the number of observations in cluster k:

$$\frac{W_k}{N_k} = \frac{1}{N_k} \sum_{i=1}^{N_k} \sum_{j=1}^{M} \left(x_{ij} - \overline{x}_j \right)^2 \tag{5.2}$$

where N_k is the number of days in cluster k, M is the number of component scores that describe each day, x_{ij} is the i^{th} member of component score j in cluster k, and $\overline{x}_j$ is the mean of component score j in cluster k. The average linkage statistic (L_{12}) is thus

$$L_{12} = \left\| \overline{\mathbf{x}} - \overline{\mathbf{y}} \right\|^2 + \frac{W_1}{N_2} + \frac{W_2}{N_2} \tag{5.3}$$

where $\overline{\mathbf{x}}$ is the vector of mean component scores that describe the weather for cluster 1 (i.e., it's centroid) and $\overline{\mathbf{y}}$ is the vector of mean component scores that describe the weather for cluster 2.

The average linkage clustering procedure has been shown to produce the most realistic grouping of similar days into synoptic climatological classes (Kalkstein *et al.*, 1987). Unfortunately, it also tends to develop many clusters that contain very few members. To remedy this, only average linkage clusters that contain more than 2% of the days of observation are retained in the present study.

Convergent k-means clustering is a partitioning technique which rearranges the days of observation after they have been assigned to initial clusters (Dillon and Goldstein, 1984). The k-means clustering algorithm seeks to find cluster definitions and memberships which minimize the total within-cluster sum of squares distance over all the clusters. Days are assigned to a cluster based on minimizing the squared distance between

component scores of the observed days and the means of the component scores for that cluster. The function being minimized is

$$\phi = \sum_{k=1}^{K} \sum_{j=1}^{M} \sum_{i=1}^{N_k} (x_{ijk} - \bar{x}_{jk})^2 \qquad (5.4)$$

where K is the total number of clusters, M is the number of variables, N_k is the number of days in cluster k, x_{ijk} is the i^{th} member of variable j in cluster k, and $\bar{x}_{jk}$ is the mean of variable j in cluster k.

The k-means procedure regroups the data and updates the means again until the number of days which change groups is below the specified tolerance level. This procedure depends greatly on the initial cluster seeds. It yields a local optimization such that no movement of an observation from one cluster to another will reduce the value of parameter ϕ.

5.3.2 Stochastic model for synoptic class

5.3.2.1 Semi-Markov model

The first step toward stochastically generating photochemical airshed model meteorological inputs in which the changes in synoptic weather class from one day to the next follow each other in a way that is typical of the location considered and the season of the year is to generate daily sequences of weather. The simplest stochastic model would be to use a first order, homogeneous Markov chain (Feller, 1968) to model the transition from one synoptic class to another. In a Markov chain the current state (synoptic class)

depends only on the previous state. Thus, the simplest Markov chain would fail to meet our needs because it would not capture the persistence of weather patterns.

A seasonal semi-Markov model can be stated that is a generalization of a Markov process. This stochastic process characterizes the period of duration of each synoptic class and the transition probabilities to the remaining classes at the end of the period. The transitions between individual states of a semi-Markov process form a simple Markov chain. However, the within-state duration (in days) within the current synoptic class is determined by a probability distribution function assigned to each state. This persistence depends only on the current state. Thus, the state of the current day depends on both the state of the previous day and the length of stay in that state.

The within-state duration (in days) is determined by a probability distribution function assigned to each synoptic class. We will see later that the distribution of persistence is described adequately by a geometric probability distribution:

$$f(t) = pq^{t-1} \quad t = 1,2,\ldots \tag{5.5}$$

where t is the duration of the state in days, p is the reciprocal of the mean duration in the state, and q equals 1 minus p.

The seasonal aspect of weather is captured by calculating the matrix of transition probabilities between synoptic weather classes separately for each month. Each entry in the transition matrix for a month represents the probability of transition from one synoptic class to another. For simulation, first the month is selected. The initial synoptic class within that month is selected by a random draw from the frequency distribution of synoptic classes within that month. The duration of persistence of the synoptic class is

determined by generating a geometrically distributed random variable based on equation 5.5. The method of Ross (1987) transforms U, a uniform random with over the range 0 to 1, to a geometric random variable, X that defines the number of days of persistence within the synoptic class.

$$P\{X = i\} = (1-p)^{i-1}p \qquad i \geq 1 \tag{5.6}$$

$$\sum_{i=1}^{j-1} P\{X = i\} = 1 - P\{X > j-1\} \tag{5.7}$$

Since $P\{X > j - 1\}$ is the probability that the first j-1 are all failures which is equivalent to

$$\sum_{i=1}^{j-1} P\{X = i\} = 1 - (1-p)^{j-1} \qquad j \geq 1 \tag{5.8}$$

The transformation of U to X occurs by setting X equal to the value j for which:

$$1 - (1-p)^{j-1} < U < 1 - (1-p)^{j} \tag{5.9}$$

Because 1 - U has the same distribution as U, this condition is satisfied by:

$$X = \min\{j: j > \frac{\log U}{\log(1-p)}\} \tag{5.10}$$

After the period of persistence of the first synoptic class is ended, the procedure is to draw the next synoptic class from the transition probability density function that describes transition to the next synoptic class given the class just exited at that time of year. Then within the next chosen synoptic class, the duration of persistence in that class is drawn as before. This technique allows the random generation of a full month sequence of synoptic class indicators representative of the chosen month. The next step is to generate local weather variables sufficient to drive the photochemical airshed model that are representative of days belonging to the selected synoptic classes.

5.3.3 Generation of meteorological fields

5.3.3.1 Key variables

Over the years, meteorologists studying the Los Angeles air pollution problem have determined several meteorological variables which exert a strong influence on air quality (Zeldin et al., 1988; Cassmassi and Bassett, 1993; Davidson, 1994). In order to transfer the large-scale information described in the synoptic class to the regional scale of interest, two additional key variables are chosen to further characterize the local meteorological situation.

The temperature at the 850 mb level (T850) of the rawinsonde sounding taken in the early morning at Loyola Marymount University (LMU) near Los Angeles International Airport (LAX) is one of the key variables used here. T850 has been found to represent the potential for photochemical air pollution (Davidson, 1994) since it is indicative of the strength of the temperature inversion.

The other key variable chosen captures the atmospheric pressure gradient between the desert and the coastal plain in the Southern California area. This variable is formed from the sum of three surface pressure gradients (ΣPG) at 0700 PST: Long Beach - Daggett, San Diego - Las Vegas, and Riverside - Victorville. ΣPG characterizes the potential for onshore or offshore flow, before the onset of the sea breeze. Strong, onshore flow potential yields positive ΣPG values. Strong, offshore flow potential yields negative ΣPG values. Values of ΣPG near zero indicate stagnation.

The distribution of T850 and ΣPG values within a given synoptic class approaches a two-dimensional joint normal distribution. In order to deal with the dependency between T850 and ΣPG, a multivariate technique is used to estimate the parameters of the underlying multivariate normal distribution. The process begins by generating two univariate normal random variables, N_1 and N_2, from two uniform random variables, U_1 and U_2 defined over the range from 0 to 1, via the method of Box and Muller (1958):

$$N_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2) \tag{5.11}$$

$$N_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2) \tag{5.12}$$

The multivariate, d-dimensional normal random vector, $\vec{M}$, can be generated from a linear transformation of independent univariate normal random variables, $\vec{N}$ (Johnson, 1987).

$$\vec{M} = \sigma A \vec{N} + \vec{\mu} \tag{5.13}$$

$$AA' = R \tag{5.14}$$

$$\mu_i = E(M_i) \tag{5.15}$$

where R is the dxd correlation matrix (non-singular), A is an dxd matrix (not unique), σ is a dxd matrix with the standard deviation of each variable on the diagonal, and $\vec{\mu}$ is a dx1 vector of the mean values. For the 2x2 case, where ρ is the correlation coefficient:

$$R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{5.16}$$

$$A = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \tag{5.17}$$

Thus, the desired multivariate normal distribution for T850 and ∑PG can be simulated by

$$T850 = \sigma_{T850} N_1 + \mu_{T850} \tag{5.18}$$

$$\sum PG = \sigma_{\sum PG} (\rho N_1 + \sqrt{1-\rho^2} N_2) + \mu_{\sum PG} \tag{5.19}$$

5.3.3.2 Generation of local weather variables

Next, values for the remaining local weather variables of interest are generated given the synoptic class and random draws from the joint distribution of T850 and ∑PG values within that synoptic class at the time of year of interest. One method to do this would be to continue to base selections purely on statistical properties of observed values of all the many weather variables such as wind speed and direction, temperature, humidity, and so forth. However, this method has two large drawbacks. First, the conservation of quantities such as the mass of air advected across the airshed would not be assured easily. Second, a large number of parameter distributions and random numbers would be required to generate the large amount of input data required for a photochemical airshed model. Instead of the purely statistical approach, we will synthetically generate the initial conditions for a mesoscale meteorological model given the chosen synoptic class, T850, and ∑PG values; and use the mesoscale meteorological model to generate the time series of meteorological input fields. This method has the clear advantage of generating wind fields, temperature fields, and humidity fields that are mass-consistent

and that obey the underlying physical laws that govern atmospheric motion. The meteorological model selected for use is the Pennsylvania State University - National Center for Atmospheric Research Mesoscale Model, a state-of-the-science three-dimensional, prognostic model developed from the mesoscale model of Anthes and Warner, 1978. The current version, MM5 (Dudhi, 1993; Seaman et al., 1995), uses prognostic equations based on non-hydrostatic dynamics to predict the three-dimensional wind components, temperature, and water vapor mixing ratio given local topography, land use, initial conditions, and boundary conditions.

Figure 5.1 shows a map of the western United States and northern Mexico with the three nested regions used for this study based on a Lambert conformal map projection. The outermost region is divided into 56x60 grid squares that are 45 km on each side, the middle region is divided into 70x40 grid squares that are 15 km on each side, and the inner region which covers the Southern California area over which the photochemical model will be used is divided into 115x64 grid squares that are 5 km on each side. The nested grids are two-way interactive. The vertical coordinate used in MM5 is a terrain-following non-dimensional pressure coordinate, referred to as sigma levels. The sigma levels are based on a hydrostatically balanced reference state and are constant in time. For this study, the vertical resolution of MM5 is 30 layers, with 9 layers below 1100 m above ground level. The top of the meteorological model is set at 100 mb.

The Blackadar Planetary Boundary Layer (PBL) parameterization (Zhang and Anthes, 1982) represents the sub-grid scale fluxes of heat, moisture, and momentum

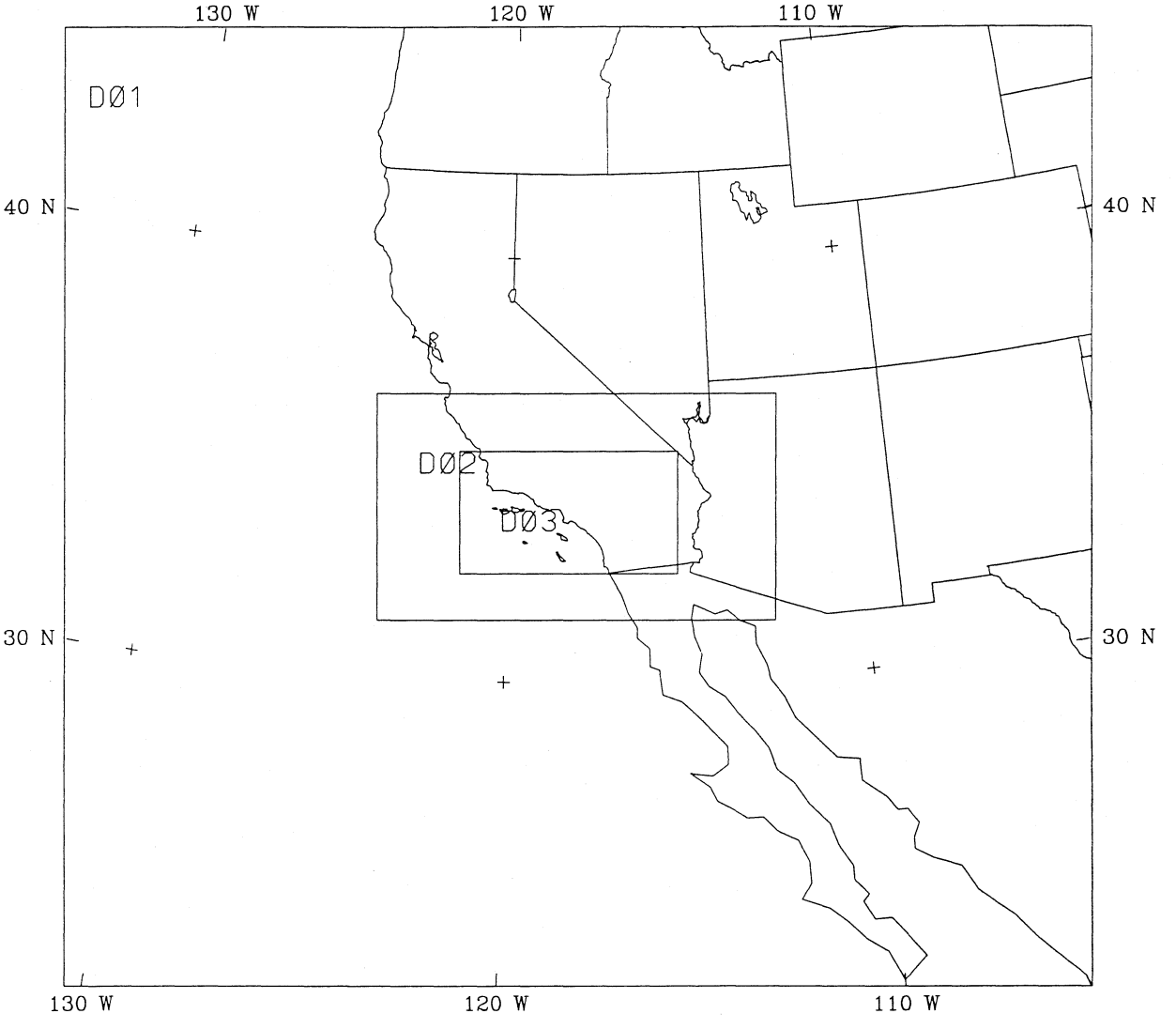

Figure 5.1 Western United States and northwestern Mexico, showing the three nested regions used for prognostic mesoscale meteorological model simulations.

within MM5 using a first-order eddy diffusivity scheme. The ground temperature is predicted by solving the surface energy budget equation.

The first step in executing MM5 is to specify the terrain which guides the air flows. Then initial conditions and boundary conditions are developed by objective analysis applied to the outermost gridded region of Figure 5.1 at 12 h intervals. The National Center for Environmental Prediction (NCEP) spectral analysis procedure (Benjamin and Seaman, 1985) is used to produce a first guess set of initial conditions defined over the outermost grid of Figure 5.1. Rawinsonde data at 12 h intervals and surface observations at 3 h intervals are then combined with the first guess fields. The initial conditions for the middle and inner grids are produced by spatial interpolation from the outermost grid. The boundary conditions that apply to the outermost grid are developed by temporal interpolation between consecutive rawinsonde and surface observations made near the outside edges of the map in Figure 5.1. The boundary conditions are applied to the 5 rows and columns of grid cells at the edges of the outermost grid through a relaxation technique. The middle and inner grid are two-way nested. The boundary conditions for the middle grid are based on values that prevail within the larger grid at the boundary where the larger grid meets the middle grid. Information from the middle grid is also feed to the larger grid. The innermost grid relates to the middle grid in an analogous fashion.

To use MM5 to generate synthetic meteorological conditions, it is first necessary to provide the model with synthetic initial conditions. The initial conditions and boundary conditions for the area outside of the fine grid are set using observed data from that

historical day in the period 1981-1992 which matches the selected values of the synoptic class, T850, and ∑PG as closely as possible. The standard weather observations at sites within the area of the innermost grid are replaced by synthetically generated values. The locations of the few meteorological sites at which synthetic initial conditions will be supplied are shown in Figure 5.2. These synthetic initial values for u and v wind components, temperature, and relative humidity are drawn from their historically observed distributions based on the synoptic class, T850, and ∑PG values drawn earlier in the process. Since the meteorological files needed as inputs to MM5 over the period 1981-1992 are already archived at the National Center for Atmospheric Research, the further processing needed to generate mass consistent meteorological fields over the grid system of Figure 5.1 using MM5 driven by synthetic initial conditions requires the further processing of only about 1% as much weather data as would be required to develop meteorological fields using direct interpolation of the historical meteorological records.

5.3.4 Air quality simulation

The output of meteorological fields from MM5 is used as the meteorological input to the CIT photochemical airshed model (Harley *et al.*, 1993ab). The specific variables of interest are the u and v components of the horizontal wind as a function of height, the temperature, the water vapor mixing ratio, the PBL height, and the surface downward radiation flux; all as a function of horizontal location in the airshed. The Lambert conformal coordinates used by MM5 are translated into the Universal Transverse Mercator (UTM) coordinate system used by the CIT airshed model. Particular care is

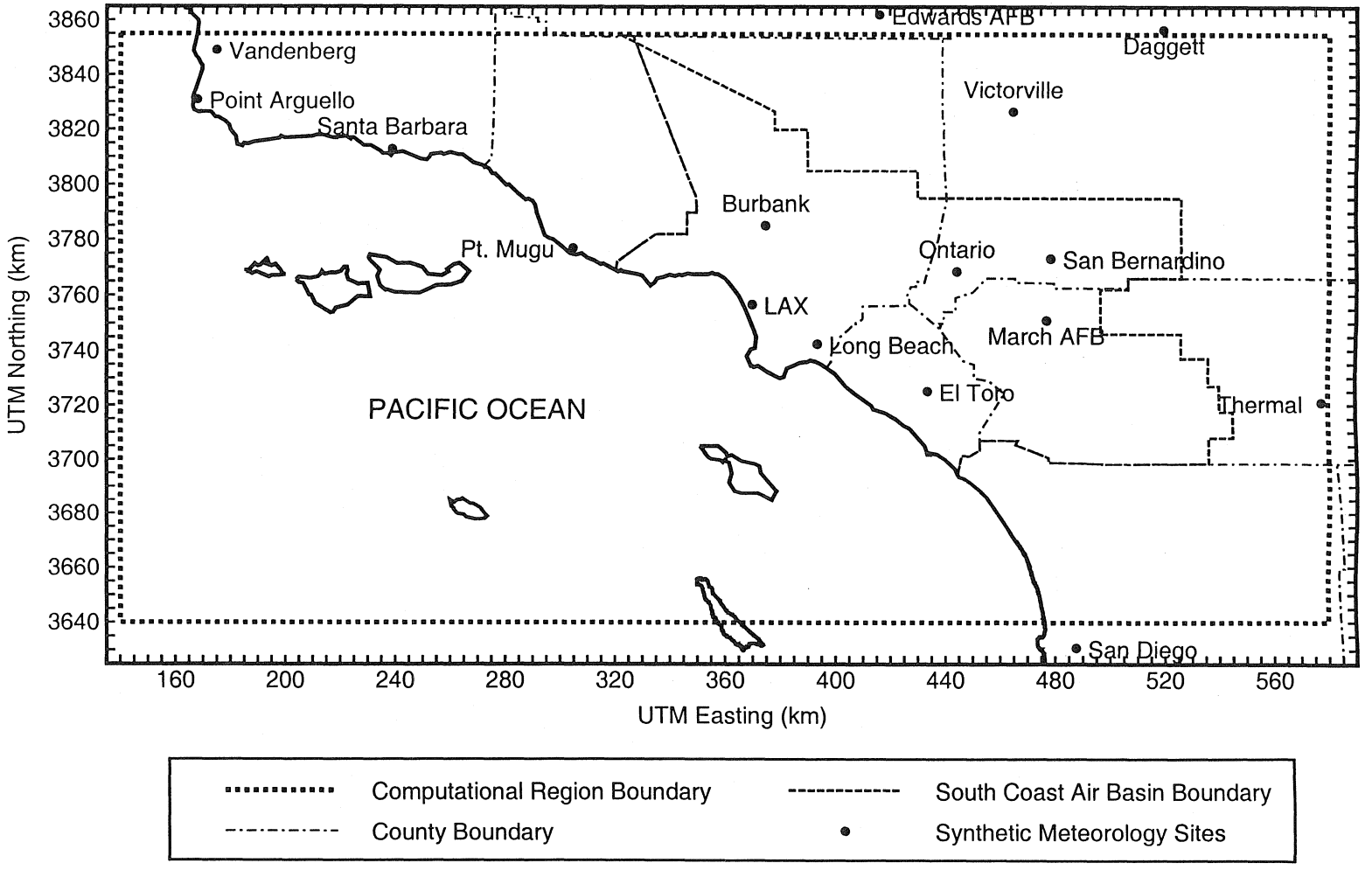

Figure 5.2 Southern California; showing the computational region for the air quality model simulations, the legal boundaries of the South Coast Air Basin that forms the air quality control region for Los Angeles, the county boundaries, and meteorological sites at which synthetic initial conditions are supplied.

taken to account for the map projection effect on the direction of the winds. Once the coordinates are transformed, the values from MM5 are interpolated to the center of each UTM grid cell used by the CIT model. Figure 5.2 shows the modeling region used by the airshed model in this case.

The CIT photochemical airshed model numerically solves the atmospheric diffusion equation (Lamb and Seinfeld, 1973; McRae *et al.*, 1982)

$$\frac{\partial c_i}{\partial t} + \nabla \bullet (\vec{u} c_i) = \nabla \bullet (K \nabla c_i) + R_i + S_i \tag{5.20}$$

where c_i is the ensemble mean concentration of species i, $\vec{u}$ is the mean wind velocity at location $\vec{x}$ at time t, K is the eddy diffusivity tensor (assumed to be diagonal), R_i is the chemical reaction rate of species i which depends on the pollutant species $c_1 \ldots c_n$ and on the temperature T, and S_i is the elevated source emission rate of species i at location $\vec{x}$ at time t.

The boundary condition at the earth's surface requires that the upward flux of each chemical species equals the ground-level emissions minus the dry deposition flux:

$$-K_{zz} \frac{\partial c_i}{\partial z} = E_i - v_g^i c_i \tag{5.21}$$

where K_{zz} is the vertical eddy diffusivity, here E_i is the ground-level emission flux of species i, and v_g^i is the deposition velocity for species i. A no-flux boundary condition is applied at the top of the modeling region.

The version of the CIT airshed model used here employs a revised dry deposition module based on surface resistance values (Russell *et al.*, 1993) and an extended version of the LCC chemical mechanism (Lurmann *et al.*, 1987). The LCC chemical mechanism

represents organic gas emissions through 8 lumped organic species classes: C4+ alkanes, ethene, C3+ alkenes, monoalkyl benzenes, di- and trialkyl benzenes, formaldehyde, C2+ aldehydes, and ketones. The LCC mechanism has been extended by Harley *et al.* (1993ab) to explicitly include the chemistry of methane, methanol, ethanol, methyl tert-butyl ether (MTBE), isoprene, hydrogen peroxide, and sulfur dioxide. This extended mechanism contains 35 differential species, 10 steady-state species, and 107 chemical reactions.

Extensive testing of the CIT model has been performed using experimental data from the August 1987 episode of the Southern California Air Quality Study (SCAQS; Harley *et al.*, 1993ab). The 24–25 June 1987 SCAQS episode has also been simulated using the CIT model (McNair *et al.*, 1996).

A baseline model evaluation study conducted for each day of an entire calendar year (Winner and Cass, 1998; see chapter 2) demonstrated that it is possible to successfully model the time series of O_3 concentrations over very long periods of time by automating the use of routinely measured air quality and meteorological data in Southern California. When supplied with the best estimates of pollutant emissions in the Los Angeles area and with motor vehicle emissions scaled upward to match the 1987 Van Nuys Tunnel study (Ingalls *et al.*, 1989; Harley *et al.,* 1993a) and calculated according to daily temperature variations, the model predicts 1 h average O_3 concentrations for the entire year 1987 that display normalized bias of +15% and an average station peak prediction accuracy of +14%. The region-wide peak O_3 concentration for each day is on average 7% above the observed value over the entire year. The predictions for the

concentrations of O_3 precursors show a normalized bias of +23% for total NO_2; daily data on the atmospheric concentrations of Reactive Organic Gases (ROG) are not available for most days, but when such data are available for a few days during the SCAQS episodes that year, most lumped and individual ROG species are reproduced to within a normalized bias of less than ±50% (Harley *et al.*, 1993b; Harley and Cass, 1995).

5.4 Results

5.4.1 Synoptic climatological classification

5.4.1.1 Data set and PCA

Rawinsonde data for the 21 stations west of the Rocky Mountains in the U.S. and Mexico for the period 1979–1992 were obtained from the USAF Environmental Technical Applications Center (USARETAC/GCOO). These soundings are taken two times per day (0000 and 1200 GMT). Observations of geopotential height, temperature, dewpoint temperature, and wind components at four pressure levels (800, 700, 500, 250 mb) are used. Because dewpoint is not observed at 250 mb, each station has 19 variables of interest per sounding for a total of 38 variables per station daily. Thus, the data set contains 798 variables each day.

Due to the statistical manipulations that will be performed on the data set, all missing data must be filled in. Three different methods are used to replace missing data. First, missing data are reconstructed by interpolation between nearby pressure levels in the same rawinsonde sounding if available. If that is not possible, then data are reconstructed by temporal interpolation between rawinsonde measurements taken at the

same station at the same time on days before and after the missing value. Third, any remaining missing data are replaced by spatially interpolating between the measured values that do exist using the SHEREMAP program (Willmott *et al.*, 1985).

As described in the methods section, principal component analysis next was used to reduce the size of the data set by combining weather variables that vary together from day to day. For this data set, the first six components were found to be significant through two criteria. First, all retained components must have eigenvalues greater than one. Because the input to PCA used here is the correlation matrix, the eigenvalues of the components describe the variance accounted for by that component. Second, a Scree plot is used to determine the point at which the eigenvalues level off.

5.4.1.2 Cluster analysis

The average linkage clustering procedure produces a solution containing 49 clusters of days from the family of 5110 days studied in the period 1979-1992. Retaining only clusters that contain more than 2% of the observations reduces the 49 cluster average-linkage solution to 13 clusters. These 13 clusters are used as initial seeds to the k-means clustering procedure which produces the final clusters.

The final 13 clusters correspond closely to the synoptic classes described by Davis and Walker (1992). The differences are due to use of a larger period of observation in the data set processed and the replacement of missing data by using values in the same sounding from nearby pressure levels. See Figure 5 of Davis and Walker (1992) for maps of the mean weather conditions at 0000 GMT for each class. A brief description of each

synoptic class along with the percent of days that fall in that class for the period 1979–1992 follows.

The summer monsoon synoptic class (SM, 16.7% frequency of occurrence) has the highest temperatures. This class occurs in the summer months and is accompanied by elevated dewpoints, especially over the Southwestern U.S. This moisture is carried by anticyclonic flow over the region caused by a Bermuda high.

The summer dry synoptic class (SD, 14.3%) characterizes summer conditions with lower temperatures and dewpoints than in the summer monsoon class.

The Rockies ridge synoptic class (RR, 8.2%) has a lower tropospheric ridge centered over western Colorado that is not as deep as in other synoptic classes with ridges. This spring cluster has warm temperatures caused by southwesterly flow over the western region.

The Great Basin trough synoptic class (TB, 6.8%) has a deep trough over the Great Basin region. Days in this class occur most frequently in April, May, October, and November. A strong polar jet combined with southwesterly flow carrying cold air over the central part of the region creates a strong north to south gradient in temperature and geopotential height.

The west coast ridge synoptic class (RC, 6.8%) has a strong coastal ridge. This warm, autumn class has the lowest wind speeds of all classes.

The Great Basin ridge–polar jet synoptic class (RB, 6.7%) has a blocking high over Utah and Wyoming. This strong, warm anticyclone makes this winter cluster relatively warm. The polar jet is further north than usual.

The continental high–subtropical jet synoptic class (CH, 6.5%) is similar to the Great Basin ridge–polar jet class. Winds are light except for the subtropical jet in the upper troposphere.

The zonal–strong jets synoptic class (ZBJ, 6.5%) has relatively strong polar and subtropical jets. This class occurs primarily in the winter and spring.

The zonal–strong polar jet synoptic class (ZPJ, 6.4%) has the strongest polar jet of any class and no subtropical jet. This class occurs in the winter and often produces storms over Washington and Oregon. A characteristic of this class is a large north to south gradient in both temperature and geopotential height.

The zonal–southern polar jet synoptic class (ZSPJ, 5.9%) has a strong polar jet which is located farther south than in any other class. This class occurs in the winter and spring and is associated with isolated showers over Nevada and Utah.

The strong subtropical jet–southwest flow synoptic class (STJ, 5.3%) has the highest wind speeds in the southern region of the map of Figure 5.1. The northern region of that map has extremely light upper-level winds. This cold weather class brings storms to much of the region.

The continental polar–north-northwest flow synoptic class (CP, 5.0%) has north-northwesterly winds at all levels. This wintertime, cold weather class has a strong ridge off the coast which steers polar air into the region. A west to east geopotential height gradient is characteristic of this class.

The arctic synoptic class (A, 5.0%) is the coldest class. This wintertime class has northerly winds along the coast at all levels. A ridge is present off the coast and the polar jet is weak for a wintertime class.

The monthly frequency of occurrence of days falling in these synoptic classes shows a strong seasonal dependence as displayed in Figure 5.3. Some synoptic classes exhibit strong persistence over many consecutive days, while other classes show almost no persistence.

5.4.2 Monte Carlo simulation of synoptic class

The within-state duration (in days) within each synoptic class is represented by fitting a geometric probability distribution to the frequency of persistence of consecutive days within each synoptic class. Table 5.1 shows the calculated values of p, the reciprocal of the mean duration in each synoptic class. Figure 5.4 shows two examples of the frequency distribution of the observed duration beyond the first day (t-1) within the same class. The least persistent state, class TB, is shown in Figure 5.4a. The most persistent state, class SM, is shown in Figure 5.4b.

5.4.3 Generation of meteorological fields

Figure 5.5 shows an example of the observed joint distribution of normalized T850 and ΣPG values for synoptic class SM compared to the simulated joint distribution using the method described in section 5.3.3. In that diagram, the mean of the joint normal distribution is located at (0, 0) and the integers along the horizontal axes are given in

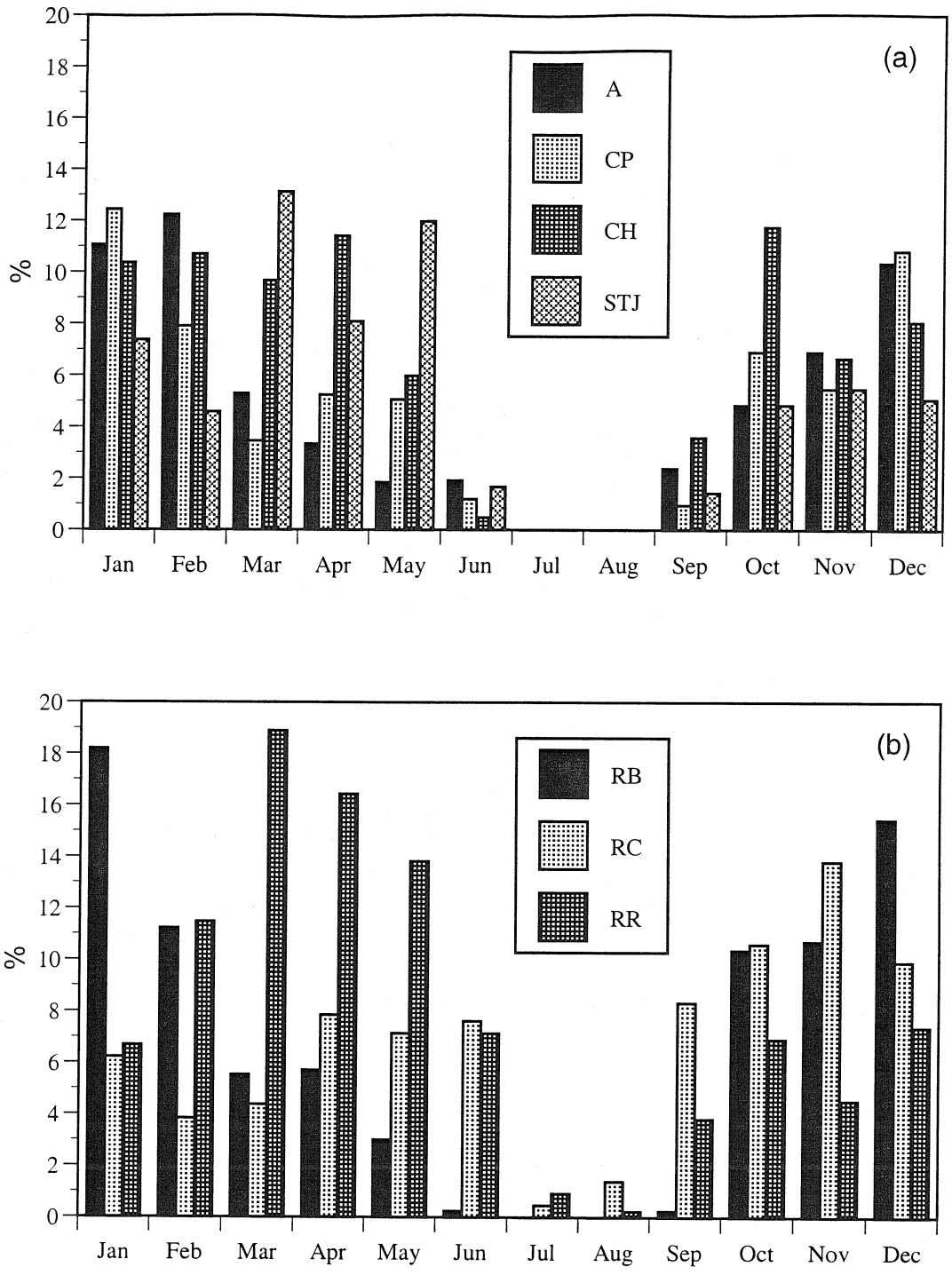

Figure 5.3ab Frequency of occurrence of days within specific synoptic weather classes for each month.

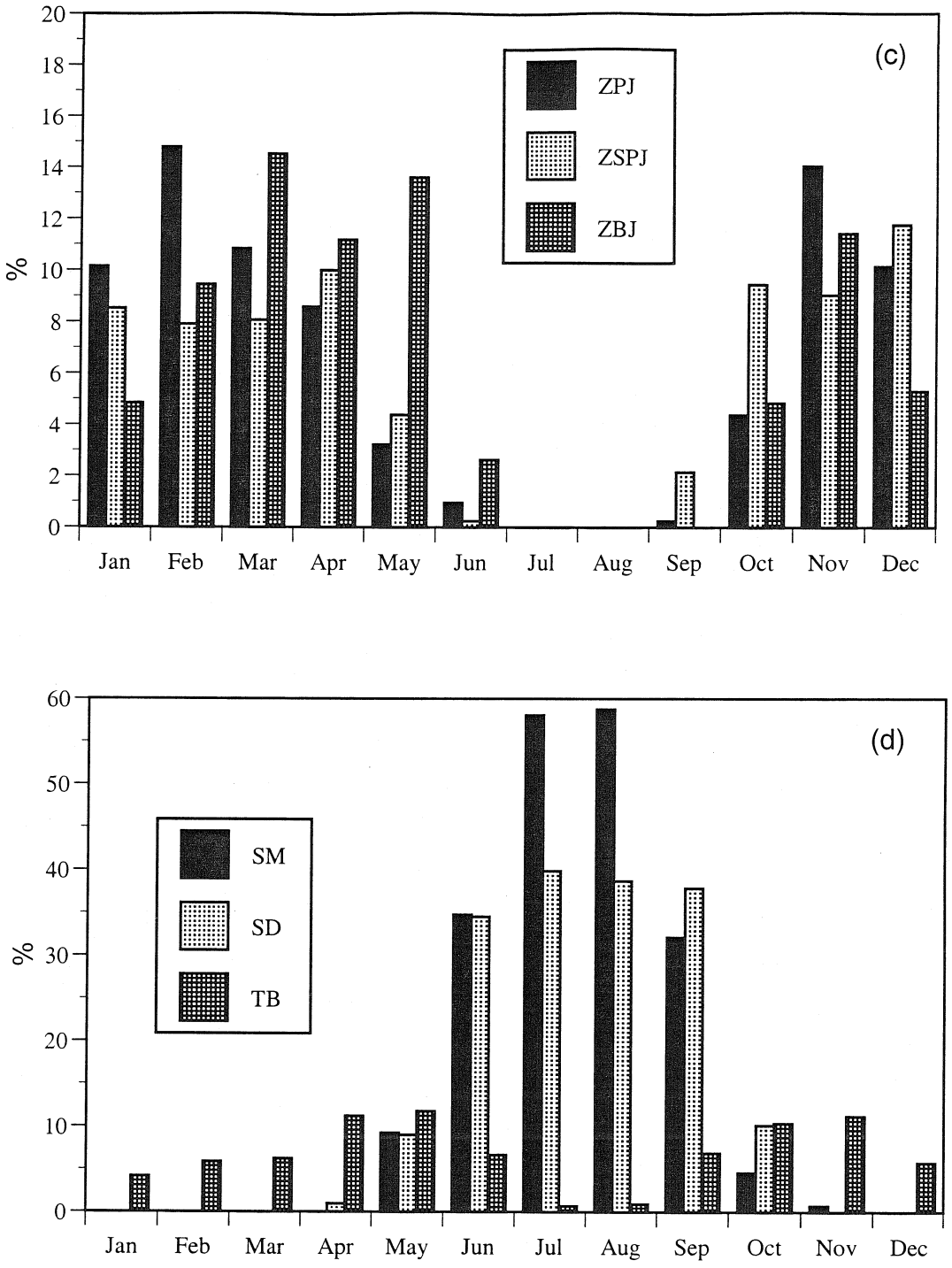

Figure 5.3cd Frequency of occurrence of days within specific synoptic weather classes for each month.

Table 5.1 Geometric distribution parameter, p, that describes the mean duration within each synoptic class.

Synoptic Class	p[(a)]
Summer monsoon	0.2856
Summer dry	0.3361
Continental high subtropical jet	0.5539
Great basin ridge polar jet	0.5627
Rockies ridge	0.5721
Arctic	0.5787
Continental polar north-northwest flow	0.5810
Zonal strong jets	0.6273
Zonal southern polar jet	0.6415
Zonal strong polar jet	0.6534
Strong subtropical jet southwest flow	0.6618
West coast ridge	0.6859
Great Basin trough	0.7320

[(a)] p is the reciprocal of the mean duration (days) of persistence of weather pattern within a single synoptic weather class before transition to a different synoptic weather class occurs.

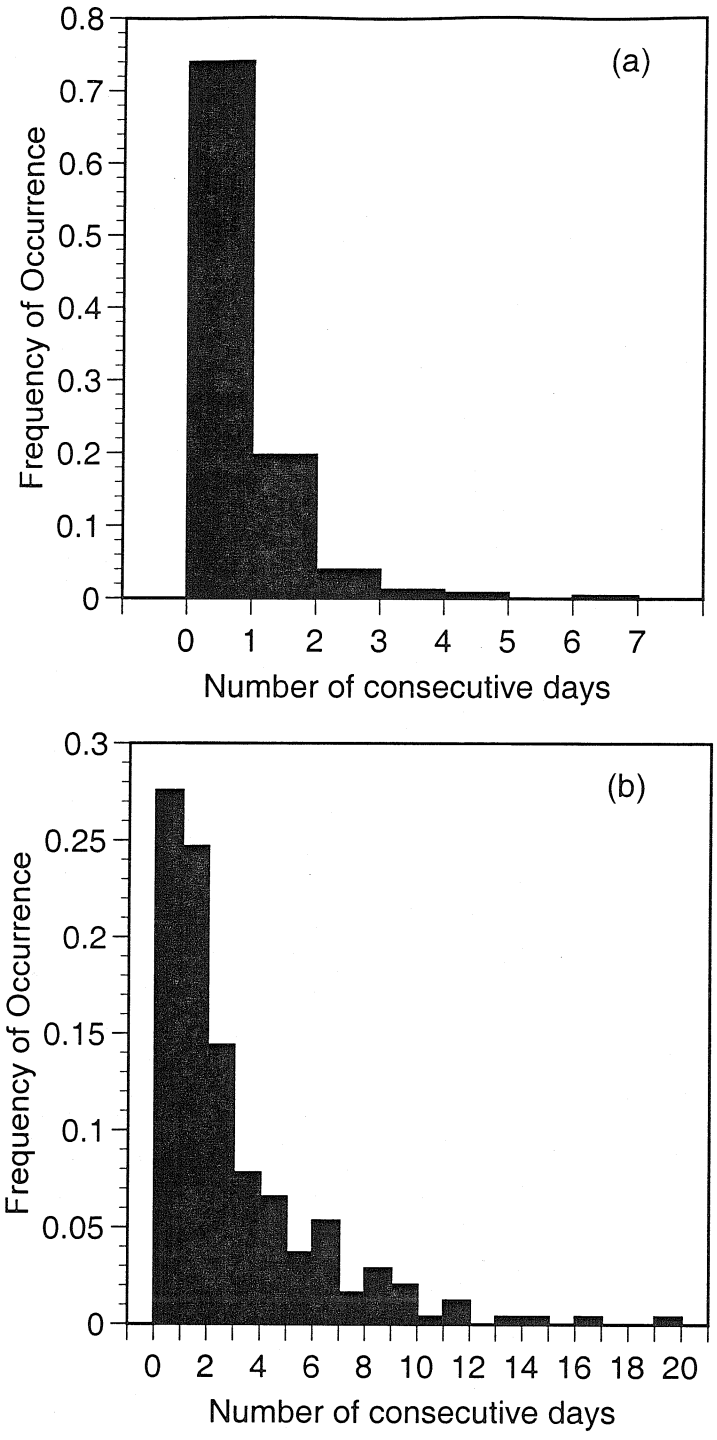

Figure 5.4 The frequency of occurrence of the duration of synoptic weather classes: (a) least persistent synoptic class TB and (b) most persistent synoptic class SM.

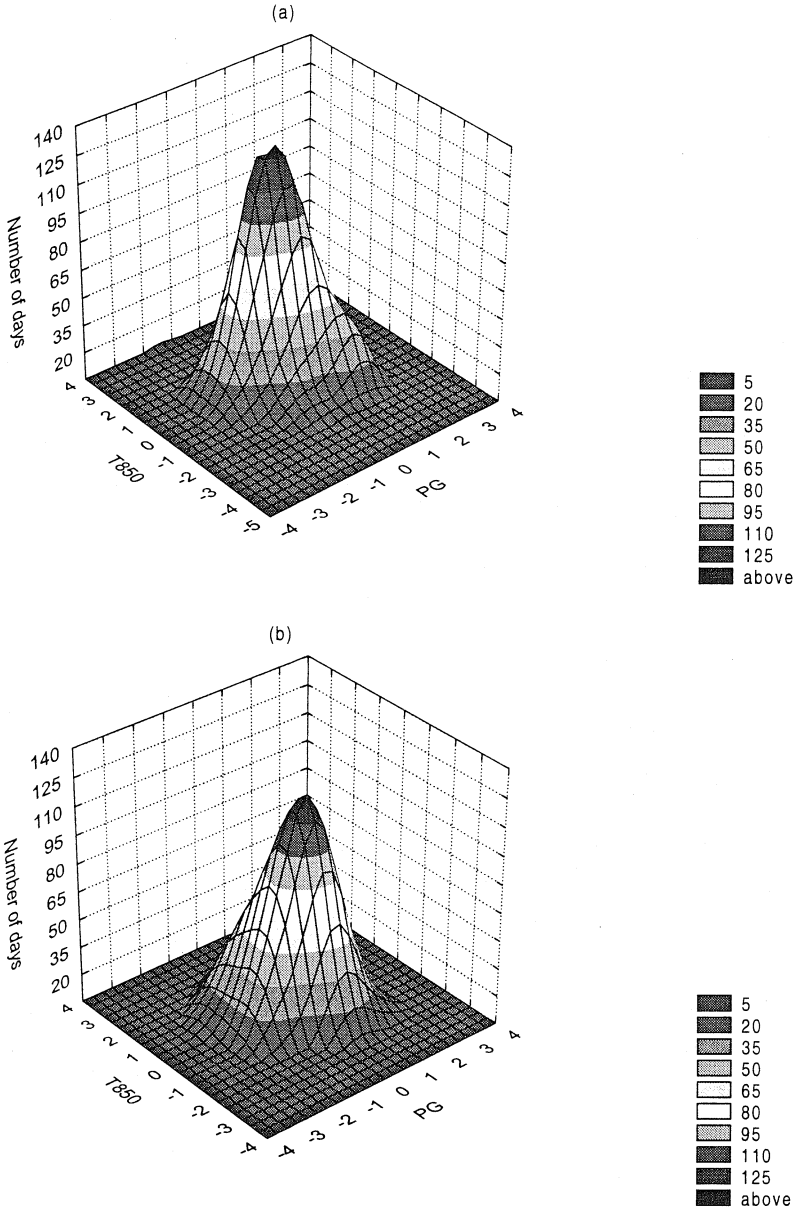

Figure 5.5 Bivariate histogram showing joint distribution of standardized T850 and ΣPG values for synoptic class SM: (a) Observed number of days in the time period 1981-1992 (b) Simulated using multivariate normal distribution. Units along the horizontal axes are standard deviations from the mean.

units of standard deviations from the mean. The actual mean value of T850 is 19.5°C with a standard deviation of 4.1°C while the mean of the ΣPG variable is 5.6 mb with a standard deviation of 5.4 mb. The full range of possible values of T850 and ΣPG next is divided into five sub-groups that contain only days falling in a narrower range of T850 and ΣPG values. Analysis of the relationship between T850, ΣPG, and early morning inversion base height data shows that low inversion base heights (thus poor mixing) are associated with lower than average values of ΣPG (i.e., stagnant conditions) and higher than average T850 values (i.e., strong temperature inversions). When the T850 and ΣPG data are subdivided into groups of days with similar characteristics, the groupings are created so that days with properties near to the mean of the joint distribution function are kept together as a group. The remaining days are grouped with higher than average T850 plus lower than average ΣPG values together. First, those days with T850 and ΣPG values falling within a radius of 0.25 standard deviation from the mean of the joint distribution are grouped together as a block of typical days. Then those days with T850 and ΣPG values falling between a radius of 0.25 and 1.0 standard deviations from the mean are divided into two groups that fall on either side of the diagonal line drawn through the distribution function of Figure 5.5 defined by normalized ΣPG equal to normalized T850. This places days with either higher than average values of T850 or lower than average values of ΣPG, or both, generally on one side of that line. The remaining two groups are defined as those falling on a common side of the diagonal line described above and outside a radius of 1 standard deviation from the mean of the joint distribution of ΣPG and T850. Within each of these 5 groups in the joint distribution of

T850 and ∑PG values within each synoptic class, a separate distribution is developed that describes the probable horizontal wind components, temperature, and humidity at 4 pm PST daily at each of the 16 monitoring sites shown in Figure 5.2. Once the synoptic class and T850 and ∑PG values are selected, then random draws from these probable horizontal wind components, temperature, and humidity distributions provide the initial values for starting the next MM5 simulation. That simulation then proceeds without any further data addition (other than the outermost grid boundary values) until the period of persistence drawn for that simulation has expired. Then the next synoptic class is drawn and the process continues.

5.4.4 Air quality simulation

Figure 5.6 shows the cumulative frequency distribution of daily peak 8 h average O_3 concentrations within the South Coast Air Basin (SoCAB) for the month of August at 1987 emissions levels. The synthetic values shown represent the frequency of occurrence of the daily maximum 8 h O_3 concentration over the SoCAB predicted by the CIT photochemical airshed model using synthetic meteorological inputs as described above along with 1987 emissions data. The observed values given are based on the daily maximum 8 h average O_3 concentrations measured at SoCAB air monitoring sites during August 1987. The deterministic values shown in Figure 5.6 relate the daily maximum 8 h average O_3 concentrations within the SoCAB as predicted during August 1987 by the CIT photochemical airshed model using deterministic meteorological inputs (Winner and Cass, 1998; this thesis chapter 2). All three frequency distributions are quite similar. The

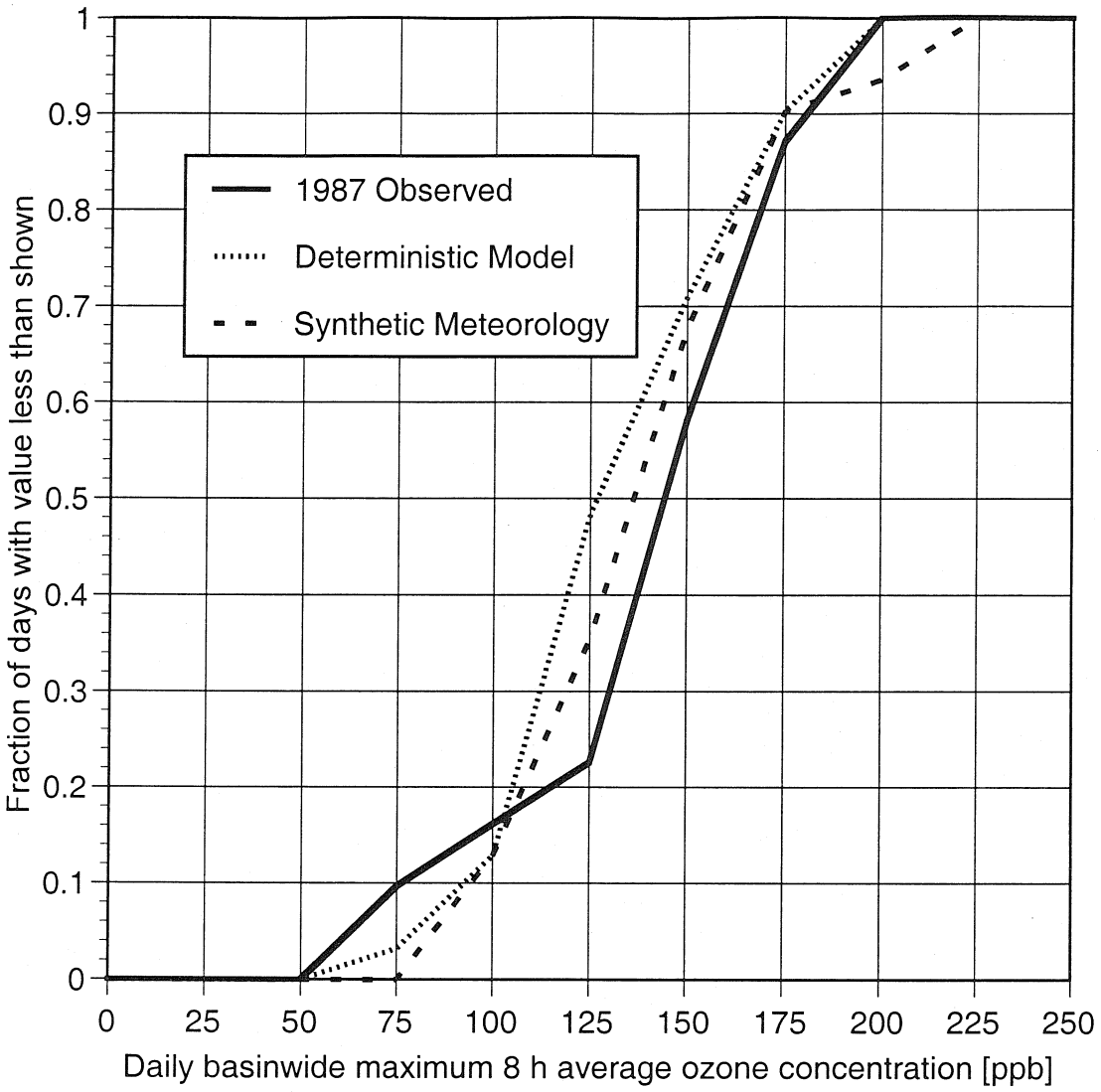

Figure 5.6 Cumulative frequency of occurrence of the daily region-wide maximum 8 h average ozone concentrations in California's South Coast Air Basin for August based on 1987 levels of air pollutant emissions.

observed and deterministic cases frequency distributions both include an event that occurred in 1987 with unusually low O_3 concentrations. Not surprisingly, the frequency distribution of peak 8 h average O_3 concentrations generated using synthetic meteorology does not include such an extreme event during this 31 day realization of the random processes. Figure 5.7 shows the same predicted distribution of the daily 8 h average basin-wide maximum ozone concentration for a generic August month in the SoCAB based on the use of synthetic meteorological data and 1987 emissions data, but this time these results are compared to the ensemble of basin-wide daily 8 h peak ozone concentrations observed during August of years surrounding 1987. Since the emissions inputs to the airshed model are based on data for the year 1987, the farther the observations are away from 1987, the larger the likely differences in emission rates. The different frequency distributions measured for each year give some indication of the influence of the variation of meteorology on air quality. It can be seen from Figure 5.7 that the frequency distribution of the daily 8 h average basin-wide peak ozone concentrations generated using synthetic meteorology falls within the ensemble of the values observed over the period 1984-1990.

5.5 Conclusions

A photochemical airshed model driven by synthetically-generated sequences of meteorological data has been used to calculate the frequency distribution of summertime peak O_3 concentrations. In this method, random draws first are made from the historically observed probability distribution of the synoptic weather patterns and two key local

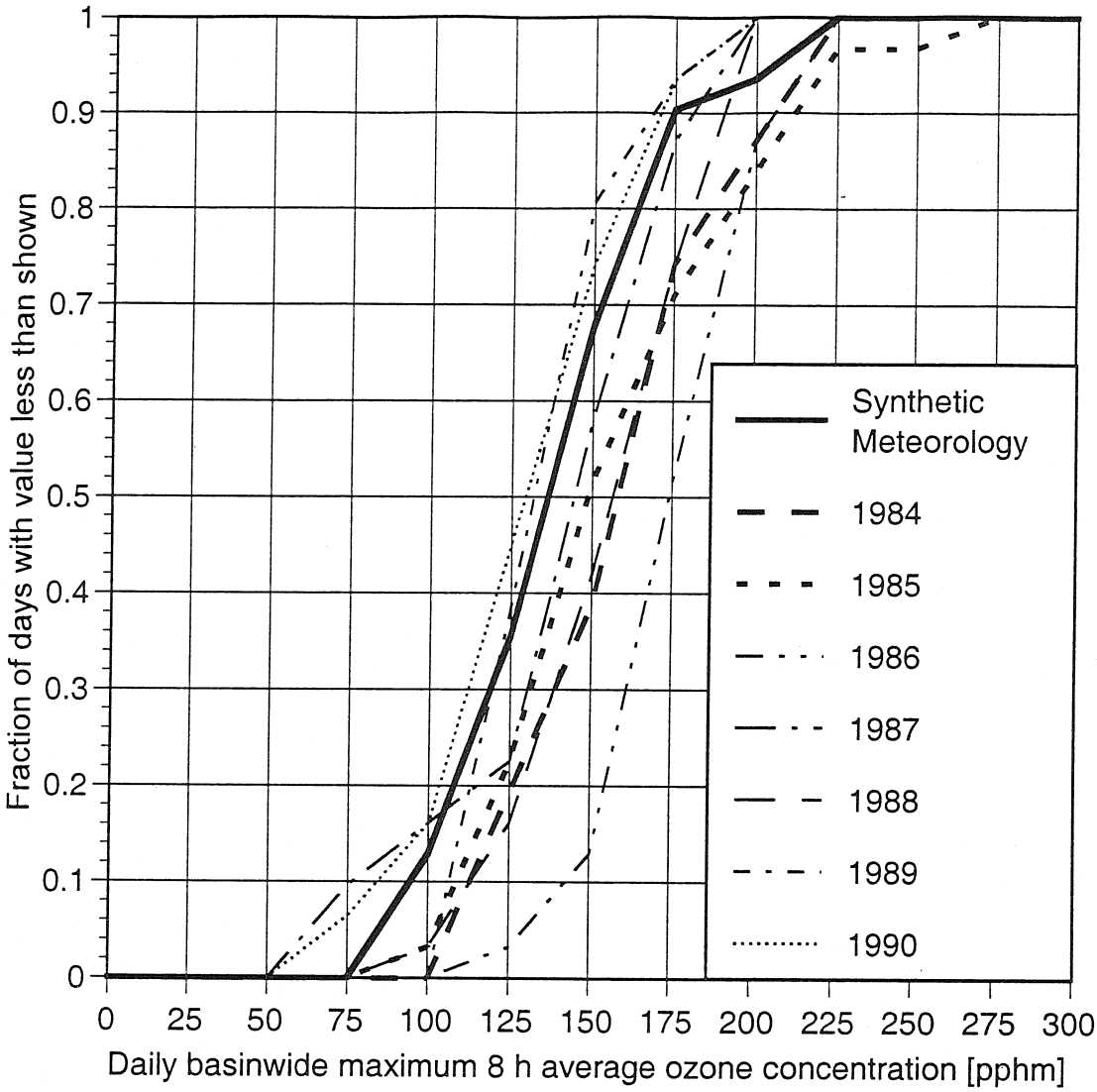

Figure 5.7 Cumulative frequency of occurrence of the daily regionwide maximum 8 h average ozone concentrations in California's South Coast Air Basin for August computed using 31 days of meteorological data and 1987 pollutant emissiondata compared to O_3 values measured during each August of the years 1984-1990.

weather variables. These parameters are utilized to generate a set of synthetic initial conditions for a prognostic mesoscale meteorological model which then proceeds to generate the detailed meteorological fields required as inputs by the CIT photochemical airshed model. The method was demonstrated by simulating the frequency distribution of the air basin-wide daily peak 8 h average O_3 concentrations in Southern California over a summertime month using emissions data from the year 1987. The use of synthetic meteorological data produces a frequency distribution of peak 8 h average O_3 concentrations that is similar to the observed values for 1987. Results also are compared to an ensemble of historically observed summertime O_3 concentration measurements made in Southern California over the years 1984-1990. Again, the results obtained by use of the synthetic meteorological data sets are within the range of the historically observed O_3 concentrations. The advantages of the use of synthetic meteorological data when calculating how the long-term frequency distribution of O_3 concentrations depends on pollutant emissions from sources is two-fold. First, the use of synthetic meteorology in principle reduces the bias that occurs when any specific historical time period is chosen to represent all conditions that can occur. Second, the generation of synthetic sequences of meteorological data by this method given the data sets archived at NCAR requires that the analyst process about 1% of the actual meteorological data otherwise necessary to conduct comparable airshed model calculations based on interpolation of all of the directly observed weather data taken in Southern California.

Chapter 6 Summary and Conclusions

6.1 Summary of results

Efficient methods were developed in chapter 2 for modeling emissions–air quality relationships that govern ozone and NO_2 concentrations over very long periods of time. A baseline model evaluation study was conducted to assess the accuracy and speed with which the relationship between pollutant emissions and the frequency distribution of O_3 concentrations throughout the year can be computed along with annual average NO_2 values using a deterministic photochemical airshed model driven by automated objective analysis of measured meteorological parameters. Methods developed were illustrated by application to the air quality situation that exists in Southern California. Model performance statistics for O_3 were similar to the results obtained in previous short-term episodic model evaluation studies that were based on hand-crafted meteorological inputs obtained by expensive field measurement campaigns. Model predictions at one of the highest NO_2 concentration sites in the U.S. indicated that measured violation of the U.S. annual average NO_2 air quality standard at that site occurred because other species such as HNO_3 and PAN were measured as if they were NO_2 by the chemiluminescent NO_x monitors in current use.

Chapter 3 illuminated the importance of assumptions made regarding boundary and initial conditions on the predicted performance of regional ozone control strategies. A computationally efficient approach to depicting the response of an air basin to emission controls was developed. The problem of ozone isopleth generation was addressed using a large Eulerian grid-based photochemical airshed model that was distributed over a grid

system of 64 different reactive organic gas (ROG) and NO_x control combination points that were run simultaneously on a parallel computer. This method was used for the Los Angeles area to examine the effect on predicted ozone concentrations of alternative assumptions about how boundary conditions at the edge of the air basin will change as a result of changed recirculation of pollutants from within the airshed plus emission decreases upwind. The results demonstrate that the method chosen to approximate the effect of emission controls on the future year boundary conditions applied to an airshed model are absolutely critical to the predictions that will be made regarding future year ozone control strategy performance. As a result of this analysis, the Southern California modeling region used for the calculations presented in chapters 2, 4, and 5 is greatly expanded when compared to the traditional South Coast Air Basin modeling region used in chapter 3. The expanded modeling region extends from nearly the Mexican border to north of Point Conception and from the Pacific Ocean to the desert east of Palm Springs, CA. The purpose of the extended modeling domain is to set the boundaries of the airshed model within the cleanest air for which air quality measurements are routinely available.

Chapter 4 combined the method of ozone control isopleth generation used in chapter 3 with the techniques developed in chapter 2 that simulated ozone concentrations throughout an entire calendar year. The response of ozone concentrations in Southern California to ROG and NO_x emissions controls over an entire calendar year was examined for both 1 h average ozone concentrations and 8 h average ozone concentrations. Model results based on 23,680 days of simulation indicated that emission controls are much more effective in reducing the frequency of occurrence of high 1 h

average ozone concentration events than in reducing lower ozone concentrations that still violate the new Federal 8 h average ozone standard. Stringent emission control efforts are predicted to reduce the frequency of violation of the former Federal 1 h average standard to approximately 25 days per year under 1987 meteorological conditions. The same model results cast very serious doubts on the feasibility of ever bringing the Los Angeles area into even approximate compliance with the new 8 h average Federal ambient air quality standard for ozone.

In chapter 5, a new method was developed to generate the meteorological input fields required for use with photochemical airshed models that seek to predict the effect of pollutant emissions on the long-term frequency distribution of peak O_3 concentrations. Instead of using meteorological fields derived from interpolation of direct weather observations as in chapters 2-4, a method that used synthetically generated meteorological data was developed. These synthetic meteorological fields were created by first constructing a semi-Markov process that generates a time series of large-scale synoptic weather conditions that statistically resemble the occurrence and persistence of synoptic weather patterns during specific months of the year. Then for each day within each synoptic weather category, local weather variables indicative of the meteorological potential for ozone formation were drawn from the approximated joint distribution of the summation of three pressure gradients across the airshed and the 850 mb temperature measured in the early morning. Synthetic initial conditions for use in a prognostic mesoscale meteorological model then were generated that conform to the chosen time series of synoptic meteorological classes and key local temperature and pressure gradient

variables combined with boundary values for that model extracted from historical days that match the chosen synoptic class, temperature, and pressure gradient values as closely as possible. The prognostic mesoscale meteorological model then generated the meteorological input fields necessary for the photochemical airshed model. The airshed model driven by synthetically generated meteorological data was executed for a 31 day period that statistically resembles weather during the month of August in Southern California using pollutant emissions data from the year 1987. The procedure produced a frequency of occurrence of peak 8 h average ozone concentrations that compared well both to that produced by the deterministic model of chapter 2 as well as to the O_3 concentrations observed over the August months of the years 1984 to 1989.

6.2 Recommendations for future research

The emission control strategy modeling technique used for ozone in chapter 4 could be combined with emerging models that simulate the atmospheric concentrations of particulate matter. The newly adopted Federal air quality standard for fine particles sets limits on annual average concentrations, and models hence are needed that are both physically realistic and that can be exercised over long periods of time as was done here for ozone formation. While the computational resources necessary exercise advanced models that predict the size and composition distribution of airborne particles over periods as long as a year would be difficult to obtain at present, a simplified model that seeks to model fine particle chemical composition but not the detailed particle size distribution might still yield useful results. Future research might utilize the techniques

developed here and demonstrated for Southern California to address air quality problems that exist in other parts of the United States as well as in other parts of the world. Of special interest would be a study of the response of ozone concentrations in the Eastern U.S. to emission controls, and the relationship of emission controls to the attainment of the new 8 h average Federal ozone air quality standard.

Another area that is ripe for further study is the impact of large-scale climate anomalies, such as the El Nino-Southern Oscillation, on regional air quality. Recent years in Southern California have had the lowest ozone concentrations on record. This could be due in large part to the lack of meteorological episodes conducive to the formation of high ozone levels during those recent years when there have been strong local weather effects induced by El Nino conditions. The synoptic weather and local key variable classification techniques developed in chapter 5 combined with modeling efforts targeted at these recent years might shed some light on the role of El Nino in clearing the skies in Los Angeles.

Further research into refining the emission inventories that are available to support ozone formation calculations is warranted. One area that deserves investigation is the comparison of the effect of temperature on motor vehicle emissions as predicted by the EMFAC7F model to observed emissions in real-world settings. In the present work, a scale factor of 2.25 was used to correct the motor vehicle hot-exhaust emissions predicted by EMFAC7F to match measurements made in the Van Nuys tunnel in August 1987. Whether or not that scale factor is constant over the entire year should be determined by a series of tunnel experiments conducted in different seasons.

References

Amdahl G.M. (1967) Validity of the single-processor approach to achieving large scale computing capabilities. In *AFIPS Conference Proceedings*, vol. 30, pp. 483–485. AFIPS Press, Reston, VA.

Anthes R.A. and Warner T.T. (1978) Development of hydrodynamic models suitable for air pollution and other mesometeorological studies. *Mon. Wea. Rev.*, **106**, 1945-1078.

Bardossy A. and Plate E.J. (1991) Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. *J. Hydrol.*, **122**, 33–47.

Barry R.G. and Perry A.H. (1973) *Synoptic climatology: Methods and applications* Methuen, London.

Beckett W.S. (1991) Ozone, air pollution and respiratory health. *The Yale Journal of Biology and Medicine* **64**, 167–175.

Benjamin M.T., Sudol M., Vorsatz D., Winer A.M. (1997) A spatially and temporally resolved biogenic hydrocarbon emissions inventory for the California South Coast Air Basin. *Atmos. Environ.* **31**, 3087–3100.

Benjamin S.G. and Seaman N.L. (1985) A simple scheme for objective analysis in curved flow. *Mon. Wea. Rev.*, **113**, 1184-1198.

Benkovitz C.M., Berkowitz C.M., Easter R.C., Nemesure S., Wagener R., Schwartz S.E. (1994) Sulfate over the North Atlantic and adjacent continental regions: Evaluation for October and November 1986 using a three-dimensional model driven by observation-derived meteorology. *J geophys. Res.* **99**, 20725–20756.

Binkowski F.S. and Shankar U. (1995) The regional particulate matter model 1. Model description and preliminary results. *J. geophys. Res.* **100**, 26191–26209.

Bishop G.A., Starkey J.R., Ihlenfeldt A., Williams W.J., Stedman D.H. (1989) IR long-path photometry, a remote sensing tool for automotive emissions. *Anal. Chem.* **61**, 671A–677A.

Bishop G.A. and Stedman D.A. (1990) On-road carbon monoxide emission measurement comparisons for the 1988–1989 Colorado Oxy-Fuels Program. *Environ. Sci. Technol.* **24**, 843–847.

Box G.E.P. and Jenkins G.M. (1976) *Times series analysis: Forecasting and control* Holden-Day, San Francisco.

Box G.E.P. and Muller M.E. (1958) A note on the generation of random normal deviates. *The Annuls of Mathematical Statistics*, **29**, 610-611.

California Air Resources Board (1987) California air quality data: Summary of 1987 air quality data: Gaseous and particulate pollutants. Vol. XIX, Sacramento, CA.

California Air Resources Board (1990) Emission inventory 1987, Sacramento, CA.

California Air Resources Board (1990b) Technical guidance document: Photochemical modeling. Sacramento, CA.

California Department of Transportation (1994) Direct Travel Impact Model 2 (DTIM2). Sacramento, CA.

Carmichael G.R., Cohen D.M., Cho S.-Y., Oguztuzun M.H. (1989) Coupled transport/chemistry calculations on the massively parallel processor computer. *Computers chem. Engng.* **13**, 1065–1073.

Cass G.R. and McRae G.J. (1981) Minimizing the cost of air pollution control. *Environ. Sci. Technol.* **15**, 748–757.

Cassmassi J.C. and Durkee K.R. (1991) Comparison of mixing height fields for UAM application generated from limited and multiple temperature sounding profiles in the South Coast Air Basin. In *Tropospheric Ozone and the Environment–Papers from an International Conference* (edited by Berglund R.L., Lawson D.R., McKee D.J.), Air & Waste Management Association, Pittsburgh, PA.

Cassmassi J.C. and Bassett M.E. (1993) Air quality trends in the South Coast Air Basin. In *Southern California Air Quality Study Data Analysis*, Air & Waste Management Association, Pittsburgh, PA.

Chico T., Bassett M.E., Cassmassi J.C., Mitsutomi S., Zhang X., Hogo H. (1993) Application of the Urban Airshed Model for a SCAQS episode in the South Coast Air Basin. In *Southern California Air Quality Study Data Analysis*, Air & Waste Management Association, Pittsburgh, PA.

Chock D.P., Kumar S., Herrmann R.W. (1982) An analysis of trends in oxidant air quality in the South Coast Air Basin of California. *Atmos. Environ.* **16**, 2615–2624.

Dabdub D. and Seinfeld J.H. (1994) Air quality modeling on massively parallel computers. *Atmos. Environ.* **28**, 1679–1687.

Danielson E.F. (1968) Stratosphere-troposphere exchange based on radioactivity, ozone, and potential vorticity. *Journal of the Atmospheric Sciences* **25**, 502–518.

Davidson A. (1986) Comment on ten-year ozone trends in California and Texas. *Journal of the Air Pollution Control Association* **36**, 597–598.

Davidson A. (1993) Update on ozone trends in California's South Coast Air Basin. *Air & Waste* **43**, 226–227.

Davidson A. (1994) The Los Angeles Aerosol Characterization and Source Apportionment Study: A meteorological-air quality analysis. *Aerosol Sci. & Technol.*, **21**, 269-282.

Davis R.E. and Kalkstein L.S. (1990) Development of an automated spatial synoptic climatological classification. *International Journal of Climatology* **10**, 769–794.

Davis R.E. and Walker D.R. (1992) An upper-air synoptic climatology of the Western United States. *Journal of Climate* **5**, 1449–1467.

Davis R.E. and Gay D.A. (1993) A synoptic climatological analysis of air quality in the Grand Canyon National Park. *Atmos. Environ.* **27A**, 713–727.

Dillon W.R. and Goldstein M. (1984) *Multivariate Analysis: Methods and Applications* John Wiley & Sons, New York, NY.

Dodge M.C. (1977) Combined use of modeling techniques and smog chamber data to derive ozone-precursor relationships. In *Proceedings of the International Conference on Photochemical Oxidant Pollution and Its Control,* Vol. II. (edited by Dimitriades B.) EPA-600/3-77-001b, pp. 881–889.

Dudhia J. (1993) A nonhydrostatic version of the Penn State-NCAR Mesosclate Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493-1513.

Environmental Protection Agency (1997) Final Report to Congress on Benefits and Costs of the Clean Air Act, 1970 to 1990. EPA 410-R-97-002.

Falls A.H. and Seinfeld J.H. (1978) Continued development of a kinetic mechanism for photochemical smog. *Environ. Sci. Technol.***12**, 1398–1406.

Falls A.H., McRae G.J., Seinfeld J.H. (1979) Sensitivity and uncertainty of reaction mechanisms for photochemical air pollution. *International Journal of Chemical Kinetics* **XI**, 1137–1162.

Fairly D. (1993) Photochemical model bias: Is it real or is it a statistical artifact? *Air & Waste* **43**, 348–351.

Feller W. (1968) *An Introduction to Probability Theory and Its Applications, Volume I* John Wiley & Sons, New York, NY.

Fiering M.B. (1967) *Streamflow Synthesis* Harvard University Press, Cambridge, MA.

Fishman J., Fakhruzzaman K., Cros B., Nganga D. (1991) Identification of widespread pollution in the Southern Hemisphere deduced from satellite analysis. *Science* **252**, 1693–1696.

Fujita E.M., Croes B.E., Bennett C.L., Lawson D.R., Lurmann F.W., Main H.H. (1992) Comparison of emission inventory and ambient concentration ratios of CO, NMOG, and NOx in California's South Coast Air Basin. *J. Air Waste Manage. Assoc.* **42**, 264–276.

Goodin W.R., McRae G.J., Seinfeld J.H. (1979) A comparison of interpolation methods for sparse data: Application to wind and concentration fields *J. Appl. Meteor.* **18**, 761–771.

Goodin W.R., McRae G.J., Seinfeld J.H. (1980) An objective analysis technique for constructing three-dimensional urban-scale wind fields *J. Appl. Meteor.* **19**, 98–108.

Gray H.A., Cass G.R., Huntzicker J.J., Heyerdahl E.K., Rau J.A. (1986) Characteristics of atmospheric organic and elemental carbon particle concentrations in Los Angeles. *Environ. Sci. Technol.* **20**, 580–589.

Haagen-Smit A.J. (1952) Chemistry and physiology of Los Angeles smog. *Indust. Eng. Chem.* **44**, 1342–1346.

Haagen-Smit A.J. and Fox M.M. (1954) Photochemical ozone formation with hydrocarbons and automobile exhaust. *J. Air Pollut. Control Assoc.* **4**, 105–109.

Hall J.V., Winer A.M., Kleinman M.T., Lurmann F.W., Brajer V., Colome S.D. (1992) Valuing the health benefits of clean air. *Science* **255**, 812–817.

Harley R.A. and Cass G.R. (1994) Modeling the concentrations of gas-phase toxic air pollutants: Direct emissions and atmospheric formation. *Environ. Sci. Technol.*, **28**, 88–98.

Harley R.A. and Cass G.R. (1995) Modeling the atmospheric concentrations of individual volatile organic compounds. *Atmos. Environ.* **29**, 905–922.

Harley R.A., Hannigan M.P., Cass G.R. (1992a) Respeciation of organic gas emissions and the detection of excess unburned gasoline in the atmosphere. *Environ. Sci. Technol.* **26**, 2395–2408.

Harley R.A., Russell A.G., McRae G.J., McNair L.A., Winner D.A., Odman M.T., Dabdub D., Cass G.R., Seinfeld J.H. (1992b) Continued development of a photochemical model and application to the Southern California Air Quality Study (SCAQS) intensive monitoring periods: Phase I. Report to the Coordinating Research Council. Carnegie Mellon University, Pittsburgh, PA, and California Institute of Technology, Pasadena, CA.

Harley R.A., Russell A.G., McRae G.J., Cass G.R., Seinfeld J.H. (1993a) Photochemical modeling of the Southern California Air Quality Study. *Environ. Sci. Technol.* **27**, 378–388.

Harley R.A., Russell A.G., Cass G.R. (1993b) Mathematical modeling of the concentrations of volatile organic compounds: Model performance using a lumped chemical mechanism. *Environ. Sci. Technol.* **27**, 1638–1649.

Hay L.E., McCabe G.J. Jr., Wolock D.M., Ayers, M.A. (1991) Simulation of Precipitation by weather type analysis. *Water Resour. Res.*, **27**, 493–501.

Heck W.W., Taylor O.C., Adams R., Bingham G., Miller J., Preston E., Weinstein L. (1982) Assessment of crop loss from ozone. *J. Air Pollut. Control Assoc.* **32**, 353–361.

Holtslag A.A.M. and Ulden A.P.V (1983) A simple scheme for daytime estimates of the surface fluxes from routine weather data. *J. Clim. Appl. Meteorol.*, **22**, 517–529.

Holzworth G.C. (1962) A study of air pollution potential for the western United States. *J. Appl. Meteor.* **1**, 366–382.

Holzworth G.C. (1967) Mixing depths, wind speeds and air pollution potential for selected locations in the United States. *J. Appl. Meteor.* **6**, 1039–1044.

Ingalls M.N., Smith L.R., Kirksey R.E. (1989) Measurement of on-road vehicle emission factors in the California South Coast Air Basin. Volume I: Regulated emissions. NTIS Document PB89220925, Report SwRI-1604 for the Coordinating Research Council, Atlanta, GA.

Jackson B.B. (1975) The use of streamflow models in planning. *Water Resources Research* **11**, 54–95.

Jackson B. (1995) Personal communication, California Air Resources Board, Sacramento, CA.

Janach W.E. (1989) Surface ozone: Trend details, seasonal variations, and interpretation. *J. geophys. Res.* **94**, 18289–18295.

Jensen A.R. (1976) Computer simulation of surface water hydrology and salinity with an application to studies of Colorado River management. PhD Thesis, Caltech, Pasadena, CA.

Junge C.E. (1962) Global ozone budget and exchange between stratosphere and troposphere. *Tellus* **14**, 363–377.

Kalkstein L.S., Tan G., Skindlov J.A. (1987) An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.*, **26**, 717–730.

Kumar S. and Chock D.P. (1984) An update on oxidant trends in the South Coast Air Basin of California. *Atmos. Environ.* **18**, 2131–2134.

Kuntasal G. and Chang T.Y. (1987) Trends and relationships of O_3, NO_X, and HC in the South Coast Air Basin of California. *J. Air Pollut. Control Assoc.* **37**, 1158–1163.

Lamb R.G. and Seinfeld J.H. (1973) Mathematical modeling of urban air pollution: General theory. *Environ. Sci. Technol.* **7**, 253–261.

Lawson D.R. (1990) The Southern California Air Quality Study. *J. Air Waste Manage. Assoc.* **40**, 156–165.

Lawson D.R., Groblicki P.J., Stedman D.H., Bishop G.A., Guenther P.L. (1990) Emissions from in-use motor vehicles in Los Angeles: a pilot study of remote sensing and the inspection and maintenance program. *J. Air Waste Manage. Assoc.* **40**, 1096–1105.

Lioy P.J. and Dyba R.V. (1989) Tropospheric ozone: The dynamics of human exposure. *Toxicology and Industrial Health.* **5**, 493-504.

Lippmann M. (1989) Health effects of ozone: A critical review. *J. Air Waste Manage. Assoc.* **39**, 672–695.

Logan J.A. (1985) Tropospheric ozone: Seasonal behavior, trends, and anthropogenic influence. *J. geophys. Res.* **90**, 10463–10482.

Lurmann F.W., Carter W.P.L., Coyner L.A. (1987) A surrogate species chemical reaction mechanism for urban-scale air quality simulation models. Volumes I and II. EPA Report Contract 68-02-4104. ERT, Inc., Newbury Park, and Statewide Air Pollution Research Center, University of California, Riverside, CA.

Lu R. and Turco R.P. (1995) Air pollutant transport in a coastal environment 2. Three-Dimensional simulations over Los Angeles basin. *Atmos. Environ.* **29**, 1499–1518.

Lyons W.A., Tremback C.J., Pielke R.A. (1995) Applications of the Regional Atmospheric Modeling System (RAMS) to provide input to photochemical grid models for the Lake Michigan Ozone Study (LMOS). *J. Appl. Meteor.* **34**, 1762–1786.

Mahlman J.D. and Moxim W.J. (1978) Tracer simulation using a global general circulation model: Results from a mid-latitude instantaneous source experiment. *Journal of the Atmospheric Sciences* **35**, 1340–1374.

Main H.H., Lurmann F.W., Roberts P.T. (1990) Pollutant concentrations along the western boundary of the South Coast Air Basin Part I: A review of existing data. Report STI-90080-1016FR, prepared for the South Coast Air Quality Management District by Sonoma Technology Inc., Santa Rosa, CA.

Main H.H., Roberts P.T., Anderson J.A., Fung K. (1991) Pollutant concentrations along the western boundary of the South Coast Air Basin Part II: Analysis of June 1990 data collected offshore. Report STI-90080-1149FR, prepared for the South Coast Air Quality Management District by Sonoma Technology Inc., Santa Rosa, CA.

McNair L.A., Harley R.A., Russell A.G., Cass G.R., Odman M.T., McRae G.J., Seinfeld J.H., Winner D.A., Dabdub D. (1992a) Continued development of a photochemical model and application to the Southern California Air Quality Study (SCAQS) intensive monitoring periods: Phase II. Report to the Coordinating Research Council. Carnegie Mellon University, Pittsburgh, PA, and California Institute of Technology, Pasadena, CA.

McNair L.A., Russell A.G., Odman M.T. (1992b) Airshed calculation of the sensitivity of pollutant formation to organic compound classes and oxygenates associated with alternative fuels. *J. Air Waste Manage. Assoc.* **42**, 174–178.

McNair L.A., Harley R.A., Russell A.G. (1996) Spatial inhomogeneity in pollutant concentrations, and their implications for air quality model evaluation. *Atmos. Environ.* **30**, 4291–4301.

McRae G.J., Goodin W.R., Seinfeld J.H. (1982a) Numerical solution of the atmospheric diffusion equation for chemically reacting flows. *Journal of Computational Physics* **45**, 1–42.

McRae G.J., Goodin W.R., Seinfeld J.H. (1982) Development of a second-generation mathematical model for urban air pollution-I. Model formation. *Atmos. Environ.* **16**, 679–696.

McRae G.J. and Seinfeld J.H. (1982) Development of a second-generation mathematical model for urban air pollution-II. Evaluation of model performance. *Atmos. Environ.* **16**, 679–696.

McRae G.J., Milford J.B., Slompak B.J. (1988) Changing roles for supercomputing in chemical engineering. *Internat. J. Supercomputer Applications* **2**, 16–40.

Milford J.B., Russell A.G., McRae G.J. (1989) A new approach to photochemical pollution control: Implications of spatial patterns in pollutant responses to reductions in nitrogen oxides and reactive organic gas emissions. *Environ. Sci. Technol.* **23**, 1290–1301.

Moore G.E., Killus J.P., Whitten G.Z. (1991) Composition of marine air offshore of the western United States. *J. appl. Met.* **30**, 707–713.

National Research Council (1991) *Rethinking the ozone problem in urban and regional air pollution.* National Academy Press, Washington, D.C.

National Research Council (1992) *Coastal meteorology: A review of the state of the science.* National Academy Press, Washington, D.C.

Niemeyer L.E. (1960) Forecasting air pollution potential. *Mon. Weath. Rev.* **88**, 88–96.

Parrish D.D., Holloway J.S., Trainer M., Murphy P.C., Forbes G.L., Fehsenfled F.C. (1993) Export of North American ozone pollution to the North Atlantic Ocean. *Science* **259**, 1436–1439.

Pierson W.R., Gertler A.W., Bradow R.L. (1990) Comparison of the SCAQS tunnel study with other on-road vehicle emission data. *J. Air Waste Manage. Assoc.* **40**, 1495–1504.

Pielke R.A. (1981) Mesoscale numerical modeling in *Advances in Geophysics*, **23**, pp. 185–344, Academic Press, New York.

Rao S.T., Sistla G., Paagnotti V., Peteren W.B., Irwin J.S. (1985) Resampling of extreme value statistics in air quality model performance evaluation. *Atmos. Environ.* **19**, 1503–1518.

Reynolds S.D., Roth P.M., Seinfeld J.H. (1973) Mathematical modeling of photochemical air pollution. I. Formulation of model. *Atmos. Environ.* **7**, 1033–1061.

Reynolds S.D., Liu M., Hecht T.A., Roth P.M., Seinfeld J.H. (1974) Mathematical modeling of photochemical air pollution. III. Evaluation of model. *Atmos. Environ.* **8**, 563–596.

Reynolds S.D. and Seinfeld J.H. (1975) Interim evaluation of strategies for meeting ambient air quality standards for photochemical oxidant. *Environ. Sci. Technol.* **9**, 433–447.

Rombout P.J.A., Lioy P.J., Goldstein B.D. (1986) Rationale for an eight-hour ozone standard. *J. Air Pollution Control Assoc.* **36**, 913-917.

Ross S.M. (1987) *Introduction to probability and statistics for engineers and scientists*, John Wiley & Sons, New York.

Roth P.M., Roberts P.J., Liu M., Reynolds S.D., Seinfeld J.H. (1974) Mathematical modeling of photochemical air pollution. II. A model and inventory of pollutant emissions. *Atmos. Environ.* **8**, 97–130.

Roth P.M., Reynolds S.D., Tesche T.W., Gutfreund P.D., Seigneur C. (1983) An appraisal of emissions control requirements in the California South Coast Air Basin. *Environ. Int.* **9**, 549–571.

Russell A.G., McCue K.F., Cass G.R. (1988) Mathematical modeling of the formation of nitrogen-containing air-pollutants 1. Evaluation of an Eulerian photochemical model. *Environ. Sci. Technol.* **22**, 263–270.

Russell A.G., McCue K.F., Cass G.R. (1988) Mathematical modeling of the formation of nitrogen-containing pollutants. 2. Evaluation of the effect of emission controls. *Environ. Sci. Technol.* **22**, 1336–1347.

Russell A.G., St. Pierre D., Milford J.B. (1990) Ozone control and methanol fuel use. *Science* **247**, 201–205.

Russell A.G., Winner D.A., Harley R.A., McCue K.F., Cass G.R. (1993) Mathematical modeling and control of the dry deposition flux of nitrogen-containing air pollutants. *Environ. Sci. Technol.* **27**, 2772–2782.

Russell A.G. and Dennis R. (1998) NARSTO critical review of photochemical models and modeling. *Atmos. Environ.*, submitted for publication.

Saylor R.D. and Fernandes R.I. (1993) On the parallelization of a comprehensive regional-scale air quality model. *Atmos. Environ.* **27A**, 625–631.

Seaman N.L., Stauffer D.R., Lario-gibbs A.M. (1995) A multiscale four-dimensional data assimilation system applied in the San Joaquin valley during SARMAP 1. Modeling design and basic performance characteristics. *J. Appl. Meteor.* **34**, 1739–1761.

Seinfeld J.H. (1988) Ozone air quality models: A critical review. *J. Air Pollut. Control Assoc.* **38**, 616–645.

Seigneur C. (1995) The status of mesoscale air quality models. in *Planning and Managing Regional Air Quality Modeling and Measurement Studies* (P.A. Solomon ed.), Lewis Publishers, Boca Raton, FL.

Shin W.-C. and Carmichael G.R. (1992) Comprehensive air pollution modeling on a multiprocessor system. *Computers chem. Engng.* **16**, 805–815.

Skarby L. and Sellden G. (1984) The effort of ozone on crops and forests. *Ambio* **13**, 68–72.

South Coast Air Quality Management District (1990) 1991 Air Quality Management Plan, Technical Report V-B, Ozone Modeling–Performance Evaluation, South Coast Air Quality Management District, Diamond Bar, CA.

South Coast Air Quality Management District (1996) Final 1997 Air Quality Management Plan. Diamond Bar, California.

Tesche T.W., Seigneur C., Oliver W.R., Haney J.L. (1984) Modeling ozone control strategies in Los Angeles *J. Envir. Engng. Div., ASCE* **110**, 208–225.

Tesche T.W., Georgopoulos P., Seinfeld J.H., Cass G.R., Lurmann F.W., Roth P.M. (1990) Improvement of procedures for evaluating photochemical models. Radian Corporation, Sacramento, CA. Final report to the California Air Resources Board under Contract A832–103.

Tesche T.W. and McNally D.A. (1991) Three-dimensional photochemical-aerosol model for episodic and long-term simulation: Formation and initial application in the Los Angeles Basin. In *Atmospheric Chemistry: Models and Predictions for Climate and Air Quality* (edited by Sloane C.S., Tesche T.W.), pp. 63–82, Lewis, Chelsea, MI.

Tesche T.W. and McNally D.E. (1991) Photochemical modeling of 2 1984 SCCCAMP ozone episodes. *J. Appl. Meteor.* **30,** 745–763.

Thomas H.A. Jr. and Fiering M.B. (1963) Operations research in water management. Report of the Harvard Water Group to the USPHS, Harvard University Press, Cambridge, MA.

Trijonis J.C. (1974) Economic air pollution control model for Los Angeles County in 1975: General least cost-air quality model. *Environ. Sci. Technol.* **8**, 811–826.

Warner T.T. (1989) Mesoscale atmospheric modeling in *Earth-Science Reviews*, **26**, pp. 221–251, Elsevier, Amsterdam.

Watson C.E., Fishman J., Reichle H.G. Jr. (1990) The significance of biomass burning as a source of carbon monoxide and ozone in the Southern Hemisphere tropics: A satellite analysis. *J. geophys. Res.* **95**, 16443–16450.

Willmott C.J., Rowe M., Philpot W.P. (1985) Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid point interpolation and contouring. *Amer. Cartographer*, **12**, 5–16.

Wilson L.L., Lettenmaier D.P., Skyllingstad E. (1992) A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *J. geophys. Res.*, **97**, 2791–2809.

Winner D.A., Cass G.R., Harley R.A. (1995) The effect of alternative boundary conditions on predicted ozone control strategy performance: A case study in the Los Angeles area. *Atmos. Environ.* **29**, 3451–3464.

Winner D.A. and Cass G.R. (1998) Modeling the long-term frequency distribution of regional ozone concentrations. *Atmos. Environ.* accepted for publication.

Zeldin M.D., Bregman L.D., Wylie B. (1988) Meteorological conditions during the 1985 Southern California nitrogen species methods comparison study. *Atmos. Environ.*, **22**, 1541-1543.

Zhang D.L. and Anthes R.A. (1982) A high-resolution model of the planetary boundary layer-Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594-1609.