Treatment of Uncertainties in Atmospheric Chemical Systems: A Combined Modeling and Experimental Approach

by

Betty Kong-Ling Pun

B.S. Chemical Engineering, California Institute of Technology (1992)
M.S.C.E.P. Massachusetts Institute of Technology (1994)

Submitted to the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology

February, 1998

Signature of Author: _____
Department of Chemical Engineering
December 5, 1997

Certified by: _____
Gregory J. McRae
Bayer Professor of Chemical Engineering
Thesis Supervisor

Accepted by: _____
Robert E. Cohen
St. Laurent Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Treatment of Uncertainties in Atmospheric Chemical Systems: A Combined Modeling and Experimental Approach

by

Betty Kong-Ling Pun

Submitted to the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy at the
Massachusetts Institute of Technology

Abstract

Uncertainty is endemic in modeling. This thesis is a two-phase program to understand the uncertainties in urban air pollution model predictions and in field data used to validate them. Part I demonstrates how to improve atmospheric models by analyzing the uncertainties in these models and using the results to guide new experimentation endeavors. Part II presents an experiment designed to characterize atmospheric fluctuations, which have significant implications towards the model validation process.

A systematic study was undertaken to investigate the effects of uncertainties in the SAPRC mechanism for gas-phase chemistry in polluted atmospheres. The uncertainties of more than 500 parameters were compiled, including reaction rate constants, product coefficients, organic composition, and initial conditions. Uncertainty propagation using the Deterministic Equivalent Modeling Method (DEMM) revealed that the uncertainties in ozone predictions can be up to 45% based on these parametric uncertainties. The key parameters found to dominate the uncertainties of the predictions include photolysis rates of NO_2, O_3, and formaldehyde; the rate constant for nitric acid formation; and initial amounts of NO_x and VOC. Similar uncertainty analysis procedures applied to two other mechanisms used in regional air quality models led to the conclusion that in the presence of parametric uncertainties, the mechanisms cannot be discriminated. Research efforts should focus on reducing parametric uncertainties in photolysis rates, reaction rate constants, and source terms.

A new tunable diode laser (TDL) infrared spectrometer was designed and constructed to measure multiple pollutants simultaneously in the same ambient air parcels. The sensitivities of the one hertz measurements were 2 ppb for ozone, 1 ppb for NO, and 0.5 ppb for NO_2. Meteorological data were also collected for wind, temperature, and UV intensity. The field data showed clear correlations between ozone, NO, and NO_2 in the one-second time scale. Fluctuations in pollutant concentrations were found to be extremely dependent on meteorological conditions. Deposition fluxes calculated using the Eddy Correlation technique were found to be small on concrete surfaces. These high time-resolution measurements were used to develop an understanding of the variability in atmospheric measurements, which would be useful in determining the acceptable discrepancy of model and observations.

Thesis Supervisor: Gregory J. McRae
Bayer Professor of Chemical Engineering

Acknowledgments

My family gave me what it takes to pursue a Ph.D. degree. Thanks to my parents, Chiu-Tung and Shirley, whose unconditional love and support I can always count on. My sisters Amy and Christine, who, in their subtle ways, show that they will be there for me.

My thesis advisor and mentor, Greg McRae, never failed to see the relevance of my work. His positive attitude and his creativity are contagious. My other mentor, Mark Zahniser, is kindness and patience personified. Because they believed, so do I. Who can ask for better people to work with? I will always look up to both of them.

Many collegues provided valuable help and discussion at Aerodyne. I appreciate having the opportunity to work with Chuck Kolb, Barry McManus, Dave Nelson, Joanne Shorter, and Doug Worsnop. My heartfelt thanks also go to the McRae group: to Jonathan Knight, Tim Resch, Menner Tatang, and Pete Wyckoff, who saw me through the turbulent times of my thesis proposal; and to Gary Adamkiewicz, Jose Jimenez, Gene Lin, and Mort Webster, who provided the academic and emotional support throughout the thesis research.

Last, but certainly not least, my gratitude goes to Kate Baty, the fairy godmother, who is always there to listen, support and encourage. Without her, my graduate life would have been quite miserable.

I would like to acknowledge the United States Environmental Protection Agency for the financial support of this research on three cooperative research grants: CR822081 Uncertainty/Air Quality Modeling, CR822067 Atmospheric Modeling, and R824970-01 Airborne Organic Compounds.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Role of Uncertainty in Model Construction and Verification

In science and engineering, computer models are a standard tool to integrate observational knowledge in a mathematical format. This representation can be analyzed to understand the system's behavior. Most importantly, models are a critical component of any process control scheme and are used in designing control strategies for complex systems. There are many different levels of sophistication in models: some are simple correlations while others are complex differential equations. No matter the complexity, models are built upon a theoretical understanding, that is itself based upon theoretical knowledge developed from observations of the system. The relationship between model and data is illustrated in Figure 1-1. Due to the role of data in the model construction and validation processes, uncertainties in measurement data have significant influence on the quality of the models.

Since models are constructed to explain observations, the modeling process places a strong emphasis on the observed (and observable) aspects of the phenomenon. Processes that are not observed or that are observed to be unimportant may be left out of the model. Therefore, uncertainties in the observation may lead to faulty a theoretical understanding of the system, and thus a misrepresentation of the model. Even when the processes are correctly represented in a model, uncertainties in data are still endemic in the simulation processes. These uncertainties can affect both the model execution and validation processes.

Measurement data are needed to define the inputs and the parameters of models. Examples are flow rates into reactors, reaction rate constants of chemical systems, and thermodynamic properties of compounds. If a model is driven by inaccurate parameters, the predictions may be inaccurate even when the theoretical understanding is sound.

Observational data, usually of the outcome of scientific or political interest, are frequently used to verify models. For many complex models, the same observations are used to deduce any adjustable parameters that cannot be measured independently (to "tune" the model, so to speak). High quality data is needed for "tuning", which has implications towards all the future predictions made by the model. What data to collect, the frequency of data, the length scale or the location of point measurements are all important questions to address in designing a data collection campaign to validate complex models. A model should not be accepted or rejected based on uncertain data without first characterizing the uncertainty associated with the data. Modelers have a tendency to treat measured data as the absolute standard to which any reasonable models ought to predict exactly. In truth, being consistent with observed data may not be a very good indication of the models' performances if the measurements are themselves uncertain or if the quantity measured is highly variable. Models are only as good as the data that is used to drive and validate them. It is critically important to understand the uncertainties of data and their implications in the modeling process.

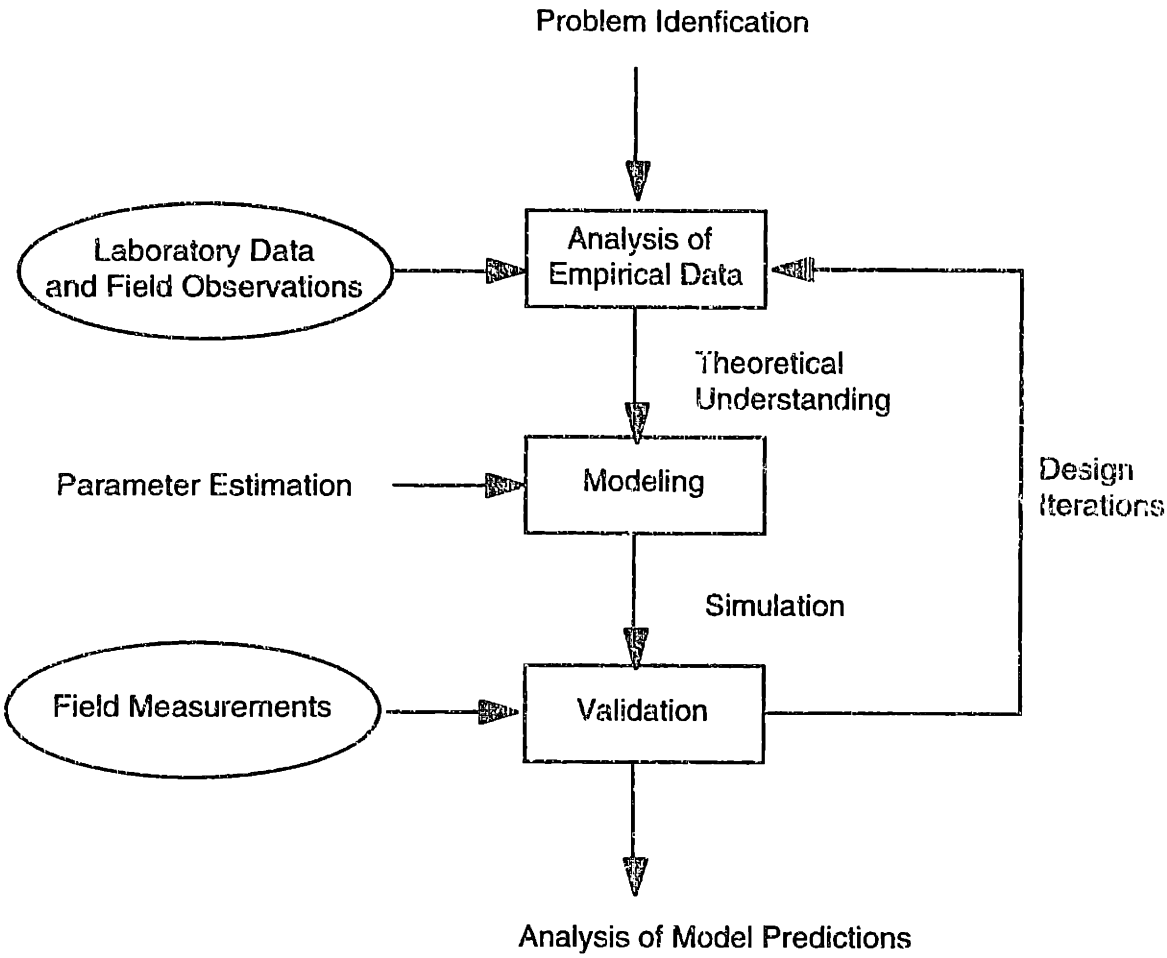


Figure 1-1. The Use of Data in the Modeling Process.

1.2 Motivating Problem — Urban Ozone Air Pollution

Ozone, one of the seven criteria pollutants, has been regulated in the Clean Air Act (CAA) since 1970. By 1990, when the new Clean Air Act amendments were written, the problem persisted, affecting more people than ever -- almost 70 million in the United States alone. Smog, which was once a fairly unique problem of a few large cities such as Los Angeles, has now become a common sight in most metropolitan cities around the world, from Paris to Mexico City. In the United States, despite almost thirty billion dollars that is spent annually on air pollution control, urban smog is still spreading from cities to suburbs, and is fast becoming a regional concern rather than a local problem (Figure 1-2).

Ozone is formed as a secondary pollutant in the atmosphere from precursors volatile organic compounds (VOC) and nitrogen oxides (NO_x). Because of the scale of the problem and the cost of controls, it is impossible to determine empirically the method to control atmospheric ozone. Instead, models are constructed to simulate the complex, non-linear relationship between ozone and the emission of its precursors. Air quality models are akin to chemical engineering models that simulate chemical reactors. Treating the atmosphere as a giant volume-varying reactor, models contain modules to simulate mixing and transport (meteorology), sources of reactants (emissions) as a function of time and space, and chemistry of the urban atmosphere. The formulation of effective abatement strategies relies on models to predict the response of ozone under a variety of control scenarios. Therefore, the efficacy of the pollution control strategies is also dependent on the accuracy of the model under a range of conditions.

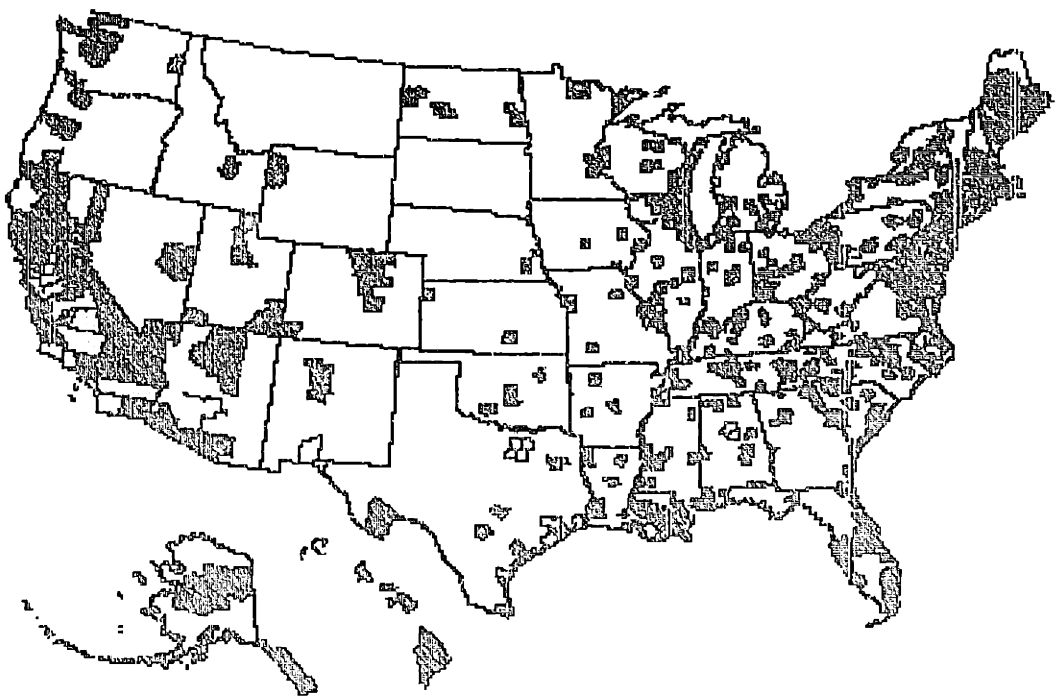


Figure 1-2. Second Daily Maximum Ozone in 1995 by County
(95 - 165 ppb ■ >165 ppb)
(from EPA's National Air Quality and Emissions Trends Report, 1995)

Unfortunately, air quality models are far from perfect. Figure 1-3 shows a comparison between a model's prediction and the observation of ozone at a downwind site of Los Angeles, from the Southern California Air Quality Study (SCAQS) of 1987. Even state-of-the-art models underpredict ozone concentrations by as much as 40% at some locations. The discrepancy between model and data is even greater for a multi-day scenario. Despite the imperfect predictions, the models are considered to be verified because they seem to be able to predict many qualitative features of the observations. Ideally, before these air quality models are used to design control strategies, the poor performance of the models against the episodic data should be understood. Unfortunately, the political urgency of the ozone problem dictates that controls be implemented. Therefore, the air quality models are used to predict the pollution scenarios at lower emissions without having been validated for a scenario with lower concentrations of ozone.

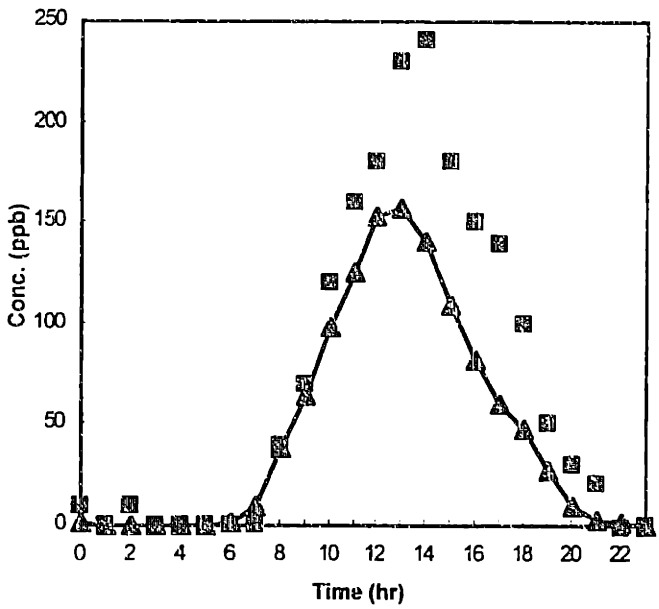


Figure 1-3. Observed vs. Simulated Ozone Concentration, Riverside, CA, August 27, 87. (Observations denoted by ■, model predictions denoted by ▲.)

Many factors can contribute to the discrepancy between model predictions and observed ozone concentration during an air pollution episode. Model predictions are uncertain because of uncertainties in the input data and the parameters used in the formulation. Observations may also be associated with errors and variability. Table 1-1 summarizes the possible sources of uncertainties / errors that cause the disagreement between the model predictions and the observations.

Before a model is rejected as wrong, it is important to verify that the model is unable to predict the observations by adjusting uncertain parameters within their ranges of acceptable values. Similarly, before an air quality model is deemed "good enough" for predicting the system's behavior under dramatically different scenarios, the model predictions and the observations should be compared together with their respective "error bars" -- the uncertainty in the model predictions due to uncertainties in the parameters, and the uncertainty in the measurements due to measurement error and spatial and temporal variability. A 40% discrepancy (Figure 1-3) may be acceptable if the model predictions and measurements are uncertain enough that there is significant probability for the two to be

consistent. Whereas the same level of agreement is not good enough if both the predictions and measurements are exact of have little uncertainty.

Category	Specific details
Emissions	source strength (stationary and mobile sources) composition of source emissions
Meteorology	wind speed and direction, Temperature mixing height, Relative Humidity radiation estimates, Topology wet and dry deposition, Turbulent Diffusion Coefficients
Chemistry	reaction rates product distribution lumping representation
Observations Verification Data	initial and boundary conditions measurement error spatial / temporal variability of pollutant concentrations

Table 1-1. Uncertainties in Data and Observations.

1.3 Thesis Objectives

This thesis focuses on several items in Table 1-1. A series of uncertainty analyses was conducted on the chemical modules to:

- characterize the effects of parametric uncertainties on the model predictions (for isolated chemical mechanisms and for three dimensional models)
- compare the effects of uncertainties in mechanistic parameters to that of precursor emissions
- compare the predictions of different chemical mechanisms with uncertain parameters
- investigate the effects of uncertain mechanism of isoprene, a key biogenic compound
- identify parameters whose uncertainties contribute to the uncertainties of air quality model predictions.

The second part of this thesis was to develop a new measurement technique based on the tunable diode laser (TDL) spectroscopy technology. The characteristics of the instrument are:

- high time resolution to study the atmospheric variability of key photochemical pollutants
- simultaneously measurements of ozone, NO and NO_2 in a single sample volume.

The goal was to compare the variability of the three pollutants, and to study the short time-scale correlations.

The specific contributions of this Ph.D thesis are:

- the development and application of a systematic methodology for conducting uncertainty analysis of large models with many parameters
- a first comparison of different chemical mechanisms used in Air Quality Modeling based on their responses to parametric uncertainties
- recommendations for strategies to reduce uncertainties of atmospheric model predictions based on parameters that affect the variance of uncertain outcomes
- the design and construction of a tunable diode laser spectrometer to measure three pollutants simultaneously
- observations of the short time-scale fluctuations of and correlations between pollutant concentrations.

1.4 Thesis Outline

The body of this document is organized in three parts. Part I presents the relevant background material, Part II documents various uncertainty analyses conducted on atmospheric chemical mechanisms and air quality models. The field experimental work to monitor ozone, NO, and NO_2 using a TDL spectrometer is described in Part III. The last two chapters highlight the conclusions of this thesis work and some directions for future research. Table 1-2 contains a brief description of each chapter in this thesis.

Chapter	Topics
1	Introduction
Part I	Background
2	brief introduction to urban atmospheric chemistry
3	data requirements for modeling urban air pollution
4	traditional analysis of atmospheric mechanisms
Part II	Uncertainty of Chemical Mechanisms
5	description of deterministic equivalent modeling method (DEMM)
6	illustrative example of applying DEMM to analyze a simple chemical mechanism
7	sources of parametric uncertainties in atmospheric chemical mechanisms
8	uncertainty analysis of the Statewide Air Pollution Research Center (SAPRC) mechanism
9	uncertainty analyses of the Regional Acid Deposition Model (RADM) mechanism and the Carbon Bond (IV) mechanism, comparison of three mechanisms using their uncertain predictions
10	analysis of the effects of uncertain parameters in the chemical mechanism on the predictions of a three dimensional air quality model
11	analysis of uncertainties in new isoprene mechanisms
Part III	Field Measurements of ozone, NO, and NO_2
12	new challenges in atmospheric measurements
13	theories of TDL operation
14	experimental design / procedures
15	brief description of the data set obtained
16	variability of atmospheric data
17	correlations of ozone, nitric oxide, and nitrogen dioxide
18	future research
19	key conclusions

Table 1-2. Organization of the Thesis

2. Background

2.1. Photochemical Oxidants

Smog, urban ozone, and photochemical oxidants are all used to describe the air pollution phenomenon that is characterized by elevated tropospheric ozone concentrations, particularly in densely populated urban and suburban areas. Smog, a term originally invented to describe the London brand of fog and smoke caused by coal burning in the winter, is applied to this new variety of air pollution caused by emissions from modern amenities such as cars, electrical power plants, and paints. It is a term used by the general public to describe the air pollution phenomena associated with "brown air", reduced visibility, and breathing difficulties. Smog was first discovered in Los Angeles, California, U.S.A. in the 1950's and have since "spread" to virtually air major cities in the world: Bangkok, Mexico City, Paris, to name a few. Ozone was regulated under the Clean Air Act of 1970 as a surrogate measure for pollutants commonly found with smog. Since then, the specified measurement method has changed such that the National Ambient Air Quality Standard (NAAQS) is defined for ozone alone. The ozone standard is controversial. There is a natural background for ozone, from 15 to 40 ppb (National Research Council, 1991), which has been increasing at a rate of about 1-2% per year over the last 30 years. By itself, the proven health effects of ozone atmospheric concentrations of ozone was limited to eye and respiratory tract irritations and additional asthma attacks (although ozone also damages vegetation and crops). However, together with other pollutants formed in an smog episode, including particulate matter and other oxidants, there may be a strong synergistic effect on human health, as suspected from epidemiological studies. In 1993, recent studies concerning the health and welfare effects of ozone were summarized (EPA, 1993) in support of a new ozone NAAQS which was issued on July 16, 1997 (EPA, 1997). The new standard is an eight-hour standard of 80 ppb, not to be exceeded more than three times in three years, which is different both in averaging time and concentration from the previous standard, which was a one-hour standard of 120 ppb.

Ozone is only one component of photochemical oxidants that are associated with urban air pollution. Photochemical oxidants are formally defined as the class of compounds which oxidizes iodide (I^-) in the KI measurement method (Finlayson-Pitts and Pitts, 1986). Ozone is considered to be a potent oxidant in the following reaction:

$$2H^+ + 2I^- + O_3 \rightarrow I_2 + O_2 + H_2O. \qquad (2.1)$$

Other oxidants, such as NO_2, PAN, and H_2O_2, also act as oxidizing agents to iodide, although ozone is by far the strongest and the most dominant of the oxidants in the atmosphere. All these oxidants are produced during a photochemical smog episode.

As oxidizing agents and precursors to reactive radicals, ozone and other atmospheric oxidants play a significant role in controlling the chemical lifetimes and reaction products of many atmospheric species, especially in the polluted urban environment.

2.2. Role of Photolysis in Photochemical Air Pollution

The formation of the photochemical oxidants is the result of a series of free radical reactions that take place in the troposphere. A number of light-initiated, or photolysis, reactions, serve as "initiation" steps to produce radicals for the system. Thus, urban smog is of "photo"-chemical origin. The primary chemical process of a photolysis reaction is the absorption of radiation by a reactant A to form the "excited" version, A',

$$A + h\nu \rightarrow A'. \tag{2.2}$$

There are several possible fate for A', i.e., fluorescence, collisional deactivation, dissociation, and direct reaction. Fluorescence is the reverse reaction of (2.2), which returns the excited A' to A, with the emission of light. Collisional deactivation is the process through which the extra energy possessed by A' is transferred to a third body, usually a nitrogen molecule or an oxygen molecule in the lower troposphere. A' is then deactivated back into A. Dissociation and direct reaction are the major processes that lead to photochemical reactions:

$$\begin{aligned} &\text{dissociation}: && A' \rightarrow D_1 + D_2 + \ldots \\ &\text{direct reaction}: && A' + B \rightarrow D_1 + \ldots \end{aligned} \tag{2.3}$$

In the lower troposphere, the predominant species, nitrogen and oxygen molecules, are relatively inert. Therefore, direct reactions are far less important than dissociation as a fate of excited molecules. The products of the dissociation reaction are usually very reactive, and some may be in electronically excited states, causing further reactions of importance in the photochemical reaction chain. The term photochemistry is used for the study of the fate of molecules excited by the absorption of radiation, as well as the subsequent reaction of their relatively reactive products.

Before the important reactions in the urban atmosphere are discussed, it is useful to identify the molecules that have the potential to absorb radiation and react photochemically in the lower atmosphere. At ground level, most of the more energetic photons have been filtered out by absorbers in the stratosphere. Therefore, the compounds that are photochemically reactive absorb Ultra Violet (UV) and visible light from 300 to 700 nm (Leighton, 1961). (The transition from UV to visible radiation occurs at 395 nm.) Table 2-1 lists the common absorbers and non-absorbers of light in this wavelength region in the urban atmosphere. Absorbing compounds undergo electronic excitations with the absorption of UV or visible radiation. Vibrational excitation may also accompany the electronic excitation (Barrow, 1979, Chapter 20).

While the energies of the electronic states determine whether the molecule absorbs light in the UV and visible regions, the bond strength and spin conservation rules determine if and how the excited molecule dissociates. By definition, bond strength is equivalent to the homolytic bond dissociation energy, which is the energy needed to pull the molecule apart at the specific bond such that each fragment receives one of the pair electron originally forming the bond. In the case of the oxygen molecule, the energy absorbed for electronic excitation falls short of the energy required for oxygen molecule to dissociate into two oxygen atoms at ground state. Excited oxygen molecules mostly undergo deactivation in the lower troposphere. All other absorbers listed in Table 2-1 are known to undergo dissociation in the lower atmosphere as a result of the excitation.

Absorbers	Non-Absorbers
Oxygen	Nitrogen
Ozone	Water
Nitrogen Dioxide, Nitrogen Trioxide	Carbon Monoxide
Nitric Acid and Alkyl Nitrates	Carbon Dioxide
Nitrous Acid, Alkyl Nitrites, Nitro Compounds	Nitric Oxide
Acyl Nitrites, Pernitrites, Nitrates	Sulfur Trioxide and Sulfuric Acid
Aldehydes	Hydrocarbons
Ketones and Dicarbonyls	Alcohols
Peroxides	Organic Acids
Sulfur Dioxide (not discussed in detail)	

Table 2-1. Absorbers and Nonabsorbers in the Region 300 - 700 nm.
(Leighton, Chapter III, Table II, 1961)

Ozone absorbs mostly in the UV region (200 - 350 nm), which corresponds to about 112 kcal/mol of energy. The dissociation energy of ground state ozone is about 25 kcal/mol to one normal oxygen molecule and one oxygen atom at ground state, labelled $O(^3P)$ because it is in a triplet state. The decomposition of ozone following the absorption of UV radiation is virtually instantaneous. Both ground state and excited products can be formed. The excited ozone is in a singlet state. Since the correlation rules forbid a singlet to decompose into a singlet product and a triplet product, either both products are in the excited state or they are both in ground state. The excited products are formed much more frequently, due to the large amount of energy absorbed. The excited oxygen molecule product deactivates, while the excited oxygen singlet atom, $O(^1D)$, may either thermally deactivate or, in moist air, undergo chemical reaction with water to form two hydroxyl radicals in the following reaction which is exothermic by 27 kcal/mole for $O(^1D)$:

$$O(^1D) + H_2O \rightarrow OH + OH. \qquad (2.4)$$

Reaction 2.4 is energetically favorable only for excited oxygen atom, whose main source in the lower troposphere is the photolysis of ozone. The ultimate formation of hydroxyl radicals is of tremendous importance because of OH's role as the main oxidizing agent for hydrocarbons and other trace gases in the polluted troposphere. The amount of hydroxyl radicals present in a tropospheric region directly reflects the "oxidative capacity" of that atmosphere. The higher the concentration of hydroxyl radicals, the shorter the lifetimes of hydrocarbons and other trace pollutants, and vice versa.

Nitrogen dioxide, NO_2, is another important species which absorbs over virtually the entire UV and visible range of the solar spectrum in the lower atmosphere (300 - 700 nm). Near the visible range, NO_2 undergoes fluorescence after absorbing radiation, which is the basis of chemiluminescence technique for detection of ozone and NO_x. At lower wavelengths, the radiation causes decomposition of NO_2:

$$NO_2 + h\nu \rightarrow NO + O(^3P). \qquad (2.5)$$

The bond dissociation energy of nitrogen dioxide about 71 kcal/mol, which corresponds to the energy of radiation at about 404 nm. Experimental results indicate that the changeover from excited molecule reaction to dissociation as a dominant process occur between 405 and

366 nm, which is consistent with the bond dissociation energy. Due to high absorption rates of NO_2, Reaction 2.5 is by far the fastest of all known photochemical primary processes in polluted air.

The oxygen atom formed in Reaction 2.5 is very reactive. The most important removal process of the (ground state) oxygen atom is the combination reaction with oxygen molecules to form ozone:

$$O(^3P) + O_2 \xrightarrow{M} O_3. \qquad (2.6)$$

The combination reaction requires a third body for collision stabilization to absorb at least 25 kcal/mol, the bond strength of the O_2-O bond. Therefore the rate of Reaction 2.6 is pressure-dependent. In polluted atmosphere, the lifetime of $O(^3P)$ is of the order 10^{-5} seconds. In polluted air, $O(^3P)$ also serves to initiate the breakdown of olefinic organic compounds by adding to the carbon-carbon double bond. However, due to the fast rate of reaction of $O(^3P) + O_2$, these other reactions involving $O(^3P)$, are probably unimportant in the urban atmospheres unless the pollutant concentration is high enough to compete effectively with oxygen molecules for the reaction with oxygen atoms.

The nitrate radical (NO_3), whose formation will be explained in the next section, plays a somewhat important role in the night-time chemistry in the troposphere. During the day, NO_3 absorbs radiation at 400 to 600 nm. Since the high electron density on this radical makes it rather unstable to start with, nitrate radical photolyzes readily during the day, breaking one of the N-O bonds in an analogous reaction to Reaction 2.5, to form NO_2 and $O(^3P)$ (Reaction 2.7a). At the high wavelength region, where the energy absorbed is limited, the nitrate radical undergoes a rearrangement reaction to form NO and O_2, two stable compounds (Reaction 2.7b). The energy released by the O-O bond is about 10 kcal/mole more than the second N-O bond in nitrogen dioxide, making the rearrangement reaction a more viable option at the low energy end of the visible spectrum.

$$NO_3 + h\nu \rightarrow NO_2 + O(^3P) \qquad (2.7a)$$
$$NO_3 + h\nu \rightarrow NO + O_2 \qquad (2.7b)$$

Two other classes of nitrogen compounds undergo photolysis reactions in the atmosphere. The first is nitric acid and nitrates (HNO_3 and RNO_3) where the nitrogen oxidation state is at +5. The second class is nitrous acid, alkyl nitrite, and nitro compounds, with a +3 oxidation state for the nitrogen atom. The discussion of the photolysis reactions of these nitrogen compounds is quite brief because of their minor importance in urban atmospheric chemistry. The weakest bond in nitric acid is the HO-NO_2 bond, which is about 53kcal/mol. This amount of energy can be supplied by the absorbed radiation, and

$$\begin{matrix} HONO_2 + h\nu \rightarrow HO + NO_2 \\ (RONO_2) \end{matrix} \qquad (2.8a)$$

can be observed, although it is not a fast reaction. By an analogous reaction to Reaction 2.8a, alkyl nitrate (RNO_3) can photolyze into RO and NO_2. By bond strength considerations alone, Reaction 2.8b, which breaks a C-O bond (weaker than the corresponding H-O bond in nitric acid) is also energetically possible.

$$RNO_3 + h\nu \rightarrow R + NO_3 \qquad (2.8b)$$

Reaction 2.8b is energetically less favorable than the reaction (analogous to Reaction 2.8a) which forms RO and NO_2 because a stronger R-ONO_2 bond (~85 kcal/mol) has to be broken and two reactive radical products are formed.

For nitrous acid, the most likely photo-decomposition reaction takes place at the HO-NO bond, which is about 60 kcal/mol, although the H-NO_2 bond can also break with 80 kcal/mol. The photolysis of HO-NO is an important source of both the hydroxyl radical and nitric oxide when the sun comes up in the morning. Photodissociation of organic nitrites may occur at the RO-NO bond, forming an alkoxy radical, RO, and nitric oxide. Another viable reaction of the nitrite is an intramolecular rearrangement to form an aldehyde product:

$$RCH_2ONO + h\nu \rightarrow RCHO + HNO. \qquad (2.9)$$

The nitrate or nitrite group can also be present in a compound with a carbonyl bond, mostly as a result of combination reactions. The terminology used for these carbonyl compounds are as follows. The RC(O)O group is the "acyl group", where (O) represents the carbonyl group with an oxygen atom double-bonded to the carbon atom. RC(O)ONO is called an acyl nitrite, and RC(O)ONO_2 an acyl nitrate. The RC(O)OO moiety is named the peroxyl acyl group. RC(O)O-ONO is a peroxylacyl nitrite, and RC(O)O-ONO_2 is peroxyl acyl nitrate, A common compound of this class is $CH_3C(O)OONO_2$, which is peroxyl acetyl nitrate or PAN. Acyl nitrite and nitrates may undergo fission at the O-N bond, the C-O bond, the C-C bond at the carbonyl carbon (the photolysis of carbonyl compounds are discussed next), or a rearrangement like Reaction 2.9. For example,

$$RC(O)ONO + h\nu \rightarrow RC(O)O + NO, \qquad (2.8a)$$
$$RC(O)ONO + h\nu \rightarrow RCO + NO_2, \qquad (2.8b)$$
$$RC(O)ONO + h\nu \rightarrow R + C(O)ONO, \qquad (2.12a)$$
$$RC(O)ONO + h\nu \rightarrow CO + RNO. \qquad (2.9a)$$

For both peroxyl acyl nitrates or nitrites, the RC(O)O-ONO_x bonds are also prime candidate sites for dissociation, although the bond is weak enough that dissociation occurs as a result of a thermal reaction.

Organic compounds containing a carbonyl group all absorb in the solar radiation region. Formaldehyde, the simplest carbonyl compound, shows clearly resolved rotational structure in its absorption spectrum, which begins at about 327 nm. Aliphatic aldehydes and ketones absorb at lower wavelengths and display less structure in the absorption spectrum. Adjacent carbonyls, such as in glyoxal, shifts the absorption into the visible region. For formaldehyde, fluorescence is observed after the absorption of radiation. This observation, combined with the highly rotationally-structured spectrum, indicate that the dissociation processes are probably rather slow, with low efficiency (or in photolysis terms, with quantum yields of less than unity). Two pathways are possible for the excited formaldehyde:

$$H_2CO + h\nu \rightarrow H + HCO \qquad (2.10a)$$
$$H_2CO + h\nu \rightarrow H_2 + CO. \qquad (2.10b)$$

The carbonyl bond weakens the C-H bond in HCHO compared to its counterpart in methane. The hydrogen atom formed in Reaction 2.10a quickly adds to an oxygen molecule to form a peroxyl radical:

$$H + O_2 \rightarrow HO_2. \qquad (2.11)$$

Although it is energetically possible to split formaldehyde into two hydrogen atoms and a carbon monoxide below wavelengths of 253 nm. Reaction 2.10b is thought to occur via a intramolecular rearrangement, especially at higher wavelengths. The energy requirement for a rearrangement reaction is generally less than that of a bond dissociation reaction because new bonds are formed that release energy. That is why they are more common at higher wavelength when less energy is absorbed from the radiation, provided that the energy expended to overcome steric hindrance to the rearrangement is small.

In addition to fluorescence and deactivation, longer chain aldehydes can undergo a larger variety of photochemical processes than formaldehyde. Decomposition can result from breaking the R-CHO bond (analogous to Reaction 2.10a) or another C-C bond within the alkyl chain; for example:

$$C_2H_5CHO + h\nu \rightarrow CH_3 + CH_2CHO. \qquad (2.12)$$

Intramolecular rearrangement analogous to Reaction 2.10b forms a straight chain alkane and carbon monoxide for alkdehydes. However, another rearrangement pathway is also possible with higher aldehydes and the formation of ethene and an aldehyde with two fewer carbon atoms is illustrated in Reaction 2.13:

$$C_2H_5CHO + h\nu \rightarrow C_2H_4 + H_2CO. \qquad (2.13)$$

The dissociation into alkyl and formyl radicals (analogous reaction to 2.10a) is the most important photochemical process for straight chain aldehydes. Rearrangement reactions can be important for some specific compounds because of stereochemical reasons. Ketones dissociate into one alkyl and one carbonyl radical, breaking at the carbon-carbon bond with the carbonyl group. Rearrangement reactions like Reaction 2.13 yield ethene and a lower ketone.

Among dicarbonyl compounds, glyoxal and methylglyoxal receive attention as products of aromatic compounds. Glyoxal is expected to react through rearrangement reactions:

$$HCOCOH + h\nu \rightarrow H_2 + 2CO \qquad (2.14a)$$

and

$$HCOCOH + h\nu \rightarrow H_2CO + CO. \qquad (2.14b)$$

Considering energetics alone, the C-C bond can also be cleaved at wavelengths less than 294 nm (95 kcal/mol). This pathway is sometimes employed in photochemical models as a radical formation pathway. Methylglyoxal is expected to decompose almost exclusively at the C-C bond separating the carbonyl groups, forming an acyl radical and a formyl radical (which reacts to form a peroxyl radical and a carbon monoxide).

The last group of compounds whose photolysis reactions are discussed here are the peroxides. The dissociation is probably at the weakest O-O bond, whose bond dissociation enthalpy is in the range of 30 kcal/mol. These compounds are difficult to measure in the polluted atmosphere because of their affinity to surfaces of sampling devices, and there is evidence to show that the concentration of peroxides may be higher than expected in some rural locations. However, the low absorption rates of the peroxides make photolysis unimportant as a source of organic radicals in the urban atmosphere.

2.3 Precursors of Photochemical Air pollution (smog)

The photolysis reactions described in Section 2.2 are responsible for the formation of a variety of radicals and other reactive species in the urban atmosphere. These include: $O(^3P)$, $O(^1D)$, alkyl radicals, alkoxyl radicals, formyl radicals, acyl radicals, OH, H (thus HO_2), O_3, RCHO, NO, and NO_2. Several are important components of photochemical smog, others allow the propagation of the processes that lead to the formation of excess pollutants *in situ*. Because of the interconversion between NO and NO_2 and the direct formation of O_3 from the photolysis of NO_2, NO_x ($NO+NO_2$) is considered an important precursor to photochemical smog. The rapid production of ozone is caused by the oxidation of NO to NO_2 by agents other ozone. In the urban air system, radicals, such as HO_2, play that important role. Many of the radicals are organic peroxyl radicals, RO_2 or $RC(O)O_2$. That is why the class of reactive organic gases (ROG), the parent of these organic radicals, is the other precursor of photochemical air pollution. The formation and reactions of these radicals will be discussed in Section 2.4.

2.4 Urban Atmospheric Free Radical Chemistry

Some of the radicals mentioned in Section 2.2 are formed chiefly in photolysis and subsequent reactions, others have stronger sources in atmospheric free radical reactions. Because of their reactivity, the radicals undergo a series of radical propagation reactions before they are quenched, or removed from the urban system. Radical propagation reactions fall into one of four main categories: substitution (or abstraction), addition to double bond, beta elimination, and rearrangement. Sometimes the reactions are a combination of the four kinds (March, 1992, Chapter 6). Propagation reactions form a new radical by removing the radical site from the reactant. When radicals combine to form a stable compound, the chain reaction is "terminated".

The discussion of photolysis and the subsequent decomposition reactions have been based on energetics consideration. In a photo-dissociation reaction, the entropy effects are always favorable for the fragmented state. Therefore, the only consideration is enthalpies, or bond energies. If enough radiation energy is absorbed to break a bond, it will be broken unless the kinetics is so slow that the excited molecule is quenched before the dissociation reaction. Energetics also explain much of the radical propagation reactions that go on in the atmosphere. In most cases, the reactions of concern are substitution reactions, where the number of reactant molecules equal the number of product molecules. One example is the hydrogen-abstraction from an organic molecule by a radical to from an organic radical and a small molecule. In these substitution reactions, the entropy effects are small, and enthalpy consideration of the bonds being broken and formed will give a good idea of whether or not the reaction will proceed. Addition to a double bond is the reverse reaction of beta elimination. Here, the entropy effects are balanced between having an additional molecule and forming a single bond which is more flexible than a double bond. Rearrangement reactions involve minimal entropy change, but is favored when a five or six membered transition state is formed.

This section provides a condensed lesson of the radical reactions that take place in the atmosphere. The discussion draws on many published compilations of organic and inorganic reactions, and the readers are encouraged to refer to them.

2.4.1 Inorganic Reactions

A set of reactions involving compounds made up of hydrogen, nitrogen, and oxygen form the backbone of tropospheric chemistry. The formation of the key radicals, $O(^3P)$, $O(^1D)$, HO and HO_2 (from hydrogen atom), are discussed in Section 2.2. Other active species include oxygen molecule (O_2), ozone (O_3), nitric oxide (NO), and nitrogen dioxide (NO_2). Both ozone and nitrogen dioxide are good oxidizing agents. The products of the oxidizing reactions are water, nitrous acid (HONO), and nitric acid (HNO_3), although HONO sometimes dissociate back into the reactants.

Nitrogen species are present in the troposphere at several oxidation states: NO_3 (+6), HNO_3 (+5), NO_2 (+4), HONO (+3), NO (+2), and N_2O (+1) (does not react in lower troposphere). In the troposphere, many of the reactions involving nitrogen species are reduction-oxidation (redox) in nature. Although nitrogen species participate in free radical reactions, they also react in redox reactions that do not involve radicals.

Reactions 2.5 and 2.6 show the ozone formation process. The titration of ozone by nitric oxide,

$$O_3 + NO \rightarrow O_2 + NO_2, \tag{2.15}$$

completes the cycle of reactions. If Reactions 2.5, 2.6, and 2.15 are in equilibrium, then the so-called photostationary state determines the concentration of ozone:

$$[O_3] = \frac{j[NO_2]}{k_{15}[NO]}. \tag{2.16}$$

Reaction 2.15 is the first example of an atmospheric redox reaction. The N-O bond formed releases about 73 kcal /mol (CRC, 1991) and O-O bond broken is, by most estimates, less than 30 kcal / mole (Perry, 1992; Nuffield, 1986), making this reaction energetically favorable. Table 2-2 lists all redox reactions in the nitrogen-oxygen-hydrogen system that involve the transfer of an oxygen atom. Although carbon monoxide is not straightly speaking an inorganic species, the last reaction is included in Table 2-2 to illustrate that oxygen transfer can occur at carbon atom sites even though hydrogen abstraction, discussed next, is a far more common reaction.

Oxidation	Bond Formed (Energy)	Reduction	Bond Broken (Energy)
$NO \rightarrow NO_2$	NO-O (73 kcal/mol)	$O_3 \rightarrow O_2$	O_2-O (<30 kcal/mol)
$NO \rightarrow NO_2$	NO-O (73 kcal/mol)	$HO_2 \rightarrow HO$	HO-O (70)
$NO \rightarrow NO_2$	NO-O (73 kcal/mol)	$NO_3 \rightarrow NO_2$	NO_2-O (51)
$O \rightarrow O_2$	O-O (119)	$NO_2 \rightarrow NO$	NO-O (75)
$NO_2 \rightarrow NO_3$	NO_2-O (51)	$O_3 \rightarrow O_2$	O_2-O (<30)
$HO \rightarrow HO_2$	HO-O (70)	$O_3 \rightarrow O_2$	O_2-O (<30)
$CO \rightarrow CO_2$	CO-O (126)	$HO \rightarrow H$	H-O (102)

Table 2-2. Inorganic Redox Reactions Involving Oxygen Atom Transfer.

As shown in Table 2-2, radicals and ozone, which possess unpaired electrons, undergo reduction by donating an oxygen atom to a reaction partner. The chemical bond formed has to release enough energy to compensate for the bond broken. For example, the NO-to-NO_2 (73 kcal/mol energy release) conversion can be facilitated by a HO_2-to-HO (70 kcal/mol requirement) reduction, while the NO_2-to-NO_3 conversion, releasing substantially less energy, cannot.

Many radicals also have the ability to abstract a hydrogen atom from another reactant. The gain of a hydrogen lowers the oxidation state by one and allows a stable molecule to be formed from the radical species. Hydrogen abstraction reactions are very common in the organic-radical system as a radical propagating step and as the first step of the breakdown of an organic compound, e.g., HO abstracting a hydrogen from a hydrocarbon, (discussed in the next section). When the pollutant concentrations are low, a radical may abstract a hydrogen from another radical. The reaction forms two stable molecules and reduces the total number of radicals in the system. This type of combination reactions is a form of "termination"

reactions because they reduce the reactivity of the system. Table 2-3 shows several examples of these hydrogen abstraction reactions.

Reaction	Bond Formed (Energy)	Bond Broken (Energy)
$^*NO_3 + HO_2 \rightarrow HNO_3 + O_2$	H-NO_3 (kcal/mol)	H-OO (48 kcal/mol)
$^*HO_2 + HO_2 \rightarrow HOOH + O_2$	HOO-H (88)	H-OO (48)
$HO + HO_2 \rightarrow H_2O + O_2$	HO-H (119)	H-OO (48)
$HO + HOOH \rightarrow H_2O + HO_2$	HO-H (119)	HOO-H (88)
$HO + HNO_3 \rightarrow H_2O + NO_3$	HO-H (119)	H-NO_3 ()

Table 2-3. Hydrogen Abstraction Reactions.
(* These reactions may be catalyzed by a third bodies, especially water)

A more common type of reaction capable of serving as a termination step is the combination of two radicals. In atmospheric chemistry, the definition is expanded to include the combination of two reactive species, because species like NO_2 are generally not considered radicals, but do have the ability to combine with a radical to form a stable species and reduce the general reactivity of the system. Since only one molecule is formed from two reactants, collision stabilization is usually needed for the successful completion of these reactions. Through the combination reaction and the reverse dissociation reaction, an equilibrium is established for many stable "sink" species. The longer the time scale for dissociation of the stable molecule, the higher the probability of removal of the sink compound by deposition or transport, thus, the more effective the sink species is at removing radicals from the system. Table 2-4 gives examples of the inorganic combination reactions in the polluted troposphere.

Reaction	Dissociation
$O + O_2 \rightarrow O_3$	by photolysis not important compared to reaction
$H + O_2 \rightarrow HO_2$	not important compared to reaction
$O + NO_2 \rightarrow NO_3$	not important compared to reaction
$NO_2 + NO_3 \rightarrow N_2O_5$	thermal equilibrium with forward reaction N_2O_5 also react with water to form 2 HNO_3
$HO + NO \rightarrow HONO$	HONO photolysis
$HO + NO_2 \rightarrow HNO_3$	no dissociation reaction
$HO_2 + NO_2 \rightarrow HNO_4$	thermal equilibrium

Table 2-4. Inorganic Combination Reactions.

Many of the combinations in Table 2-4 are favored over the reverse dissociation, because of the strong enthalpy stabilization effect. Nitric acid is a particularly good sink for HO and NO_x, since the only chemical reaction with HO radicals is slow, and it is removed effectively by wet and dry deposition. Others, like HNO_4, are poor sinks, because weak bonds are formed. The dissociation reaction happens in the same time scale as the formation. For some species like N_2O_5, the removal of the product by other reactions helps push the equilibrium towards the product side.

2.4.2 Primary Organic Reactions

Section 2.2 presents photolysis reactions as a breakdown mechanism for some oxygen and nitrogen-containing organic compounds. A far more effective channel for removing organics, especially hydrocarbons, from the lower troposphere is radical reactions. The inorganic radicals of concern here are hydroxyl, peroxyl, nitrate radicals, oxygen atom, and, to the extent ozone can be considered a radical, ozone. In addition, there is a whole series of organic radicals: organic peroxyl (RO_2), acyl peroxyl ($RC(O)O_2$), and alkoxy radicals with many functional groups. This section focuses on the initial attack by an inorganic radical on an intact organic molecule to form organic radicals. The reactions of the organic radicals will be discussed in Section 2.4.3.

The hydroxyl radical is the main species responsible for removing organic compounds from the urban air. As seen previously, it has strong affinity for hydrogen via the abstraction reaction,

$$R - H + HO \rightarrow R + H_2O, \qquad (2.17)$$

because the formation of water releases almost 120 kcal/mol (HO-H bond energy). Another "hydrogen-deprived" radical is the nitrate radical, which accumulates only at night when the photolysis removal process is switched off. The H-NO_3 bond formed in

$$R - H + NO_3 \rightarrow R + HNO_3 \qquad (2.18)$$

releases about 101 kcal/mol.

In abstraction reactions, entropy effects are not important. Therefore, insights can be gained regarding the favorable reactions by looking at the bond strengths of C-H bonds in an organic compound. C-H bond strengths of typical hydrogen donors are listed in Table 2-5.

C-H Bond	Bond Strength (kcal / mol)
Methane	105
Alkane -- Primary Carbon (RCH_2-H)	100
Alkane -- Secondary Carbon (R_2CH-H)	95
Alkane -- Tertiary Carbon (R_3C-H)	92
Alkene (e.g., H_2C=CH-H)	106
Alkyne (e.g., HC≡C-H)	132
Aromatic (e.g., C_6H_5-H)	111
Aldehyde (e.g., RC(=O)-H)	86
Ether (e.g., $ROCH_2$-H)	93
Alcohol (e.g., RC(OH)-H)	93
Alcohol (OH bond, e.g., RCH-O-H)	104

Table 2-5. Bond Strengths of C-H bonds in Organic Compounds (alcohol O-H included for comparison)

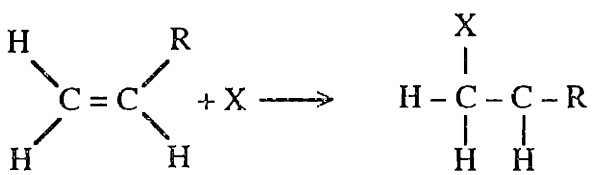
Carbon-hydrogen bonds in alkanes are between 90 and 100 kcal/mol, depending on how "shielded" the carbon is. A carbon site crowded with alkyl groups hold on less strongly to the hydrogen. Therefore, weaker C-H bonds are found on tertiary carbons than secondary than

primary carbons. In addition, the higher the number of alkyl group surrounding the carbon atom, the better the radical can be stabilized (by the delocalization of electrons). Therefore, tertiary carbons are more preferable radical sites than secondary than primary carbons. On a double bond carbon, the C-H bond strength increases to about 106 kcal/mol due to the tighter carbon (sp2-hybridized) orbitals. On a carbonyl carbon atom, the C-H bond is weakened by more 10 kcal due to the electronegativity for the carbonyl bond. This affect is also observed to a smaller extent on alcohols, whose C-H bond strengths at the alcohol site is only about 94 kcal/mol compared to methane's 105 kcal/mol.

With the formation of an O-H bond ot 120 kcal/mol, OH radical can abstract hydrogen from many alkane, aldehyde, alcohol, and ether species, with high enthalpy gains. Hydrogen abstraction from methane is known to be slow due to the inertness of the species, as reflected by the strong C-H bonds. Hydroxyl radicals probably do not abstract hydrogen from alkene and alkyne sites, because orbital hybridization in these molecules increases the C-H bond strength to over 100 kcal/mole. Addition reactions, discussed next, are much more common for these compounds due to the reactivity of the pi bond system. For alcohols, the hydroxy group is a less likely target for hydrogen abstraction than the C-H groups because of the stronger O-H bond, as shown in Table 2-5. The nitrate radical is a weaker hydrogen abstractor than the OH radical, due to lower enthalpy gains of the H-NO_3 bond formed (101 kcal/mol). The hydrogen abstraction by nitrate from alkanes are expected to be less favored due to the reduced thermodynamic driving force. In fact, the night-time Reaction 2.18 with various organics are expected to be two orders of magnitude less important a sink for organics than the day-time reactions with hydroxyl radicals (Atkinson, 1992). The peroxyl radical (HOO) does not normally abstract hydrogen from hydrocarbons, because the HOOH bond formed is only 88 kcal/mol, which is not enough to offset the enthalpy of dissociation of most C-H bonds.

Hydrogen abstraction is one example of a larger class of reactions called substitution reactions. Small radicals of concern in the lower troposphere also add readily to a double or a triple bond of an organic compound. This discussion will focus on the double bond, which is much more common and much more reactive in the urban atmosphere. The addition reaction is driven thermodynamically by the cleavage of a relatively weak pi bond and the formation of a strong sigma bond. An organic radical is formed on the carbon atom adjacent to the addition site. The apparent entropy disadvantage of forming one molecule from two is compensated by the increased flexibility of the sigma bond relative to the pi bond. Again, the enthalpy consideration dominates the reaction thermodynamics, and bond strengths can provide insights as to how these reactions proceed.

The difference between the energy of dissociation of a double bond and a single bond is a good estimate of the bond dissociation energy of the pi bond. The bond energy of a C-C single bond is usually about 90 kcal/mol. A C=C double bond is less than twice that, at 172 kcal/mol. Therefore, about 82 kcal needs to be supplied to break the pi bond completely in the following reaction:

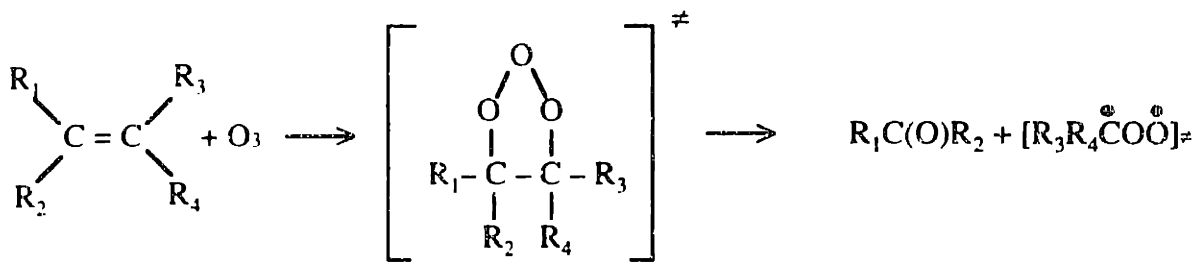


(2.18)

Bond strength information for radicals is less readily available than for stable compounds. Recent advances in quantum chemistry calculations allow good estimates of the bond

dissociation energies to be made without conducting difficult experiments (Jungkamp and Seinfeld, 1996), although this data base is still small. From the information available for the OH radical addition reaction, Reaction 2.18 is exothermic by 32 kcal/mol. If the pi bond is broken and only a sigma bond remains for the C-C system, this implies that the C-OH bond formation releases 114 kcal/mol (Streiweiser and Heathcock, 1985, Chapter 11), compared to the 94 kcal/mol for a C-OH bond in an alcohol molecule. The C-OH bond in a radical is probably stronger because of resonance structures required to stabilize the radical on the neighboring carbon atom. On the other hand, the radical center may also retain some pi (double bond) characteristics, so that the C-C bond in the adduct radical is more than 90 kcal/mol. If this is the case, the C-OH bond strength would be less than 114 kcal/mol. (It is expected that both the C-C bond and the C-OH bond are stronger next to a radical center than in a regular alcohol. Experiments and / or quantum chemistry calculations should help address this issue in the future.) NO_3 radical is also active towards the double bond, since the C-ONO_2 bond is about 80 kcal/mol (Benson, 1976, Appendix A11) in a stable compound and is expected to be higher in a radical.

Like most addition reactions, the radical addition to the pi system generate an excess of vibrational energy. (^3P) atoms cause instantaneous dissociation of adduct after adding to an alkene molecule. Ozone, which has unpaired electrons like a radical, can also add to the double bond to form an ozonide, which decomposes readily to form a carbonyl compound and a Craigee bi-radical (Reaction 2.19).



(2.19)

Alkenes with conjugated pi bonds (e.g., 1,3-butadiene and isoprene) are expected to be much more reactive towards radical additions than their one double bond counterpart, because of available resonance structures that stabilize the radical site. However, radicals are expected to add less readily to the benzene ring, which is a delocalized pi bond system, than to an alkene double bond, because the C=C bonds are stronger in the aromatic system due to resonance energy. Nonetheless, the hydroxyl radical is the main decomposition pathway for aromatic compounds (Reaction 2.20). Due to the delocalization of electrons in the benzene ring, the radical center is stabilized by the available electron density and the hydroxycyclohexadienyl radical is usually represented as an OH group associated with the ring, as shown in Reaction 2.20. No C–C bond is truly broken by the addition of hydroxyl radical. Although hydrogen abstraction from the side chains can occur in a similar manner as in regular alkanes and aldehydes, the hydroxyl mostly adds to the ring's pi system. Other radicals, such as the nitrate radical, are not expected to be reactive towards the ring, but can abstract hydrogen atoms from the side chain.

(2.20)

Another double bond system is the carbonyl. The energy difference between a carbonyl (C=O) bond and a single C-O bond is 86 kcal/mol. Due to oxygen's electron-donating effect, most radicals preferably abstract the weakly-bonded hydrogen from the carbon center instead of adding to the double bond. The peroxyl radical is one exception to the rule. The C-OOH bond releases only about 20 kcal/mol, not enough to break the C-H bond or the C=O bond. The adduct quickly isomerizes, breaking the C=O bond and forming an O-H bond, and forms a hydroxy-peroxyl radical.

2.4.3 Secondary Organic Reactions

The initial breakdown of atmospheric organics by radicals, discussed in the previous section, forms a variety of organic radicals. These organic radicals are less reactive compared to the inorganic ones such as hydroxyl and nitrate discussed in the previous section. Nonetheless, their reactions with oxygen, nitrogen oxides, and other radicals fuel the urban tropospheric chemistry and the formation of ozone and other air pollutants. This section discusses the reactions of the organic radicals.

The hydrogen abstraction and addition reactions discussed in the previous section form mostly alkyl radicals, radicals with an unpaired electron at a carbon center. All four basic types of reactions: substitution, combination, rearrangement, and elimination, are possible with alkyl radicals. In a flame, where the concentrations of organic compounds are high, the alkyl radical may abstract a hydrogen atom from another alkane, or add to a alkene double bond. Elimination reactions, because of the intrinsic endothermicity, is much more common at higher temperature situations than at atmospheric conditions. Rearrangement is unlikely because internal radicals are formed preferably due to energetics consideration in the first place. Under atmospheric conditions, alkyl radicals predominately add to oxygen molecules to form peroxyl radicals:

$$R + O_2 \rightarrow RO_2 \qquad (2.21)$$

Like the formation of HO_2 (Reaction 2.11), the high concentration of oxygen molecules makes this reaction extremely fast. Since oxygen is a better carrier of an unpaired electron than carbon, RO_2 is more stabilized compared to R. For example, the enthalpy of formation of CH_3O_2 is 2.5 kcal/mol while that of a methyl radical is 35 kcal/mol (CRC, 1994, Section 9). The entropy effect is probably small compared to the enthalpy gain in this reaction (2.21) at room temperature.

Organic peroxyl radicals do not abstract hydrogen, either inter- or intra- molecularly, from C-H bonds (90 kcal/mol), because the ROO-H bond is weaker (84 kcal/mol). With its high electron density, RO_2 can undergo substitution reactions by oxidizing other species in the atmosphere, much like its inorganic counterpart, the hydroperoxyl radical. In the urban troposphere, a common reducing agent is NO:

$$RO_2 + NO \rightarrow RO + NO_2. \qquad (2.22)$$

Reaction 2.22 is analogous to the second redox reaction in Table 2-2. The NO bond formed in the product NO_2 releases 73 kcal / mol. The RO_2 to RO transition requires an energy input

of roughly 62 kcal/mol, based on enthalpy of formation for radicals given in CRC (1994). In the urban environment, where the nitric oxide concentration is more than a few ppb's, Reaction 2.22 is the main transformation route for RO_2 radicals.

As mentioned previously, RO_2 radicals do not normally abstract hydrogen from stable organic compounds such as alkanes and alkenes. When the NO concentrations are low, and the lifetime of RO_2 radicals increase, the probability of an RO_2 radical reacting with another radical increases. Table 2-3 indicates that a HO_2 radical can abstract a hydrogen from anther HO_2 radical, forming hydrogen peroxide and oxygen. The H-OO bond in a hydroperoxyl radical is only 48 kcal/mol, and a hydrogen abstraction by an RO_2 (Reaction 2.23) is exothermic by more than 30 kcal/mol.

$$ROO + HOO \rightarrow ROOH + O_2 \tag{2.23}$$

Organic peroxyl radicals are also known to react with other peroxyl radicals. However, the reaction pathway is complex and involves multiple bonds being broken and formed. For example,

$$\begin{aligned} CH_3OO + CH_3OO &\rightarrow CH_3O + CH_3O + O_2 \\ &\rightarrow HCHO + CH_3OH + O_2 \end{aligned} \tag{2.24}$$

(Carter, 1990). These reactions cannot be predicted by enthalpy considerations alone.

Addition reactions involving RO_2 can also be important in some atmospheric systems. Organic peroxyl radicals may add to NO and NO_2 to form organic nitrate and pernitrate. The RO_2-NO_2 bond is expected to be weaker than the HOO-NO_2 bond, which is 23 kcal/mol. Therefore, the decomposition reaction back to RO_2 and NO_2 is fast enough to make the pernitrate an ineffective sink for the reactants. This combination and the reverse dissociation reactions are ignored in most atmospheric mechanisms used to simulate urban pollutant concentrations. The RO_2 + NO addition reaction is more common with larger RO_2 radicals. Reaction 2.25 leads to the formation of organic nitrate, RNO_3, which results from a intramolecular rearrangement:

$$RO_2 + NO \rightarrow [ROONO]^{\neq} \rightarrow RONO_2. \tag{2.25}$$

The acyl peroxyl radical, $RC(O)O_2$, undergoes substitution reactions with NO and peroxyl radicals just like paraffinic peroxyl radicals. However, because of the electron withdrawing effect of the carbonyl group, the radical center on the oxygen atom is not as well stabilized as the RO_2 counterpart. As a result, the $RC(O)O_2$ radicals tend to be more reactive. It oxidizes NO to NO_2. $RC(O)O_2$ does not add to NO because the resulting product would be too unstable, but it forms a stronger bond with NO_2 than its alkyl counterpart in the addition reactions. For example, the $CH_3C(O)OO$-NO_2 bond is exothermic by 29 kcal/mol. The product formed is a peroxyl acyl nitrate:

$$RC(O)OO + NO_2 \rightarrow RC(O)OONO_2. \tag{2.26}$$

When R is a methyl group, the adduct is peroxyl acetyl nitrate (PAN), a key photochemical oxidant found in urban atmospheres. PAN and PAN-like compounds serve as temporary sinks of both NO_x and organic radicals, as they are relatively stable and thermally decompose back to the reactants in a time scale of 30 minutes to a few hours.

Like RO_2, RC(O)OO reacts with fellow peroxyl radicals when the concentration of NO_x is low. The reaction with HO_2 can proceed through two pathways: one forming a peroxyl carboxylic acid (RC(O)OOH) and oxygen molecule, and the other forming a carboxylic acid (RC(O)OH) and ozone. Reacting with other RO_2 radicals, RC(O)OO behaves like a RO_2 in Reaction 2.23.

Reaction 2.22 shows that when an organic peroxyl radical oxidizes NO, NO_2 is formed together with an alkoxy radical (RO). Alkoxy radicals are generally very reactive. It can abstract hydrogen to form a strong RO-H bond if a source of hydrogen is available. Alternatively, because of the electron-withdrawing effect of the oxygen at the radical site, the other bonds on the carbon atom is weakened. A second possibility is a beta scission-type decomposition of the radical to form a carbonyl compound and a alkyl radical. If a hydrogen atom is present on the carbon, the hydrogen is activated towards abstraction by another radical or by an oxygen molecule because of the weaker C-H bond. For example, a five-carbon alkoxy radical can undergo the following reactions:

$$CH_3CH(-O)CH_2CH_2CH_3 \rightarrow CH_3CH(OH)CH_2CH_2CH_2 \quad (2.27a)$$
$$CH_3CH(-O)CH_2CH_2CH_3 \rightarrow CH_3CHO + CH_3CH_2CH_2 \quad (2.27b)$$
$$CH_3CH(-O)CH_2CH_2CH_3 + O_2 \rightarrow CH_3CH(O)CH_2CH_2CH_3 + HO_2. \quad (2.27c)$$

For longer chain alkoxy radicals, a convenient hydrogen source is from the CH groups within the molecule. Reaction 2.27a is an intramolecular hydrogen abstraction, or an isomerization, via a six-membered transition state. This reaction is possible because the OH bond is formed is more than 10 kcal/mol stronger than the CH bond broken. The primary radical product in Reaction 2.27a is likely to isomerize again such that the radical site moved to an internal carbon. Here, that is the carbon atom stabilized by the electron-donating OH group. The hydrogen at the OH group can then be abstracted in reaction similar to Reaction 2.27c to form a carbonyl compound. Reaction 2.27b is a decomposition by beta scission. A carbonyl bond is formed and a C-C bond is broken. The energy gained forming the C=O bond over the C-O bond is probably about 86 kcal/mol, which is roughly equal that of the C-C bond broken. Although the enthalpy gain is small, the entropy effects are expected to make this a thermodynamically favorable reaction. The alkoxy radical can also give up a hydrogen to form a stable compound. In most atmospheric circumstances, the hydrogen receiver is an oxygen molecule, and a HO_2 radical is formed from Reaction 2.27c. Although the energy released by the HO_2 bond is only 48 kcal/mol, Reaction 2.27c is actually quite exothermic ($CH_3O + O_2$ --> $HCHO + HO_2$ is exothermic by 27 kcal/mol), due to the energy gain by forming a carbonyl bond. This reaction is much more important for smaller alkoxy radicals than for larger ones.

Decomposition seems to be the dominant reaction for the acyl-oxy radical. The products are an alkyl radical and a carbon dioxide:

$$RC(O)O \rightarrow R + CO_2. \quad (2.28)$$

In reaction 2.28, a C-C bond is broken and a C=O bond is formed (over a C-O bond of a radical). The thermodynamics gain of such a reaction is not huge, only about 10 kcal/mol calculated from the heats of formation information (CRC, 1994). However, the combined electron-withdrawing effects of two unshared oxygen atoms on the carbon center may favor the kinetics of the dissociation reaction. A RC(O)O-H bond is about 105 kcal/mol, strong enough to cause isomerization by hydrogen abstraction. However, isomerization into a carboxylic acid is not observed, most likely due to the kinetic dominance of Reaction 2.28.

Addition reactions of alkoxy and acyl-alkoxy radicals to NO and NO_2 are not expected to be of importance at atmospheric conditions. Although the RO-NO bond is only about 42 kcal/mol, collision stabilization is required. These reactions happen at a time scale that is at least three orders of magnitude slower than the decomposition and oxygen reactions of alkoxy radicals (Atkinson, 1992).

The Craigee bi-radical is formed from the breakdown of ozonides in Reaction 2.19 is vibrationally excited. The excited bi-radical can decompose into a carbonyl and a radical or rearrange into other excited state molecules (e.g., an ether). On the other hand, the excited biradical may undergo collision stabilization which is still extremely reactive. The energetics and bond energies of the biradical has not been well-characterized. The stabilized biradical is known to react with aldehydes, SO_2, CO, H_2O, and NO_2. Among the products are carboxylic acids, H_2O_2, organic hydroperoxides. Again the reactions likely involve multiple steps and cannot be elucidated using simple thermodynamic considerations.

2.4.4 Representation of Organic Compounds in Atmospheric Mechanisms

With the hundreds, if not thousands, of organic compounds being emitted into the urban atmosphere, it is quite impossible to keep track of detailed reactions of each compound in mathematical models that simulate the transport and fate of atmospheric pollutants. Furthermore, there is insufficient information concerning the rates of reactions and the stoichiometric coefficients of products to support such an attempt. Leone and Seinfeld compiled an "explicit" mechanism for twelve species in the atmosphere (Leone and Seinfeld, 1984). This mechanism contained over 200 reactions, even though a number of slower reactions were neglected and isomers were not distinguished from one another. Even for just 12 species, the information requirement for explicit mechanisms is formidable. Many reaction rate constants were calculated by correlation to other rate constants, and there is a lack of knowledge in terms of the temperature dependence of many reactions. A detailed description such as this one is almost unimaginable for the organic mix found in the urban atmosphere.

A process know as "lumping" is used in most atmospheric mechanisms to group organic compounds into classes and to simplify the description of the chemistry. Three approaches are commonly used: surrogate species, lumped structure, and lumped molecule. The surrogate species approach allocates entire classes of compounds to single species for representation in the model. In Dodge (1977), The total amount of reactive hydrocarbons was represented by 75% n-butene and 25% propene. Only the reactions of the two alkenes and the related aldehydes were included in the mechanism. However, this approach was obviously insensitive to subtle changes in the composition of the organic mix. In the lumped structure approach (e.g., Gery *et al.*, 1989), individual molecules were not represented. Instead, each molecule is represented by its structure, namely, by a set of bonds. Reactions were written for C-C, C=C, C=O, ... groups. The advantage was that only a few classes of bonds can represent virtually all the organic compounds present in the atmosphere. Finally, the lumped molecule approach groups organics based on the homologous series they belong to: alkanes, alkenes, etc. Each type of molecules may be further divided according to the size of the compound. For example, alkanes are represented in three lumped classes: C3 or smaller, C4 to C7, C8 and bigger in the RADM mechanism (Stockwell *et al.*, 1990). There are subtle variations within the lumped molecule representation. Lurman *et al.* (1986) used a straight mole-to-mole approach from individual species to lumped class, but allowed the product coefficients to vary according to the composition of the organic mix. Stockwell *et al.* (1990) employed a reactivity lumping method, where individual organic's contribution to the lumped class was

weighted by the reaction rates of the organic compound. Carter (1990) allowed the mechanism user to select the representation of alkanes, alkenes, and aromatics. The reaction rate and products of each lumped class were determined as a weighted average of all the individual organic compounds represented in the class, by reactivity lumping or by a mole-to-mole lumping. Further details will be presented for several lumping strategies in later discussions in the uncertainty analysis section.

In addition to primary organic compounds, lumping is also carried out for secondary organics formed in the atmosphere. Most mechanisms employ only two or three species of aldehydes. All higher aldehydes are represented by propinaldehyde, as in the case of Carter's mechanism (1990). Other functional groups, such as -OH formed by isomerization (Reaction 2.27a) are typically represented using the carbonyl functional group as in a ketone. Sometimes, these representations result in lost carbon. Organic radicals, like secondary organics, are lumped in mechanistic representation. Carter (1990) chose to use several classes of radicals by their reactions: one class carry out NO-to-NO_2 conversions, while another combine with NO_2. Stockwell used size-dependent RO_2 radicals. These issues will also be discussed further in Part II of the thesis.

3. Data Need for Modeling Urban Air Pollution

Modeling the formation and transport of urban photochemical smog is a data-intensive activity (See Table 1-1 and Figure 1-1). The data requirement falls roughly into three categories. First, laboratory data are needed as parameters in the models. For example, reaction kinetics and thermodynamics data used for modeling the chemical transformations, wet and dry deposition characteristics of various pollutants are needed to describe removal processes. Second, data are required as inputs and boundary and initial conditions to models which predict the chemical dynamics. This may include pollutant emissions and composition and various meteorological measurements, such as mixing height and temperature. Finally, field measurements are needed to compare against simulation results so that the accuracy of the models can be assessed before they are used for predicting the effects of different control strategies. The following discussion focuses mostly on the data needed to model chemical kinetics of the urban atmosphere.

3.1. Laboratory Measurements

In Chapter 2, the basic chemistry of the urban troposphere was discussed. The modeling of this chemical system requires kinetic measurements for each and every chemical reaction involved. Photochemistry and gas phase chemical kinetics are well-developed experimental branches, with difficult radical-molecule and radical-radical kinetic and reactive photochemical studies now becoming routine, using techniques such as flash photolysis (Barrow, 1979, Chapter 20). The results of these photochemical and kinetic studies are compiled and periodically evaluated by NASA (with a stratospheric focus) (NASA, 1994) and IUPAC panels (Baulch *et al.*, 1992). Virtually all inorganic reactions, including photolysis reactions, used in tropospheric photochemical mechanisms are documented in these compilations. The elementary reactions of simple organic compounds, such as methane, formaldehyde, acetaldehyde etc., are also recorded. The panels also provide an expert opinion on the uncertainties of the kinetic data based on the accuracy of the experimental procedures and possible artifacts in the measurements.

The gas phase reactions kinetics and mechanisms of organics have also been reviewed. Atkinson (1992) provides a comprehensive review of tropospheric reactions of many organic compounds. A detailed review of organic nitrates has been published by Roberts (1990). There are also radical-specific reviews, such as those by Atkinson and Carter (1984) for ozone reactions and Atkinson (1991) for NO_3 radical reactions.

The reactions of more complex organic compounds, such as aromatic compounds and multifunctional species, whose fundamental mechanisms are not well-studied, are based on a body of empirical data from smog chamber experiments. A smog chamber or an environmental chamber is usually a big Teflon bag which serves as an enclosed reactor for photochemical reactions. Specific amounts of precursors are injected into the chamber. A variety of instruments, including gas chromatograph, mass spectrometer, chemiluminescence instruments, and aerosol impactors, are employed to monitor the course of the photochemical reactions: the disappearance of the reactants and the formation of products,

either *in situ* or by collecting and analyzing time samples. Smog chamber experiments provide valuable information for two main purposes. First, product yields can be measured in a smog chamber even when the mechanism of reactions of a species is unknown. From this information, mechanisms can be postulated for complex compounds. When the elementary reactions are not elucidated, parameterized mechanisms can be formulated to explain the product formation in the smog chamber and provide a starting point for modeling the breakdown of compounds in the ambient atmosphere. Second, a smog chamber provides a batch reactor environment, which isolates the chemical reactions from any transport effect. Chamber data are uniquely useful for testing chemical mechanisms. Experimental data from hundreds of these chambers form a data base against which new chemical mechanisms are tested (Jeffries *et al.* , 1982, 1985a, b; Carter *et al.*, 1985). The ability to reproduce smog chamber experimental observations is a first test for new and revised gas phase mechanisms. Unfortunately, the utility of smog chambers is limited to relatively high concentrations of pollutants. Some might argue that the results from the high concentration chamber experiments should not be directly extrapolated to modeling the ambient atmosphere, where the concentrations are much lower. Another problem is the well-known chamber dependent artifacts, such as wall reactions (Carter and Atkinson, 1982), which may affect the reliability of chamber results.

From the information available regarding the reaction mechanisms of smaller alkanes series, reaction mechanisms are suggested for the larger members of the alkane homologous series and reaction rates are estimated (Carter and Atkinson, 1985). For other compounds, surrogate representations are usually employed, such that all higher aldehydes are modeled by propinaldehyde and all substituted aromatics are modeled using 1,3,5-trimethyl benzene.

3.2. Input Data

In order to predict pollutant concentrations, a photochemical transport and transformation model requires accurate initial concentrations, boundary concentrations, as well as source information (including aloft concentrations). Of the three, the initial conditions have the smallest effect on the predictions of an episode that lasts several days. Aloft concentrations have a strong affect on the dynamics of urban chemistry especially in the morning when the mixing height increases rapidly, entraining ozone and other precursor species that are preserved from the previous day. For a multiday episode, entrainment is an important process not to be ignored. Concentrations above the mixing layer can be measured using aircrafts and balloons, although such measurements are not routinely done. Accurate source information, however, is both important and difficult to obtain.

There are two kinds of sources that emit pollutants that form ozone: stationary sources and mobile sources. In each urban area, there can be thousands of stationary point sources. The approach to preparing an emissions inventory for these point sources is to assign each one to a source category which is associated with a emissions profile (Walters and Saeger, 1990). The source profile, usually based on limited measurements, contains information on the relative amounts of various pollutants and the composition of organic emissions. According to the activity level of the source, a emission factor is determined for the particular source (e.g., EPA Air Chief Software, 1995). The product of the emission factor and the emission profile, summed over all sources, determines the stationary source strength represented in the model. Although some large sources are equipped with continuous emissions monitor, this information has seldom been incorporated into the emissions inventory preparation process. Mobile source emissions are determined in a similar fashion. Much of the emissions data from mobile sources were obtained by monitoring the exhaust levels of a small sample of

motor vehicles running on dynamometers (CFR 1983, Stump *et al.*, 1992a, b) and extrapolated to the entire fleet of vehicles. These tests were later found to be inadequate, and the sample was not representative of on-road vehicles. Indeed, based on source receptor studies, many authors have found inadequacies in the emissions inventory of both stationary and mobile sources (Harley *et al.*, 1992; NRC, 1991)

3.3. Ambient Measurements

The Clean Air Act (CAA) of 1970 mandates that monitorings sites be set up to measure criteria pollutants. A network of stationary sites have been set up under this provision and the ambient data recorded for ozone, NO_x, CO, non-methane hydrocarbons (NMHC), PM_{10} etc. are recorded in a central EPA database called AIRS (EPA, 1995). Most of the data are recorded in an hourly average format, with the exception of NMHC which is recorded as a 6 a.m. to 9 a.m. concentration. The CAA Amendments of 1990 provided for Photochemical Assessment Monitoring Stations (PAMS) -- sites established upwind and downwind of ozone non-attainment areas to measure detailed hydrocarbon composition in addition to ozone, nitrogen oxide concentrations and meteorological data. The PAMS sites are only now starting to generate reliable data, and these data have not been fully analyzed.

Once every couple of years, an intensive measurement campaign is undertaken somewhere in the United States to collect a set of comprehensive ambient measurements of air pollutants, usually during ozone episodes. Recent measurement campaigns include the North America Research Strategy for Tropospheric Ozone (NARSTO) Northeast of 1995-96 (Roberts *et al.*, 1995), the Southern Oxidant Study (SOS) of 1994 (Chameides and Cowling, 1995), SCAQS (Southcoast Air Quality Study) of 1987 (Blumenthal *et al.*, 1989; Seinfeld, 1989; Harley *et al.*, 1993) and 1997. In addition to NO_x, ozone, and meteorological data, speciated organic profiles and aerosol size and composition are recorded in a much denser network than routine monitoring. Coupled with ground measurements are a variety of aircraft and balloon measurements aimed to provide vertically-resolved concentration and meteorology information within the mixed layer and aloft.

These studies, although tremendously expensive, are important for two purposes. First, the analysis of field data may provide insights into the physical and chemical processes that may be important in a given region. Second, data from these field campaigns are used to validate three dimensional transport and transformation models. Once validated, these models are used to analyze control strategies to reduce the concentration of pollutants in reduced-emission scenarios.

Since intensive measurement campaigns takes place infrequently, the number of scenarios that can be used to validate any model is limited. Most complex models are now tuned to the available scenarios, and therefore to the episodic conditions that are chosen for measurement campaigns. Although the models are used to predict lower-ozone scenarios, the readers should be reminded that they are not truly validated in the low pollution regime.

3.4. The effects of Data Uncertainties on the Modeling Process

In the field of urban air pollution modeling, sophisticated models, such as UAM (Morris and Myers, 1990) , CIT Airshed (McRae, 1981), and CALGRID, are constructed to compile and represent the state of the knowledge in the field. In essence, modeling is an attempt to mimic, on the computer and in mathematical terms, what happens in the natural atmospheric system. The parallel relationship between the natural system and the simulated system is illustrated in Figure 3-1. Emissions, anthropogenic and biogenic, drive the natural system. Accurate estimations of these sources are necessary as inputs for the simulation. Without accurate input information, the model cannot be considered validated even if it predicts the observed concentrations exactly.

The natural system of atmospheric transport and transformation is modeled using a sequence of partial differential equations. The atmospheric diffusion equation (ADE) (Equation 3.1) includes pollutants accumulation, advection, turbulent diffusion, and reaction terms. Entrainment at the top of the mixing layer is treated as a volume source term in some models. The boundary condition describes the surface loss and emissions processes.

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (u c_i) = \nabla \cdot (\mathbf{K} \nabla c_i) + R_i[c_1, ..., c_i; T, t]$$
$$v_g^i c_i - \mathbf{K} \nabla c_i = E[\mathbf{x}, t] \qquad (3.1)$$

Many parameters are needed to define Equation 3.1 properly. Meteorological measurements are needed to describe the advection, although sometimes the predictions from a meso-scale meteorological model are used instead. Turbulent diffusion coefficients and deposition fluxes are also parameters that are difficult to measure. This research focuses mostly on chemistry. Although it is contained in only one term of the equation, the chemistry subroutine is frequently the most time-consuming to solve in an air quality model, and the description of the chemical processes that go on in the atmosphere require perhaps the largest number of laboratory-determined parameters. Parameters for the reaction rates, temperature and pressure dependence, and product coefficients are required in the mechanisms. Even if all the fundamental understanding encoded in the ADE is correct and no important physical or chemical process has been left out of the model, the model predictions will not be correct if the parameters are incorrect. Therefore, errors and uncertainties in the laboratory measurements of these parameters contribute to errors and uncertainties of the model predictions.

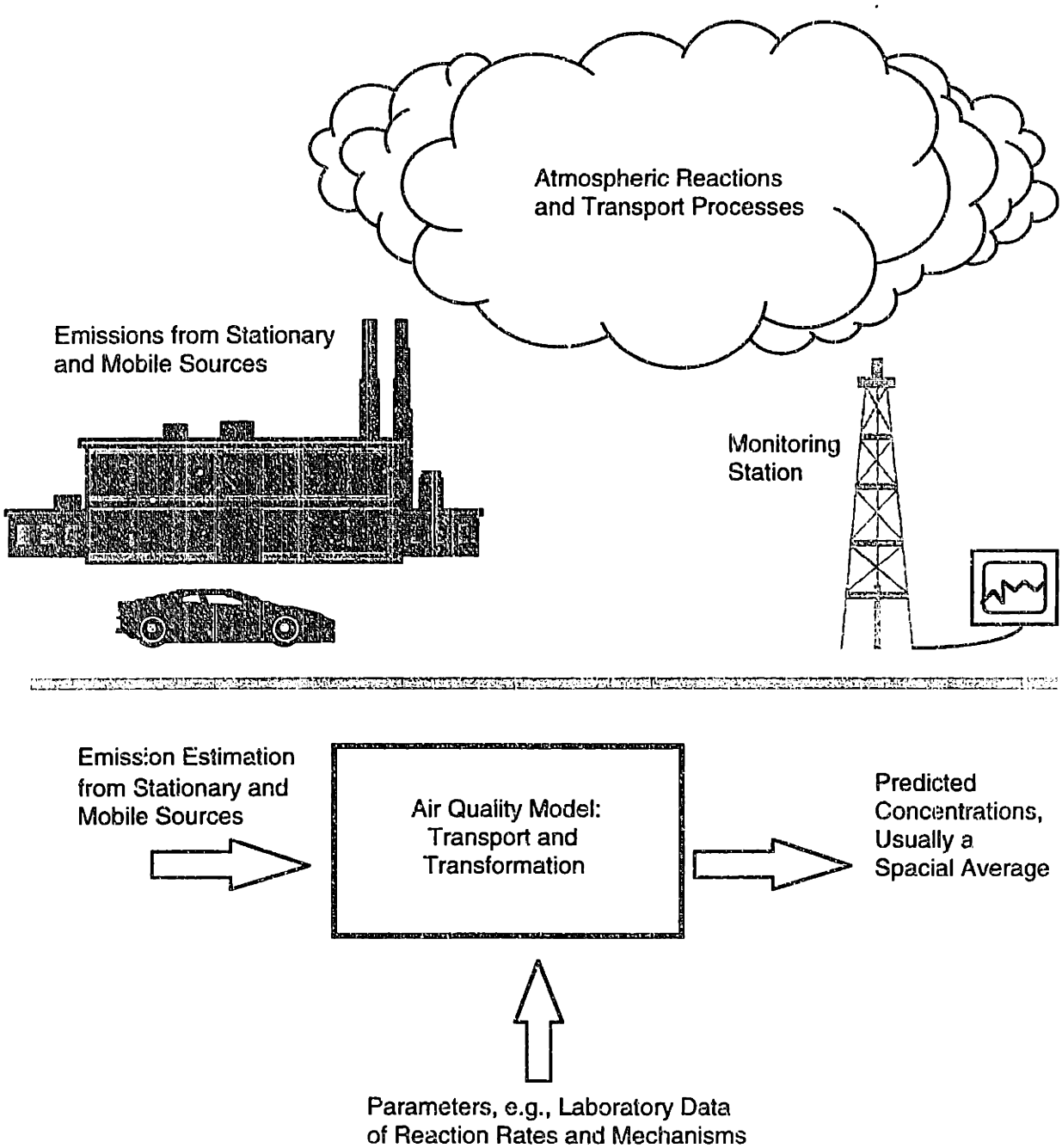


Figure 3-1. Parallel Relationship between the Atmospheric System and Air Quality Modeling

Field data is the only means by which models can be validated. Although many of the atmospheric measurements are routine, the experimental data are not perfect because of measurement error. The validation process itself is problematic. Ambient measurements are usually taken at a stationary point, while mode predictions are spatially averaged on the scale of a grid cell. Even with high-powered computers, and a variety of techniques to enhance the resolution of models (e.g., nested grid techniques), the scale the model grid will always be larger than the observations at the point measurement. Each point measurement is affected by the micrometeorological conditions and sources surrounding it. Due to this spatial variability of pollutant concentrations, model predictions and point ambient measurements

are not directly comparable, even when they are averaged for the same time scale. Furthermore, since the atmosphere is a turbulent system, the concentrations of pollutants exhibit temporal fluctuations. The air quality model can only model the mean behavior of a fluctuating system. Turbulent effects introduce another caveat into the validation process. Figure 3-2 depicts an idealized comparison between a model and measured data.

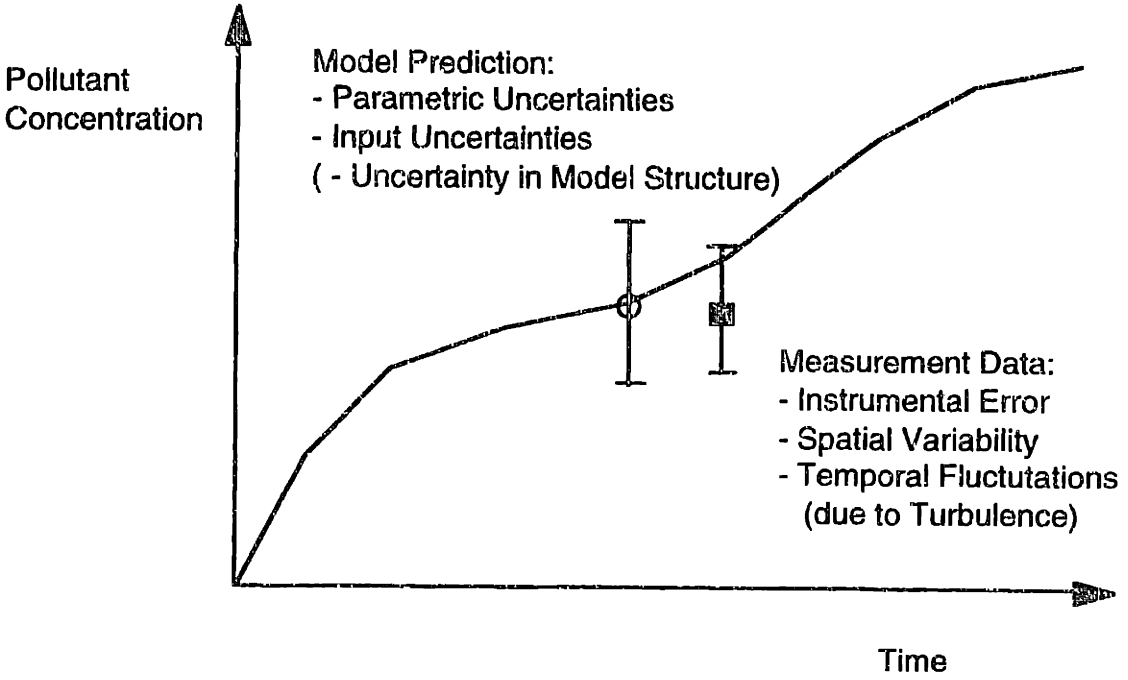


Figure 3-2. Idealized Comparison between Model Predictions and Measurement Data.

Ideally, both the grid cell average of the air quality model results and the ambient measurements should be compared in the presence of their respective error / uncertainty estimates. If the goal of the study is to determine if the model is right or wrong, or whether the understanding of the underlying physics and chemistry of the system is correct of flawed, the range of model predictions due to uncertainties in model input and parameters need to be identified. To compare oranges to oranges, measurements need to be taken that correspond to the quantity being predicted. In the case of atmospheric chemistry, unfortunately, there are no good *in situ* methods for taking spatial averages of many of the pollutants concerned. To conduct a meaning comparison between point measurements, some measure of the spatial variability is needed. One estimate is to use the temporal variability of the data as a measure of the spatial variability. The limitation of such a method is that the spatial scale observed probably does not correspond to the grid scale used in the models. Furthermore, temporal fluctuation corresponds to the spatial variability only in the ideal case where there is little mixing between air parcels from different origins. Because turbulent processes are not modeled explicitly in the model, the data variability needs to be included, together with the instrumental error, in the comparison process. With the "error estimates" in the data and model predictions, the model validation process can be conducted in a more statistically-reliable manner. A model should be rejected only when the model, using reasonable parameters, has a negligibly small probability of reproducing the observations within the range of measurement error and spatial / temporal variability.

A complete treatment of model uncertainties and data variability in the atmospheric system is beyond the scope of this thesis. The work described in the next chapters describe the first steps taken to address several key issues that were raised in Figure 3-2. On the model uncertainty side, parametric uncertainties and their effects on model predictions are the focus of Part II of the thesis. An experimental study is described in Part III where the first attempts were made to characterize the variability of atmospheric measurements of key pollutant species. No attempt was made to relate the temporal fluctuations to spatial variability. Many more pieces of the problem need to be addressed before carrying out revised kind of comparison between atmospheric data and air quality model as depicted in Figure 3-2.

4. Traditional Analysis Methods for Atmospheric Models

Long before three-dimensional models were used to predict urban ozone concentrations, mechanisms were developed and validated using environmental chamber measurements. They were then incorporated into the new three-dimensional air quality models and used to predict control strategies. This practice continues today, making air quality model development a modular activity. One of the major conclusions from the analysis of air quality models in the past ten years is that NO_x control may be needed in addition to VOC (volatile organic compounds) control to bring ozone down to the level deemed healthful by the National Ambient Air Quality Standards (NAAQS) (Milford *et al.*, 1989), especially in downwind locations. This finding came as a surprise to the community who believed then in the efficacy of VOC control. Another observation was that different mechanisms predicted substantially different degrees of emissions controls to achieve the same desired air quality under identical conditions (Leone and Seinfeld, 1985, and references therein). Due to the immense cost of emissions controls, this discrepancy needs to be understood, so that cost-effective control strategies can be formulated.

Different authors chose different strategies to balance chemical details with mechanism length. Some traditional methods of analyses of chemical mechanisms are reviewed in this chapter. The methods described here fall into two main classes. First, an understanding of the mechanisms can be gained through an analysis of their predictions, e.g., ozone concentrations, under a variety of circumstances. Second, the fundamental differences of the mechanisms can be inferred from their responses to changes in the parametric values. Comparison studies can be used to study the impacts of various assumptions made to simplify the chemical representation by comparing results of models with different levels of sophistication. Counter species is a technique to attribute the production of photochemical pollutants to the precursor organic compounds. Sensitivity and uncertainty studies provide information regarding the parameters and inputs that have the largest effects on the concentration predictions of interest.

4.1 Comparison Studies

Although all chemical mechanisms were written based on the same body of knowledge and were validated against a common set of environmental chamber experiments, they differ in the lumping of primary and secondary organic compounds. A study was recently conducted to test the reduction hypotheses made in photochemical mechanisms in a systematic manner (Aumont *et al.*, 1996). This study identified some of the key differences between different mechanisms and systematically analyzed their effects at different VOC/ROG scenarios. Some of the possible sources of errors considered were the omission of peroxyl-peroxyl reactions, lumping of secondary species by using surrogate species, and the use of "chemical operators" to describe organic peroxyl chemistry, as in the SAPRC mechanism (Carter, 1990). None of these assumptions were found to affect ozone predictions significantly. The omission of peroxyl-peroxyl reactions introduced errors in NO when NO concentrations were below 2 ppb. The lumping of secondary compounds affected the predictions of PAN when compared to the non-comporessed reaction mechanisms.

4.2 Counter Species

The concept of the counter species was designed to elucidate from the mechanisms the relative importance of the organic processes in the production of radicals or ozone. Leone and Seinfeld (1984, 1985) analyzed several mechanisms for the relative contributions of several organic classes to the number of NO-to-NO_2 conversions by adding a fictitious product (the counter species) to the reactions of interest. The reaction dynamics predicted by the mechanism did not change, but the number of species increased. Since there are no removal processes for the counter species, the accumulation of the counter species indicated the cumulative productivity of the particular reaction in which they are formed. In other words, the counter species "counted" the number of times a reaction took place within the simulation. From the concentration of counter species, the number of NO-to-NO_2 conversions, and the total amount of ozone produced during the simulation which was attributable to each of the initially present organic species can be calculated.

A similar idea to the counter species is the use of the integrated reaction rates and mass balance analysis (Jeffries, 1993, Wang, 1995) to track the progress of each reaction. This group also used this technique to compare the sources of ozone in different mechanisms (Jeffries and Tonnesen, 1995).

The concept of counter species was extended in 1994 (Bowman and Seinfeld, 1994). The effects of secondary organic compounds, like aldehydes, that originated from the breakdown of primary compounds, such as alkanes, were included. The method was generalized to analyze other species of importance in photochemical smog production, such as radical, PAN, and nitric acid, in addition to ozone. The analysis was carried out using three different VOC - NO_x scenarios to simulate the different characteristics of ozone formation at different locations of a metropolitan area. Aldehydes, alkenes, and heavy aromatics were found to be productive for all the species investigated. The high productivity was attributable to the fraction of each VOC that reacted.

4.3 Sensitivity Studies of Ozone to Volatile Organic Compounds (VOC)

The amount of ozone formed in the atmosphere depends on both the kinetic reactivity (the fraction of the emitted VOC which reacts in the time-scale of interest) and the mechanistic reactivity (the amount of ozone formed for each reacting organic molecule). For example, in the alkane family, the larger the alkane, the faster the breakdown reaction, and the higher the kinetic reactivity. The mechanistic reactivity also increases with the size of alkanes, because the alkane can cycle through several aldehyde intermediates during the breakdown process and propagate radical production.

Another technique used to obtain information concerning the contribution of each individual organic compound to ozone formation is to study the sensitivity of ozone to an incremental change of the initial condition or source strength of the compound of-interest. Carter (1991) calculated the Maximum Incremental Reactivity (MIR) scale based on the Statewide Air Pollution Research Center (SAPRC) mechanism (Carter, 1990). The MIR is a sensitivity coefficient of peak ozone to individual organic compounds calculated based on a scenario (organics/NO_x ratio) in which the peak level of ozone is the most sensitive to organic concentrations. To calculate the MIR, the model was run twice, once with a base case scenario and the second time with an additional amount of the organic compound of interest. The MIR is defined as:

$$MIR|_{\text{scenario}} = \frac{\max[O_3]_{\text{basecase+organic}} - \max[O_3]_{\text{basecase}}}{\text{Amount of VOC added}}. \qquad (4.1)$$

Though the MIR has not been proven to be accurate for large scale programs, such as gasoline reformulation, that affect the composition of the organic mixture in the atmosphere, it provides a systematic basis through which the relative effects of different organics on ozone can be compared, and has formed the basis of many newer regulations in ozone pollution control.

4.4 Sensitivity Analysis and Uncertainty Analysis of Parameters in the Mechanisms

In addition to the effects of organic precursors on ozone formation, the atmospheric community is also concerned about the uncertainties in ozone predictions due to parameters used in the mechanisms that are not well-known. A sensitivity coefficient is defined as:

$$\text{sensitivity coefficient} = \left.\frac{\partial c_i}{\partial x_j}\right|_{\text{nominal}} . \qquad (4.2)$$

An uncertainty analysis, sometimes known as a "global sensitivity analysis", takes into account the uncertainty associated with each parameter x_j in the estimates of uncertainties. A parameter to which the model predictions are sensitive to may have relatively small effects on the uncertainties of the predictions if it is well-known. On the other hand, another parameter may have such a wide range of uncertainty that its influence on the model output is substantial even when the sensitivity of is small. The information gained from an uncertainty analysis is used to identify the parameters whose uncertainties cause uncertainties in the predictions. In 1979, Falls *et al.* conducted an uncertainty analysis on a sixty-step mechanism using the Fourier amplitude sensitivity test (FAST) method. The procedure involves a simultaneous variation of all the parameters over their individual ranges of estimated uncertainty. This was the first sensitivity / uncertainty study to identify the radical yield of alkoxyl radicals and the reaction rate of HO + NO_2, in addition to the photolysis rates of NO_2 and aldehydes, to be important parameters contributing to uncertainties in ozone predictions.

Several recent studies have investigated the sensitivities and uncertainties of the MIR scale and the Regional Acid Deposition Model (RADM) chemical mechanism (Yang *et al.*, 1995, Gao *et al.*, 1995, 1996, Russell *et al.*, 1995). The authors used two descriptors to describe the effects of parametric uncertainties on model predictions. Uncertain parameters were described as a probability distribution associated with a mean and variance (σ_j^2). The measure of the uncertainty effect is the contribution of the uncertain parameter to the variance of the output. One estimate of the variance contribution (VC) is simply the product of variance of the uncertain input and the squared sensitivity coefficient as given in (4.2).

$$\text{VC of uncertainty in } x_j \text{ to output } c_i = \sigma_j^2 \left(\left.\frac{\partial c_i}{\partial x_j}\right|_{x_{base\,j}} \right)^2 . \qquad (4.3)$$

Alternatively, the uncertain parameters can be sampled at random to generate a collection of model predictions. Based on the simulation results, another metric for the VC is:

$$\text{VC of uncertainty in } x_j \text{ to output } c_i = \frac{\left(\sigma_j / x_{j,\text{nominal}}\right)^2}{\left(\sigma_i / c_{i,\text{nominal}}\right)^2} \cdot \beta_{ij}^2, \qquad (4.4)$$

where β_{ij}^2 is the linear regression coefficient of concentration c_i with respect to input x_j. The rational and derivation of these measures of VC will be discussed in the next chapter. A range

of parameters were identified to affect ozone concentration, including the reactions rates of PAN and nitric acid, as well as those of the photolysis reactions. Gao *et al.* (1996) and Russell *et al.* (1995) also found that the choice of effective control strategies (i.e., VOC vs. NO_x control) was insensitive to uncertainties in the kinetic parameters in the chemical mechanisms.

4.5 This Work

Uncertainty analyses of several mechanisms and a three-dimensional model were conducted as part of this research and are discussed in the Part II of this thesis. (See Table 1-2 for the topics of each chapter.) A new method called the Deterministic Equivalent Modeling Method (DEMM), introduced in Chapter 5, was used for these studies. DEMM is superior to the methods described so far for its computational speed advantage and its ability to represent non-linear systems. These uncertainty studies provide a comprehensive picture of the effects of uncertainties in the chemical kinetics to the urban atmospheric models' abilities to predict accurately the concentrations of pollutants. The key questions addressed are:

- Which parameters have uncertainties that contribute to uncertainties in the model predictions?
- How different are the various mechanisms that were constructed for different purposes using different amount of detail?
- Do chemical parameters have a strong influence on the uncertainties of the predictions of three-dimensional models?
- How important is it to incorporate the newest mechanistic information into mechanisms?

5. Uncertainty Analysis Using Deterministic Equivalent Modeling Method (DEMM)

Models are frequently used to inform decisions when actual experiments are expensive or impractical. When the model depends on parameters that are not perfectly known, it is important to quantify the uncertainties in the outputs before they are used for making decisions that involve huge commitment of resources. As defined by Morgan and Henrion (1992), an uncertainty analysis encompasses the activities of uncertainty description, propagation, and analysis of implications. Uncertain representation of parameters involves analyzing the sources of uncertainties, be it experimental error or disagreement in measurements, or geographical variability. When the uncertainty is not well-documented or cannot be determined based on the data source, expert opinion is drawn upon to provide mathematical descriptions for uncertain parameters. Uncertainty analysis is a "global sensitivity" analysis of the model response within the range of values the parameters can take on. Uncertainty propagation is defined as the calculation of uncertain model outputs induced by uncertainties in the inputs. Most methods of uncertainty propagation in continuous system involves sampling the scenarios that are possible within the range of uncertainty of the parameters. Uncertainty analysis sheds insights regarding the comparative importance of the uncertain inputs in terms of their effects on the uncertain outputs. Through the identification of parameters that contribute to the uncertain outputs, priorities can be set regarding possibilities of reducing uncertainties in the model predictions.

This chapter explains the mathematical concepts of the Deterministic Equivalent Modeling Method (DEMM). Chapter 6 contains an example to illustrate the application of DEMM in an uncertainty analysis.

5.1 Representation of Uncertain Parameters

The probabilistic approach is a common way to represent uncertainties in mathematical models (Morgan and Henrion 1992). In addition to truly stochastic disturbances and variable conditions, random variables are used to represent inputs whose true values are unknown or uncertain. A real-valued random variable, $\xi(\omega)$, maps the stochastic event (ω) onto the real space. A cumulative density function (CDF), $P_\xi(x)$, is defined as follows:

$$P_\xi(x) = Pr(\xi(\omega) < x). \tag{5.1}$$

This function gives the probability that $\xi(\omega)$ is less than or equal each real value x. The corresponding probability density function (PDF), $p_\xi(x)$, is simply the derivative of $P_\xi(x)$ with respect to x. In other words,

$$p_\xi(x)dx = Pr(x < \xi(\omega) < x+dx). \tag{5.2}$$

The CDF and the PDF are sometimes labeled as $P_\xi(\xi)$ and $p_\xi(\xi)$ to eliminate the use of the dummy variable x. Any random variable $\xi(\omega)$ is described completely by its probability

distribution, which can be empirical or estimated. Common forms for continuous PDF's include Gaussian, lognormal, uniform, beta, and exponential distributions. Gaussian distributions are frequently used to describe unbiased random experimental error. It is completely described by the mean and variance. The exponential of a Gaussian variable is distributed lognormally. Lognormal distributions are useful for describing random variables that can only take on positive values, such as aerosol size distributions, concentrations, and reaction rate constants. The lognormal distribution is completely described by the central value and the uncertainty factor, which are related to the mean and standard deviation of the Gaussian variable. A uniform distribution should be used when the range is only available information concerning a random variable. Equal probability is assigned to all values within the range. Another useful distribution with finite support is the beta distribution, which displays a mode value within a range. Four parameters are needed to describe the shape and the range of a beta distribution. The exponential distribution describes the frequency of random events. It is characterized by a single decay rate of the distribution. Other useful PDF's include the Poisson, Weibull, triangular, Gamma distributions. Further descriptions of these probability distributions can be found in Chapter 5 of Morgan and Henrion (1992).

Distributions of uncertain parameters can be estimated from experimental data using maximum entropy principle (Kapur, 1989). Given the range of values the random variable can take, and any available statistical measurements (e.g., mean or variance), a distribution can be derived which is the least biased and which is consistent with the available information. For example, the normal distribution is found to be the maximum entropy probability density function when the range of the random variable is $(-\infty,\infty)$, and the second moment is prescribed about a fixed point.

5.2 Traditional Methods for Uncertainty Propagation and Analysis

Since the input uncertainties propagate though the model to the output, when uncertain model inputs are represented by random variables, the model responses are corresponding random functions of the these variables in the probabilistic framework, as illustrated in Figure 5-1. In fact, the more uncertain inputs there are in the model, the higher the dimension of the function describing the output variables.

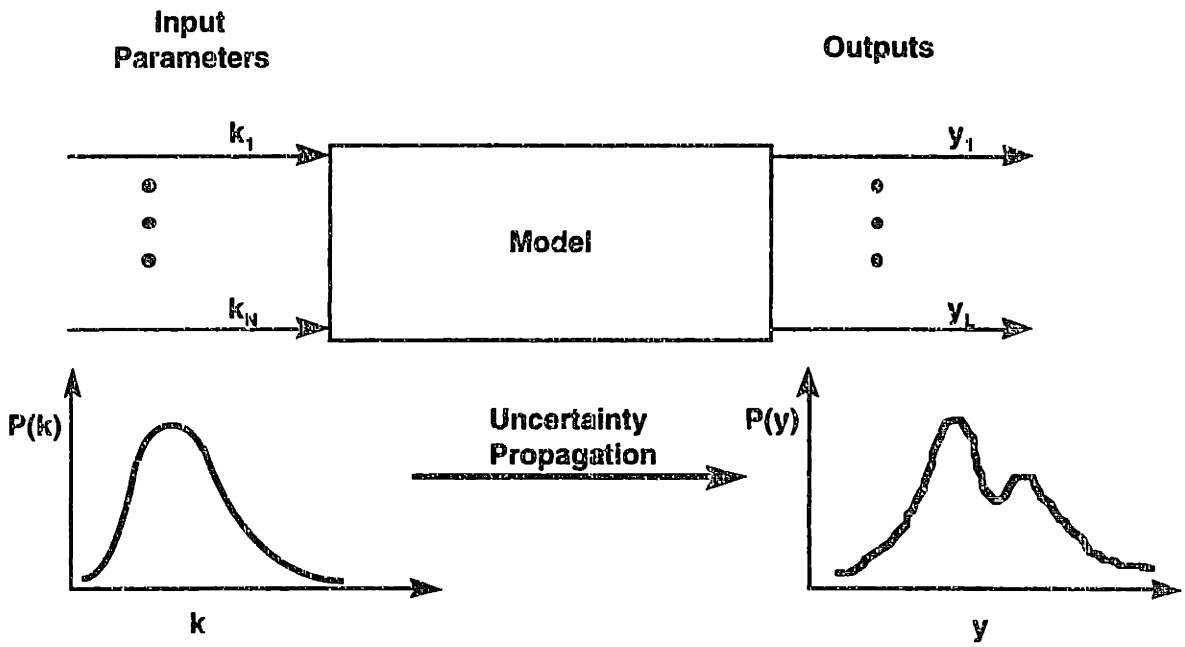


Figure 5-1. The Relationship between Uncertain model Inputs and Outputs.

The exact analytical functional form of the uncertain output variable in terms of the input random parameters is only solvable for the simplest systems. Most other analytical methods approximate. One common method used in uncertainty analysis is based on Taylor series expansion:

$$y - y_0 = \sum_{i=0}^{N} (k_i - k_{i0}) \left[\frac{\partial y}{\partial k_i} \right]_0 + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (k_i - k_{i0})(k_j - k_{j0}) \left[\frac{\partial^2 y}{\partial k_i \partial k_j} \right]_0$$
$$+ \frac{1}{3!} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{m=1}^{N} (k_i - k_{i0})(k_j - k_{j0})(k_m - k_{m0}) \left[\frac{\partial^3 y}{\partial k_i \partial k_j \partial k_m} \right]_0 + \ldots \quad (5.3)$$

From Equation 5.3, the moments of y can be derived. Though this approach is clear and relatively easy to perform numerically, the algebra gets messy when higher order terms are required for approximating complex models. Furthermore, this is basically a local approach, whose applications are limited to models with small uncertainties. Reliable estimates of the tails of the output cannot be obtained. Yang *et al.*(1995) and Gao *et al.* (1995) used the first

order Taylor series estimate (first term in Equation 5.3) as a measure of the variance contribution (VC) in uncertainty analyses:

$$VC_i = \left|\frac{\partial y}{\partial k_i}\right|^2 \cdot \sigma_i^2, \tag{5.4}$$

where σ_i^2 is the variance of the parameter k_i.

Methods such as Monte Carlo (MC) are used to obtain the output distributions (Morgan and Henrion, 1992, Chapter 8). In MC simulations, independent scenarios are formed by drawing at random from the distribution of each input. The model is run for each of these scenarios to generate a random sample of outputs induced by the sets of random inputs chosen based on their probability distribution. The properties of the output distribution, such as the moments and the confidence intervals, can be calculated from this sample of outputs. Standard statistical techniques can be used to estimate the accuracy of the PDF and other parameters. The number of model runs is a function of the variance of the output, although the output variance cannot be known *a priori* without adequate sampling of all the parameters. Straight MC may not guarantee random sampling of all parameters in a fixed number of sampling points. It is sometimes desirable to sample the parameter space more evenly, especially when good estimates of the mean / variance are needed from a small number of samples. In these cases stratified sampling can be employed. One example is the Latin Hypercube Sampling. Each input distribution is divided into N equally probable intervals. A single value is sampled at random from each of these intervals. A scenario is generated by selecting one value at random from each of the inputs, without replacement. Hence N scenarios are formed. Since the sample scenarios tend to be more evenly spread out over the input domain, the sample from each input will represent the mean, variance, and parameters more accurately than with truly random sampling, especially when the number of samples is small. However, the exact shape of the distributions cannot be represented accurately with a limited number of samples, whatever the sampling technique.

Yang *et al.* (1995) and Gao *et al.* (1996) combined the linear regression coefficient obtained from a stratified sampling experiment and the variance of the input to measure the contribution to the uncertain output (5.5). The limitations of this approach was its assumption that the model response was linear within the range of variation of the uncertain parameters.

$$VC_i = \frac{\left(\sigma_i / k_{i,\text{nominal}}\right)^2}{\left(\sigma_y / y_{\text{nominal}}\right)^2}\beta_i^2. \tag{5.5}$$

For large, computationally expensive models, response surface methods may be the most viable method because only a small number of runs is needed to provide a "reduced model" fit to the surface. The uncertain output characteristics can then be derived from the reduced model. The difficulty of using a reduced model is to determine how best to represent a model using a small number of samples of its response surface.

The Deterministic Equivalent Modeling Method (DEMM) focuses on the approximation of the multidimensional hypersurface by a set of orthogonal polynomials — polynomial chaos expansion (PCE). DEMM uses the Method of Residuals (MWR) to evaluate the coefficients of the expansion by transforming the model equation into a system of "deterministic

equations" that are equivalent in format to the original model equation, as explained in the next section. When a large number of parameters are involved in the model, a screening procedure can be superimposed on the standard DEMM to identify uncertain inputs that need to be modeled in detail. This procedure will be explained in Chapter 8 when DEMM is applied to atmospheric chemical mechanisms.

5.3 Problem-Specific Polynomial Chaos Expansion

The Deterministic Equivalent Modeling Method (DEMM) (Tatang, 1994; Tatang *et al.*, 1997) was developed as an efficient procedure for uncertainty analysis. It entails the application of polynomial approximation and the method of weighted residuals to systems of equations with stochastic variables. In Sections 5.3 and 5.4, the method will be described together with the solution procedure using the methods of weighted residuals (MWR), Galerkin and collocation.

For a single-parameter deterministic equation $N(u,x)\cdot u(x) = f(x)$, series expansions $\hat{u}(x) = \sum_{i=0}^{M} u_i g_i(x)$ based on different choices of orthogonal functions $g_i(x)$ form the backbone of "spectral methods" for solving equations. DEMM expands the use of spectral methods to stochastic systems. Consider the following single uncertain parameter problem:

$$N(u,\xi)\cdot u(\xi) = f(\xi), \tag{5.6}$$

where N operates on u, the dependent variable. ξ is a stochastic input variable described by $p_\xi(\xi)$, the probability distribution function. The solution for $u(\xi)$ can be approximated by a finite series $\hat{u}(\xi)$ of orthogonal functions, $g_i(\xi)$, $i = 0, ..., M$,

$$\hat{u}(\xi) = \sum_{i=0}^{M} u_i g_i(\xi). \tag{5.7}$$

In the stochastic representation, the set $\{g_i(\xi)\}$ ($i = 0, 1, ..., \infty$) forms a complete basis for the space containing all real random functions. A specific choice of $g_i(\xi)$ forms the polynomial chaos expansion (PCE) from Equation 5.7. Let $\hat{\Gamma}_i$ be the space of orthogonal polynomials in $\xi(\omega)$ of degree less than or equal to i. The set of all polynomials in $\hat{\Gamma}_i$ that are orthogonal to $\hat{\Gamma}_{i-1}$ is the i-th order polynomial chaos. The i-th order polynomial chaos is labeled as Γ_i. By choosing $g_i(\xi)$ to be the i-th order polynomial chaos, $\hat{u}(\xi)$ becomes the polynomial chaos approximation for the response variable $u(\xi)$ (Ghanem and Spanos, 1991). Polynomial chaos expansions are "problem specific" because of the definition of orthogonality in stochastic systems. Two stochastic functions $g_i(\xi)$ and $g_j(\xi)$ are said to be orthogonal when their inner product, defined using the distribution of the stochastic variable ξ as the weighting function, vanishes:

$$\int g_i(\xi) g_j(\xi) p_\xi(\xi) d\xi = 0 \qquad i \neq j. \tag{5.8}$$

For any probability density function, an algorithm called ORTHPOL (Gautschi, 1994) is available for the automatic generation of orthogonal polynomials. The underlying idea is

$g_{-1}(\xi) = 0$,
$g_0(\xi) = 1$, and
$$\int g_i(\xi) \cdot g_j(\xi) \cdot p_\xi(\xi) d\xi = 0, \qquad (5.9)$$

which permits $g_1(\xi)$, $g_2(\xi)$, $g_3(\xi)$ . . . to be generated successively. A recursive relationship can be used to generate a series of orthogonal polynomials using any given probability density function as the weighting function in the inner product definition (Equation 5.8).

Several properties of the polynomial chaos expansion make it useful in the solution of stochastic problems (Tatang, 1995). Given the PDF, $p_\xi(\xi)$, the set of orthogonal polynomials $g_i(\xi)$ is unique. The orthogonal expansion represented in (5.7) is also unique for the response variable $u(\xi)$ for the set of $\{g_i(\xi)\}$. Finally, the series is guaranteed to be convergent as long as $u(\xi)$ is continuous and polynomial approximatable, i.e., $\hat{u}(\xi) \dashrightarrow u(\xi)$ as $M \longrightarrow \infty$.

This method, and the associated properties are completely generalizable to systems with many stochastic parameters, with the condition that the parameters are independent, i.e.,

$$p(\xi_1, \xi_2, \ldots, \xi_N) = p(\xi_1)\, p(\xi_2) \ldots p(\xi_N). \qquad (5.10)$$

Assuming u is a function of N independent random variables, $u = u(\xi_1, \xi_2, \ldots, \xi_N)$, an M-th order polynomial chaos approximation $\hat{u}(\xi_1, \xi_2, \ldots, \xi_N)$ of u is written as:

$$\hat{u}(\xi_1,\xi_2,\ldots,\xi_N) = u_0 + \sum_{i=1}^{N} u_{i1} g_1(\xi_i) + \sum_{i=0}^{N} u_{i2} g_2(\xi_i) + \sum_{i=0}^{N}\sum_{j<i} u_{i1j1} g_i(\xi_i) g_j(\xi_2) + \sum_{i=0}^{N} u_{i3} g_3(\xi_i)$$

linear — second order — bilinear — third order

$$+ \sum_{i=0}^{N}\sum_{j<i} u_{i2j1} g_2(\xi_i) g_1(\xi_j) + \sum_{i=0}^{N}\sum_{j<i} u_{i1j2} g_1(\xi_i) g_2(\xi_j)$$

second order in ξ_i, first in ξ_j — first order in ξ_i, second in ξ_j

$$+ \sum_{i=0}^{N}\sum_{j\neq i}\sum_{k\neq j\neq i} u_{i1j1k1} g_1(\xi_i) g_1(\xi_j) g_1(\xi_k) + \text{ higher order terms.}$$

trilinear

(5.11)

To evaluate the response variable $\hat{u}(\xi)$, the deterministic expansion coefficients (e.g., u_i in Equation 5.7 for single random variable systems) need to be evaluated. In the DEMM, the method of weighted residuals is employed to transform the stochastic equation into a set of deteministic equations, the solution of which provides estimates for the coefficients u_i.

5.4 Methods of Weighted Residuals

The methods of weighted residuals (MWR) are very well developed for solving the coefficients of orthogonal basis function expansions in spectral methods (Villadsen and Michelsen, 1978). Least squares, Galerkin, collocation, finite differences can all be classified as variations of MWR. When a finite number of terms is used in the series expansion, the residual is defined as the difference between the exact solution and the series expansion. For the one-parameter case, described in Equation 5.6,

$$R(\{u_i\},\xi) = N(\hat{u},\xi)\cdot\hat{u}(\xi) - f(\xi). \tag{5.12}$$

If the expansion $\hat{u}(\xi)$ satisfies Equation 5.12 exactly, $R(\{u_i\}, \xi) = 0$. To evaluate the expansion coefficients using MWR, the inner product of the residual and a weighting function, $W_k(\xi)$, is set to zero:

$$\int R(\{u_j\},\xi)\cdot W_k(\xi)\cdot p_\xi(\xi)d\xi = 0, \tag{5.13}$$

where $p_\xi(\xi)$ is the probability distribution of ξ. In other words, the expected value of the product vanishes.

Two choices of weighting functions have been used in uncertainty applications. In the Galerkin approach, (Tatang, 94) the orthogonal trial functions, $g_k(\xi)$, are used as the weighting functions in Equation 5.13,

$$W_k(\xi) = g_k(\xi). \tag{5.14}$$

Therefore, the coefficients u_i are determined by forcing the residual to be orthogonal to the basis functions used in the expansion. A integral equation is set up for each of the M+1 basis polynomials used in Equation 5.7, including $g_0(\xi) = 1$. Thus, a system of M+1 deterministic equations results for M+1 unknowns. This system of equations can then be solved simultaneously for the coefficients u_i.

In the collocation approach (Tatang and McRae, 1995; Tatang *et al.*, 1997),

$$W_k(\xi) = \delta(\xi - c_k), \tag{5.15}$$

where $\delta(\xi - c_k)$ are Dirac delta functions at c_k (k = 1, ..., M+1). This forces the residual to vanish at c_k, the collocation points

$$R(\{u_j\},c_k) = 0, \tag{5.16}$$

thus satisfying the model exactly at $\xi = c_1$, $\xi = c_2$, ..., $\xi = c_{M+1}$, the collocation points:

$$N(\hat{u},c_k)\cdot\hat{u}(c_k) = f(c_k). \tag{5.17}$$

For an M-th order polynomial approximation, the collocation points $\{c_k\}$ are the roots of $g_{M+1}(\xi)$, the (M+1)-th order polynomial. Collocation points are chosen in a manner analogous to the Guassian quadrature method for evaluating integrals. This choice provides the best approximation to the response surface $u(\xi)$ with M+1 terms in the PCE approximation. The higher the order of the polynomial, the larger the range covered by the

collocation points. Equation 5.17 is exactly analogous to the model in Equation 5.6. Therefore, the deterministic model can be reused M+1 times to solve for $\hat{u}$ at each c_k, the k-th collocation point. The result is a set of M+1 deterministic equations for different c_k:

$$\hat{u}(c_1) = \sum_{i=0}^{M} u_i g_i(c_1)$$

...

$$\hat{u}(c_k) = \sum_{i=0}^{M} u_i g_i(c_k) \quad . \tag{5.18}$$

...

$$\hat{u}(c_{M+1}) = \sum_{i=0}^{M} u_i g_i(c_{M+1})$$

From these equations, u_0, ..., u_M can be solved. The collocation procedure can be applied to "black box" type models where the model equations are not known explicitly because it requires only the solution of the model at the collocation points.

The above discussion on the Galerkin and collocation methods for solving coefficients of the polynomial chaos expansion is generalizable to multiple dimensions representing different uncertain parameters. The choice of collocation points for higher order system warrants further discussion. Unless all the cross product terms are included, only selected pairs (sets for higher dimensions than two) of collocation points will be used to determine the PCE coefficients. A scheme was devised to choose systematically the collocation points used in the solution procedure. The procedure for a two parameter, first order system is illustrated in Figure 5-2. The collocation points for each parameter is placed in order of decreasing probability. In the case when the probability is equal (e.g., in a uniform distribution), the points are organized in increasing distance from the mean. The first pair of points, which contains the most probable values for all the parameters among the collocation points (c_1, c_3), is termed the anchor point. For each increasing order of approximation, the corresponding variable's collocation point is perturbed. Therefore, the pairs of points (c_1, c_3), (c_2, c_3), and (c_1, c_4) are chosen for an approximation which has a constant term and the first order terms in ξ_1 and ξ_2, as shown in Figure 5-2. If the bilinear term $H_1(\xi_1)H_1(\xi_2)$ is used in the approximation, the point (c_2, c_4) will also be used.

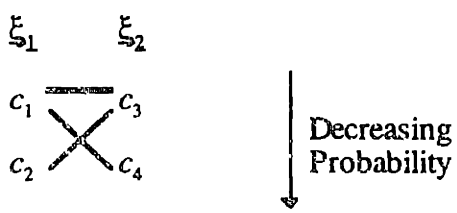


Figure 5-2. Choice of Collocation Points for a Two Parameter, First Order System.

5.5 Error Evaluation

If the response surface is smooth and polynomial-approximable, the truncation error of the convergent polynomial chaos series may be approximated by lowest order term excluded. Therefore, one estimate of the truncation error of the polynomial chaos series approximating the response surface is to compare the M-th order prediction to the (M+1)-th order prediction. A systematic way to evaluate errors of approximation is to evaluate the model at the collocations points corresponding to the (M+1)-th order approximation (i.e., at the roots of the (M+2)-th order orthogonal polynomial) and compare the exact results to the predictions obtained from the M-th order approximation at those points. The error at each higher order collocation point is defined as the square of the distance between the exact solution and the approximation:

$$\begin{aligned} &\varepsilon_i = d_i^2, \\ &\text{where } d_i = \hat{u}_i - u_i. \end{aligned} \tag{5.19}$$

Tatang (1995) defined the sum square root (*ssr*) error and the relative sum square root (*rssr*) error as:

$$\begin{aligned} ssr \text{ error} &= \sqrt{\frac{\sum_{i=1}^{N} f_\xi \varepsilon_i}{N f_\xi(\xi_{anchor})}}, \\ rssr \text{ error} &= \frac{ssr \text{ error}}{E(\hat{u})}. \end{aligned} \tag{5.20}$$

Here, a short hand notation is used such that ξ represents the set of uncertain parameters (ξ_1, ξ_2, ...). N is the number of collocation points for the higher order, f_ξ is the joint probability function at each collocation point, ξ_{anchor} is the anchor point, and $E(\hat{u})$ is the expected value of $\hat{u}$ calculated using the lower order approximation.

The error measures in Equations 5.20 can be used to guide the decision of whether more terms are needed in the PCE. However, the "acceptable level" of error cannot usually be pre-determined. The accuracy required for the response surface approximation depends on the goal of the analysis. For example, a larger error may be acceptable for the determination of the mean than the variance of an output variable. When the acceptable error can be determined, e.g. based on the convergence of the variance or another parameter-of-interest, it can be used as a target variable for an iterative procedure in the application of DEMM, as shown in Figure 5-3.

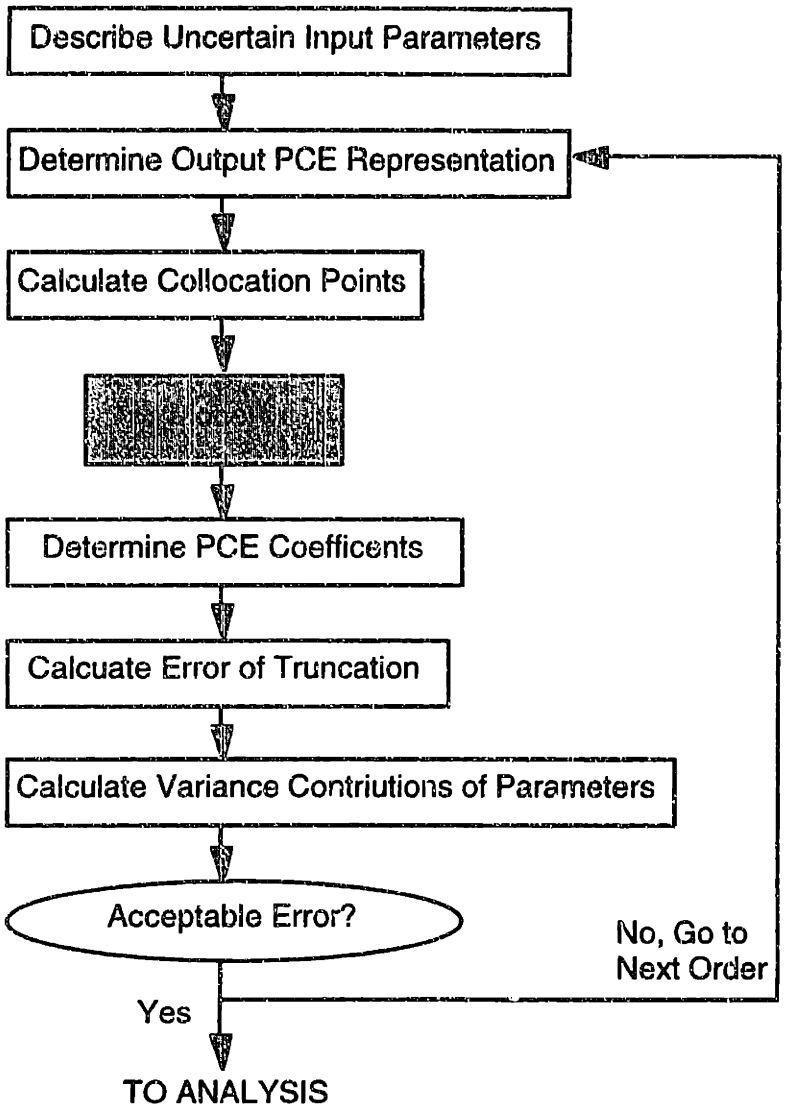


Figure 5-3. Recursive Procedure for Collocation Based on Error Evaluation.

This procedure is implemented in the Demmucom program by Tatang. Ideally, each increased order is "complete" in the sense that it includes all the cross terms. Complete representations guarantee convergence of the PCE series. Interactions between the parameters can also be elucidated. The order of approximation is increased until the improvement in error is negligible. However, the excessive number of model runs involved sometimes makes this approach infeasible, particularly when there are more than a few parameters. Furthermore, the difficulties in interpreting the data also makes such a comprehensive approach unattractive. It is possible to analyze the error contribution from each of the variables by evaluating the individual terms in Equation 5.19, and select variables that contribute to error as targets for improvement. Physical insights can also be used to guide the selection and use of cross product terms.

5.6 Analysis of the Polynomial Chaos Expansion

Properties of the stochastic output variable can be determined from the PCE approximation. Examples of properties useful for describing model outputs in an uncertainty analysis include the probability distribution, confidence intervals, moment information, and variance apportionment.

Probability Distribution and Confidence Intervals of the Output Variables. In theory, the PDF and the CDF of the response variables can be solved analytically from the PCE approximation. However, as the number of random inputs and the order of polynomial approximation increase, the analytical solution becomes impractical and involves multi-dimensional integrals. The simplest way to obtain the probability distribution of a response variable from the PCE representation is by Monte Carlo (MC) sampling of the PCE. In this regard, the PCE approximation can be viewed as a reduction of the original output variable. Where MC sampling of the original complex model is prohibitively expensive, MC sampling of a linear combination of algebraic terms containing the random input variables provides a viable alternative for understanding the behavior of the random output variable.

Both the CDF and PDF are generated in Demmucom by sampling the PCE empirically. To generate a CDF, the output samples are arranged in ascending order and the limits of each fractile are recorded. For a PDF, the range of the response variables is divided into bins or intervals and the frequency of occurrence in each interval is recorded based on the same procedures used to generate histograms.

The confidence intervals can also be determined from the list of ascending Monte Carlo empirical samples. For example, a 90% confidence interval will be the range of the empirical samples after the highest and lowest 5% of the samples are discarded.

Moments of the Output Distribution and Variance Contribution of Each Input Parameter. Two methods are available for determining the moments of the distribution. First, they can be determined empirically from the MC samples. Second, the moments can be calculated directly from the PCE coefficients, using the definition of the n-th central moment (cm_n):

$$cm_n = \int \cdots \int (u - m_1)^n f(\xi_1 \cdots \xi_N) d\xi_1 \cdots d\xi_N , \tag{5.21}$$

where m_1 is the first moment, the expected value of the distribution. The evaluation of moments is simplified by the orthogonal properties of the polynomials. For example, if

$$u(\xi_1, \xi_2) = u_0 + u_1 g_{11}(\xi_1) + u_2 g_{21}(\xi_2), \tag{5.22}$$

the mean value is equal to u_0, and the variance of the random variable is described by

$$\sigma^2 = A_1 u_1^2 + A_2 u_2^2, \quad \text{where}$$
$$A_1 = \int g_{11}^2(\xi_1) f(\xi_1) d\xi_1, \quad A_2 = \int g_{21}^2(\xi_2) f(\xi_2) d\xi_2 . \tag{5.23}$$

Higher moments can also be determined from the coefficients of higher order terms.

The relationship between the PCE coefficients and the variance suggests the utility of the PCE approximation for variance analysis. The contribution of each input parameter can be determined from the relevant terms in the approximation. In Equation 5.23, the variance

contribution (VC) from ξ_1 is $A_1\ u_1^2$, while the VC from ξ_2 is $A_2\ u_2^2$. Any cross terms are apportioned among the variables involved. This kind of analysis is particularly useful for identifying input variables whose uncertainties have strong effects on the uncertain outputs.

The next chapter presents a sample calculation in detail. The goal is to provide mathematical details on the DEMM (collocation) framework and to illustrate the uncertainty analysis methodology.

6. Illustrative Example: A —> B —> C

6.1 Parametric Uncertainties in Chemical Reaction Mechanisms

Consider the reactions series:

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \tag{6.1}$$

described mathematically by the following model:

$$\begin{aligned} \frac{dA}{dt} &= -k_1 A \\ \frac{dB}{dt} &= k_1 A - k_2 B \\ \frac{dC}{dt} &= k_2 B. \end{aligned} \tag{6.2}$$

Here, A, B, and C are used to denote the concentrations of the species. Initial conditions are needed to define the problem completely.

$$\begin{aligned} A(t=0) &= A_{in} \\ B(t=0) &= B_{in} \\ C(t=0) &= C_{in}. \end{aligned} \tag{6.3}$$

The concentrations of A, B, and C can be solved explicitly in terms of k_1, k_2, and initial concentrations A_{in}, B_{in}, and C_{in}. The solutions of the deterministic system are:

$$\begin{aligned} A(t) &= A_{in}\exp(-k_1 t) \\ B(t) &= B_{in}\exp(-k_2 t) + \frac{k_1 A_{in}}{k_1 - k_2}\left(\exp(-k_2 t) - \exp(-k_1 t)\right) \\ C(t) &= (A_{in} + B_{in} + C_{in}) - A_{in}\exp(-k_1 t) - B_{in}\exp(-k_2 t) - \frac{k_1 A_{in}}{k_1 - k_2}\left(\exp(-k_2 t) - \exp(-k_1 t)\right). \end{aligned} \tag{6.4}$$

In any mathematical model of chemical kinetics, uncertainties exist in the parameters. For example, the reaction rate constants, k_1 and k_2, may be independently measured or estimated, or they may be deduced from fits to concentration measurements based on the postulated model. In either case, k_1 and k_2 may not be perfectly known. Uncertain rate constants may be represented using random variables with finite or semi-infinite support. Since rate constants are strictly positive, common choices are lognormal, defined by the central value and uncertainty factor, or beta, with finite support.

6.2 Uncertainty Propagation Procedure

Applying DEMM-Collocation to a 2-Parameter System:

$A \xrightarrow{k_1} B \xrightarrow{k_2} C$

Consider the example where k_1 and k_2 are uncertain parameters with mean (nominal) values of 0.5 and 2.0 respectively. The initial conditions used in this example are exact:

$$\begin{aligned} A\ (t = 0) &= 100 \\ B\ (t = 0) &= 0 \\ C\ (t = 0) &= 0 \end{aligned} \qquad (6.5)$$

The concentrations of A, B, and C are uncertain because of the uncertain rate constants. This section illustrates how DEMM – collocation is used to find the approximations of A, B, and C in terms of the uncertain parameters, k_1 and k_2, and time using the procedure illustrated in Figure 5-3 in the Chapter 5.

Step 1. Specify Uncertain Parameters. The uncertain parameters are represented using random variables of specified probability distributions. In this example, the probability distributions of k_1 and k_2 are assumed to be independent and Gaussian. Gaussian distributions are not the best choice for describing parameters that are strictly positive, but are used in this illustration because of preferable properties of their orthogonal polynomials over those of lognormal or beta distributions. A Gaussian distribution is completely defined by the mean and standard deviation. In Equation 6.6,

$$\begin{aligned} k_1 &= k_{10} + k_{11}\xi_1 \\ k_2 &= k_{20} + k_{21}\xi_2, \end{aligned} \qquad (6.6)$$

k_{10} and k_{20} are the mean values of k_1 and k_2 respectively; and k_{11} and k_{21} are the standard deviations. ξ_1 and ξ_2 are standard normal distributions chosen as the independent random variables. If the standard deviations are arbitrarily assumed to be 0.1 and 0.5 for k_1 and k_2, the uncertain reaction rate constants can be expressed as:

$$\begin{aligned} k_1 &= \text{normal } (0.5,\ 0.1) \\ k_2 &= \text{normal } (2.0,\ 0.5), \end{aligned} \qquad (6.7)$$

where $k_{10} = 0.5$, $k_{20} = 2.0$, $k_{11} = 0.1$ and $k_{21} = 0.5$.

Step 2. Generate Problem-specific polynomial chaos expansions. Since the explicit forms of distributions of k_1 and k_2 are known, orthogonal polynomials $\{g_i\}$ can be generated such that the inner products, defined by $\int_{-\infty}^{\infty} g_i(\xi) g_j(\xi) p_\xi(\xi) d\xi$, are zero, where $p_\xi(\xi)$ is the PDF of the uncertain variable ξ. For standard normal distributions, the orthogonal Hermite polynomials define the polynomial chaos:

$$\begin{aligned} H_0(\xi) &= 1, \\ H_1(\xi) &= \xi, \\ H_2(\xi) &= \xi^2 - 1, \\ H_3(\xi) &= \xi^3 - 3\xi, \end{aligned} \qquad (6.8)$$

and higher order orthogonal polynomials $H_i(\xi)$.

Step 3. Approximate Uncertain Outputs Using Polynomial Chaos Expansion. The model outputs, concentrations $A(\xi_1,t)$, $B(\xi_1, \xi_2,t)$, $C(\xi_1, \xi_2, t)$, are expressed as linear combinations for the orthogonal polynomials determined in Step 2. These expressions are known as the polynomial chaos expansions (PCE) for the uncertain outputs:

$$A = A_0 + A_1H_1(\xi_1) + A_2H_1(\xi_2) + A_3H_2(\xi_1) + A_4H_2(\xi_2) + A_5H_1(\xi_1)H_1(\xi_2) + \ldots$$

constant linear terms second order terms bilinear term

$$B = B_0 + B_1H_1(x_1) + B_2H_1(x_2) + B_3H_2(x_1) + B_4H_2(x_2) + B_5H_1(x_1)H_1(x_2) + \ldots$$

$$C = C_0 + C_1H_1(\xi_1) + C_2H_1(\xi_2) + C_3H_2(\xi_1) + C_4H_2(\xi_2) + C_5H_1(\xi_1)H_1(\xi_2) + \ldots \quad (6.9)$$

The higher order terms are not listed explicitly in Equation 6.9. The concentrations of A, B, and C, and the coefficients, A_0, A_1, ..., B_0, B_1, ... , C_0, C_1, ... are all functions of time. The number of coefficients, hence the number of equations required to solve for them, is determined by the order of the polynomial approximation. In general, the higher the order of the approximation, the better the approximation. It is advisable to start with a low order expansion and to increase the order iteratively as needed. Linear PCE representations for A, B, and C, using Hermite polynomials, are listed in Equation 6.10:

$$A(\xi_1,\xi_2,t) = A_0(t) + A_1(t)\xi_1 + A_2(t)\xi_2$$
$$B(\xi_1,\xi_2,t) = B_0(t) + B_1(t)\xi_1 + B_2(t)\xi_2 \quad (6.10)$$
$$C(\xi_1,\xi_2,t) = C_0(t) + C_1(t)\xi_1 + C_2(t)\xi_2.$$

Step 4. Find Collocation Points. The objective here is to find the best coefficients, A_0, A_1, A_2, B_0, B_1, B_2, C_0, C_1, and C_2, for the approximation (6.10) using the fewest number of model runs. The values of ξ_1 and ξ_2 chosen for these runs are the collocation points. For a linear approximation, the collocation points are determined by the roots of the second order polynomial, as explained in Chapter 5. $H_2(\xi) = \xi^2 - 1$; therefore, $\xi = \pm 1$.

Since ξ_1 and ξ_2 are normal gaussians and are symmetric about zero, the four pairs of collocation points (± 1, ± 1) are equal in probability and equal in distance to the mean (0, 0). The point (ξ_1, ξ_2) = (1, 1) is arbitrarily designated as the anchor point, which should normally be the point with the highest probability. The points (-1, 1) and (1, -1) are also chosen, where one coordinate is perturbed, corresponding to the linear term in each uncertain variable. These correspond to (k_1, k_2) pairs of ($k_{10}+ k_{11}$, $k_{20}+ k_{21}$), ($k_{10}- k_{11}$, $k_{20}+ k_{21}$), and ($k_{10}+ k_{11}$, $k_{20} - k_{21}$), as listed in Table 6-1.

	k_1	k_2
anchor point	0.6	2.5
first perturbation of k_1	0.4	2.5
first perturbation of k_2	0.6	2.0

Table 6-1. Collocation Points for First Order Approximation

Step 5. Solve the Model at the Collocation Points. The model is formulated to take k_1 and k_2, in addition to the initial conditions as inputs. Solutions for $A(t)$, $B(t)$, and $C(t)$ are evaluated by substituting the pairs of (k_1, k_2) found in Step 4 into model equations (Equations 6.2 or 6.4). Due to the property that the model is exactly satisfied at the collocation points, the original model solver can be used "as is" to calculate the solution as functions of the parameters. Thus DEMM – collocation is ideal for large, black box type models with implicit model equations or complex model structure.

Step 6. Solve for the Coefficients of Expansion from Model Results. The model solutions (6.4) for $A(t)$, $B(t)$, and $C(t)$ are then equated to simple algebraic equations in (6.10) for each of the collocation points (k_1, k_2) listed in Table 6-1. The resulting time-dependent equations are transcendental in nature, so the coefficients A_0, A_1, A_2, B_0, B_1, B_2, C_0, C_1, and C_2 are evaluated numerically at selected time points. At each time point, 9 algebraic equations are solved simultaneously for the 9 unknowns. The solutions of the example are shown in Figure 6-1 for the first order PCE coefficients. Since the concentration of A does not depend on the uncertain rate constant k_2, the coefficient $A_2(t)$ in Equation 6.10 is exactly zero at all times.

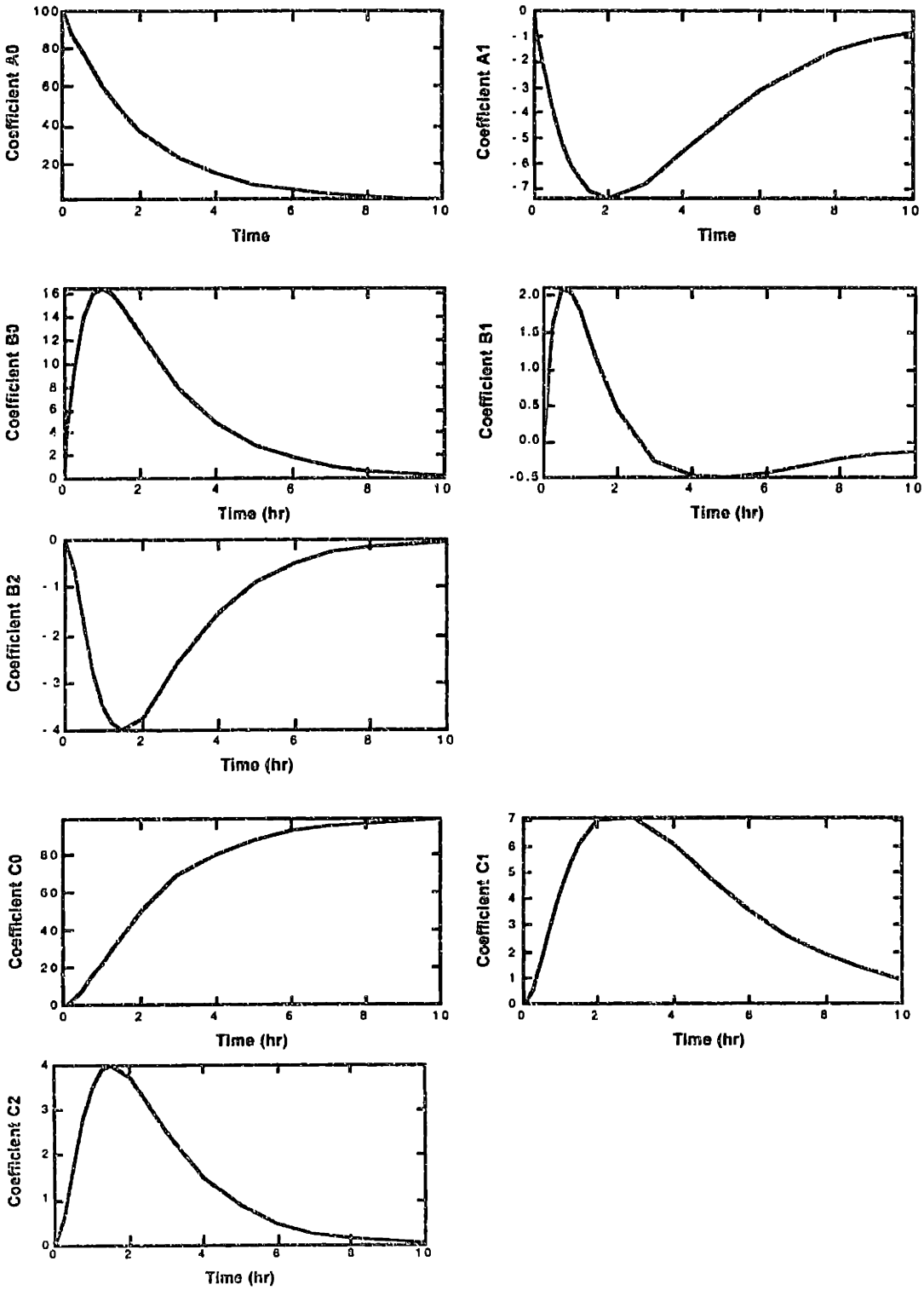


Figure 6-1. Plots for Coefficients A_0, A_1, B_0, B_1, B_2, C_0, C_1, C_2 as Functions of Time

Step 7 Estimate the Error of Approximation. The error of the linear PCE can be evaluated for each species at any given time based on the solutions at the roots of the third order Hermite polynomial (collocation points for the second order PCE). From Equation 6.8, the roots to the third order approximation are:

$$\xi = 0,$$
$$\xi = \sqrt{3}, \qquad (6.11)$$
$$\xi = -\sqrt{3}.$$

Therefore, the collocation points for the second order evaluation are:

	k_1 (ξ_1)	k_2 (ξ_2)
(P1) anchor point	0.5 (0)	2.0 (0)
(P2) first perturbation of k_1	0.67 (1.7)	2.0 (0)
(P3) second perturbation of k_1	0.33 (-1.7)	2.0 (0)
(P4) first perturbation of k_1	0.5 (0)	2.86 (1.7)
(P5) second perturbation of k_2	0.5 (0)	1.13 (-1.7)

Table 6-2. Second Order Collocation Points

Since a PCE is calculated at each time point for each species, the error of the approximation as calculated using Equation 5.20 depends on the species and the time of the simulation. A selected list of errors for species A, B, and C at various times throughout the simulation is given in Table 6-3.

Time	Error of *A* (ssr)	Error of *A* (rssr)	ssr (B)	rssr (B)	ssr (C)	rssr (C)
0.5	0.0783	0.10%	0.2832	2.02%	0.2816	3.52%
1.0	0.2446	0.40%	0.5989	3.64%	0.5833	2.58%
2.0	0.5990	1.60%	0.8180	6.49%	0.8094	1.62%
4.0	0.9150	6.25%	0.5602	11.52%	0.9044	1.12%
6.0	0.8056	13.65%	0.2743	15.49%	0.8328	0.90%
7.0	0.6921	18.26%	0.1884	17.48%	0.7368	0.77%
10.0	0.3673	35.33%	0.0703	27.34%	0.4153	0.42%

Table 6-3. Errors for A, B, and C at Various Times of the Simulation.

An example error calculation is shown in Table 6-4 for species B at time 1.0 units (when the concentration of B is maximum). The approximation of B is given by Equation 6.12 using the coefficients plotted in Figure 6-1.

$$B\,(t = 1.0) = 16.4583 + 1.7671\ \xi_1 - 3.4867\ \xi_2. \qquad (6.12)$$

If the peak concentration of B is the output parameter of interest, Table 6-3 shows that the relative error of the linear approximation of the response surface is about 4%. It should be noted, however, that such a percentage number by itself is sometimes not a good indication of the "goodness" of the approximation. When the expected value is close to zero, the rssr given by Equation 5.20 tends to grow unbounded. Therefore, caution should be used in interpreting the error estimates.

	P1	P2	P3	P4	P5
Exact solution (Eqt. 6.4)	15.7065	18.9602	11.5320	11.6369	22.4998
Approximate Solution (Eqt. 6.12)	16.4583	19.5191	13.3975	10.4192	22.4974
Error at Point (d_i)	-0.7518	-0.5589	-1.8655	1.2177	0.0024
Probability of Point	0.1591	0.0356	0.0356	0.0356	0.0356
Expected Value of B	16.4583				
Error (ssr)	0.591				
Error (rssr)	0.036				

Table 6-4. Calculation of Error for B (t = 1.0)

Step 8. Increase the Order of Approximation. One way to decrease the error of the approximation, and hence of the uncertainty estimates, is to increase the order of the polynomial chaos approximation. Including higher order terms and cross product terms have the obvious utility of capturing curvature of the response surface better. There is an additional advantage. Based on the choice of collocation points, as described in Step 4, increasing the number of terms also increase the "spatial coverage" of the collocation points, making the estimate valid over a wider range of values of the uncertain inputs.

In the example A —> B —> C, the rssr error gives a reasonable error estimate for the peak value of the intermediate B. Table 6-5 presents the errors associated with different orders of approximation for this variable. Note that the timing of the peak may change as a result of different orders of approximations, although finer time resolution is needed to observe this effect.

Approx. Order	No. of Runs (incl. error evaluation)	Error (rssr)	Error (ssr)
1	8	3.64%	0.599
2 (square terms only)	12	1.69%	0.271
2 (complete)	14	0.87%	0.139
3 (complete)	22	0.12%	0.019
4 (complete)	32	0.02%	0.004
5 (complete)	44	0.003%	0.0005

Table 6-5. Errors of B (t = 1) for Different Orders of Approximation

6.3 Comparison with Other Uncertainty Analysis Methods

One goal of uncertainty analyses is to characterize uncertain outputs of the models. The results of DEMM – collocation (linear approximation) are discussed here for a sample case where $k_1 = 0.5 \pm 0.1$, $k_2 = 2.0 \pm 0.5$. Initial conditions are exact: $A_0 = 100$, and $B_0 = C_0 = 0$. In this section, the results of DEMM – collocation will be compared to the analytical solution (Section 6.3.1) and the Monte Carlo method (Section 6.3.2), which provide standards to which the new uncertainty analysis method is compared. A brief comparison will also be made with the results from DEMM – galerkin in Section 6.3.3. The predictions of the mean value (Equation 6.13), variance (Equation 6.14), and probability density functions (PDF) of the response variables by different methods will be discussed. PDF's are obtained from the results of 5000 Monte Carlo runs of the PCE, choosing values of k_1 and k_2 based on their respective normal distributions. The range covered by the Monte Carlo results are divided into 50 bins, and the frequency of occurrences in each bin is counted, as in the procedure used to determine a histogram. The PDF is obtained by nomalizing the area under the histogram to one.

$$E(A) = \int_{-\infty}^{\infty} A \cdot p_A(A) dA. \tag{6.13}$$

$$Var(A) = \int_{-\infty}^{\infty} (A - E(A))^2 p_A(A) dA. \tag{6.14}$$

6.3.1 Comparison: Analytical Solution

When a random variable is a function of one uncertain input variable, the procedure for deriving an analytical stochastic solution of the output in terms of the input is relatively simple. From Equation 6.4, $A = A_{in} \exp(-k_1 t)$. Since A is a monotonic function of only one uncertain parameter, k_1, there is a one-to-one correspondence between k_1 and $A(t_1)$ at any time t_1:

$$k_1 = \frac{1}{t_1} \ln \frac{A_{in}}{A}. \tag{6.15}$$

The analytical solution for the PDF of $A(t_1)$ in terms of the PDF of k_1 is readily available at any time t_1. Consider the cumulative distribution function (CDF) $P_A(A)$. The probability of A being less than a is given by:

$$P_A(a) = P(A \le a). \tag{6.16}$$

From Equation 6.15, for $A(t_1)$ to be less than a, k_1 has to be greater than the corresponding value k_{1a}, which is equal to $\frac{1}{t_1} \ln \frac{A_{in}}{A}$. Therefore,

$$P(A \le a) = P(k_1 \ge k_{1a}) = P(k_1 \ge \frac{1}{t_1} \ln \frac{A_0}{a}). \tag{6.17}$$

Using the definition of CDF, Equation 6.17 can be rewritten as:

$$P_A(A) = 1 - P_k(k_1)\Big|_{k_1=\frac{1}{t_1}\ln\frac{A_0}{A}}, \text{ such that}$$

$$P_A(A) = 0 \qquad \text{for } A \le 0$$
$$P_A(A) = 1 \qquad \text{for } A \ge A_0. \tag{6.18}$$

The probability density function (PDF) is equal to the derivative of the cumulative density function. As discussed in Papoulis, 1991, Chapter 5, for a single valued function, such as $A = A(k_1)$ at time t_1, the probability distribution $p_A(A)$ of the response variable A is expressed in terms of the probability distribution $p_{k1}(k_1)$ of the input variable k_1 as follows.

$$\begin{aligned} p_A(A) &= \frac{dP_A(A)}{dA} \\ &= -\frac{dP_k(k_1)}{dA} \\ &= -\frac{dP_k(k_1)}{dk_1}\cdot\frac{1}{\left.\frac{dA}{dk_1}\right|_{t=t_1}} \\ &= -p_{k1}(k_1)\cdot\left(-A_0 t_1 \exp(-k_1 t_1)\right)^{-1} \\ &= \frac{1}{At_1}\,p_{k1}\left(\frac{1}{t_1}\ln(\frac{A_0}{A})\right). \end{aligned} \tag{6.19}$$

The mean and variance of A at any given time can be calculated using the definition of the moments, Equations 6.13 and 6.14. The standard deviation is the square root value of the variance.

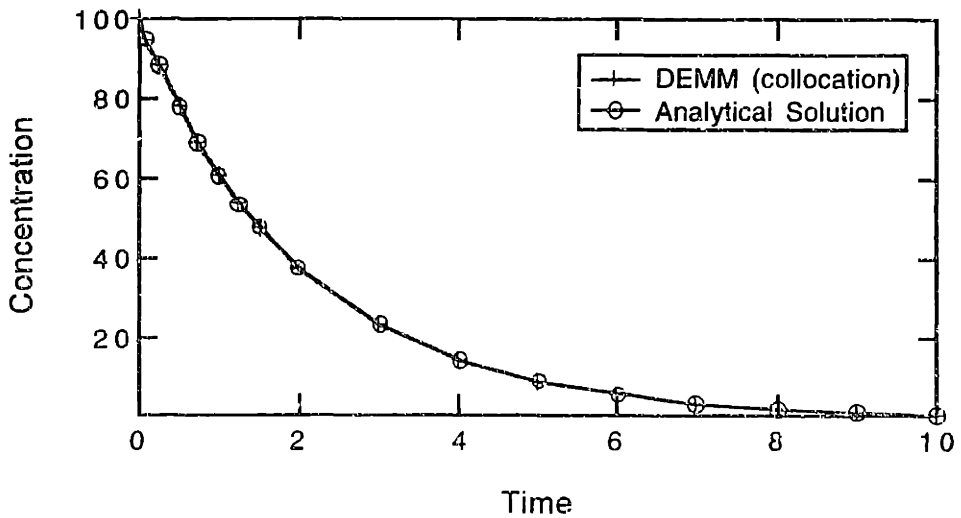


Figure 6-2. Comparison between Predictions of DEMM – Collocation (Linear) and Analytical Solutions of the Expected Value of A.

With just two model runs, DEMM can estimate the first central moment (mean) with remarkable accuracy (Figure 6-2). Figure 6-3 compares the predictions of the standard

deviation time series. DEMM tends to underpredict the standard deviation of the probability distribution, especially at the later times when the concentration of B is low. The largest error in the standard deviation estimates is roughly 30%, indicating the limitations of a first order approximation to the surface. However, the qualitative behavior of the system response is captured well.

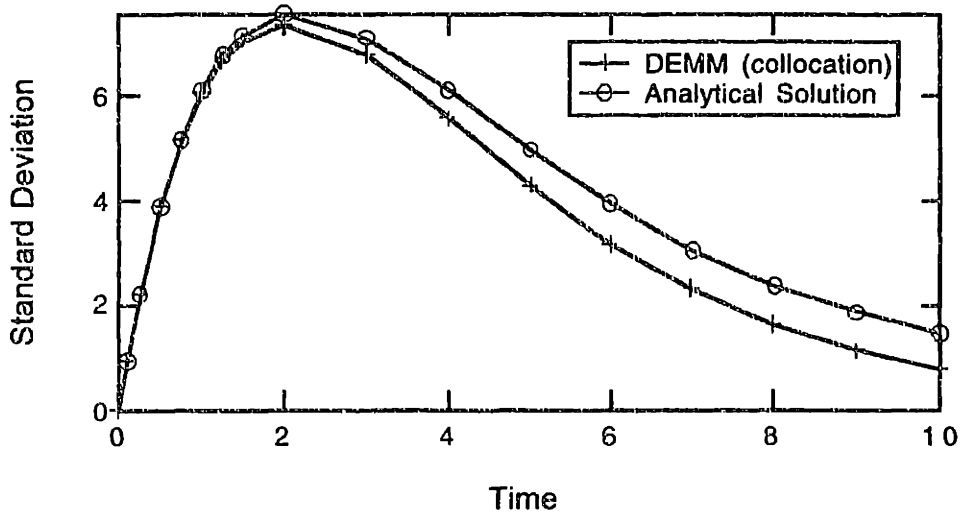


Figure 6-3. Comparison between Predictions of DEMM – Collocation (Linear) and Analytical Solutions of the Standard Deviation of A.

Several probability distribution functions are plotted in Figures 6-4, 6-5 and 6-6 for the concentration of A at time of 1, 2, and 6 units. In each figure, the analytical solution is plotted in smooth line and the PDF approximation from DEMM, generated using Monte Carlo sampling on the PCE representation, is shown in "cityscape".

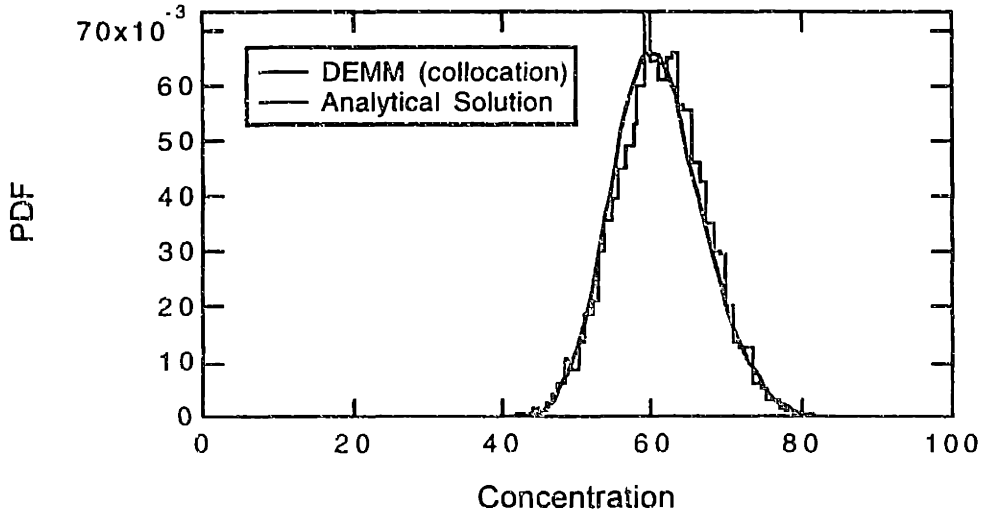


Figure 6-4. Comparison between Predictions of DEMM – Collocation (Linear) and Analytical Solutions of the PDF of A at time = 1.

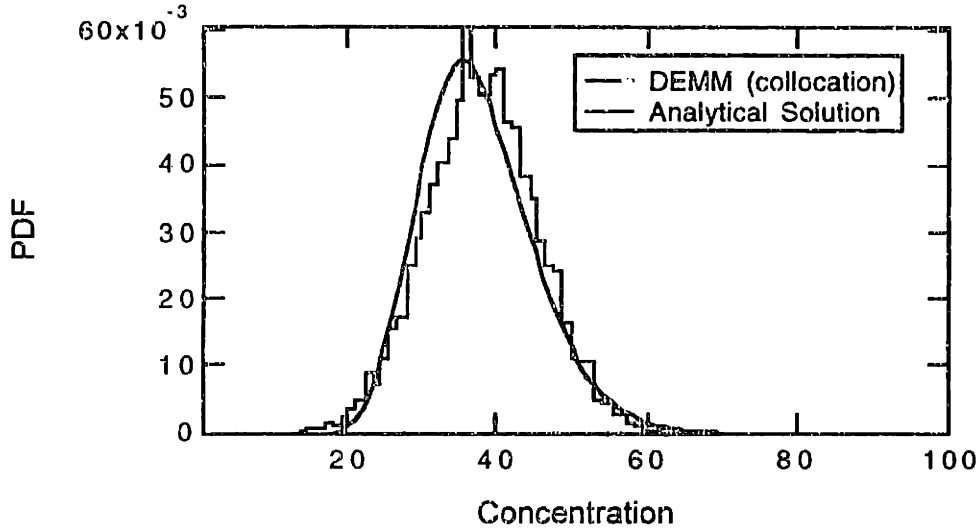


Figure 6-5. Comparison between Predictions of DEMM – Collocation (Linear) and Analytical Solutions of the PDF of A at time = 2.

At the first time point shown (Figure 6-4), the distribution is very symmetric and well-approximated by the linear polynomial chaos expansion of gaussian random functions used in DEMM. At time = 2, DEMM predicts a more symmetric probability function than the analytical solution, even though the width of the distribution, hence the standard deviation, still matches the analytical solution reasonably well. At time = 6, the probability density function of A is highly skewed, resulting in a relatively large second moment associated with the small mean value. The linear approximation of the DEMM does not predict highly skewed distribution satisfactorily (Figure 6-6). The modes of the distributions are obviously different. DEMM underestimates the probability density in tail of the of the distribution. The lack of accuracy is also reflected in the high relative error (rssr) values listed in Table 6-3. The selection of collocation points close to the center of the distributions for a linear PCE approximation is likely to be responsible for poor approximations of distributions with long tails.

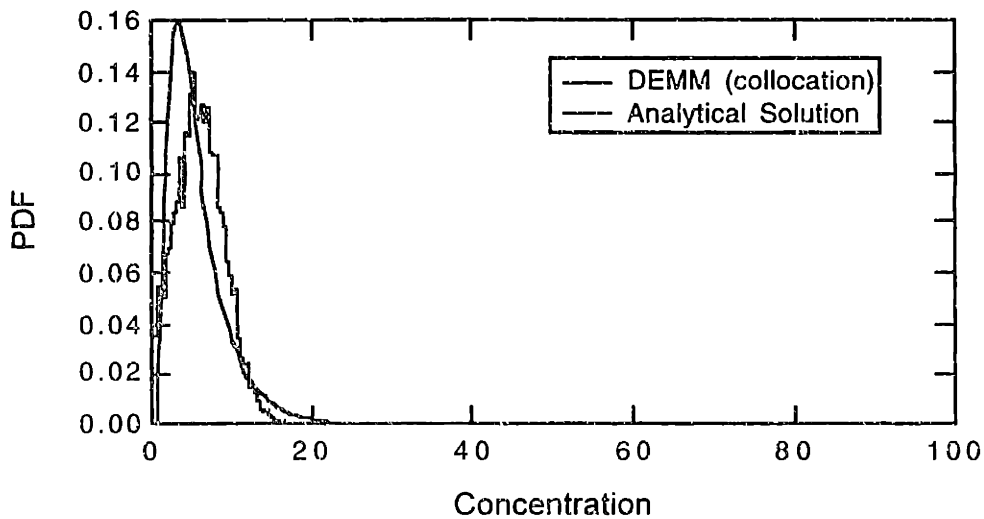


Figure 6-6. Comparison between Predictions of DEMM – Collocation (Linear) and Analytical Solutions of the PDF of A at time = 6.

As explained in the previous section, the error of the approximated response surface can be reduced by increasing the order of the polynomial used in the PCE representation. This approach is illustrated as follows. In Figure 6-7, the analytical probability density function for A at time = 2 is compared to the collocation results using higher orders of approximation. While the linear collocation underestimates the skewedness of the density function, a third order PCE (with four terms) approximates the analytical response very closely. Further increase in the number of terms used in the approximation results only in small improvements of the response variables. The same comparison is made in Figure 6-8 for the concentration of A at time = 6. The asymmetric PDF is poorly approximated by a linear polynomial. However, its shape is captured by a third order approximation, with runs at four collocation points for A, which is a function of k_1 only. The fourth order approximation, with 5 degrees of freedom, improved the approximation further.

Comparing DEMM solution to the analytical solution for the concentration of A, it is concluded that DEMM can be used to approximate uncertain outputs based on expansions of input parameters. A linear representation can predict the expected values of the uncertain concentration of A reasonably well. Higher orders of approximation are needed for higher moments and for probability density functions, especially skewed PDF's.

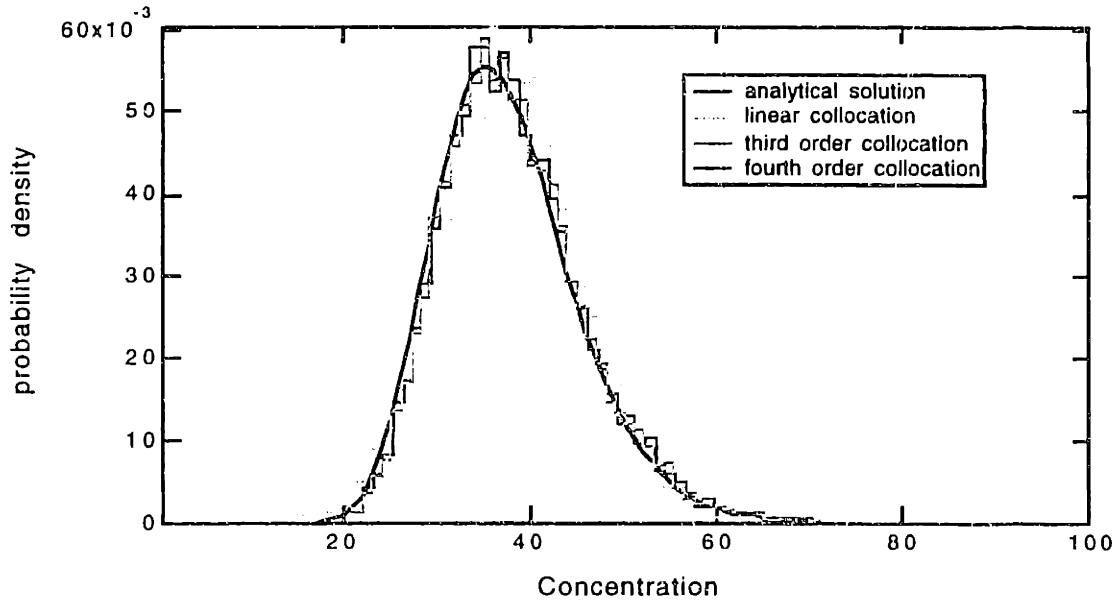


Figure 6-7. Comparison between Predictions of DEMM – Collocation (Linear and Higher Orders) and Analytical Solutions of the PDF of A at time = 2.

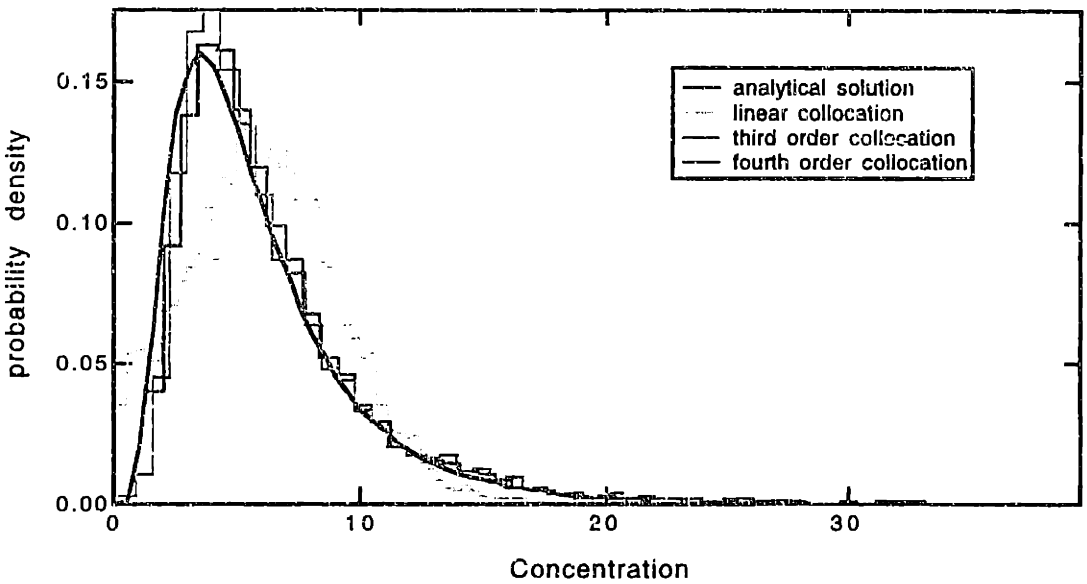
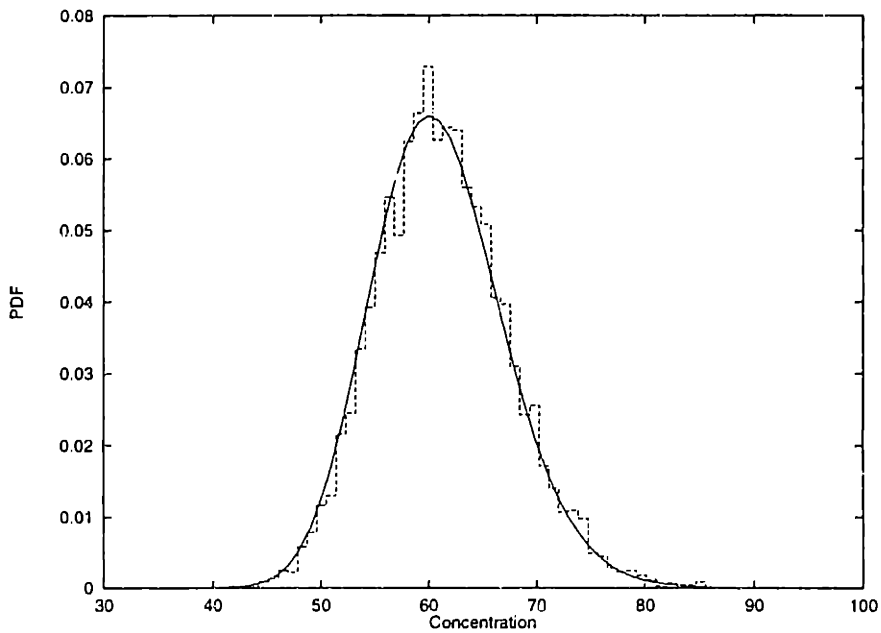


Figure 6-8. Comparison between Predictions of DEMM – Collocation (Linear and Higher Orders) and Analytical Solutions of the PDF of A at time = 6.

6.3.2 Comparison to Monte Carlo

For complex systems with multiple random variables, analytical solutions for uncertain outputs cannot be easily obtained. The traditional technique for uncertainty analysis in these complex models is the Monte Carlo (MC) simulation. The model is run repeatedly using different values of the uncertain parameters chosen at random based on the distribution. The collection of results is then analyzed to extract the moment information and the distribution of the response variables. It is not difficult to imagine how expensive Monte Carlo experiments are for larger models and for models with more uncertain inputs, as the number of runs grows as N^N to provide adequate sampling of each of N random variables. This section focuses on species B and C, for which analytical solutions are not as readily available. The results of a 5000 point Monte Carlo simulation are compared to that of the DEMM solved by collocation with a much smaller number of runs. A MC experiment with 5000 points achieved convergence in the shape of the predicted PDF. The MC results agree very well with the exact (true) PDFs for species A (See Figure 6-9 for different time points), and is expected to predict the PDFs of B and C with similar accuracies, assuming k_2 is adequately sampled and the variances in the concentrations of B and C are no bigger than that of A.



(Figure 6-9a Comparison of Analytical Solution and Monte Carlo for PDF of Species A at Time = 1)

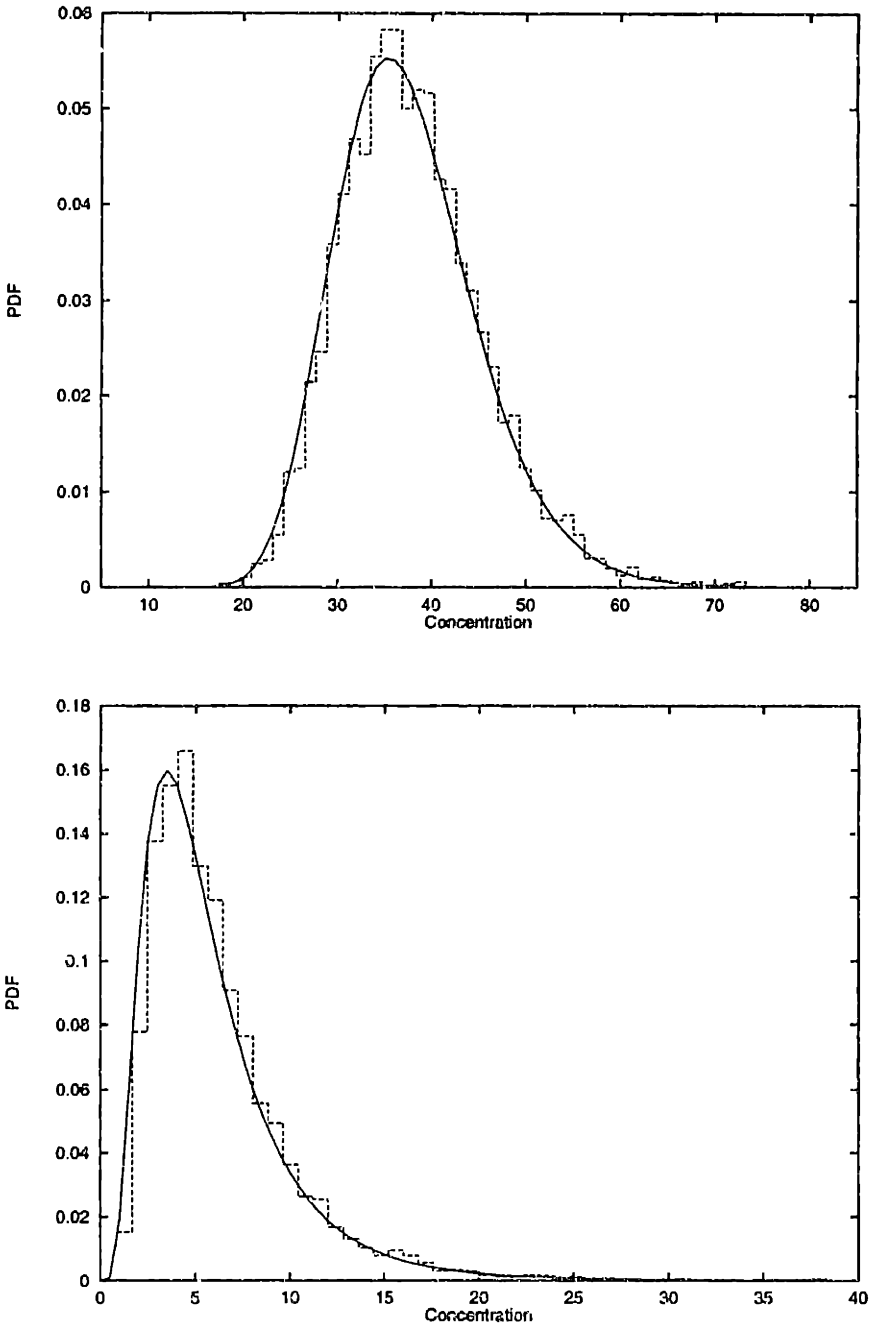


Figure 6-9. Comparison of Analytical Solution and Monte Carlo for PDFs of Species A (analytical solution = line; MC = histogram) (a) time = 1, (b) time = 2, (c) time = 6.

The mean values of $A(t)$, $B(t)$, and $C(t)$ calculated using DEMM and MC are plotted in Figure 6-10. The first moment predictions are found to be almost indistinguishable between the linear approximation of DEMM – collocation and the Monte Carlo simulation. DEMM predicts a slightly slower decay of B from the peak value. Since the MC and the analytical results for $A(t)$ are similar, the MC-collocation comparison of standard deviation predictions for species A is very similar to Figure 6-3, with DEMM underpredicting the standard deviation of A at later times. Figure 6-11 shows the predictions of the standard deviations of the two methods for species B. DEMM underpredicts the standard deviations associated with the responses $B(t)$ with respect to MC, as it does in the case of $A(t)$. The collocation method

also missed the point of highest absolute uncertainty. This underprediction at and after the point of highest uncertainty is also true for species C, as shown in Figure 6-12. The poor approximation of distributions with long tails is consistent for the concentrations of species A and B when they are close to depletion, and C when it approaches 100 asymptotically.

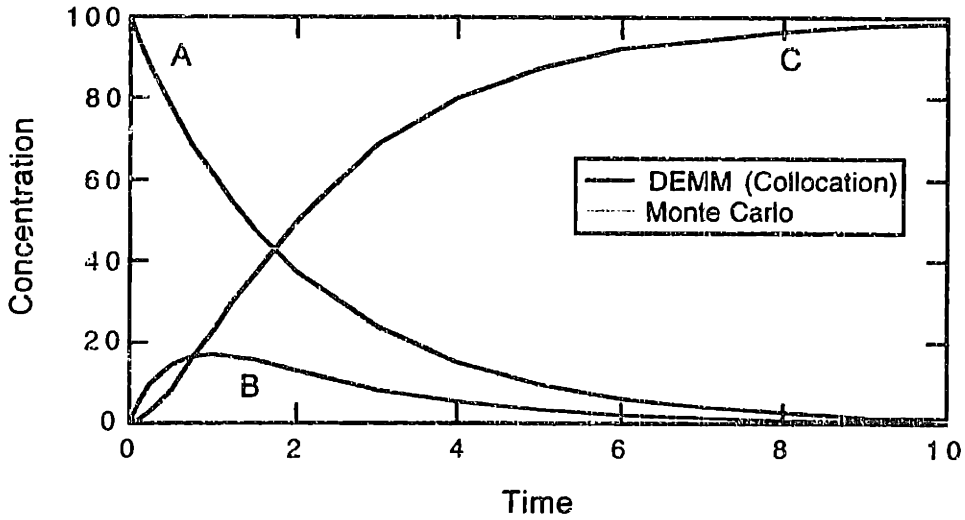


Figure 6-10. Comparison of Monte Carlo and DEMM – Collocation Predictions of Mean Values of A, B, and C.

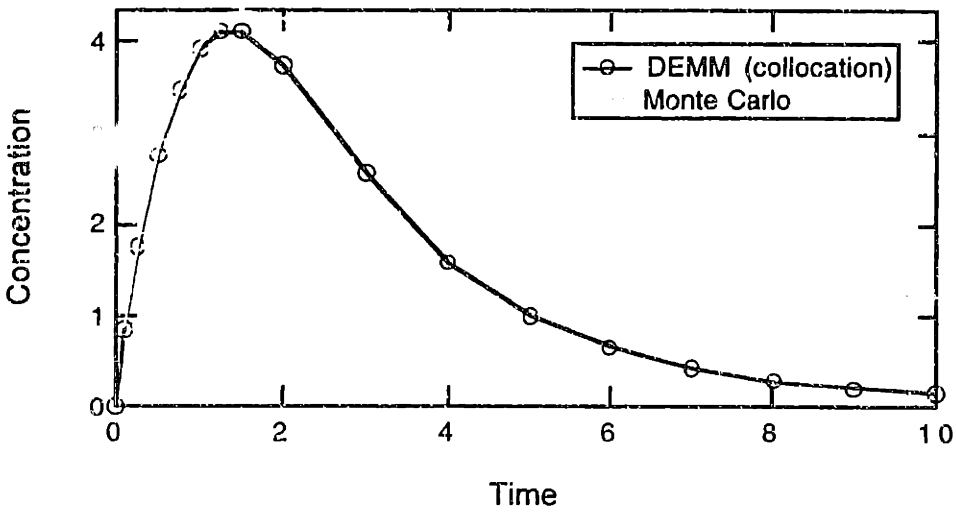


Figure 6-11. Comparison of Monte Carlo and DEMM – Collocation Predictions of Standard Deviations of B.

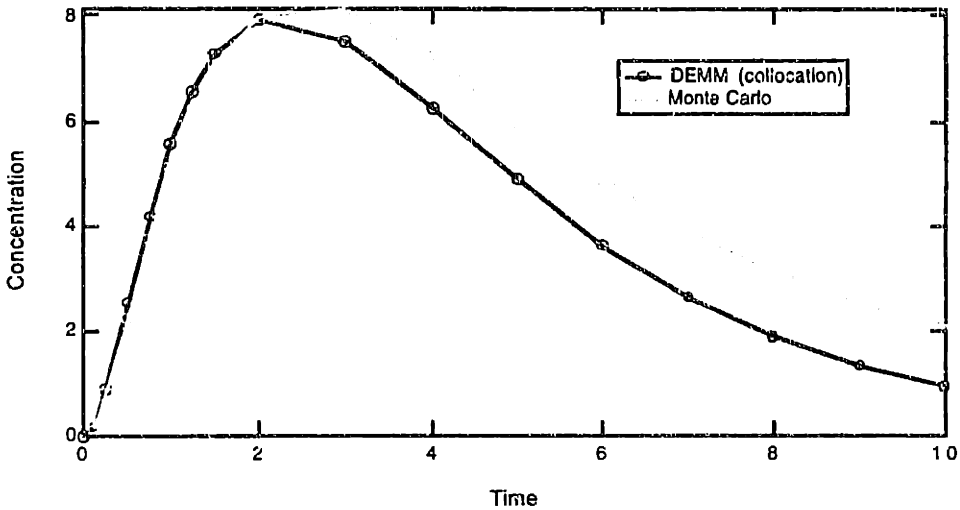


Figure 6-12. Comparison of Monte Carlo and DEMM – Collocation Predictions of Standard Deviations of C.

Figures 6-13, 6-14, and 6-15 show comparisons of the probability density functions predicted by the Monte Carlo method and the DEMM for species B and C at different times. The comparison for species A is very similar to those described in Section 6.3.1, and is not repeated here.

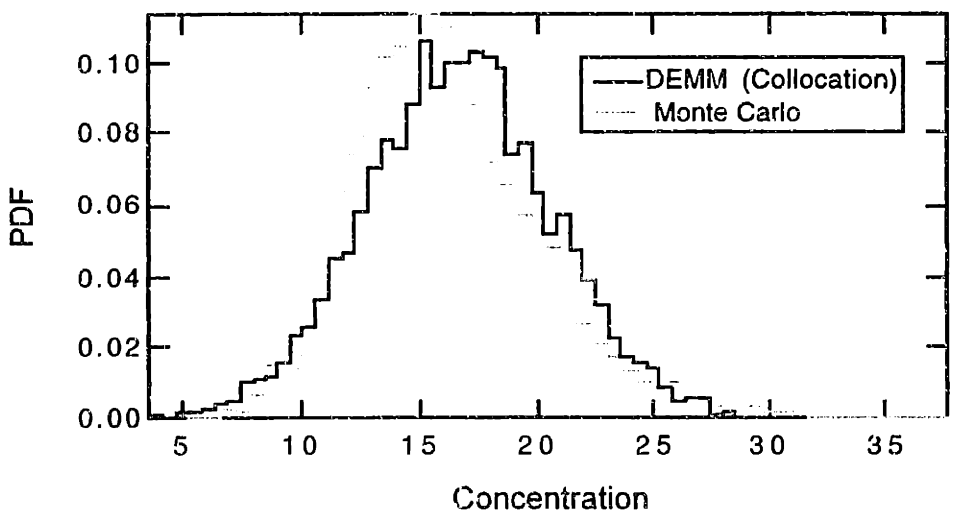


Figure 6-13. Comparison of Monte Carlo and DEMM – Collocation (linear) Predictions for PDF of B (time = 1).

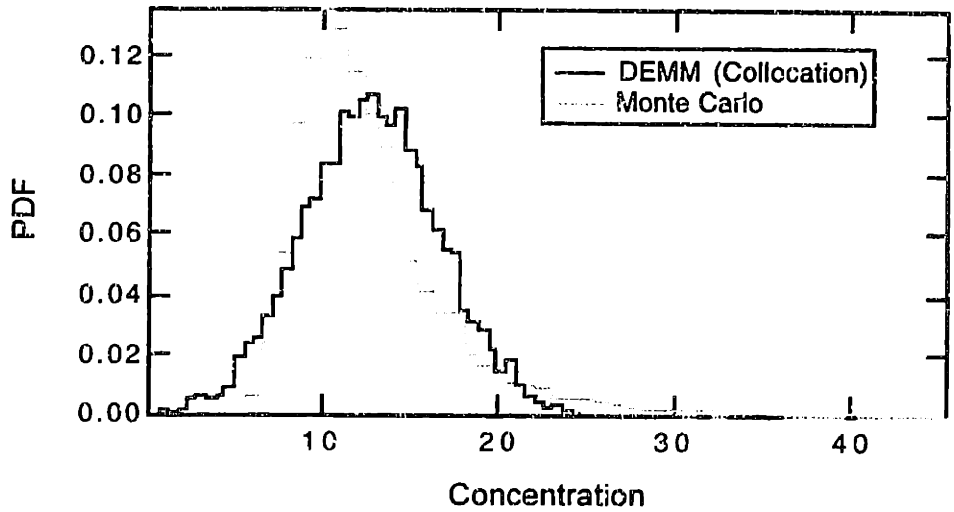


Figure 6-14. Comparison of Monte Carlo and DEMM – Collocation (linear) Predictions for PDF of B (time = 2).

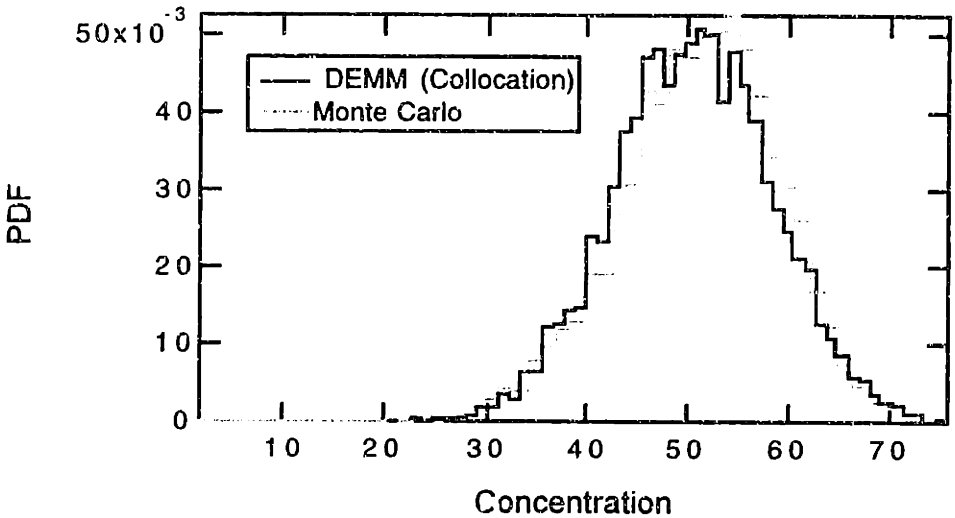


Figure 6-15. Comparison of Monte Carlo and DEMM – Collocation (linear) Predictions for PDF of C (time = 2).

At time = 1, the distributions of B, shown in Figure 6-13, show similar range and width. However, the mode value predicted by DEMM is slightly higher than MC and MC predicts a more skewed distribution. The probability densities of B predicted by MC and DEMM at time = 2, when the concentration of B is most uncertain, are shown in Figure 6-14. The range of values covered by the Monte Carlo simulation is clearly larger than that of DEMM. The long tail of the PDF prediction by MC also results in a higher value of standard deviation. The modes of the distributions are quite different, as are the shapes. Figure 6-15 shows the distributions of $C(t = 2)$. The linear collocation representation again underestimates the skewedness of the PDF. The long tail predicted by MC towards the lower edge of the distribution, already evident at this early time, mirrors the tails of distributions of A and B.

The previous section illustrates that a linear approximation (with two terms for A) does not capture the shape of the PDF's of $A(t)$ at increasing time. The comparisons of the PDF's

confirm that linear PCE representations are not adequate for predicting $B(t)$ and $C(t)$ at several time points. As in the case of species A, increasing the order of the polynomial expansion improves the approximation of highly skewed probability density functions. Figure 6-16 shows that the third order collocation solution approximates the 5000-point Monte Carlo very well for species B at time = 2, where a first order approximation proves unsatisfactory in Figure 6-14. Increasing the approximation to fourth order results in only slight improvements. Similar improvements are observed for species C and at other time points as the order of PCE approximation increases.

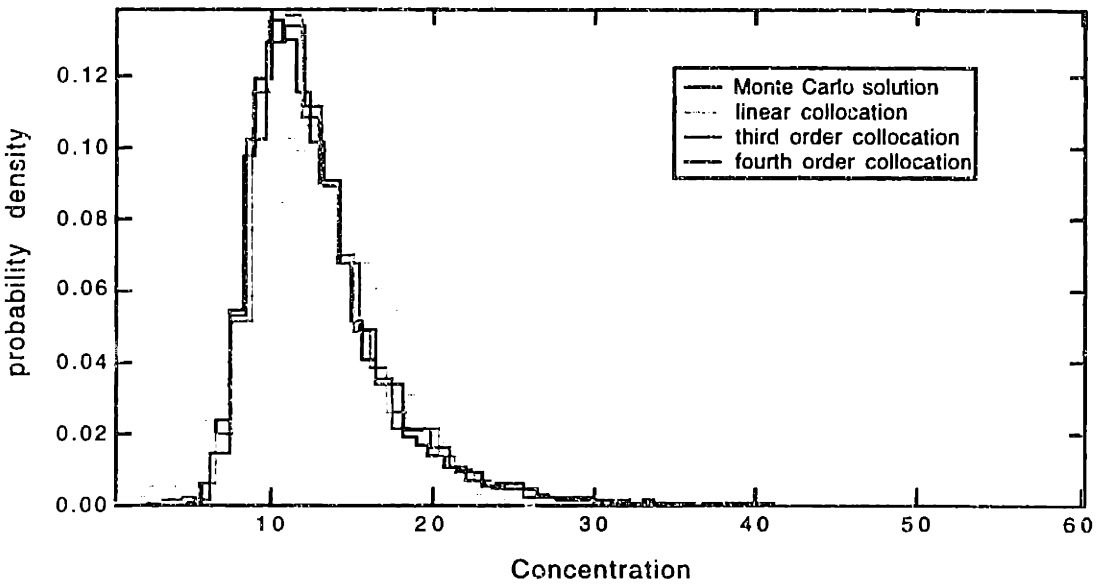


Figure 6-16. Comparison of Monte Carlo and DEMM – Collocation Predictions (Linear and Higher Orders) for PDF of B (time = 2).

6.3.3 Comparison: DEMM — Galerkin

The coefficients of the polynomial chaos expansion (Equation 6.10) can also be solved using the Galerkin method instead of the collocation method (Tatang, 1995). As explained in Chapter 5, the Galerkin method is based on minimizing the distance between the residuals and the surface spanned by the orthogonal polynomials by ensuring that the inner products of the residuals and the polynomials used in the PCE are identically zero. Unlike the collocation method, which can be used to treat a "black box" model, knowledge is required of the model equations to use the DEMM – Galerkin method. The solution procedure of the DEMM – Galerkin method is presented for the first order PCE approximation.

The residuals of the system are defined in Equation 6.10 into Equation 6.2:

$$R_1 = \frac{d(A_0(t) + A_1(t)\xi_1 + A_2(t)\xi_2)}{dt} + k_1 \cdot (A_0(t) + A_1(t)\xi_1 + A_2(t)\xi_2)$$
$$R_2 = \frac{dB}{dt} - k_1 A + k_2 B \qquad (6.20)$$
$$R_3 = \frac{dC}{dt} - k_2 B,$$

where the approximations for A, B, and C are not listed for R_2 and R_3. Since the residuals are perpendicular to the polynomials used in the approximation, three inner product equations can be formed with each of R_1, R_2, and R_3 as follows:

$$\iint R \cdot 1 \cdot p_{\xi_1}(\xi_1) p_{\xi_2}(\xi_2) d\xi_1 d\xi_2 = 0$$
$$\iint R \cdot \xi_1 \cdot p_{\xi_1}(\xi_1) p_{\xi_2}(\xi_2) d\xi_1 d\xi_2 = 0 \qquad (6.21)$$
$$\iint R \cdot \xi_2 \cdot p_{\xi_1}(\xi_1) p_{\xi_2}(\xi_2) d\xi_1 d\xi_2 = 0.$$

Since both ξ_1 and ξ_2 are normal variables with infinite support, the ranges of the integrals are from $-\infty$ to ∞. A system of 9 differential equations can be set up using Equation 6.21 in terms of the uncertain coefficients, A_0, A_1, A_2, B_0, B_1, B_2, C_0, C_1, and C_2. The parameters are k_{10}, k_{11}, k_{20}, k_{21}, and the initial conditions of A_0, B_0, and C_0. Assuming that initial conditions are perfectly known, that for A_1, A_2, B_1, B_2, C_1, and C_2 are exactly zero,

$$\frac{dA_0}{dt} + A_0k_{10} + A_1k_{11} = 0$$

$$\frac{dA_1}{dt} + A_1k_{10} + A_0k_{11} = 0$$

$$\frac{dA_2}{dt} + A_2k_{10} = 0$$

$$\frac{dB_0}{dt} - A_0k_{10} - A_1k_{11} + B_0k_{20} + B_2k_{21} = 0$$

$$\frac{dB_1}{dt} - A_1k_{10} - A_0k_{11} + B_1k_{20} = 0 \qquad (6.22)$$

$$\frac{dB_2}{dt} - A_2k_{10} + B_2k_{20} + B_0k_{21} = 0$$

$$\frac{dC_0}{dt} - B_0k_{20} - B_2k_{21} = 0$$

$$\frac{dC_1}{dt} - B_1k_{20} = 0$$

$$\frac{dC_2}{dt} - B_2k_{20} - B_0k_{21} = 0$$

Although the equations in (6.22) are more complex, they still have the basic ordinary differential equation format as Equation 6.2, the model equation, from which they are derived. These equations can be solved symbolically in Mathematica™ to find A_0, A_1, A_2, B_0, B_1, B_2, C_0, C_1, and C_2 as functions of time. The PCE coefficients solved by the Galerkin method are plotted as functions of time in Figure 6-17 together with the PCE coefficients determined using the collocation method for the first order representation (6.10). The Galerkin method and the collocation method yield very similar results for the PCE coefficients. Although they do not predict the same coefficients for B_1, B_2, and C_2 at the times of peak variance, the agreement is acceptable.

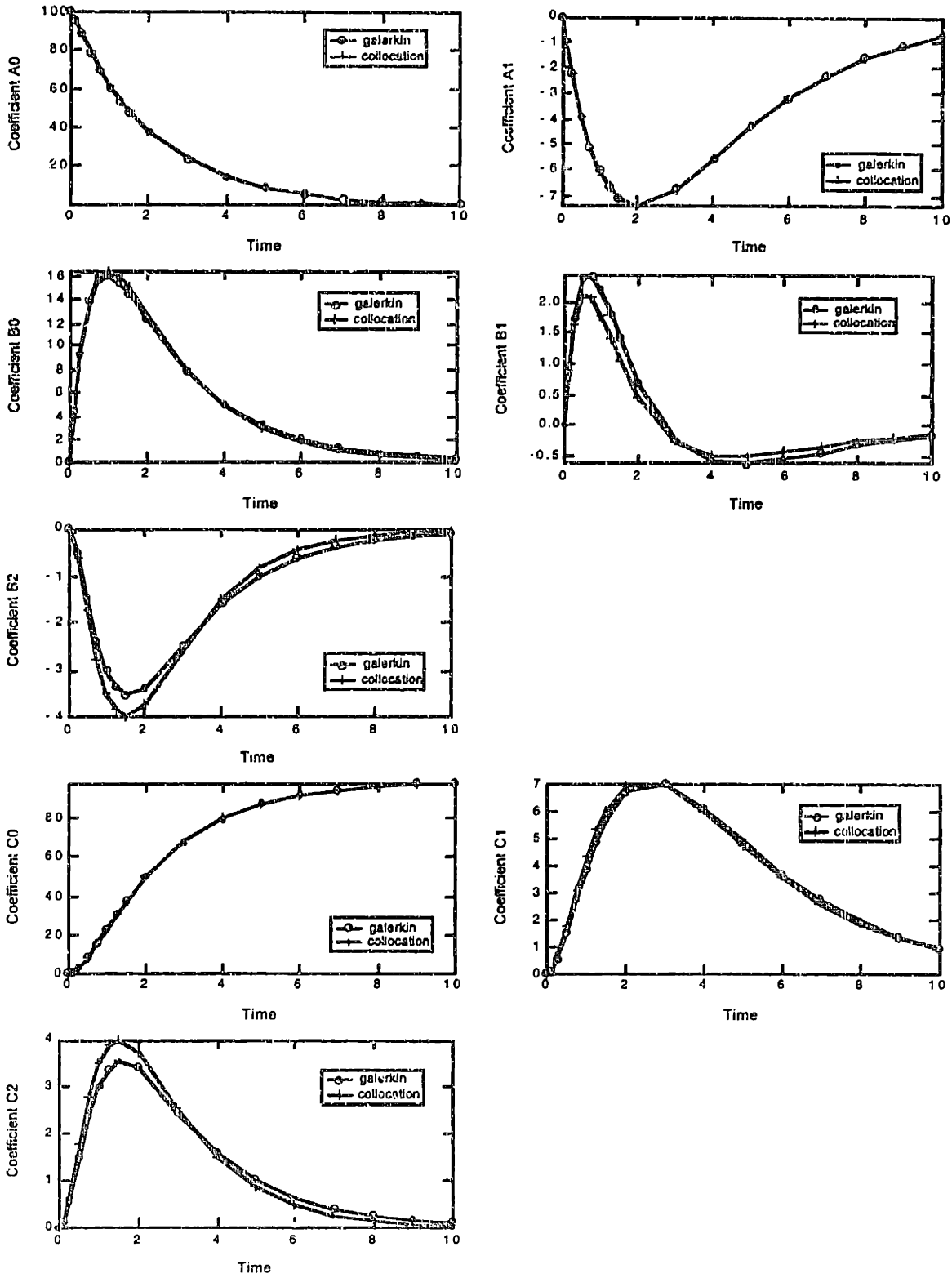


Figure 6-17. Comparison Plots for Coefficients A_0, A_1, B_0, B_1, B_2, C_0, C_1, C_2 Derived by Galerkin and Collocation Methods

6.4 Discussion and Observations

Mean and Variance. As described in Section 5.6, the moments of the probability density functions of the responses can be calculated from the coefficients of the polynomial chaos expansions. The first and the second central moments are of particular interest. They characterize the mean and the variance, or the spread, of the response. Figure 6-18 shows the expected values with markers. The error bars about the markers represent the standard deviation (positive square root of the variance) of the PDF estimate. The solid lines are the nominal solution, that is, the deterministic solution calculated using $k_1 = 0.5$ and $k_2 = 2.0$.

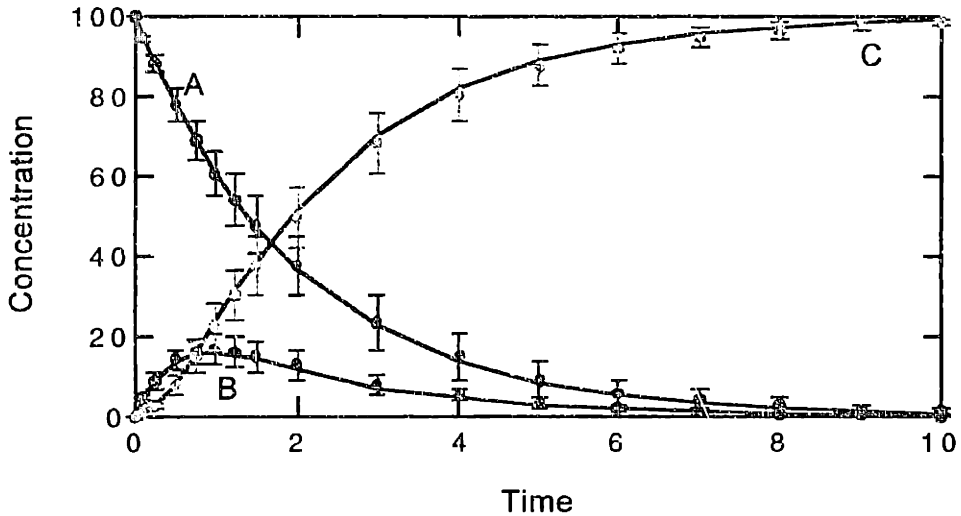


Figure 6-18. Deterministic Solution, Expected Values, and Standard Deviation (line = deterministic solution; markers = expected value; error bars = standard deviation)

Several observations are noted. The expected values are not always equal to the nominal solution based on the best estimates of k_1 and k_2. In fact, the expected solution in an uncertainty analysis can deviate significantly from the nominal solution in complex and highly non-linear mechanisms, as discussed in Chapter 8. Another observation is that uncertainties do not increase without bound. They decrease as the reactions continue, beyond a point of maximum uncertainty. The initial condition are certain, and uncertainties of the concentrations at the beginning of the simulation are small. Since all reactions are irreversible, the end point is also certain: A and B disappear, and C asymptotically approaches the initial concentration of A due to the conservation of mass. This explains the decrease of uncertainties towards the end of the simulation. The transient portion of the reaction is most uncertain for all three species, indicating the uncertainties of the exact timing of the reactions and the concentration profiles. Uncertainties of the three species are inter-related, because total mass is certain and conserved. When compounds A and B are depleted, the concentration of C is certain to approach the initial condition of A. When the concentrations of A and B are uncertain, C is bound to be uncertain as well.

From the individual PCE coefficients, the contribution of any particular uncertain parameter to the variance can be calculated. The PCE coefficients give information regarding the "global sensitivity" of the response variable to the parameter. Figure 6-19 depicts the variance analysis for the intermediate species B. Both k_1 and k_2 contribute to uncertainties in the concentration of B. Although the uncertainty of k_2 is higher than that of k_1, both in absolute and relative terms, the variance contributions of k_2 is not always dominant. In the

beginning of the reaction, k_1 dominates the variance, reflecting the sensitivity of the concentration of B to the rate constant of the A —> B reaction when the concentration of B is low. As B builds up, the rate of the B —> C reaction increases. The concentration of B becomes more sensitive to k_2 than k_1. At this point, uncertainties in k_2 translate to concentration uncertainties of species B. This kind of analysis proves to be useful in identifying input parameters that affect the uncertainties of the model predictions.

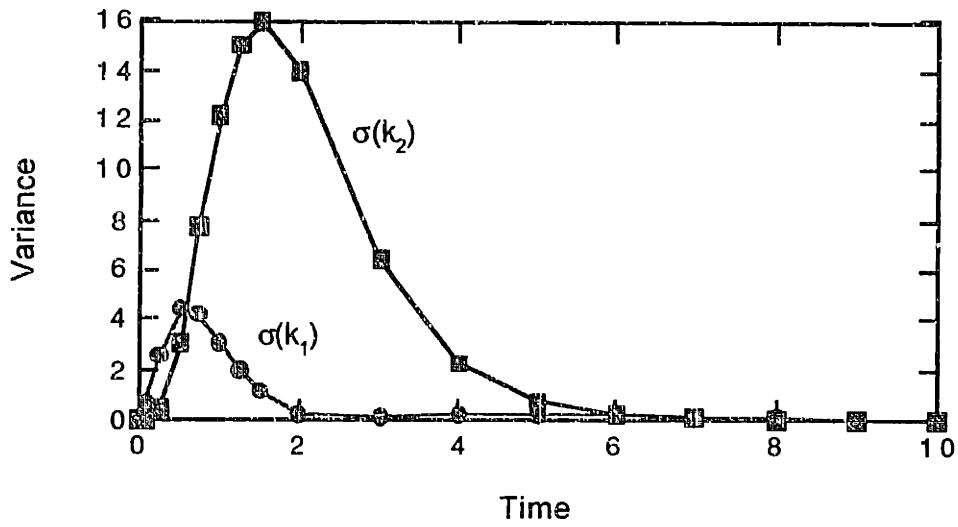


Figure 6-19. Variance contribution of species B as a function of time.

Probability Density Functions . The calculated polynomial chaos expansion represents the distribution of the response variable as a functional of the distributions of uncertain input parameters. Monte Carlo sampling procedures are applied to the polynomial chaos expansion to obtain the PDF. When probabilistic collocation is applied to a complex model, the overhead computer resource utilized to run a Monte Carlo analysis on the PCE is small compared to the time taken to solve the model at the collocation points. For example, a 39-term PCE takes 1.8 seconds to solve and sample. If the model takes two hours to run, a overhead time of several seconds is negligible, and the time savings of using DEMM instead of Monte Carlo is the ratio of the number of model runs needed for the two methods. Figures 6-20, 6-21 and 6-22 show the time series probability density functions for A, B, and C respectively.

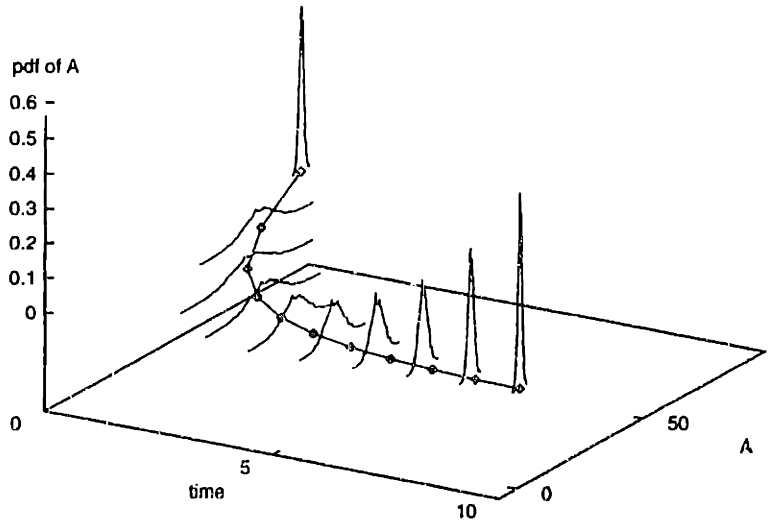


Figure 6-20. PDF Time Series of Species A

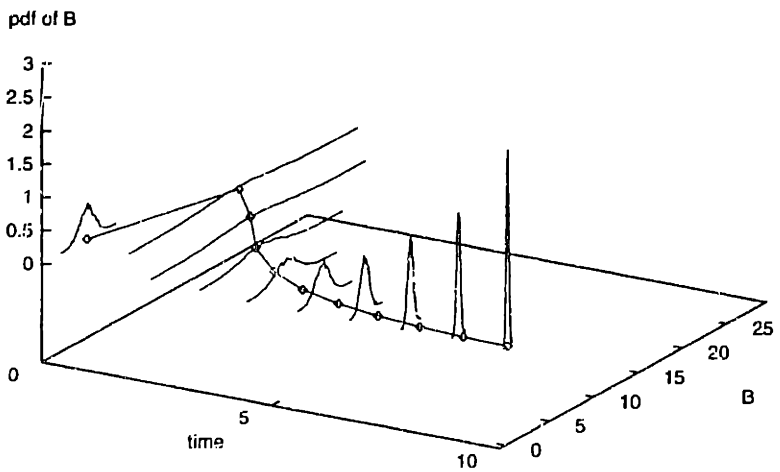


Figure 6-21. PDF Time Series of Species B.

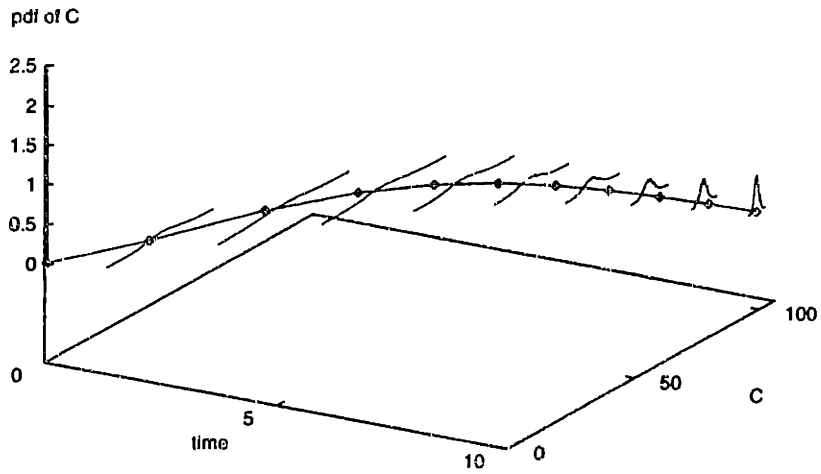


Figure 6-22. PDF Time Series of Species C.

Figures 6-20 to 6-22 confirm the observation in Figure 6-18 that the distributions of the species broaden and then sharpen again as time goes on, reflecting the maximum uncertainties of concentrations at an intermediate time during the simulation. In addition, the distribution can be very skewed even for simple systems. Therefore, some uncertainty methods such as stratified sampling that result only in information about the first and second moments of the uncertain response variable are not ideal for kinetic models. In this example, even though both uncertain input parameters are distributed normally, the probability functions of the output parameters are by no means normal. For example, the PDF for the concentration of A at time = 6 (Figure 6-6) the probability density function is asymmetrically distributed, with a relatively sharp edge on the low concentration side and a long tail on the high concentration side.

Uncertainty does not defy the physics of the problem. In this case, the concentration cannot be less than zero. The rate of the reaction A —> B decreases as A is depleted, resulting in the sharp edge of the distribution at small concentrations. The same is true as C approaches the limit dictated by mass balance. The skewed distributions at the later time points can be explained by the physics of the problem and is captured by the collocation procedure.

6.5 A Word on Uncertain Product Coefficients and Initial Conditions

If the chemical reactions modeled are elementary in nature, stoichiometric coefficients are derived from material balances. However, some reactions represent competitive pathways and product splits. Consider a modified system where

$$\begin{aligned} &A \xrightarrow{k_1} pB + (1-p)D \\ &B \xrightarrow{k_2} C \\ &\mathrm{IC}(A) = A_{ni}. \end{aligned} \tag{6.23}$$

In this system, $k_1 A$ is the rate of disappearance of A. D is a competitve product for B, but the relative rates of formation of B and D are not well known, i.e., the product split coefficient, p, is uncertain. Due to mass balance constraints, p lies between 0 and 1. Without further information, a uniform distribution can be assumed for p between 0 and 1 or between some lower and upper bounds. When more information is available, the principle of entropy maximization (Kapur, 1989) can be applied to determine the least biased distribution for the variable.

The initial conditions (A_{in} in Equation 6.5) for the model may also contain measurement error. This error is expected to be relatively small for controlled laboratory experiments unless the compound concentrations are particularly difficult to control or measure. However, the initial conditions used in three-dimensional modeling can be highly uncertain because of limited spatial coverage of field measurements.

Steps 1 through 8 in Section 6.2 can be applied to "mixed" systems where uncertain parameters are described by more than one type of probability distributions. For the system described in Equation 6.23, assume that p, the uncertain product split coefficient, is described by a uniform distribution. Since there is no standard transformation for uniform distributions, a specific range will be used in this illustration,

$$\xi_3 = \text{uniform } (0.8,\ 1.0). \tag{6.24}$$

The orthogonal polynomials associated with ξ_3, as generated by ORTHPOL (Gautschi, 1994), are:

$$\begin{aligned} H_0 &= 1 \\ H_1 &= \xi - 0.9 \\ H_2 &= \xi^2 - 1.8\xi + 0.806 \\ H_3 &= \xi^3 - 2.7\xi^2 + 2.43\xi - 0.72... \end{aligned} \tag{6.25}$$

Assuming that the uncertain initial condition is described by a normal distribution with mean A_{in1} and standard deviation A_{in2}, the uncertain initial condition is represented as $A_{in} = A_{in1} + A_{in2}\,\xi_4$, where ξ_4 is a standard normal distribution. The linear approximations of $A(t)$, $B(t)$, and $C(t)$ in Equation 6.10 are replaced by:

$$A(t) = A_0(t) + A_1(t)\xi_1 + A_2(t)\xi_2 + A_3(t)(\xi_3 - 0.9) + A_4(t)\xi_4$$
$$B(t) = B_0(t) + B_1(t)\xi_1 + B_2(t)\xi_2 + B_3(t)(\xi_3 - 0.9) + B_4(t)\xi_4 \qquad (6.26)$$
$$C(t) = C_0(t) + C_1(t)\xi_1 + C_2(t)\xi_2 + C_3(t)(\xi_3 - 0.9) + C_4(t)\xi_4.$$

At each time point, 5 model runs are required to determine the coefficients of the linear expansion. The collocation points are determined by solving the second order orthogonal polynomials, in a procedure similar to that depicted in Figure 5-2.

The results of the mixed system is discussed next for a specific case. The competitive product D is formed from A with first order kinetics of A and an uncertain stoichiometric coefficient p uniformly distributed (Equation 6.24) between 0.8 and 1.0. The initial concentration of A, A_{in}, is also taken to be 10% uncertain, distributed normally.

Figure 6-23 is analogous to Figure 6-18. The dotted line represent the same nominal result of A, B, and C when A_{in} = 100 and p = 1 (no competitive product). In general, the variance increases with the number of independent parameters for the same order of PCE representation. This four-parameter system has higher second moments than the previous case when only the two rate constants are uncertain. Due to the uncertain initial condition, the uncertainty of the initial condition of A in the beginning is significant. However, this uncertainty does not translate directly to the concentrations of B and C. The mean value of B as predicted by probabilistic collocation is below the deterministic value, because of the presence of a side product D. Interestingly, the concentration of B is only slightly more uncertain, i.e., the variance associated with B is only slightly bigger than shown in Figure 6-18, with the competitive product of uncertain stoichiometric coefficient. Towards the end of the simulation, the expected asymptotic concentration of C is less than the initial condition of A, because a fraction of A reacts to form D. $C(t)$ is also very uncertain. Since both A and B are depleted, the uncertainty in the initial condition of A and the uncertain amount of D formed are translated into an uncertain final concentration of C.

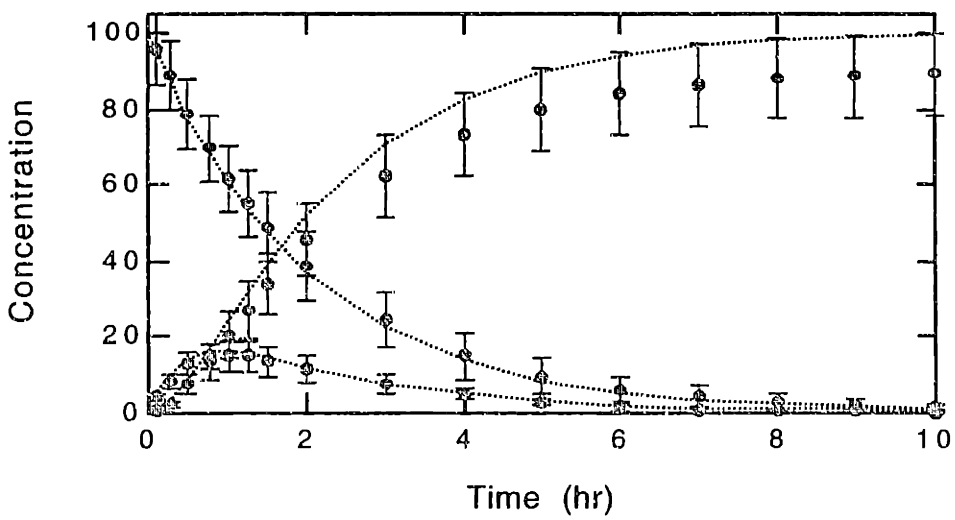


Figure 6-23. Deterministic solution, Expected Values, and Standard Deviation (dotted line = deterministic solution; markers = expected value; error bars = standard deviation)

An analysis of the variance contribution to the concentration predictions confirms the effect of uncertain initial conditions on A (early times, Figure 6-24) and on C (late times, Figure 6-26). The only other parameter that affects the uncertainty of A is k_1, as p and k_2 exert their influence downstream of the irreversible reaction A —> B. Figure 6-25 shows that the two rate constants dominate the uncertainty associated with B. The effect of uncertain initial condition A_{in} is small and limited to the early part of the simulation for species B. The variance contribution of p, the uncertain stoichiometric coefficient of the competitive product D, is also small, and the competition of D does not seem to increase the uncertainty of $B(t)$ by very much.

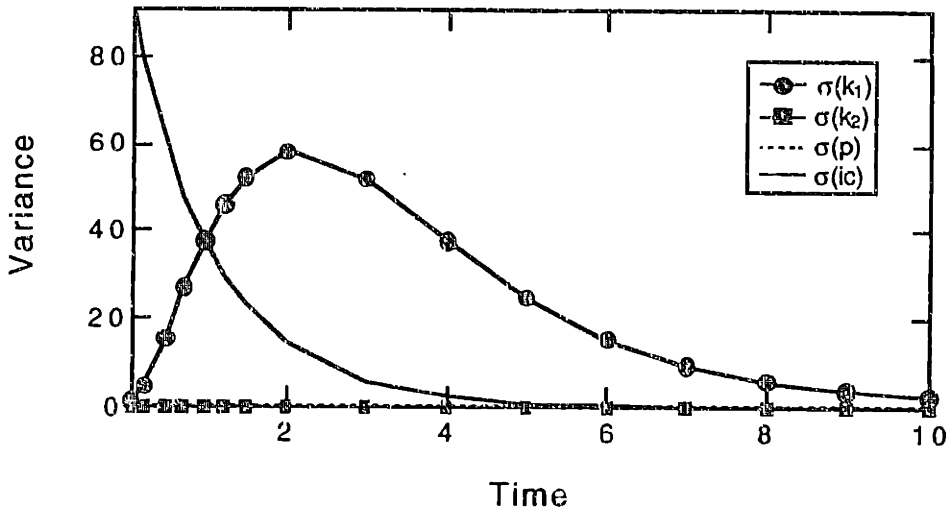


Figure 6-24. Variance Contribution of Species A as a Function of Time.

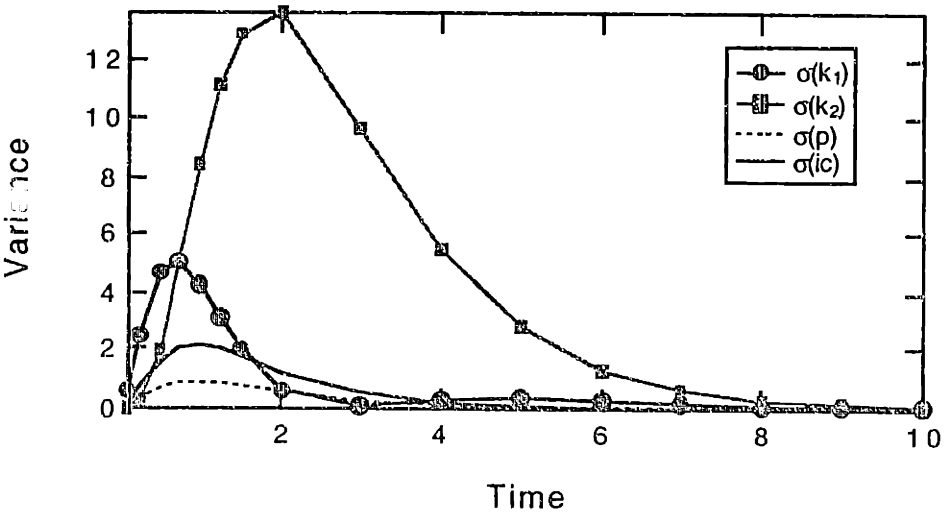


Figure 6-25. Variance Contribution of Species B as a Function of Time.

The variance of $C(t)$ is also dominated by the uncertainties of k_1 and k_2 at the beginning of the simulation. However, the influences of the uncertain initial condition and the product coefficient increase continuously. Instead of reaching an asymptotic value with certainty, C approaches a very uncertain final concentration because of the uncertain initial uncertain value A_{in} and by the uncertain coefficient of formation of the side product D.

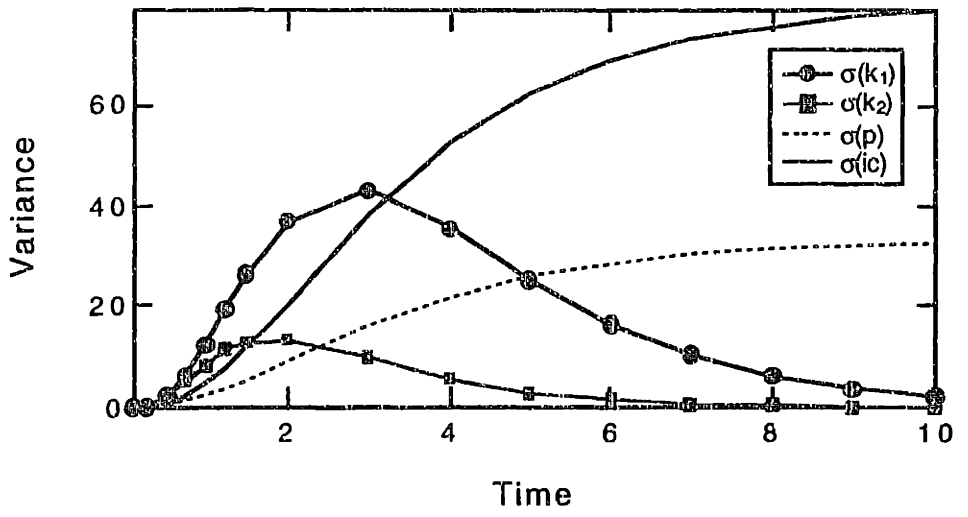


Figure 6-26. Variance Contribution of Species C as a Function of Time.

7. SAPRC Mechanism and Uncertainties

7.1 The SAPRC Mechanism

The Statewide Air Pollution Research Center (SAPRC) mechanism is one of the most detailed gas phase mechanisms designed to-date to simulate urban atmospheric photooxidation. There are 42 inorganic reactions and 106 organic product reactions built into the mechanism. Intermediate organic species (radicals and stable products), including nitrates, are represented by 19 model species. In addition, over 100 organic compounds commonly found in emission sources of polluted urban atmospheres can be represented in the mechanism, either explicitly or in lumped model classes. A brief description of the SAPRC mechanism is given in this section. The lists of species and reactions are reproduced in the Appendix. For a detailed discussion of the formulation of the mechanism, readers are encouraged to refer to the original publication (Carter, 1990). Key features of the SAPRC mechanism include the use of lumped molecule representation of primary organics, variable product parameters of organic reactions, and the use of radical operators to describe organic peroxyl radical reactions.

Inorganic Reactions. Inorganic reactions describe the transformations of oxygen, nitrogen and hydrogen species, including oxygen atoms ($O(^1D)$ and $O(^3P)$), hydroxyl (HO), peroxyl (HO_2), and nitrate (NO_3) radicals. The set of inorganic reactions used in the SAPRC mechanism is similar to that in many other tropospheric mechanisms. Of the 42 inorganic reactions, there are 7 photolysis reactions (NO_2, NO_3(2 sets of products), O_3 (2 sets of products), HONO, and HO_2H). The original SAPRC mechanism includes reactions of sulfur dioxide to form sulfuric acid. Sulfur reactions are not included in this work, which focuses on urban smog (ozone) formation.

Organic Products Reactions. There are 106 reactions in the SAPRC mechanism that describe the inter-conversions of various organic products and radicals. Secondary organic products (19 model species) include aldehyde and other carbonyl and di-carbonyl species, organic nitrates, phenolic and nitro aromatic compounds. Several of the species are explicit, others are lumped to represent the higher members of a homologous series. For example, RCHO is employed in the SAPRC mechanism to represent propionaldehyde and higher aldehydes. Due to this lumped representation, mass is not conserved in this mechanism. Lost carbon and nitrogen are explicitly accounted for using LOSTC and LOSTN species. Secondary products are usually quite reactive. Modeled reactions of the secondary products include photolysis (13 reactions), and reactions with HO, HO_2, and NO_3. Two product species, AFG1 and AFG2, are used to represent unknown aromatic fragmentation products in the reactions of aromatic compounds that have not been thoroughly studied.

Organic Radical Reactions. Organic peroxyl and acyl radicals are not represented by their explicit structures. Instead, a series of "operators" is used to represent different reaction pathways of these radicals. For example, consider the following reaction sequence for $(CH_3)_3COO\bullet$, the t-butyl peroxyl radical:

$(CH_3)_3COO\bullet + NO \longrightarrow NO_2 + (CH_3)_3CO\bullet$ (7.1)

$(CH_3)_3CO\bullet \longrightarrow CH_3C(O)CH_3 + CH_3\bullet$ (7.2)

$CH_3\bullet + O_2 \longrightarrow CH_3OO\bullet$ (7.3)

$CH_3OO\bullet + NO \longrightarrow NO_2 + CH_3O\bullet$ (7.4)

$CH_3O\bullet + O_2 \longrightarrow HCHO + HO_2\bullet$. (7.5)

Reactions 7.1 – 7.5 suggest the following representation for $(CH_3)_3COO\bullet$:

$$(CH_3)_3COO\bullet = CH_3C(O)CH_3 + HCHO + (NO \longrightarrow NO_2) + (NO \longrightarrow NO_2 + HO_2) + RO_2\bullet. \quad (7.6)$$

$RO_2\bullet$ is an accounting species for total peroxyl radicals. In each reaction where a $(CH_3)_3COO\bullet$ is formed, a $RO_2\bullet$ is also listed as a product. In the same reaction, the stable products of subsequent reactions are also represented, including CCHO (representing $CH_3C(O)CH_3$) and HCHO. Organic peroxyl operators are used to represent the reactions of organic radicals: R_2O_2 represents a NO-to-NO_2 conversion, and RO_2R represents a NO-to-NO_2 conversion with the formation of a HO_2 radical. Other peroxyl operators used in the SAPRC mechanism include RO_2N (alkyl nitrate formation), RO_2XN (unreactive nitrates), RO_2H (HO_2 yield), and a similar set for aromatic compounds. In total, 15 radical species and operators are employed in the SAPRC mechanism. Acyl radicals, with carbonyl groups, are not represented using operators. Organic peroxyl operators and acyl radical species are assumed to react with NO, NO_2 (acyl radicals only), HO_2 (to form organic peroxide), and fellow RO_2 and RCO_3 radicals (to form termination products).

Primary Organic Reactions. The kinetic and mechanistic parameters are defined for over 100 alkanes, alkenes, and aromatic species. Alkanes and aromatics are assumed to react only with HO radicals. Rate and mechanistic parameters are defined for the reactions of alkenes with HO, ozone, oxygen atom ($O(^3P)$), and nitrate radicals. Product coefficients for alkanes are estimated by a model which extrapolates empirical product information available for the lower alkanes (Carter and Atkinson, 1985). Alkene product parameters are estimated based on a structure-mechanistic relationship. All aromatics are assumed to react like a few species that have been studied experimentally in environmental chambers. Multifunctional products are assigned to RCHO (aldehyde functional group) and MEK (hydroxy and carbonyl groups). If this assignment results in extra carbon in the products than in the reactants, the yield of MEK is adjusted.

Lumping. In most model applications, the amount of computing time dictates the number of reactions that can be represented in a model. Lumping representation is required for the primary organic mixture for efficient execution of the model. The SAPRC mechanism is designed to give the user flexibility in the level of detail used to describe the primary organic reactions. The user defines the characteristics of lumped classes and the number of explicit species used to represent the primary organic mix, and the reactions of the classes are calculated according to the definition. Smog chamber simulations usually employ more classes of organics so that key species can be represented explicitly.

Different lumping strategies can be used such that the rates and product yields of lumped classes are weighted averages of the parameters of the individual member species, e.g.,

$$k_{Class} = \sum_i w_i k_i, \quad (7.7)$$

where i represent an individual species lumped into the organic class. The standard lumping procedure calls for three alkane/aromatic (AAR) classes representing compounds with different reaction rates with hydroxyl radicals ($k_{AAR1} < 5e3$, $5e3 < k_{AAR2} < 1.5e4$, $k_{AAR3} > 1.5e4$ (ppm-min)$^{-1}$. For each AAR class, the weights are determined by the amount of the individual species expected to react during the course of the simulation. Therefore,

$$w_i = \frac{\left(1-\exp(k_i[HO]_{int})\right)\times c_i}{\sum_i \left(1-\exp(k_i[HO]_{int})\right)\times c_i}, \tag{7.8}$$

where $[HO]_{int}$ is the total amount of HO integrated over time, k_i is the HO reaction rate of species i, and c_i is the initial concentration. Carter (1990) recommended an $[HO]_{int}$ value of 5e-5 ppm-min (50 ppt-min) for use in the basic lumping procedure. A value of 82 ppt-min was used in simulating the Maximum Incremental Reactivity (MIR). This lumping strategy similar to the "reactivity lumping" idea used in RADM modeling (See Chapter 9). The behavior of the class mimics those species within the class that react in the most significant amounts. In addition, the initial condition for the lumped class is calculated based on the class reaction rate and the total amount reacted. This representation sometimes result in an initial condition that is quantitatively different from the amount of organics physically present in the system.

The reactions of alkenes are represented by Ethene and two lumped OLE classes in the standard lumping procedure. The assignment is based on the HO reaction rates of the individual species. The more reactive class contains most of the internal alkenes and have a cut off HO reaction rate of 70000 (ppm-min)$^{-1}$. Due to the reactive nature of alkenes, however, the weights w_i can be simplified to the fraction of initial concentration:

$$w_i = \frac{c_i}{\sum_i c_i}. \tag{7.9}$$

Reaction Rate Constants. Reaction rate constants are specified according to types of reaction. Photolysis reaction rates are calculated based on integrating the product of actinic flux $I(\lambda)$, absorption cross section $\psi(\lambda)$, and quantum yield $\phi(\lambda)$ over the range of wavelengths (λ) where absorption takes place:

$$j = \int I(\lambda)\psi(\lambda)\phi(\lambda)d\lambda. \tag{7.10}$$

Due to the resolution of the solar spectrum and other data, the integration is usually approximated by a finite sum over wavelength intervals of 5 nm or larger. The absorption cross section and quantum yield are provided with the SAPRC mechanism. For atmospheric simulations, Peterson's actinic flux estimates (Peterson, 1976) for cloud-free conditions are used to calculate ground level photolysis rates.

The majority of the reactions are bimolecular and have reaction constants that obey a modified Arrehenius form:

$$k = A(T/300)^{B}\exp(-E_a/RT) \qquad (7.11)$$

where
A = pre-exponential constant
B = exponential factor for temperature
E_a = activation energy
R = gas constant (0.001987 kcal / deg K/ mole).

These constants are evaluated at T = 300K, the simulation temperature used in this work.

Combinations and the reverse dissociation reactions are pressure-dependent because non-reacting bodies (with concentration *M*) are required for collision stabilization for these reactions. Since Troe developed the functional form of this type of reaction rate constants, the pressure-dependent reactions are referred to as Troe reactions in this work. Equation 7.12 shows the fitted rate expression of these Troe reactions, which are characterized by the "fall-off" behavior at intermediate pressures.

$$k = \frac{k_0 M}{\left[1+(k_0 M/k_\infty)\right]} \times f^{g}, \qquad (7.12)$$

where

$$g = \left[1+\left(\frac{\log_{10}(k_0 M/k_\infty)}{n}\right)^2\right]^{-1}$$

k_0 = low pressure limit of rate constant
k_∞ = high pressure limit of rate constant
M = concentration of "third body"
f, n = fitted constants.

Many of the combination reactions characterized by Equation 7.12 are reversible. It should be noted that the reverse reactions are defined using the equilibrium constants, even though the format of the rate constants is the same as Equation 7.11.

Mechanism Validation. The SAPRC mechanism has been successfully validated against more than 500 smog chamber experiments. The peak ozone predictions are within 30% of experimental values in all the experiments simulated. Since much of the variability can be attributable to uncertainties in the characterization of chamber effects, the predictions are considered to be consistent with available experimental data. Other species, such as NO and NO_2, have also been successfully simulated using the SAPRC mechanism.

Mechanism Implementation. In this study, the SAPRC mechanism is implemented in a box model using the ordinary differential equation (ODE) solver LSODE (Hindmarsh, 1983). A brief description of the chemap compiler (Wyckoff, 1995), which generates the ordinary differential equations (ODE's) in FORTRAN code from the mechanism description, is included in the Appendix. This implementation has been validated against Carter's mechanism-solving software available from University of California, Riverside's ftp site.

7.2 Compilation of Rate Constant Uncertainties

There are four types of reactions in the SAPRC mechanism: photolysis, thermal (two-body), Troe (three-body), and equilibrium reactions. The uncertainties in rate constants were compiled by Stockwell *et al.* (1994) using uncertainty factors in most cases. The use of uncertainty factors suggested the use of lognormal distributions to describe the probability distributions of rate constants. Lognormal distributions have semi-infinite support, which provides a realistic description of the non-negative nature of reaction rate constants. Many of the uncertainty factors in Stockwell *et al.* were adopted from NASA or IUPAC panels' estimates. Other sources included reviews of organic chemistry by Atkinson (1990). This section summarizes the uncertainties of some key reaction rate constants in the SAPRC mechanism. A comprehensive list of uncertainty factors is given in the appendix.

Photolysis reaction rates are defined by Equation 7.10. The experimental uncertainties for the absorption cross section and quantum yields are wavelength dependent, as are the uncertainties associated with the actinic flux estimates. Furthermore, some jagged-shaped absorption spectra are not well-approximated by a finite sum. Uncertainties of photolysis rates are likely to be wavelength dependent. A complete assessment of the uncertainties in photolysis rates was not undertaken by Stockwell. Instead, uncertainty factors were estimated for the action spectra, the product of the absorption cross section and the quantum yield. As a first estimate of the uncertainties of the photolysis rates, these uncertainty factors were applied to the actual photolysis rate constants. No uncertainty was assigned to the actinic flux. A selected list of uncertainty factors (UF) is provided in Table 7-1.

Photolysis Reaction	UF	Photolysis Reaction	UF
NO_2 + hν	1.3	ACET + hν	1.4
O_3 + hν —> O + O_2	1.1	MEK + hν	1.5
O_3 + hν —> $O(^1D)$ + O_2	1.4	MGLY + hν	1.6
HCHO + hν —> 2 HO_2 + CO	1.4	BALD + hν	1.5
HCHO + hν —> H_2 + CO	1.4	AFG + hν	3.0

Table 7-1. Uncertainty Factors of Selected Photolysis Reactions.

Uncertainty factors in the photolysis rates ranged from 1.1 to 3.0. Simple molecules that have been studied more thoroughly and have lower uncertainties. The photolysis rates of AFG1 and AFG2, unknown fragmentation products of aromatics, were empirically determined to fit smog chamber data. The high uncertainty of the photolysis of the unknown aromatic fragmentation products led to the assignment of 3 as their uncertainty factors.

Most of the available measurements of bimolecular reaction rate constants (Equation 7.11) were made at room temperature. Therefore, bimolecular reaction rates tend to be more uncertain at temperatures that differ from 298K. Most compilations of uncertainties took this into account by using temperature-dependent uncertainty factors, such as:

$$UF(T) = UF(298)\exp\left|\frac{\Delta E}{R}\left\{\frac{1}{T} - \frac{1}{298}\right\}\right|, \qquad (7.13)$$

where $UF(T)$ is the uncertainty factor at temperature T, $UF(298)$ is the uncertainty factor at 298K and $\frac{\Delta E}{R}$ is the uncertainty in the Arrhenius temperature coefficient. For the purpose of this study, the uncertainty in the temperature dependence was neglected, as all simulations were at constant temperature. The uncertainty factors at 298K were used, even though the simulation temperature is at 300K. The error in the uncertainty factors at this termperature was expected to be small.

Table 7-2 contains a selected list of thermal reactions and their uncertainty factors. Even though many aromatic mechanisms are unknown, the SAPRC mechanisms formulated such that aromatic reaction rates and parameterized products were consistent with chamber experiments. In fact, many parameters were adjusted to fit chamber data. Stockwell imposed a maximum uncertainty assignment of 2.0 for secondary aromatic reactions. Some reaction rates and product spectra in SAPRC were derived by analogy when no experimental data were available. In those cases, the reaction rates of the analogous species were assumed to equal to that of the species for which data were available and the same uncertainty factor applied to both reactions. Furthermore, when data is limited, the uncertainties of the rate and parameters may be assigned to be the same as a species that react similarly. For example, Stockwell (1994) assigned an uncertainty factor to the rate of the phenol-OH reaction that is equal to that of the xylene-OH reaction.

Intuition might suggest higher uncertainty factors for reactions with limited experimental data. However, there appeared to be no systematic way to assess these uncertainties. Assigning uncertainty in an analogous manner to other species that are expected to react similarly is a viable alternative supported by expert judgment. Several uncertainty factors for lumped reactions were determined by analogy to simple reactions, e.g., uncertainty in RO_2 rates analogous to CH_3O_2 reactions. The assessed uncertainty did not represent the range of possible values for the homologous series.

Thermal Reaction	UF	Thermal Reaction	UF
$NO + O_3$	1.2	ACET + HO	1.58
$O(^1D) + H_2O$	1.26	GLY + HO	2.0
$RO_2 + NO$	2.0	PHEN + HO	1.25
HCHO + HO	1.26	BALD + HO	1.41
CCHO + HO	1.41	BALD + NO_3	2.0
RCHO + HO	1.41	AFG + HO	2.0

Table 7-2. Uncertainty Factors of Selected Thermal Reactions.

There are seven Troe reactions in the SAPRC mechanism, as listed in Table 7-3. The combination reaction rate of $RCO_3 + NO_2$ is used in the formation of PAN and PAN-like products. At intermediate pressures, the reaction rates are dependent on both the high pressure limit (k_∞) and the low pressure limit (k_0). Therefore, the uncertainties of these reaction rate constants at any given temperature and pressure are functions of the uncertainties of k_0, k_∞, and the other parameters that are used in Equation 7.12. The PDF of a Troe reaction rate constant can be determined using Monte Carlo simulation, based on lognormal distributions of k_0 and k_∞. It was found that many of the Troe reaction rate constants could be well-approximated by lognormal distributions. The uncertainty factors of the lognormal fits to the Monte Carlo results are given in Table 7-3. These uncertainty factors were used to describe the uncertain inputs in the uncertainty analysis.

Rxn Number	Reaction	UF
2	$O + O_2$	1.1
4	$O + NO_2$	1.1
9	$NO_2 + NO_3$	1.53
19	HO + NO	1.29
21	$HO + NO_2$	1.28
26	$HO_2 + NO_2$	1.12
43	$RCO_3 + NO_2$	2.0

Table 7-3. List of Troe Reactions and Uncertainty Factors Determined by Fitting Monte Carlo Results.

A few reactions were assigned uncertainty factors in non-standard ways. The reactions $NO_2 + NO_3 = N_2O_5$ and $HO_2 + NO_2 = HNO_4$ were assumed to be in thermal equilibrium. Each reverse reaction rate was defined as the product of a equilibrium constant and the forward reaction rate. The uncertainty factors assigned to the equilibrium constants at 298K were 1.5 and 5.0 respectively.

The reaction rate constant of the Troe reaction $RCO_3 + NO_2$ was not applied to all acyl radicals. The high pressure limit rate constant recommended by Atkinson (1990) for the acetyl peroxyl + NO_2 reaction was used for the reactions of propionyl (C_2COO_2) and benzoyl peroxyl ($BZCOO_2$) reactions. The independent uncertainty factors used for these reactions were both 2.0, corresponding to the uncertainty of the high pressure limit of $RCO_3 + NO_2$ reactions.

While the decomposition of PPN was treated as a regular thermal reaction, the decomposition of PAN was assumed to be pressure-dependent. Since the reaction rate constant at atmospheric pressure approaches the high pressure limit, an uncertainty factor of 2 (equal UF of k_∞) was applied to the decomposition of PAN.

For $N_2O_5 + H_2O$, Stockwell *et al.* recommended a discrete probability function derived from a range of studies. This probability function covered three orders of magnitude and was fitted to a lognormal distribution with a central value of 1.0e-24 and an uncertainty factor of 10.0.

Even though Carter used a two-parameter system to describe the reaction rate of HO + HNO_3, other studies suggested that this reaction rate should be a complicated function of three reaction rate terms. Since all three terms can be described in the Arrhenius form and the uncertainty factors assigned are 1.3 by Stockwell *et al.*, the Arrhenius rate constant used in SAPRC for this equation was assumed to have an uncertainty factor of 1.3 as well.

In addition to explicit reactions used in the mechanism, Stockwell also characterized the uncertainties of a number of individual primary reactions that were lumped into classes before modeled in the mechanism. Uncertainty factors of selected organic-HO reaction rates are listed in Table 7-4. The assignment of uncertainty factor to other reactions followed the guidelines recommended by Stockwell, as listed in Table 7-5.

	UF		UF
Ethane	1.2	n-Butane	1.2
2-Methyl Pentane	1.25	Methyl Cyclopentane	1.3
Ethene	1.15	2-Methyl-1-Butene	1.2
1,3-Butadiene	1.2	Isoprene	1.2
Benzene	1.3	Toluene	1.2
m-Xylene	1.25	1,2,4-Trimethyl Benzene	1.35

Table 7-4. Uncertainty of Selected Primary Organic-Hydroxyl Reactions.

Reaction	Recommended UF
HO + AARn (≤C6)	1.2
HO + AARn (>C6)	1.4
HO + AARn (Aromatics)	1.3
HO + OLEn (Anthropogenic)	1.2
HO + OLEn (Biogenics)	1.25
O_3 + OLEn	1.5
O + OLEn	1.2
NO_3 + OLEn	2.0

Table 7-5. Assignment of Uncertainty to Primary Reactions.

7.3 Characterization of Uncertainties in Product Coefficients

An additional source of uncertainty in complex reaction mechanisms is the coefficients of reaction products. Many organic reactions do not represent elementary reactions. Multiple pathways may be condensed onto one reaction, resulting in uncertain product coefficients. It is recognized that exiting knowledge concerning reaction pathways and product splits are limited for many complex compounds, especially aromatics and biogenic compounds. A comprehensive study of uncertainties in atmospheric chemistry must evaluate the effects of these mechanistic uncertainties in addition to uncertainties in rate constants. This section presents a compilation of uncertainties in product coefficients of primary and secondary organic reactions.

7.3.1 Primary Organic Reactions

The reactions of over 100 alkanes, alkenes, aromatics, alcohols, ethers, and other organic compounds that are commonly emitted into the urban atmosphere can be described in the SAPRC mechanism. Lumping of individual primary species (refer to Section 7.1) decreases the computational requirement of the model. Uncertainties in the chemical fate of individual species can be represented by corresponding uncertainties in the stoichiometric coefficients of the product species represented in the model, although this uncertainty representation does not address missing products or misrepresentation of products.

Carter (Carter, 1991) classified the most common alkanes, alkenes, alcohols, carbonyls and aromatic compounds into 8 categories according to the availability of experimental data and mechanistic details. These categories and example compounds of each are listed in Table 7-6.

Category		Example
1	Least uncertain, tested by simulations of chamber data	carbon monoxide, n-butane, ethene, methanol
2	Probably not uncertain, but not tested by simulations of chamber data	methane, ethane, propane
3	Laboratory data available for mechanism, but not tested against chamber data	glyoxal, methylglyoxal
4	Uncertain portions adjusted or parameterized to fit chamber data	propene, 1-butene, 1-hexene, benzene, toluene, o-xylene, m-xylene, acetaldehyde
5	Uncertain, only limited or uncertain data available to test mechanism	n-pentane, n-nonane, 2,3-dimethyl butane, methyl cyclohexane, 2-butenes, a-pinene
6	Mechanism not optimized to fit existing chamber data	isoprene
7	Mechanism is estimated and not tested against chamber data	n-decane, branched C4 to C9 alkanes, C5 - C7 cycloalkanes, 1-pentene, ethylbenzene
8	Estimated mechanism is highly uncertain, and not tested against chamber data	branched >C10 alkanes, >C8 cycloalkanes, >C7 alkenes, >C10 substituted benzene

Table 7-6. Uncertainty categories for mechanistic details (Carter, 1991 pg. 26 Table 3).

Most organic reactions in SAPRC used non-integer product coefficients to incorporate different reaction pathways for the compound into one reaction. For example, the reactions of propane with hydroxyl radical is depicted in Figure 7-1. Even for this simple molecule, knowledge of the product splits p_1 and p_2 is required to determine the yields of the stable products. p_1 describes the probability of radical attack at the internal carbon vs. a terminal carbon. p_2 is the fraction of organic peroxyl radicals that combines with NO_2. For larger compounds, another set of product coefficients is needed to describe the reaction pathways of the alkoxyl radicals that are formed when peroxyl radicals oxidize NO to NO_2. Small (e.g., propyl) radicals react exclusively with oxygen molecule, while larger radicals can react with oxygen, decompose, or isomerize (as discussed in Chapter 2). Instead of listing individual steps explicitly, the series of reactions depicted in Figure 7-1 is summarized by one chemical reaction:

$$CH_3CH_2CH_3 + HO \longrightarrow 0.96\ RO_2R + 0.04\ RO_2N + 0.30\ RCHO + 0.66\ ACET,$$
$$k = 1.22e\text{-}12\ molecule/cm^3/s, \qquad (7.13)$$

where k is the reaction rate constant for the HO attack. Uncertainties in the pathway splits are reflected as uncertainties in the product coefficients of the stable products and radical operators (e.g., RO_2R and RO_2N).

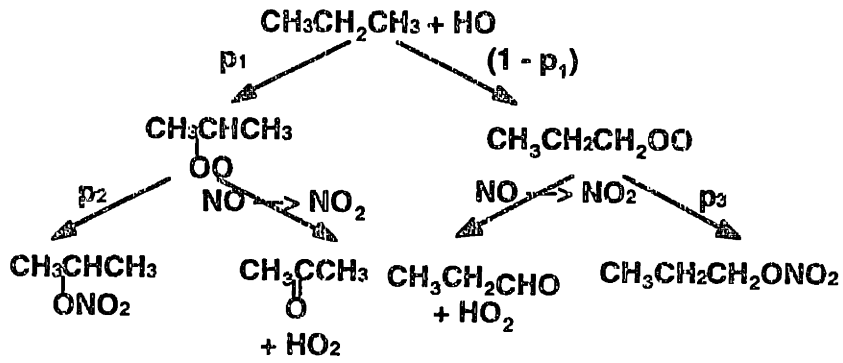


Figure 7-1. Reaction Mechanism of Propane (Seinfeld, 1986)

The uncertainties associated with product coefficients should be described by distributions with finite support, like beta or uniform distributions, due to the mass balance constraint of most reactions. In this study, uniform distributions were chosen to describe the uncertainties of product coefficients to provide the maximum spread of sampling points within a predetermined range. The qualitative descriptions in Table 7-6 provided the basis to characterize the ranges of the uncertainty distributions of product coefficients.

Gao *et al.* (1995) estimated the uncertainties of about 20 product coefficients in the RADM mechanism using uniform distributions. A different method was used in this study to estimate the ranges of uncertainties, resulting typically in higher uncertainties than the previous estimates. Each category in Table 7-6 was characterized by two numerical factors that were used to describe the ranges of the uncertain product coefficients (Table 7-7), with larger factors assigned to categories with higher uncertainties. Mechanisms verified by laboratory data and those postulated to be consistent with smog chamber data were assumed to be equally valid / certain. For alkanes, a factor (fr) was applied to the ratio of RO_2N to RO_2R, which was usually determined as (1 - nitrate yield RO_2N). This representation was chosen so that both RO_2R and RO_2N ranged between 0 and 1, and that the uncertainty would be independent of which product was used in the numerator of the ratio. The yields of the various stable carbonyl compounds were dependent on the position of the initial radical attack and the competition between three different alkoxy radical reaction pathways, as well as on lumping strategies applied to secondary compounds. Instead of re-determining the exact relationships between the product coefficients (Carter and Atkinson, 1985), they were treated as independent uniform variables whose ranges were determined by the factors fp listed in the third column of Table 7-7 (nominal value x/÷ fp).

Uncertainty Category (see Table 7-6)	Product split factors (fr) (p1/p2 = nominal value x/÷ fr)	Product range factors (fp) (p3 = nominal value x/÷ fp)
1	1.25	1.1
2	1.5	1.25
3	1.75	1.5
4	1.75	1.5
5	2.0	1.75
6	2.0	1.75
7	2.5	2.0
8	3.0	2.5

Table 7-7. Factors applied to determine the range of product coefficients.

For example, for n-butane, the radical operators RO_2N and RO_2R were competitive products. The coefficients of RO_2R (RR) and RO_2N (NR) are complementary and correlated, such that RR = 1 - NR. The product split factor (fr) was used in determining the uncertainty range for these two parameters. The nominal value of the ratio NR/RR is 0.0824. Since n-butane is of uncertainty category 1, fr = 1.25, and the ratio NR/RR was between 0.0659 (0.0824÷1.25) and 0.103 (0.0824*1.25). Therefore, the distribution of NR was uniform in the range (0.0618, 0.0934). The relationships between the yields of stable product and R_2O_2 were not determined explicitly. The factor fp = 1.1 was used to determine their uncertainty ranges. The uncertainties associated with the products of the HO-butane reaction are listed in Table 7-8.

Product	Nominal Coefficient	Uncertainty Description	Notes
RO_2N	0.07613	Uniform (0.0618, 0.0934)	1.
RO_2R	0.92387 = (1 - nr)	--	1.
R_2O_2	0.39654	Uniform (0.3605, 0.4362)	2.
HCHO	0.00071	Uniform (0.000645, 0.000781)	2.
CCHO	0.57105	Uniform (0.5191, 0.6282)	2.
RCHO	0.13981	Uniform (0.1272, 0.1538)	2.
MEK	0.53331	Uniform (0.4848, 0.5864)	2.

Table 7-8. Uncertainty assignment to product coefficients of n-butane.
(Notes: 1. product ratio uncertainty, range of uniform distribution described by fr = 1.25.
2. independent product uncertainty, range described by fp = 1.1)

Only a few reaction mechanisms were available for aromatic species, including benzene, toluene, xylenes, and 1,3,5-trimethyl benzene. Most aromatics compounds were assumed to react like one of these few which have been studied in smog chamber experiments. While some products were measured in smog chambers, others were not identified. A parameterized representation was postulated such that product coefficients were adjusted to fit the chamber results. Parametric uncertainties of "unknown compounds", such as ring opening compounds AFG1 and AFG2, were described in an analogous manner as alkane parameters using uniform distributions whose ranges were nominal value x/÷ factor. Product coefficients for measured products were described according to compiled information (Atkinson, 1990) in the form of nominal value +/- uncertainty. The uncertainty estimates were based on error estimates or

the discrepancies between reported values, whichever was larger. Normal distributions were used for these coefficients, with upper and lower bound applied where necessary to ensure that the yield was >0 and that mass was not created. For example, the uncertainties of the product coefficients in the Toluene + HO reaction:

$$\text{Toluene} + \text{HO} \longrightarrow 0.085\ \text{BALD} + 0.26\ (\text{CRES} + \text{HO}_2) + 0.118\ \text{GLY} + 0.131\ \text{MGLY} + 0.41\ \text{AFG2} + 0.74\ \text{RO2R} \quad (7.14)$$

are described in Table 7-9.

Product	Uncertainty Description	Product	Uncertainty Description
BALD	Normal (0.085, 0.013)	MGLY	Normal (0.131, 0.019)
CRES	Normal (0.26, 0.017)	AFG2	Uniform (0.273, 0.615)
HO_2	none (= CRES)	RO_2R	none (= 1.0 - HO2)
GLY	Normal (0.118, 0.016)		

Table 7-9. Product Coefficient Uncertainties of Toluene-HO Reaction

Except for the nitrate yield of alkene-HO reactions, products coefficients of alkene reactions were correlated to structures of reactants. The coefficients of the correlations were treated as uncertain, with uniform distributions. Mechanistic uncertainties of terminal alkenes (uncertain category 4) were assumed to be lower than that of internal alkenes (uncertain category 5).

7.3.2 Secondary Organic Reactions

Secondary species included aldehydes, ketones, organic nitrates (e.g., PAN, and RNO_3), dicarbonyls, phenolic compounds, organic peroxides, and unknown aromatic reaction products. Several of these compounds represented lumped classes of higher compounds, even though their reactions were modeled as those of the smaller member of the group. For example, RCHO was used to represent all aldehydes with more than three carbons, even though the reactions were exactly those of propionaldehyde. Therefore two levels of uncertainties were associated with secondary organic reactions: uncertainties related to the representation of higher secondary compounds by smaller members of the homologous series, and parametric uncertainties of the reaction rate and product distribution of the compounds modeled (such as propionaldehyde).

The parametric uncertainty of product coefficients in secondary organic reactions were assumed to be uniformly distributed within the range of values reported in the literature, or compiled by Atkinson, 1990. Radical yields that were not based on mass balance were described by lognormal distributions. Selected uncertain coefficients are listed in the Table 7-10. A list of uncertain product coefficients of secondary reactions is included in the Appendix.

Reaction	Uncertainty Description
$RO_2R + RO_2 \rightarrow 0.5^{uc}\ HO_2 + RO_2$	C(HO_2) = Uniform (0.35, 0.62) (Madronich and Calvert, 1990)
$ACET + HO \rightarrow 0.8^{uc}\ (MGLY + RO_2R) + 0.2^{uc}\ (R_2O_2 + HCHO + CCOO_2)$	C(ratio) = Uniform (0.8, 20) (nominal value = 4; UF = 5, Atkinson, 1990)
$MEK + HO \rightarrow 0.5\ (CCHO + CCOO_2) + 0.5\ (HCHO + C_2COO_2) + RCO_3 + 1.5^{uc}\ (R_2O_2 + RO_2)$	C (RO_2) = Lognormal (1.5, 1.5)

Table 7-10. Selected Uncertain Product Coefficients of Secondary Organic Reactions. (^{uc} denotes the uncertain coefficient described in column 2)

The uncertainty introduced into model predictions due to the design decision to lump secondary organics using a surrogate species approach was a source of "structural" uncertainties intrinsic in the SAPRC model. Structural uncertainty resulted from model decisions rather than from imperfect knowledge of the data required in the model. Such simplifying assumptions are necessary to lower the demand for mechanistic data requirement to formulate the model. For example, even though many of the intermediate products of aromatic reactions are not characterized, aromatic reactions can still be parameterized in the SAPRC mechanism by fitting to chamber data. Structural uncertainties cannot be investigated systematically in the DEMM framework. In some cases, it may be possible to investigate the effects of structural uncertainties by relating the structural uncertainty to some parametric uncertainties, or by comparing to an alternative model. (Examples of mechanism comparison studies for the purpose of testing assumptions were discussed in Chapter 4).

8. Uncertainty Analysis of Large Models: Application to the SAPRC Mechanism

8.1 Method of Approach

Complex models, like the SAPRC mechanism, can literally have hundreds of uncertain parameters. Even with an efficient method like DEMM, it is impossible to evaluate simultaneously the effects of all these parameters in one comprehensive study at any reasonable level of accuracy. Most uncertainty methods to-date rely on some form of sensitivity estimation (e.g., Yang *et al.*, 1995) to identify a small subset of parameters to investigate in an uncertainty analysis (see Chapter 7 for discussion). The drawback of this method is that the screening procedure considers only the local sensitivity, while the output uncertainty is a function of both the uncertainty of the input and the sensitivity of the model within the range of the uncertain parameter. The first order approximation of uncertainty, or the Gaussian approximation

$$VC_{ij} \equiv \left(\left. \frac{dy_j}{dx_i} \right|_0 \right)^2 \cdot Var(x_i) \qquad (8.1)$$

suggests that the variance contribution (VC) of a model input (x_i) to the uncertainty of a model output (y_j) is proportional to the product of the square of the sensitivity coefficient at the nominal value $\left. \frac{dy_j}{dx_i} \right|_0$ and the uncertainty of input x_i (Morgan and Henrion, 1992). This approximation of uncertainty effects is only good for evaluating the local effects of uncertainty in the vicinity of the nominal scenario, or for linear models or models with small uncertainties. For complex, non-linear models, a better procedure is needed to identify those input parameters that have a significant effect on the uncertainties of the outputs.

Variance contributions should be used as the criteria to limit the number of parameters considered. Instead of using sensitivity coefficients or approximated VCs calculated by Equation 8.1 for parameter selection, DEMM allowed the calculation of the variance contributions of individual parameters directly from the response surface representation. The coefficients of the polynomial chaos approximation contained information about the variance contribution, as discussed in Chapter 5. In this study, variance-contributing inputs (VCIs) were identified and were used to guide the DEMM iterative procedure shown in Chapter 5. The revised method is illustrated in Figure 8-1.

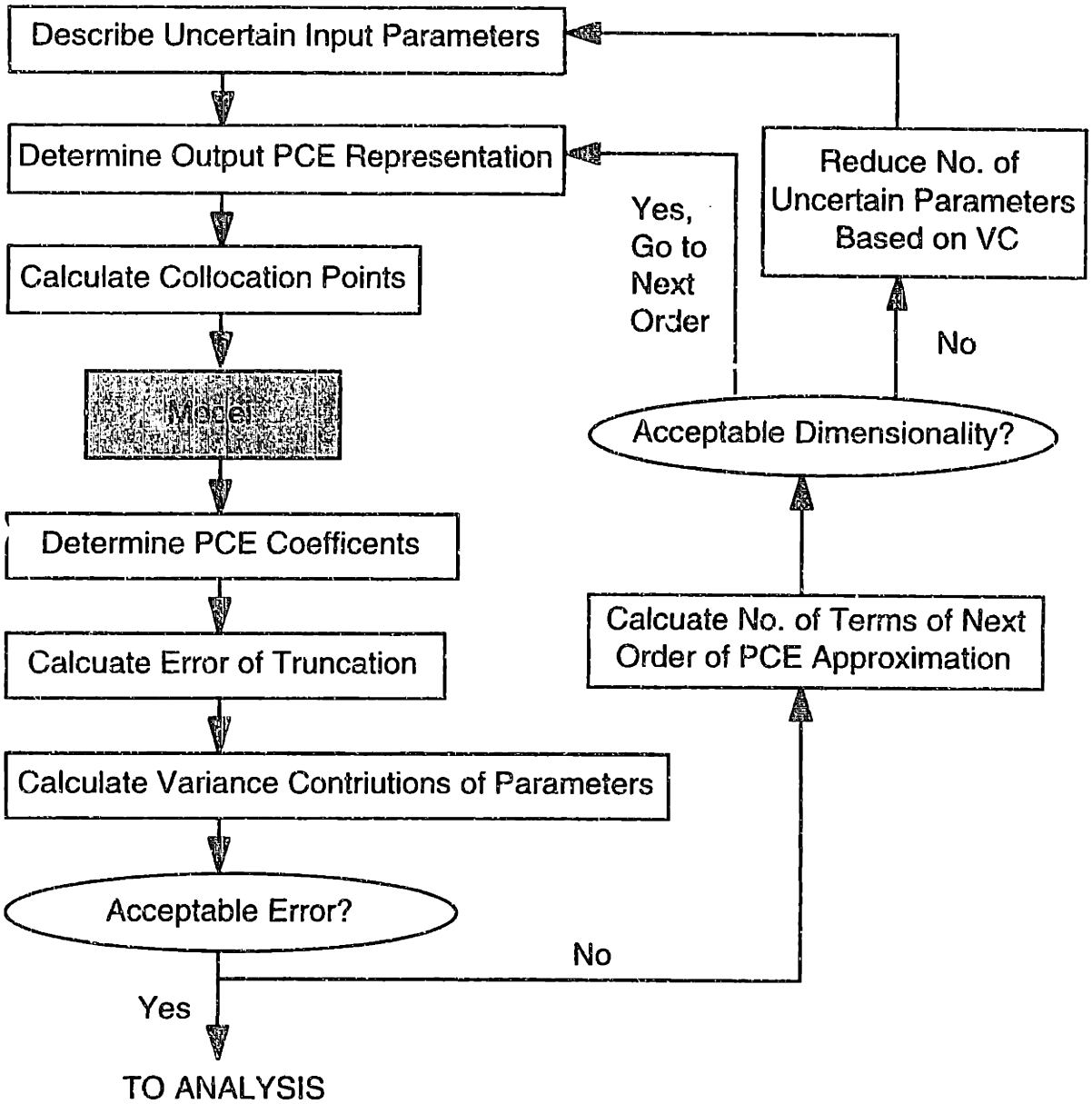


Figure 8-1. Procedure for Uncertainty Analysis Using DEMM.

The standard DEMM procedure called for increasing the order of approximation to reduce the truncation error of the polynomial chaos expansion (PCE) approximation of the response surface. Since the dimension of the problem grew with the number of parameters and the order of approximation, it was usually not practical to increase the order of approximation beyond first or second order when dealing with systems of more than 20 parameters. In the revised procedure, the number of uncertain variables was reduced when increasing the approximation order led to an unacceptable number of model runs to evaluate the PCE coefficients. After each DEMM experiment, the input parameters were ordered according to their contribution to the output variance. The least important parameters were eliminated when the solution to the PCE coefficients becomes too expensive with the original number of uncertain inputs. Guidelines for the elimination of parameters were established in this study as follows: (1) each eliminated parameter contributed no more than 0.1% to output variances; (2) the remaining parameters accounted for at least 90% of the

variance of the output parameters. This procedure was applied until the maximum relative error was less than 1% or until the variance of the uncertain output and the VC of key inputs converged from one approximation to the next order of approximation.

A "divide and conquer" approach was used in this study to facilitate the analysis of the large number of parameters used in atmospheric chemistry modeling. The SAPRC mechanism parameters were organized into four categories: (1) inorganic reaction rate constants, (2) primary organic reaction rates and product parameters, (3) secondary organic (product species) reaction rates and product parameters, and (4) initial conditions. The primary organic reaction category was divided further into four experiments: alkane, alkene, aromatic reactions and composition. Each category was investigated separately to identify the variance contributing inputs, or VCIs, which were then evaluated jointly in one uncertainty analysis to elucidate synergistic effects and to characterize overall output uncertainties. Descriptions of uncertain parameters can be found in Demmucom input scripts included in the appendices.

8.2 Inorganic Reactions

Forty parameters were investigated in this uncertainty experiment of inorganic reactions in the SAPRC mechanism, including photolysis rates, thermal and combination reactions rates, and equilibrium constants. As shown in the input file in the appendix, all uncertain reaction rates were described using the lognormal probability density function. Except for N_2O_5 + H_2O —> 2 HNO_3, all reaction rate constants were assigned uncertainty factors that were less than 2.

Using the iterative procedure depicted in Figure 8-1, 9 VCIs were identified for ozone, NO, NO_2, and HO concentrations. These parameters, listed in Table 8-1, accounted for over 99% of the variance in these concentrations for the 3rd order PCE approximation. Even though the list of VCIs did not include reaction parameters that were highly uncertain, their combined effects on the output concentrations were significant. The standard deviations associated with uncertainties in the predicted ozone were as high as 40%, with mean values that were about 10% higher than the nominal results, which were the predictions obtained using the best estimate parameters. The main purpose of this experiment was to identify VCIs, and so the order of the PCE representation was increased to optimize the representation of the response surface and to reduce errors. Rather, the iterative procedure was followed until the relative variance contributions of the most important VCIs converged with increasing number of terms. In this case, the relative importance of the VCIs did not change from the second order polynomial response surface representation to the third order. The maximum error in the representation of ozone was 0.61%, and was 6.3% in hydroxyl radical

Photolysis Rate Constants	Reaction Rate Constants
j (NO2)	k (O (1D) + H2O)
j (HONO)	k (O (1D) —> O(3P))
j (O3)	k (HO + NO2)
	k (HO + NO)
	k (HO + CO)
	k (O3 + NO)

Table 8-1. Variance-Contributing Inorganic Rate Constants for Ozone, NO, NO_2, and HO Concentrations.

8.3 Secondary Organic Reactions

The group of secondary organic reactions included reactions involving organic radicals and carbonyl compounds that were the second generation products of alkanes, alkenes, aromatics, etc. Uncertainties in product coefficients and rate constants were assessed for the surrogate compounds of lumped groups. For example, mechanistic uncertainties of propionaldehyde were assessed and applied to the lumped species RCHO. Uncertainties due to the lumped presentation using surrogate species were not evaluated.

The uncertainties in the product coefficients did not significantly affect the predictions of ozone, NO, NO_2, and hydroxyl radical. Identified VCIs were the reaction rates of acyl radicals and secondary carbonyl products from the decomposition of emitted species, as listed in Table 8-2.

Photolysis Rate Constants	Reaction Rate Constants
j (HCHO —> radicals)	k (RCO_3 + NO)
j (HCHO —> stable products)	k (RCO_3 + NO_2)
j (CCHO)	k (HCHO + HO)
j (RCHO)	k (CCHO + HO)
j (MGLY)	k (MGLY + HO)

Table 8-2. VCIs from Secondary Organic Reactions

The effect of kinetic and mechanistic uncertainties in secondary organic reactions was smaller compared to that associated with inorganic reactions. The difference between the mean and nominal predictions of the model were negligible. The 10 VCIs listed in Table 8-2 resulted in a maximum standard deviation of 28% in ozone predictions using a full third order representation of the response surface. The relative errors of the response surface approximations for the concentrations of O_3, NO_x, and HO were less than 1%.

8.4 Primary Organic Reactions

The primary reactions of SAPRC were built by lumping the reaction rates and products of individual species based on the composition of the organic mix. The large number of parameters involved required the division of primary organic parameters into smaller subgroups to facilitate analysis. The uncertainty analyses described next were carried out for alkane, alkene, aromatic parameters, and composition of the organic mixture.

8.4.1 Reactions Rates and Parameters – Alkanes, Aromatics, and Alkenes

Alkanes. Alkanes are the most studied of the organics, and are frequently modeled with the greatest detail. There were close to 30 individual alkane species in the all city averaged organic profile used in this work. (Acetylene was included in this group because of its low reactivity.) The only modeled atmospheric destruction route of these compounds was the reaction with the hydroxyl radical. The uncertainties of the HO-rate constants were determined based on the quality and availability of laboratory data using Stockwell's guidelines (Table 7-5). Mechanistic uncertainties of over 100 product species were described using uniform distributions, as shown in the input file in the appendix. 53 VCIs were selected after 6 iterations and are listed in Table 8-3. Many of the VCIs were associated with compounds that were abundant and that reacted within the simulation time scale. Key parameters included reaction rate constants, carbonyl yields, and the split between organic nitrate (RO_2N, yield coefficient represented by NR) and peroxyl radicals (1-NR).

Compound	Parameter	Compound	Parameter	Compound	Parameter
Ethane	k	N-hexane	k	3-methylhexane	K3
Acetylene	k	2-methylpentane	k		K4
	HO_2 (RH)		NR	Me-cyclohexane	k
	glyoxal (GL)		R2	4-methylheptane	k
Propane	k		A3		NR
N-butane	k		K4		R2
2-methylpropane	k	3-methylpentane	k		A3
N-pentane	k		R2		K4
	RO_2N (NR)		A2	4-ethylheptane	k
	R_2O_2 (R2)	2,3-dimebutane	NR		NR
	RCHO (A3)		acetone (K3)		A3
	ketone (K4)		A3	N-decane	k
2-methylbutane	k	Cyclohexane	k	4-propylheptane	k
	NR		NR		NR
	R2		A3		A3
	CCHO (A2)	3-methylhexane	k		K4
	A3		NR	3,5-diethyloctane	k
	K4		R2		

Table 8-3. VCI of Alkane Parameters

Even though alkanes are the most abundant of the organics in an urban atmosphere, the collective contribution of alkane parameters to uncertainties in the model predictions was found to be much smaller than that of the inorganic and organic parameters described previously. Two reasons were postulated for this observation. First, SAPRC predictions of ozone, NO, NO_2, and HO were probably not very sensitive to alkanes because of their relatively low reactivity. Second, according to the central limit theorem, the weighted sum used in lumping resulted in smaller uncertainties of the lumped alkane reactions than the individual species (on a percentage basis). Due to the large number of alkanes lumped together in each AAR class, the small uncertainty contribution could be a direct result of the lumping procedure.

Aromatics. Atmospheric reactions of aromatic compounds are frequently cited as a glaring gap in the knowledge of the chemical transformation of airborne organics. The HO reaction rates of quite a few simple aromatics have been characterized. However, the mechanistic understanding is incomplete, even for simple aromatics. Current mechanisms adopted a parametric representation that allowed the simulation of the concentration profiles of key species like ozone in smog chamber experiments. However, even these parameterized reaction mechanisms were only available for a small number of simple aromatics, such as benzene, toluene, xylene, and trimethylbenzene. Most species were assumed to react in an analogous manner to one of these few. Since the parametric representations were verified against smog chamber data, the uncertainties assessed for the parameters involved were relatively small. Nonetheless, significant structural uncertainties are associated with the lack of mechanistic information and the assumptions made about the parametric representations.

Ten reaction rates and 16 mechanistic parameters were used to describe the reactions of all aromatics present in the all city organic mixture. Using DEMM to propagate these uncertainties through to the predictions, it was found that uncertainties in aromatic reaction rates and product parameters had a larger effect on the predictions of the model than alkane parameters. In particular, parameters that contributed to the variance in the predictions of ozone, NO, and NO_2 were the reaction rate constants, the product yields of methyl glyoxal, and the coefficients of the unknown ring-opening products, represented by AFG1 and AFG2. These ring opening products were also shown to be important contributors to uncertainties of hydroxyl radical concentrations.

Compound	Parameter	Compound	Parameter
Benzene	k	M-xylene	k
	AFG1 (U1)		MG
Toluene	k		U2
	Methylglyoxal (MG)	O-xylene	k
	AFG2 (U2)	P-xylene	k
Ethylbenzene	k	1,3,5 Trimethylbenzene	k
I-propylbenzene	k		MG
N-propylbenzene	k		U2
		1,2,3 Trimethylbenzene	k

Table 8-4. VCI of Aromatic Parameters

Alkenes. This study focused on anthropogenic alkenes, even though it is now believed that biogenic alkenes, like isoprene, can be a significant source of organic radicals in some parts of the country. Alkenes are the most reactive component of the organics present in an urban atmosphere. In the SAPRC description, four uncertain rate constants were associated with each alkene for its reactions with HO, O_3, NO_3, and $O(^3P)$. Except for the organic nitrate yield, the product yields of these reactions were assumed to correlate with the structure of the compound. The uncertainty analysis of alkene parameters involved 46 rate parameters, 11 non-zero nitrate yields, 3 coefficients for ethene reactions, and 13 coefficients used in structure-product correlations.

The response parameters seemed to be less sensitive to uncertainties in alkene rates and parameters than to the aromatic parameters investigated. The maximum variance in ozone, NO_x, and HO were only a few percent, insignificant compared to the effects of other organic parameters. Ethene parameters seemed to contribute more that the other alkene parameters. Possible explanations can be that ethene was modeled explicitly, or that it was more abundant than any other alkenes. None of the NO_3 reaction rates contributed to uncertainties, even though they are more uncertain than the other rate parameters. Nitrate reactions were not important during the day, when photolysis prevented the accumulation of the radical. Oxygen atom reaction rates only contributed to variance in the output in the very beginning of the simulation. Variance-contributing alkene parameters are listed in Table 8-5.

Compound	Parameter	Compound	Parameter	Structure-Product Correlation for Ozone Rxn
Ethene	k_{OH}	1-Pentene	nitrate (PN)	U2A1
	k_{O3}		k_{OH}	U2RR
	k_{NO3}	2-Pentene	PN	U2RH
	k_{O3P}		k_{OH}	U2OH
	CCHO (HO rxn)		k_{O3}	U4OH
	HO_2 (O_3 rxn)	1-Hexene	PN	
	HCHO (NO_3 rxn)		k_{OH}	
Propene	k_{OH}	2-Hexene	PN	
1-Butene	k_{OH}	1-Heptene	PN	
T-2-Butene	k_{O3}	1-Octene	PN	
C-2-Butene	k_{O3}	1-Nonene	PN	
		1-Undecene	PN	

Table 8-5. VCIs of Alkene Parameters

Uncertain Predictions of Alkane, Aromatic, and Alkene Parameters. Of the organic classes investigated, akanes are the most abundant in urban air, aromatics have the most uncertain fate, and alkenes are the most reactive. The relative contributions of the parameters from these classes to uncertainties in ozone predictions are shown in Figure 8-2. The strongest effects come from aromatic kinetic and product parameters. Alkanes are probably too unreactive to affect the reaction dynamics by contributing to the formation of radicals in the time scale of concern. The influence of uncertain alkene parameters is probably small because alkenes are very reactive. Most of the VCIs from the alkene group are reaction rates. As shown in Chapter 6, uncertain rate parameters have small effects on model predictions if the reactants are depleted within the time-scale of concern.

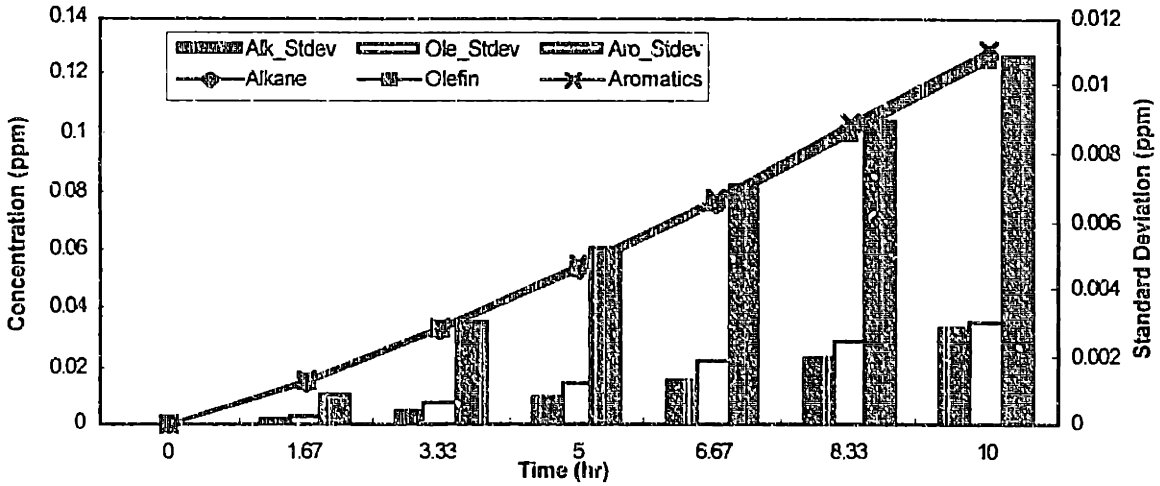


Figure 8-2. Ozone VC Analysis Results of Organics Experiment.

8.4.2. Organic Composition

The composition of the organic mixture affects both the reaction characteristics (rates and products) and the initial conditions for the different lumped classes of organics given the current SAPRC lumping procedure. In practice, the organic composition used in atmospheric modeling is determined from source profiles, for which few measurements are available. The assessment of uncertainties in organic source profiles was beyond the scope of this project. An uncertainty factor of 20% was assigned to each of the 61 independent concentrations found in the all city profile. (The all city profile originally contained 78 species, but several of them were equivalent to other compounds in the profile.) This uncertainty factor was determined loosely based on the errors of the experimental procedures used to measure atmospheric concentrations of common organics, such as alkanes and alkenes. This UF served well to illustrate the effects of uncertain composition, as it was also found that a larger uncertainty factor (e.g., UF = 1.5) only increased the total variance of the output, but did not affect the relative importance of individual concentrations for any given order of PCE representation.

The maximum standard deviation associated with the concentrations was less than 15%. The change in the expected value relative to the nominal values was more pronounced in this case than in experiments where the reaction parameters were varied. Compounds that were modeled explicitly, i.e., ethene, HCHO, and CCHO, accounted for a significant fraction of the variance in the response variables of interest, possibly indicating the effect of lumping to reduce the variance of the lumped class relative to explicit species. Uncertainties in the initial conditions of the two aldehydes accounted for an average of 50% of the variance in ozone. The other explicit species, ethene, contributed a maximum of 12% to the uncertainties in the predictions of ozone. Of the lumped species, individual concentrations found to be relatively important included m-xylene, trimethylbenzenes, toluene, 1-butene, and propene. Variance contributions from individual alkanes were less than 1% to uncertain outputs. These results indicated that the effects of uncertain organic concentrations on predictions were related to the time-scale of interest and the reactivity of organic compounds. An uncertain initial concentration of a species did not result in uncertainties of the outcomes if the species was unreactive within the time scale of urban transport. Less obvious was the reason for the lack of variance contribution from highly reactive compounds such as internal alkenes. These compounds contributed to uncertainties in the very beginning of the simulation. Once they were depleted, the uncertain initial condition did not have lasting effects on the predictions of ozone, NO, and NO_2.

8.4.3 All Primary Organic Parameters

After short-listing primary organic parameters in group-specific experiments described in this section, 140 parameters uncertain parameters were studied together to elucidate any synergistic effects between the parameters. Cross terms were used in the PCE approximation for concentration, reaction rate, and product yields of the each compound being considered. The procedure depicted in Figure 8-1 was used to improve the accuracy of the representation while reducing the number of parameters to be considered. A sixth order representation with first order cross terms of 70 parameters provided convergence in the error as well as in the variance of ozone, NO, and NO_2 concentrations. 23 of the parameters were found to account for over 90% of the registered uncertainties. These parameters are listed in Table 8-6. The m-xylene parameters were particularly important for explaining the variance in ozone, NO, NO_2, and HO. An explanation for the importance of xylene parameters was that many other aromatic compounds were modeled after xylene (a correlation effect). Xylene was also quite reactive in the time scale of the simulation. The total amount of xylene and xylene-like compounds that reacted during the course of the simulation exceeded any other individual species.

Compound	Parameter	Compound	Parameter
2-Methylbutane	conc.	Toluene	conc
3-Methylhexane	k		k
4-Methylheptane	k		AFG2 (U2)
		M-xylene	conc
ethene	conc		k
	k_{OH}		U2
propene	conc	P-xylene	conc
1-butene	conc		k
2-pentene	conc	O-xylene	k
	k_{O3}	1,3,5 Trimethylbenzene	conc
			k
			U2
		1,2,3 Trimethylbenzene	conc
			k

Table 8-6. Primary Organics VCI.

8.5 Initial Conditions

8.5.1 IC ROG

In the composition uncertainty studies conducted so far, individual organic compounds were assumed to be measured separately and were treated as independent variables that affect both the initial conditions and the parameters of the reactions. In reality, however, total organic is routinely measured, and the speciation of organics is frequently fixed for modeling purposes (e.g. a source profile). This provides a motivation to look at the effects of different treatments of compositional uncertainties on the resulting uncertainties of the predictions. The following approaches were compared:

	Individual Species Amount	Total NMOC	Reaction Parameters
a	Vary	Renormalize to Fixed Total	Vary
b	Equal fraction of total	Vary (*)	Fixed
c	Vary	Equal sum of species (*)	Vary

Table 8-7. Experiments to Compare the effects of Composition on IC and on Parameters. (*Distribution determined in a used in c.)

When species were treated as independent variables, the uncertainties associated with the output variables (Figure 8-3a and 8-3c) were smaller than in the fixed composition case (Figure 8-3b). The correlation forced onto the individual species by the fixed composition resulted in higher variance in the predictions due to uncertainty of the total amount. An observed 18% standard deviation was roughly constant between the two approaches where the composition of each species was allowed to vary freely (independently), although the fixed total case predicted a mean value much closer to the nominal values than the variable total case. The fixed composition case resulted in a 23% standard deviation and a much larger deviation from the nominal value even though the same distribution of total IC NMOC was used in both cases b and c in Table 8-7.

It is possible to extrapolate this result to the case of an organic source with an uncertain profile and uncertain source strength. As long as some information is available about the speciation, the output uncertainties due to uncertain composition is limited, and reducing the uncertainty of source strength will reduce uncertainties of the predictions. On the other hand, having perfect knowledge of the organic profile will not help reduce uncertainties in model predictions if the source strength remains poorly known.

These observations show that correlation of the input uncertain parameters can have strong effects on the predictions. The problem is: correlations between parameters are not always easy to elucidate. In this case of organics, total NMOC is routinely measured by FID, independent of the composition. Composition information is gathered by a series of in-situ and sample measurements, which may have correlated or independent errors. The effects of correlation are also not easily generalizable. A complete treatment of the correlations is beyond the scope of this thesis, but the issue should be revisited when mathematical tools become available to deal with correlated variables efficiently.

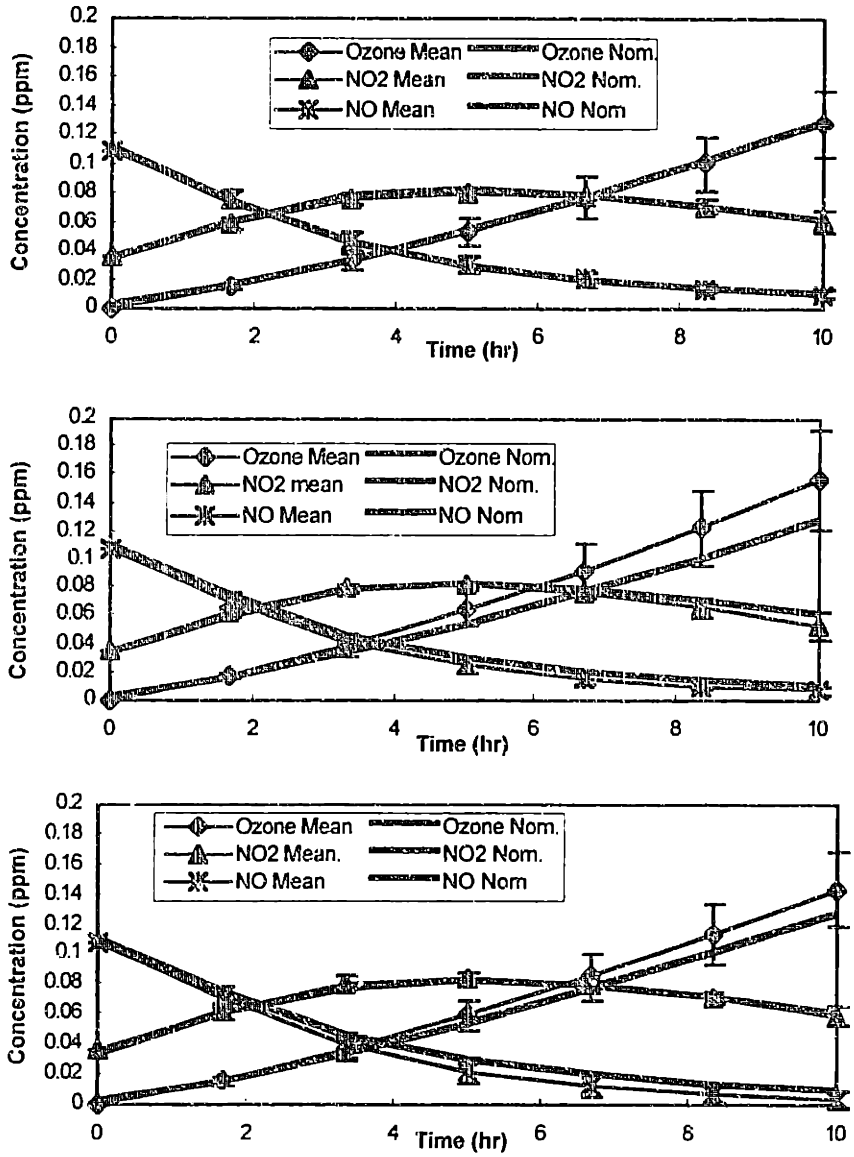


Figure 8-3a,b,c. Uncertainty Analysis Results of Composition Experiment. (a) Composition Varied, Total IC NMOC fixed. (b) Composition Fixed, Total IC NMOC varied. (c) Composition Varied, Total IC NMOC varied accordingly.

8.5.2 IC NO_x and IC ROG

The effects of uncertainties in the initial conditions were expected to depend on the absolute amounts of precursors (NO_x and ROG), i.e., whether the chemistry is NO_x limited or radical limited. Since this base case was chosen to simulate a situation with maximum incremental organic reactivity, the model behavior was expected to be sensitive to the amount of organics present. Furthermore, the base case also corresponded to a situation where the amount of ozone formed was negatively related to the amount of NO_x present, any uncertainty in the initial condition of NO_x was expected to contribute to uncertainties in the predictions.

Two approaches were used in the investigation of initial conditions to elucidate the effects of composition correlation at a less detailed level than in Section 5.1. When the initial condition of each modeled species (8 initial conditions for ROG and 2 ICs for NO_x) was treated as an independent uncertain variable with an uncertainty factor of 1.2, uncertainties in the predictions of ozone were about 40% of the mean predicted value of ozone, comparable to the variance caused by uncertain inorganic parameters (Section 8.2). Much of the uncertainties originated from uncertainties in the initial conditions of NO_x. However, another experiment, in which total ROG and total NO_x were the only uncertain parameters (the composition was assumed to be fixed), showed that the variance contribution from IC ROG exceeded that of IC NO_x. The predictions of ozone was much more uncertain when the composition of ROG and NO_x were fixed. In fact, the variance contribution of both uncertain initial conditions exceed that of the 10-parameter case. These results are compared in Figures 8-4a and 8-4b.

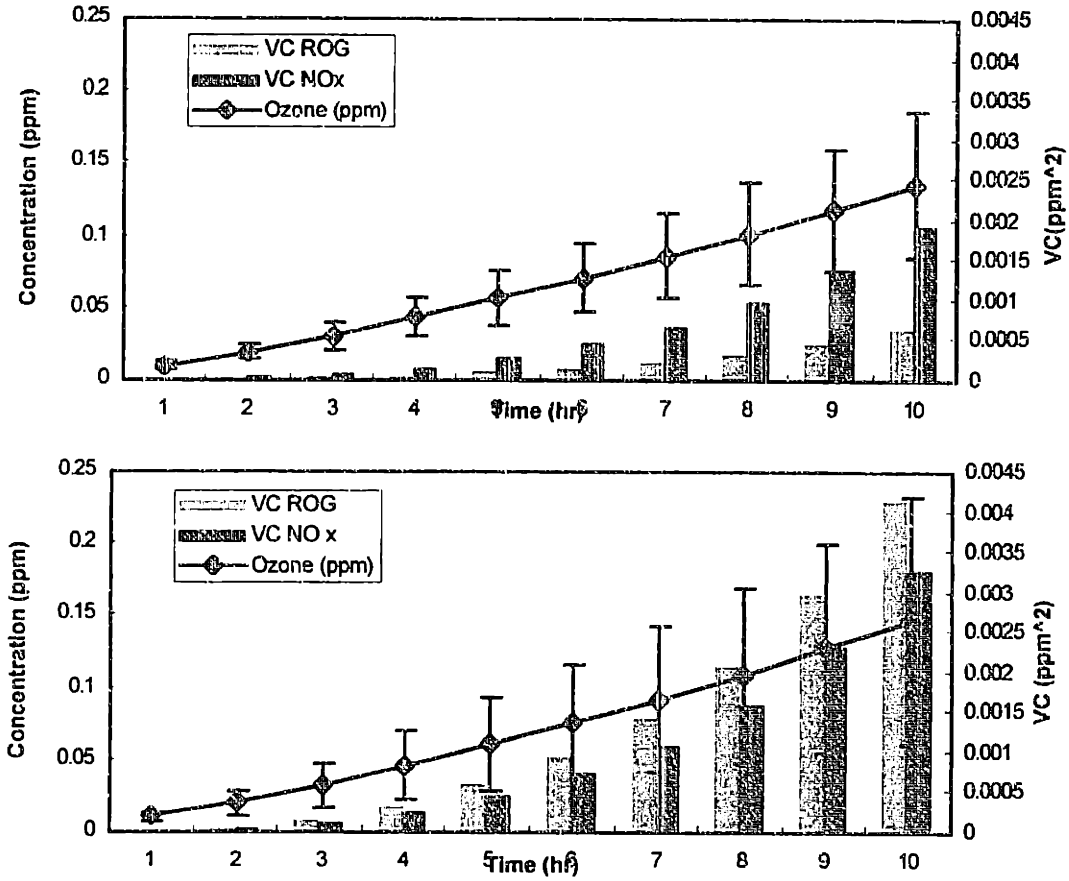


Figure 8-4a, b. Ozone Mean +/- Standard Deviation and VC from Initial Conditions. (Standard deviation represented by Error Bars)
(a) 10 independent parameters, 8 ROG and 2 NO_x species. (b) 2 independent parameters, IC ROG and IC NO_x, individual ICs determined as fractions of total.

The correlation among the initial concentrations of individual species in ROG and NO_x had significant effects on the uncertain model predictions. In both the correlated and the uncorrelated cases, the individual classes of ROG varied within a 20% range. However, the variation of total ROG was smaller in the case where the component species varied independently and much larger in the case with correlation. This was a result of the central limit theorem. Correlation of ROG concentrations resulted in a significant increase of the output variance attributable to ROG initial conditions. It was obvious that the effects of synergistic parameters such as initial concentrations of ROG species were enhanced by positive correlation. As discussed in the previous section, uncertainties in the total amounts of ROG (as a class) present in the system had significant implications for the uncertainties of model predictions in ozone, NO, and other species. The same observations could be made for NO_x (NO plus NO_2), to a lesser extent.

8.6 Combination Results

A comprehensive uncertainty analysis was conducted for all VCIs identified from uncertainty analyses of their respective subgroups. The purposes of this experiment were three fold:

- evaluate the total uncertainty of input parameters on the model predictions,
- compare the uncertainty contributions of input parameters, and
- elucidate any synergistic effects between parameters from different subgroups.

A list of parameters can be found in the Demmucom input file included in the appendix. Several inputs that were not identified as VCIs (e.g., k(HO+NO), jAFG) were also included as a spot check to confirm the small variance contributions of these parameters relative to the VCIs.

Figure 8-5 shows the results uncertainty propagation of the 37 VCIs identified in previous experiments. There was a significant increase (30%) in the prediction of the expected value of ozone concentrations over the nominal values. For NO and NO_2, uncertainties in the input parameters had a smaller, but observable, effect on the mean value of the predictions. Since the nominal value was calculated based on the median of most input parameters, the median output, as calculated based on Monte Carlo sampling of the PCE, was also compared to the nominal value. For ozone, the median was slightly lower than the mean, as shown in Figure 8-5. Both the mean and median predictions of ozone were more than 25% higher than the nominal value (see Figure 8-6). The significant increases of the mean and median values were due to the interaction between uncertain inputs and non-linearity in the response surface. For NO_2, the deviation from the nominal value was in the other direction, with the median and mean values lower than the nominal results.

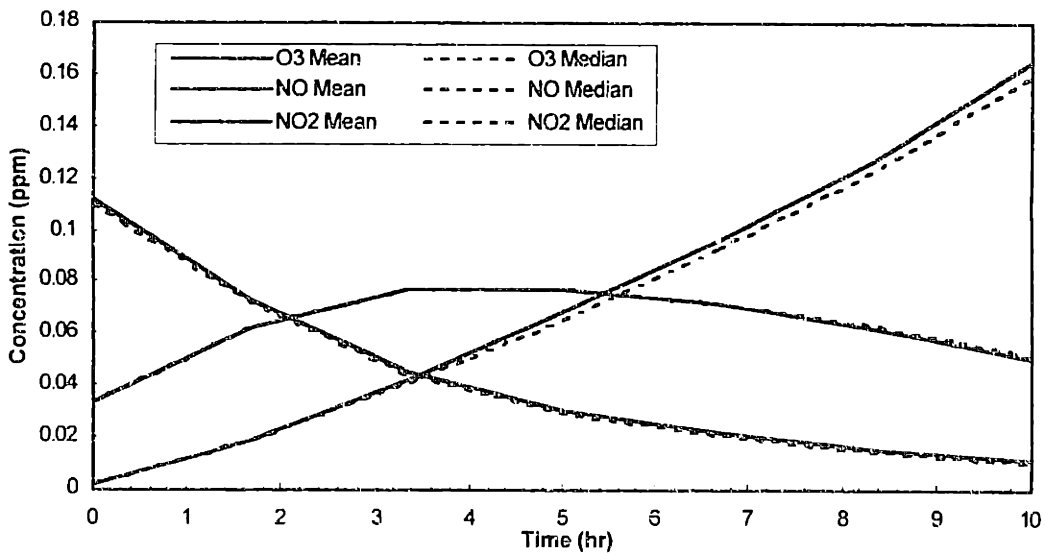


Figure 8-5. Mean and Median Values of Predictions of O_3, NO, and NO_2.

The uncertainties in ozone predictions (as shown by the standard deviations, represented by error bars in Figure 8-6) were found to be 45 - 50% of the mean value. The uncertainties of NO predictions reached a peak within the first two hours and then decreased as NO was

depleted, unlike the uncertainties of ozone and NO_2, which increased throughout the simulation.

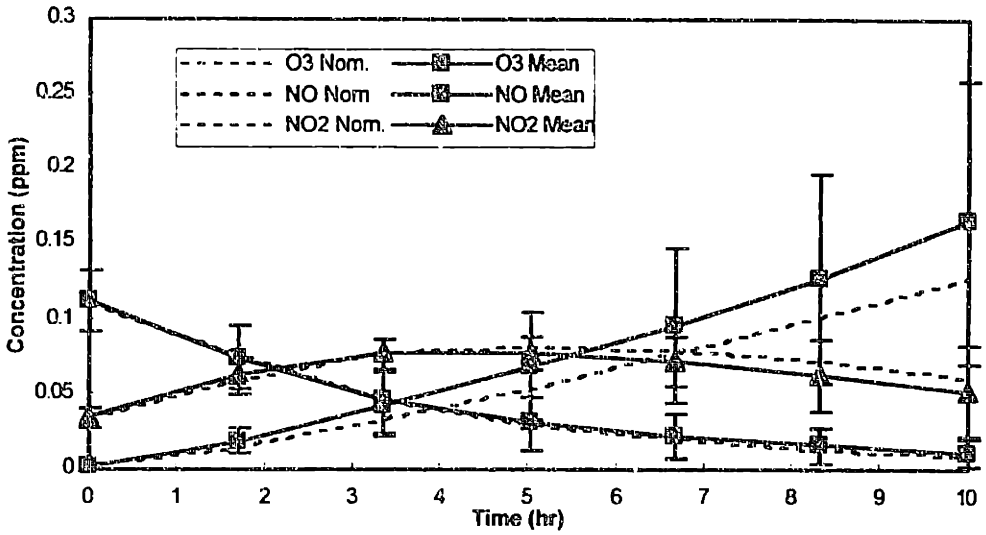


Figure 8-6. Uncertainty Analysis Results of All Parameters – Mean and Standard Deviation.

The uncertainty results for these species show that the effects from individual groups of input parameters (discussed in Sections 8.2 through 8.5) were not additive. This was probably a result of the interactions between the parameters, even though imperfect representation of the response surface could also lead to errors in the estimates of the variance in the various experiments. At the third order level, the maximum variance and the maximum error associated with each species are summarized in Table 8-8.

Species	Maximum Standard Deviation, % of Mean (Time)	Maximum Relative Error (Time)
O_3	94 ppb, 57% (10 hr)	3.1% (10 hr)
NO	22 ppb, 32% (1.3 hr)	3.3% (10 hr)
NO_2	33 ppb, 65% (10 hr)	2.5% (10 hr)
HO	0.23 ppt, 79% (10 hr)	7.1% (10 hr)
HCHO	2.6 ppb, 21% (5 hr)	1.2% (10 hr)
HOOH	0.46 ppb, 309% (10hr)	49.9% (10 hr)
HNO_3	24 ppb, 34% (10 hr)	1.9% (3.3 hr)
PAN	4.6 ppb, 80% (10 hr)	5.3% (10 hr)
RCO_3	2.2 ppt, 126% (10 hr)	22.0% (10 hr)
RO_2	6.6 ppt, 108% (10hr)	20.2% (10 hr)

Table 8-8. Maximum variance and error associated with SAPRC predictions (3rd Order).

Selected probability density functions at the 10^{th} hour are plotted in Figure 8-7. When sampled randomly, the third order PCE approximation resulted in probability density at negative concentrations for NO and NO_2. These samples were ignored in the PDF generation

procedure. The results indicated that the distributions of the response parameters were not symmetric. The distributions of ozone typically had long tails towards higher concentrations, while those of NO_2 had mirror tails at lower (and negative) concentrations.

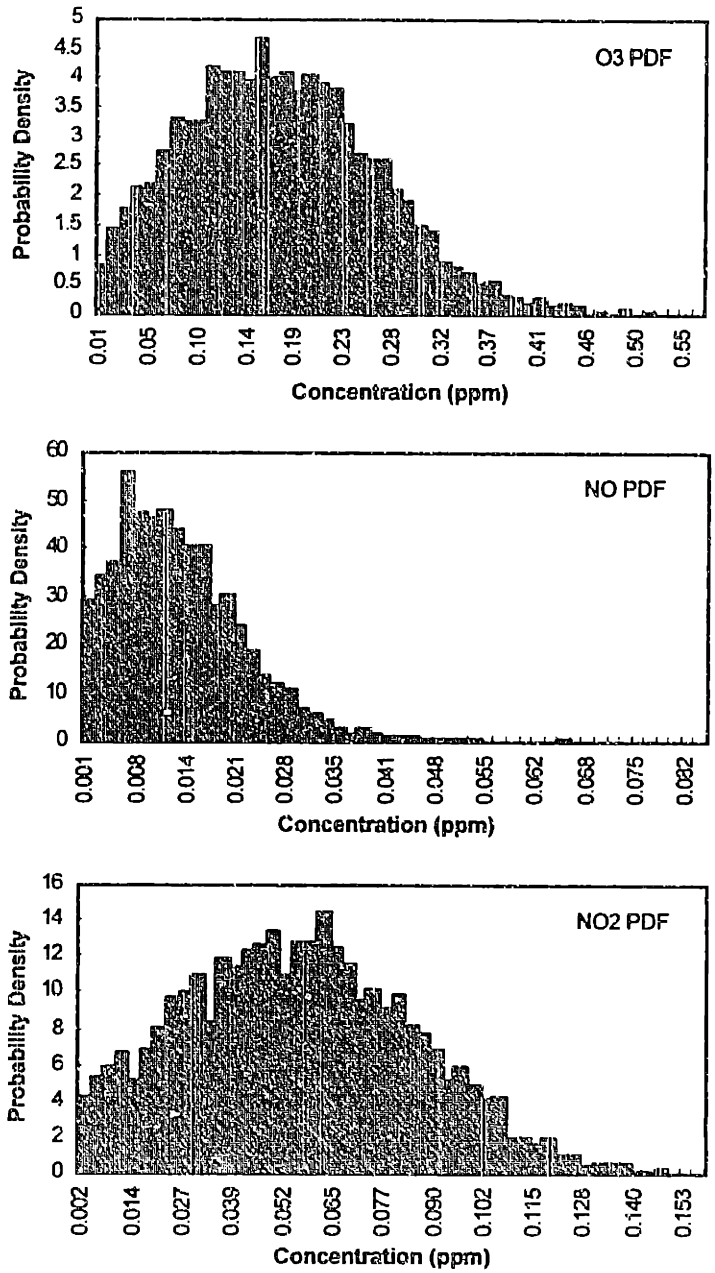


Figure 8-7. PDFs of Ozone, NO, NO_2 at the 10th hr.

The variance contributions of the various parameters to ozone are shown as a function of time in Figure 8-8. The initial conditions of NO and NO_2 had strong effects on the uncertainties in ozone. Consistent with the findings in Section 8.5.2 for independent initial conditions, the uncertainty contribution from uncertain organics species was smaller, but still quite significant. Of the kinetic parameters, the reaction rates of HO + NO_2 and O_3 + NO,

and the photolysis rates of HCHO, O_3, and NO_2 also contributed to the variance in ozone, with jNO_2 being particularly important in the beginning of the simulation.

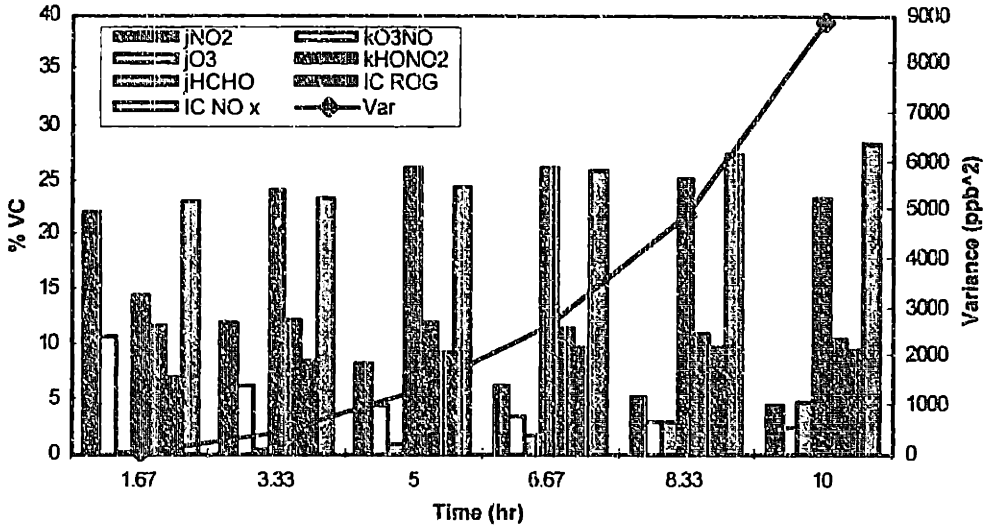


Figure 8-8. Ozone Variance Contribution

Figure 8-9 shows the variance contribution as a function of time to hydroxyl radical concentration. Many of the parameters that contributed to uncertainties in ozone concentrations also contributed to uncertainties in the HO concentration. In particular, the variance contributions from the reaction rate k(HO + NO_2) and the initial condition of NO_x were found to dominate the uncertainties in the HO concentration, indicating the importance of the sink reactions in controlling the dynamics of the system. The relative contributions of other parameters were quite different from Figure 8-8 (VC for ozone). For example, the photolysis rate of O_3 to $O(^1D)$ explained a large fraction of the variance of HO; the contributions of $O(^1D)$ reaction rates were also more significant to the concentrations of HO, especially towards the end of the simulation, than to any other response parameter.

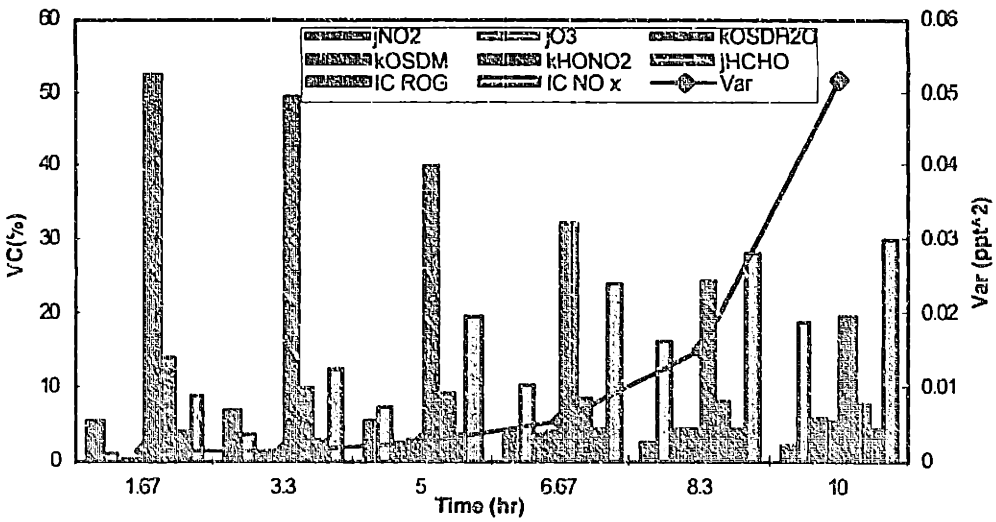


Figure 8-9. HO Variance Contribution

Formaldehyde, an intermediate product of organic decomposition in urban air, seemed less sensitive to the uncertainties in input parameters than most other product species investigated in this study. The variance reached a maximum roughly four hours into the simulation, and stayed within 4 - 6 ppb throughout the simulation. The VCIs identified for HCHO were IC HCHO in the beginning and jHCHO (to molecular products) during the latter part of the simulation. In addition, jHCHO (to radicals) and k(RCO_3 + NO_2) also had strong influence on the uncertainties of HCHO. This list was quite different from those for ozone and hydroxyl radical which were sensitive to the formation of sink products, such as nitric acid. The concentration of HCHO seemed to be most sensitive to the parameters specifically controlling the abundance of organic radicals (e.g., RCO_3). The variance contributions to formaldehyde by key input species are shown in Figure 8-10.

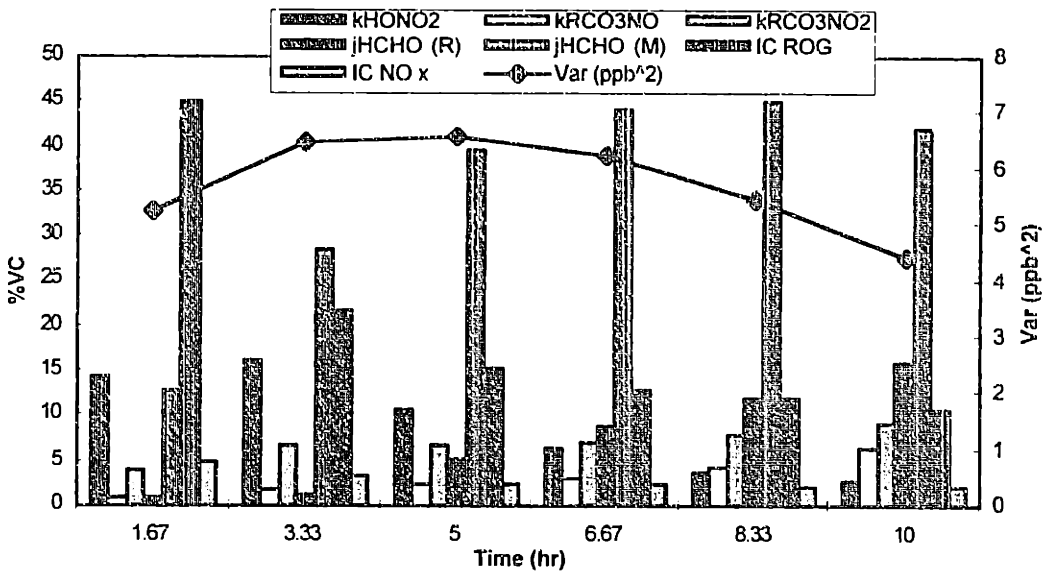


Figure 8-10. HCHO Variance Contribution

The main VCIs to NO_2 uncertainties were the initial conditions of NO_x, but a large number of parameters contribute at small but significant levels. The top ten VCIs combine to account for only 80% of the variance of NO_2 at some time points.

Another important nitrogen species was PAN, which functioned mostly as a sink of both organic radicals and NO_x in the base case simulation. PAN could also be a source of radicals and NO_2 in other environments, like downwind locations of urban areas. Even though the concentrations of PAN were only a few ppb in the simulation, uncertainty in the prediction of PAN could have significant implications due to its short life-time. The predictions of PAN varied by as much as 80% in the SAPRC simulations. Since PAN acted as a sink throughout the simulation, the uncertainties in PAN were affected by uncertainties in the formation rate of PAN (RCO_3 + NO_2) and in the reaction rate of RCO_3 + NO. These two rate parameters were particularly important in explaining the uncertainty of PAN concentrations at the beginning of the simulation. Later on, the same parameters that controlled the reactivity of the system and the availability of NO_x and radicals (such as the initial conditions, the reaction rate of HO + NO_2, and the photolysis rate of HCHO) contributed to uncertainties in PAN.

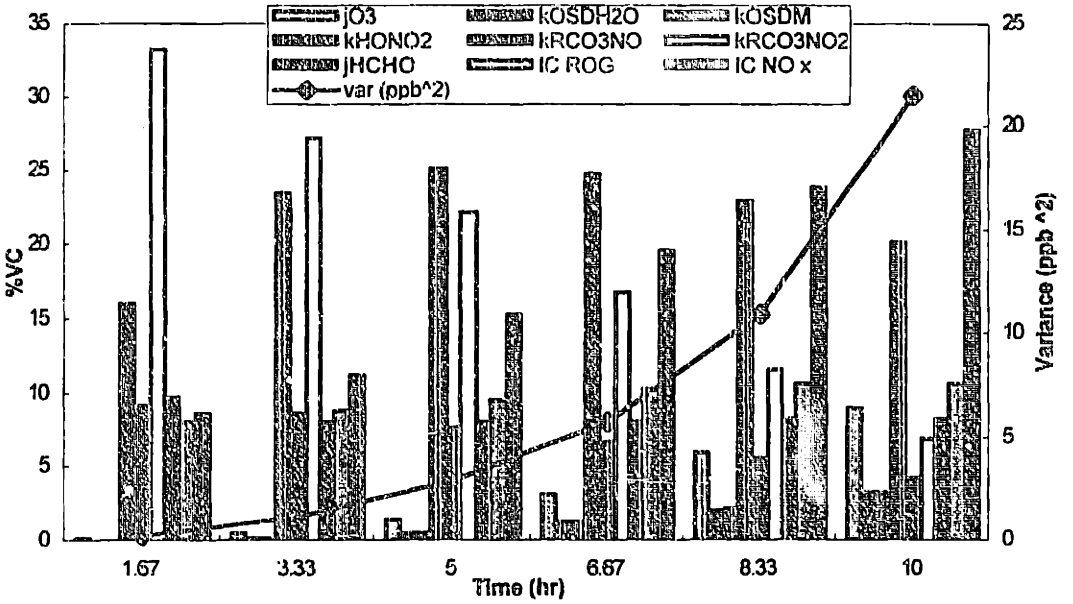


Figure 8-11. PAN Variance Contribution.

8.7. Discussion

The utility of Demmucom to conduct uncertainty analyses for large systems with many uncertain parameters was demonstrated in the SAPRC mechanism case study. A systematic procedure was used to eliminate successively parameters that did not contribute significantly to uncertainties of the outcome while improving the statistics of the estimation of the uncertain output by increasing the order of the approximation for the response surface. Key results from the analysis of the SAPRC mechanism are summarized as follows:

- The expected values of uncertain ozone concentration predicted by the SAPRC mechanism were always elevated from the nominal predictions using the best estimates of the parameters.
- The uncertainties in the predictions of ozone were significant (45 - 50 % of the mean predictions) due to parametric uncertainties in the inputs. The absolute variance for ozone was higher than other those in the predictions of NO, NO_2, and formaldehyde. PAN predictions from the SAPRC mechanism were extremely uncertain.
- Even though organic reactions compose the bulk of the SAPRC mechanism, most parameters identified to affect uncertainties in the model predictions of ozone, NO_2, etc. were inorganic reaction rate constants and initial conditions.
- HO + NO_2 is a chemical sink for both radicals and NO_x. Uncertainties in the rate of this combination reaction was found to contribute critically to concentrations of ozone, NO, NO_2, and radicals.
- Several photolysis reactions, jNO_2, jHCHO, jO_3, etc., were found to contribute to variance of predictions. These results indicated the importance of photolysis as a driving force to the formation of smog due to the radical formation potential of these reactions.
- In light of the dominating effect of uncertain photolysis reaction rates on the uncertainties in the predictions, a case study was conducted to investigate the effect of uncertain actinic flux by applying a multiplicative factor to all uncertain photolysis rates. (Uncertainty factors used in the study so far reflected only uncertainties in the activity spectrum.) It was found that uncertainty in actinic flux contributed at least as much to the variance as the uncertain activity spectrum of key photolysis parameters.
- Of the three classes of organic compounds, uncertainties in the aromatic reactions contributed the most to uncertainties in predictions of ozone. Uncertainties in the reactions of alkenes, the most reactive class of compounds modeled, did not have significant effect on the uncertainties in the predictions, except at very small time scales (localized effects). The reactions of alkenes were so rapid that all alkenes were broken down during the simulation and the uncertainty in alkene reaction rates did not contribute to uncertainties in the amount of products formed. Alkanes, though abundant, did not seem to have strong effects on the uncertainties of model predictions. However, they may contribute to uncertainties of predictions in locations much further downwind.
- Primary reactions of aromatics are not a great source of radicals, as evident from the lack of variance contributions from the radical yields of AROM + HO reactions. Using simplified parametric representation, secondary products were shown to be reactive and the product coefficients of these "second generation reactions" contribute to uncertainties in ozone and radicals.
- With the current formulation of mechanisms, uncertainties in organic rate and composition parameters contribute much more to uncertainties in the predictions than that in product coefficients, with the exception of reactive aromatic products.
- Correlation of the uncertainties in organic composition have strong effects on the resulting uncertainties in predictions.

- Initial concentration and source information are needed for species of intermediate reactivity, such as terminal alkenes and poly-substituted aromatics. These compounds have strong contribution to predictions at the urban time scale of the order of 10 hours. More reactive compounds have localized effects (short time scale) in uncertainties of predictions.
- The lumping representation may have decreased the possible variance contribution of uncertain organic parameters to uncertain predictions of the SAPRC mechanism.

In additional to parametric uncertainties, structural uncertainties may also contribute significantly to uncertainties in the predictions of the SAPRC mechanisms. Structural uncertainties arise from the hypothesis and assumptions used in formulating the model. In many cases, the only way to test the effects of these design decisions is to compare the predictions with a different formulation of the model (See Chapter 4, Aumont *et al.*, 1996).

Two areas warrant special attention in the SAPRC mechanism. One is the parameterization of aromatics. The parametric uncertainty analysis in Section 8.3 showed that the reactions of the second generation compounds from aromatic reactions have significant effects on the uncertainties of the predictions. Ozone and radical concentrations were particularly sensitive to the product coefficients of AFG, the unknown aromatic fragmentation products. It is therefore important to characterize the products currently represented by AFG1 and AFG2. The structural uncertainty in the representation needs to be resolved before the parametric uncertainties of the product coefficients can be fully evaluated.

A second area is the lumped representation of organic compounds. The representation of secondary organics by surrogate species that are usually small members of the homologous series may also be a source of uncertainties of output uncertainties, because multiple cycling between peroxyl radicals and carbonyls are ignored. A simple experiment was conducted to capture the effects of these uncertainties using uncertain product coefficient parameters. The amount of RO_2 formed from secondary organics was treated as uncertain to represent additional NO-to-NO_2 conversion from higher species. The range of uncertainties was determined by the number of NO-to-NO_2 conversions resulting from the mechanistic breakdown of the modeled species and that from the next member of the homologous species. Figure 8-12 shows the comparison between the nominal value of the ozone and the expected value of the prediction with extra NO-to-NO_2 conversions. The mean prediction of ozone would be elevated by as much as 25%.

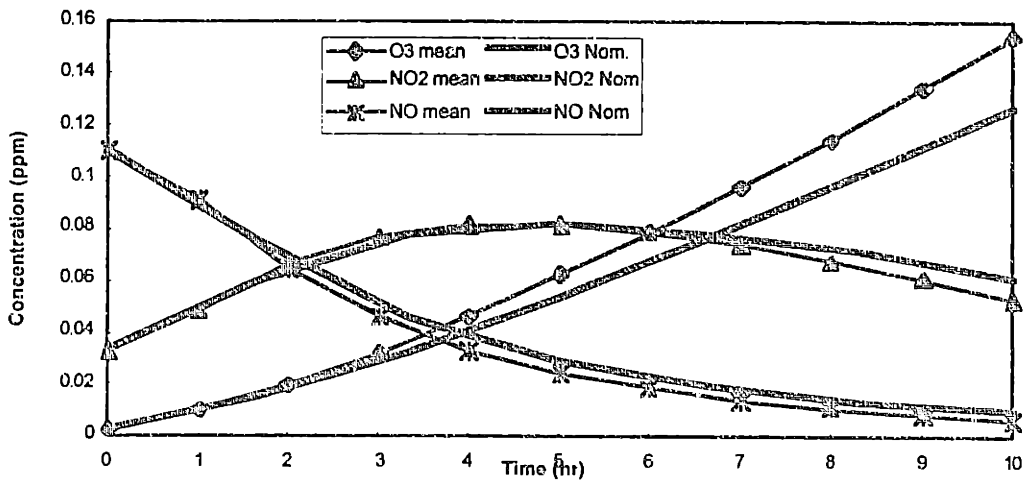


Figure 8-12. Uncertainty Analysis of Lumping of Secondary Organics.

8.8. Implications of Results for Field and Laboratory Measurements

Uncertainty analysis provides a valuable tool to help set priorities for laboratory and field studies that will improve the predictions of air quality models. Based upon the findings of this work, the following priorities can be inferred in the myriad of needs in the field of atmospheric modeling.

- The photolysis of HCHO to form radicals is a key VCI. Since there are two pathways to the photolysis of HCHO (one forming peroxyl radicals and CO, the other forming H_2 and CO) and the quantum yields are not very well characterized, laboratory experiments are recommended measure the uncertain quantum yield and to reduce the uncertainty of this photolysis rate.
- Uncertainties in the actinic flux translates into uncertainties in all photolysis rates. Field measurement of actinic flux should be taken at all measurement campaigns to provide better estimates of photolysis reaction rates for modeling studies.
- Some have suggested that the HO + NO_2 reaction rates does not follow the Troe formula (Donahue *et al.*, 1997). Due to its important contribution in the predictions of current chemical mechanisms, a re-evaluation is recommended for this reaction rate constant. New laboratory data may also help provide a better estimate of the pressure-dependent reaction rate constant.
- Organic speciation data are completely lacking in some parts of the country. There is no doubt such information is needed. However, once some information is available concerning the organic profile, additional reduction in uncertainty in model prediction is gained by better characterization of source strength rather than a perfect knowledge of the profile. In any given geographic area, spatial coverage of total ROG measurement may be more valuable than detailed speciated measurements at a few sites for modeling purposes.
- Since the concentration of ozone is sensitive to the parameterization of aromatic reactions, especially to the reaction parameters of the "unknown aromatic fragmentation products", it is recommended that this parameterization be constantly reviewed as new information becomes available regarding the reactions and reaction products of aromatic compounds. New information should be incorporated into existing mechanisms to reduce structural uncertainties. Ultimately, better characterization of aromatic reactions and products would provide the basis for a better and more meaningful basis for parametric uncertainty analysis.
- Better knowledge of inventory and initial conditions seem to be the best way to increase the confidence of the predictions of atmospheric models. This is particularly important for species that are both abundant and reactive.

9. Mechanism Comparison in the Presence of Parametric Uncertainty: Uncertainty Analyses of RADM / CB4 / SAPRC

9.1 Introduction

In any model, uncertainties in input parameters lead to uncertainties or errors in the predictions. These errors may have significant implications on the model validation process, because competitive candidate models may become indistinguishable in the presence of uncertainties in the predictions. This is illustrated in Figure 9-1.

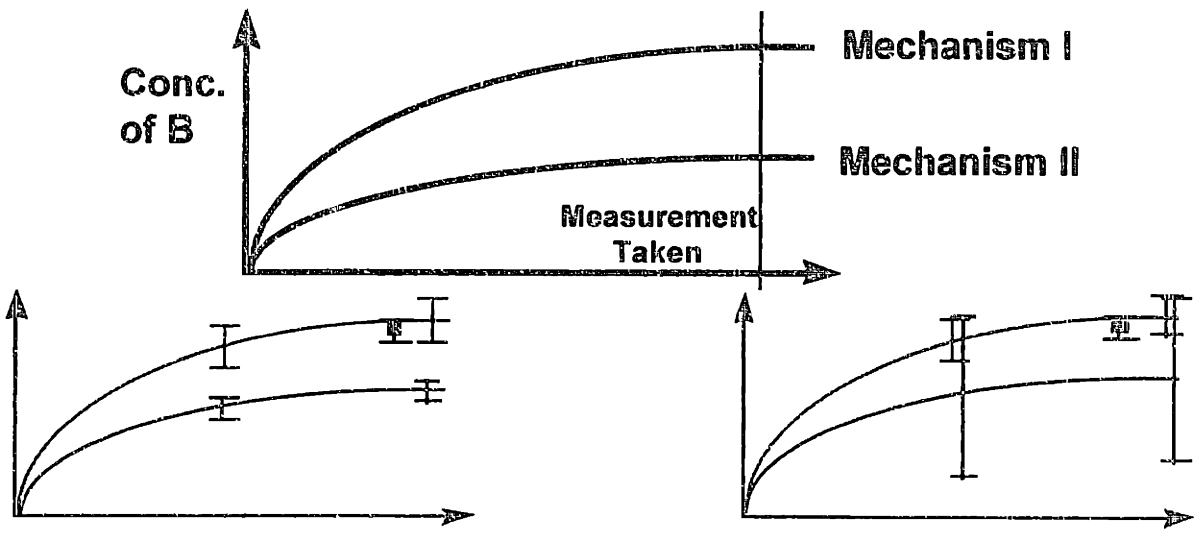


Figure 9-1. Mechanism Discrimination in the Presence of Uncertainties.
Mechanism I is A→ B; Mechanism II is A→ pB + (1-p)C

Suppose two candidate mechanisms, A → B and A → pB + (1-p) C, are proposed for a particular sequence of reactions. Suppose further that B is the quantity being measured. If there is no uncertainty in the parameters such as rate constants and product coefficients, a measurement can be taken at the end point, when the concentration of B approaches an asymptotic value. Provided that the measurement is reasonably accurate, it can be used to discriminate the correct mechanism from the incorrect one. If B approaches the initial concentration of A (A_0), then Mechanism I is correct. If B approaches a fraction p of A_0, then Mechanism II is consistent with the data. However, when parametric uncertainties are present, drawing conclusions in same exercise can become more complicated, because predictions are uncertain. If the uncertainties in the predictions are small, as in the bottom left panel of Figure 9-1, measurements may still be used to discriminate between the competing mechanisms. However, sometimes parametric uncertainties result in large

uncertainties in the predictions. For instance, if the product coefficient is poorly know in this example, uncertainties can be quite large in the predictions. This scenario is shown in the right bottom panel for Mechanism II. In that case, a measurement of B may not discriminate between the mechanisms, because both sets of predictions are consistent with the measurement. Without better characterization of the parameters in the model, the relative merits of the two mechanisms cannot be evaluated. The range of the uncertainty bounds should not be used as a measure of the quality of the model, because uncertainties in the model structure are not captured in the parametric analysis.

Several chemical mechanisms are currently being used in air quality or atmospheric pollutant dispersion models. Although all chemical mechanisms developed to-date have been validated against the same existing body of smog chamber measurements, their intended applications were different, and thus have different assumptions and focus. In this study, three mechanisms were compared in the presence of parametric uncertainty. The key questions to address are:

- How do different mechanisms behave in the presence of parametric uncertainty?
- Are the mechanisms distinguishable within the current accuracy limits of the input data?
- Do the same parameters the contribute to variance of the predictions of different mechanisms?
- What is the most efficient resource allocation strategy? Should experiments be done to define better the parameters used in current models or should research focus on improving the theoretical understanding behind these models?

9.2. Descriptions of Mechanisms

The three mechanisms compared in this study are the Regional Acid Deposition Model (RADM), Carbon Bond Mechanism IV (CB4), and Statewide Air Pollution Research Center (SAPRC) chemical mechanisms. Table 9-1 presents a comparison of the key features of RADM, CB4, and SAPRC mechanisms. RADM was first created in 1986 (Stockwell, 1986). In 1990, the second generation mechanism was published (Stockwell, 1990) to include details for simulating organic chemistry, especially aromatic chemistry. The mechanism contains 158 reactions of 63 model species.

In the Carbon Bond Mechanisms (CBM) (Gery *et al.*, 1989), organic compounds are modeled according to the bond types, e.g., carbon single bonds, carbon double bonds, and carbonyl bonds. The reactions of the lumped structures were determined by condensing reactions that represented different groups of bonds within that lumped structure, as explained in Gery *et al.*, 1988. (For example, the PAR reactions were condensed from reactions describing primary, secondary, and tertiary C-H bonds). This representation of reactions is highly parameterized and is therefore less transparent than the lumped molecule approach used in the RADM and SAPRC. The version used in this study is the Carbon Bond Mechanism IV (CB4) which utilizes 81 reactions of 32 species. The SAPRC mechanism was described in detail in Chapter 7 of this document.

The inorganic reactions used in the RADM, CB4, and SAPRC mechanisms are comparable in detail and are typical of urban and regional scale tropospheric chemical models. This group of reactions is represented explicitly in all mechanisms. The rate and product information is compiled by the IUPAC panal (Atkinson *et al.*, 1989).

	RADM	CB4	SAPRC
Active Inorganic Sp.	15	14	15
No. Inorganic Rxns.	33	36	42
Inorganic Photolysis Reactions	9	6	7
Organic Product Sp.	14	9	19
Aromatic Products	4	3	10
Organic Peroxides and Acids	5	none	1
Organic Products Rxns.	106	30	106
Organic Photolysis Reactions	12	5	13
Organic Radicals	16 Peroxyl 5 alkane-derived 3 alkene-derived 1 alkene-NO3 adduct 2 aromatic radicals 3 peroxyl carbonyl 2 "operators" to account for additional organic nitrates and NO-to-NO2 conversions by lumped organic species	2 Peroxyl (C2O3, TO2) 2 Peroxyl Operators (XO2, XO2N) 2 Alkoxy (ROR, CRO)	15 Peroxyl 2 total peroxyl species (RO2, RCO3) 5 acyl radical species 2 phenoxy radicals 6 "chemical operator" species to represent aspects of RO2 reactions
Mass Balance	no	yes, -ve product coef.	LOSTC, LOSTN sp.
Primary Organic Reactions	19 CH4 4 for alkanes 3 rxns each for OL2, OLT, OLI 2 for aromatics 3 for ISO	15 CH4 PAR 7 rxns for ETH/OLE TOL XYL 4 rxns for ISO	user defined This study uses 15 3 classes of alkanes and aromatics (1 rxn each) ethene, 2 classes of alkenes (4 rxn each)
Lumping of Primary Organics	Parameters of reactions fixed. Reactivity weighting to vary initial concentrations	Lumped structure approach	Rates and product coef. Calculated as weighted average of all compounds in the lumped class. Initial condition also defined by lumping.

Table 9-1. Key Features of the RADM, CB4, and SAPRC mechanisms.

The bulk of urban/regional scale chemical mechanisms, is comprised of reactions of organic compounds. The foci of organic reactions in RADM, CB4, and SAPRC mechanisms are slightly different. Aromatic reactions are represented in the greatest detail in the SAPRC mechanism. On the other hand, RADM provides detail on radical combination reactions and products such as peroxides and organic acids. CB4 provides a reasonable amount of details of aromatic and biogenic reactions. However, it does not simulate organic peroxide and organic acids at all. These differences reflect the intended applications for SAPRC and CB4 in urban environments and RADM in a regional scale model. However, the major distinctions between the mechanisms lie in their authors' different philosophies in organic lumping and the representation of organic radical reactions.

Organic lumping. A major difference between the mechanisms is the lumping of individual organic compounds into modeled classes. In RADM, the modeled molecules and reactions are pre-determined, and each individual species is mapped onto a model species according to its homologous series and reaction rate. In the RADM implementation, organic lumping is typically accomplished in two steps. First, individual species are assigned to one of 32 emission categories (Middleton *et al.*, 1990) on a mole-to-mole basis. Second, a reactivity lumping procedure is applied to lump the 32 emission categories into modeled species such as HC3, HC5, etc. An implicit assumption in the procedure is that the effect of VOC emissions on the model simulation results is approximately proportional to the amount of the compound that reacts in a given period of time (a day). An emitted compound is represented by a model species which reacts at a different rate using a weighting factor that is applied to the compound emissions. The weighting factor (*WF*) is the ratio of the fraction of the emitted compound which reacts to the fraction of the model species which reacts:

$$WF = \frac{1 - \exp(-k_{HO-Emit} \cdot \int [HO]dt)}{1 - \exp(-k_{HO-Model} \cdot \int [HO]dt)}, \qquad (9.1)$$

where $k_{HO\text{-}Emit}$ and $k_{HO\text{-}Model}$ are the hydroxyl reaction rates of the emitted species and the modeled species respectively. $\int [HO]dt$ is the daily average integrated HO radical concentration. A value of 110 ppt-min was recommended by the author of RADM (Stockwell, 1990). Note that this value is larger than the value used in SAPRC modeling, reflecting the longer time-scale considered in regional modeling, RADM's principle application. The three classes of alkanes are HC3, HC5, and HC8. HC3 represents alkanes with HO reaction rate constants < 5e3 $ppm^{-1}min^{-1}$, and HC3 reacts with HO at a rate of 3.84e3 $ppm^{-1}min^{-1}$ in the mechanism. HC5, representing alkanes with HO reaction rate constants between 5e3 and 1e4 $ppm^{-1}min^{-1}$, reacts in the model at 7.14e3 $ppm^{-1}min^{-1}$ with hydroxyl radicals. HC8 represents the most reactive alkanes. The average rate constants for reactions of HO with HC8 is 1.5e4 $ppm^{-1}min^{-1}$. In addition to product yields that are representative of the species in each class, the radical operator XO_2 is also "produced" in HC5 + HO and HC8 + HO reactions to accomplish additional NO-to-NO_2 conversions by the higher members of the lumped class.

The RADM mechanism models aromatic reactions using two species: TOL and XYL. Reactivity weighting is used only for lumping benzene and haloaromatic compounds into TOL. Other aromatic species are lumped on a mole-to-mole basis. Alkenes are represented by four species: OL2 for ethene, OLT for terminal alkenes, OLI for internal alkenes, and ISO for isoprene. Alkenes are lumped on a mole-to-mole basis, on account of their high reactivity.

CB4 employed the "lumped structure" approach to assign organic compounds to the model species, which represent different types of bonds. The structures represented explicitly were PAR (single bond), OLE (double bond), ALD2 (carbonyl bond), TOL (mono-substituted benzene), and XYL (di-substituted benzene). The central assumption of the lumped structure approach is that the overall reactivity of a branched molecule is the same as that of the corresponding straight chain isomer. The main advantage of the lumped structure approach over the lumped molecule approach, used in the both the RADM and SAPRC mechanisms, is that fewer modeled species are needed to represent bond groups. Therefore, the resulting mechanism is smaller and more easily implemented in large air quality models. Ethene, isoprene, and formaldehyde are represented explicitly, even though ethene is also used to simulate isoprene products. A few examples are listed in Table 9-2 to illustrate the lumping procedure.

Organic Species	Representation
n-Butane	4 PAR
2,2,5-Trimethylpentane	8 PAR
Propene	1 OLE + 1 PAR
trans-2-Butene	2 ALD2
Ethylbenzene	1 TOL + 1 PAR
Propionaldehyde	1 ALD2 + 1 PAR

Table 9-2. Representation of Selected Organics in Carbon Bond Mechanisms (Gery et al, 1989)

As indicated in Table 9-1, the SAPRC mechanism employs a flexible organic lumping scheme. Users can define the number of classes used to represent alkanes/aromatics and alkenes based on the desired level of detail in the simulation. A detailed explanation of the lumping procedure is provided in Chapter 7. All reaction parameters of the lumped class are defined as weighted averages of the parameters of the individual species in each lumped SAPRC class. Therefore, the requirement for mechanistic information in using SAPRC exceeds that of RADM and CB4. The standard procedure for SAPRC lumps alkanes and aromatics together, while the other two mechanisms model them separately. The current lumping strategies used for SAPRC and RADM mechanisms result in very similar classes for alkenes both lumped molecule mechanisms.

Primary organic reaction sequence (from the formation of peroxyl radicals to the production of carbonyls). The sequences of primary organic reaction modeled in RADM, CB4, and SAPRC are different. Figure 9-2 illustrates the reactions as they are modeled in the three mechanisms. RADM contains a set of pre-determined primary organic species: CH4, ETH, HC3, HC5, HC8, OL2, OLI, OLT, ISO, TOL, and XYL. There are 11 corresponding peroxyl radicals, each with its specific reactions. Even though the mass is not conserved in some of these reactions, the reactions are representative of compounds of the specific functional class and size. The reactions of the alkoxyl radicals are lumped into the reactions of peroxyl radicals with NO. Therefore, secondary carbonyl products are formed in the RO_2 + NO reaction step.

The lumped structure approach allowed CB4 to model the reactions of organics with fewer reactions than either SAPRC or RADM. In the CB4 mechanism, the initial reaction of the primary species and the subsequent reactions of the peroxyl radicals with NO are condensed into one reaction, forming alkoxyl radicals and RO_2 operators (which represent NO-to-NO_2 conversions and organic nitrate formation). The production of alkoxyl radicals is "ahead of schedule." The subsequent reactions of the alkoxyl radical are modeled explicitly with fixed branching ratios for isomerization, decomposition, and reaction with oxygen.

SAPRC yet another different strategy in representing primary organic reactions. The peroxyl radicals formed from the initial hydroxyl attack of alkanes, aromatics, and alkenes are represented by operators and the products of reaction in the presence of NO_x, with an RO_2 species to count the total peroxyl radical concentration. The operators reacted with NO, NO_2, HO_2, and other peroxyl radicals which produced NO_2, alkyl nitrates, organic hydroxides, and other stable products. The carbonyl compounds that are usually produced as a result of the reactions of alkoxyl radicals (formed from the reaction of peroxyl with NO) are included as products of the primary organic + HO step. The use of peroxyl radical

operators has the advantage of flexibility when used with mechanisms allowing variable lumping strategies, such as SAPRC. The disadvantage of this representation, however, is that the sequence of product formation is unrealistic. Specifically, the carbonyl products are formed as soon as the primary organic compound reacts with HO radicals, but the NO-to-NO_2 conversion takes place later when the peroxyl radical operator reacts with NO.

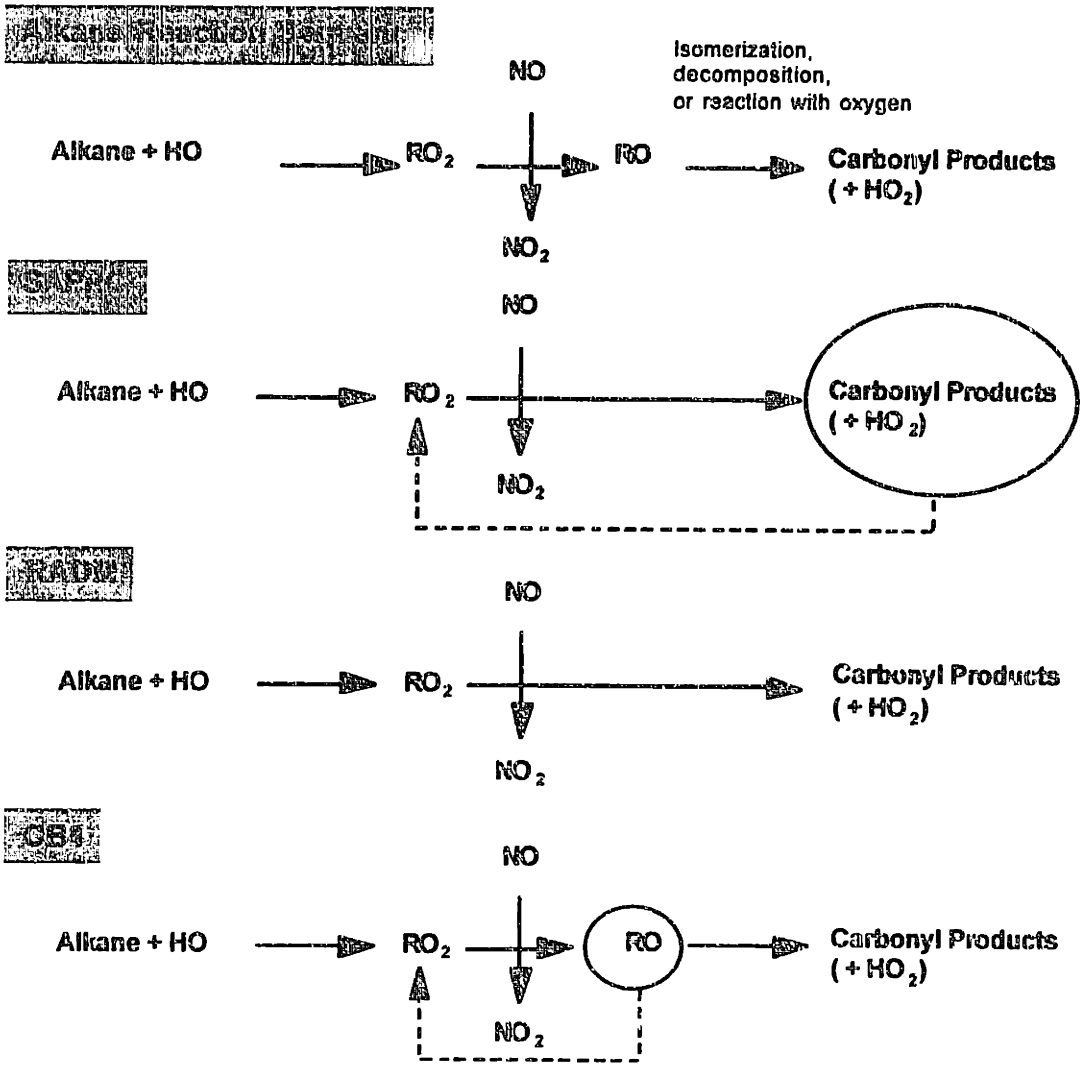


Figure 9-2. Illustration of Primary Reaction Sequence in SAPRC, RADM, and CB4 Mechanisms.

All three mechanisms, RADM, CB4, and SAPRC, describe aromatic reactions using a parameterized representation with "unknown aromatic ring fragmentation products." SAPRC models aromatic species with more detailed product species than either RADM or CB4. For example, SAPRC employs glyoxal, methylglyoxal, and two species of unknown ring opening products (different reactivity) while RADM and CB4 use only methylglyoxal and one unknown ring opening product to describe the reactions of aromatic compounds. SAPRC represents Benzaldehyde chemistry explicitly, but both RADM or CB4 use specific radical species for toluene reactions.

9.3 Base Case Comparison

The base case used in this study (see the Appendix of the SAPRC mechanism), simulated the chemistry of a high incremental reactivity (MIR) scenario (low VOC-to-NO_x ratio) in a box model environment. The base case scenario was simulated using SAPRC, CB4, and RADM mechanisms. Figure 9-3 shows that the predictions of key species such as ozone, NO, and NO_2 were very similar. CB4 seemed to produce slightly more ozone and less NO_2 in 10 hours. The similarity of the predictions from the three mechanisms provided a good basis for comparison of the mechanisms' behaviors with uncertain parametric inputs. A series of case studies were performed on the RADM, CB4, and SAPRC mechanisms. Case studies were chosen based on the current understanding of the key similarities and differences among these mechanisms, and on key uncertain reaction parameters found to cause uncertainties in the SAPRC mechanism in Chapter 8. The uncertain predictions were analyzed to compare the effects of input uncertainties in the mechanisms and the parameters contributing to output uncertainties in each mechanism were identified.

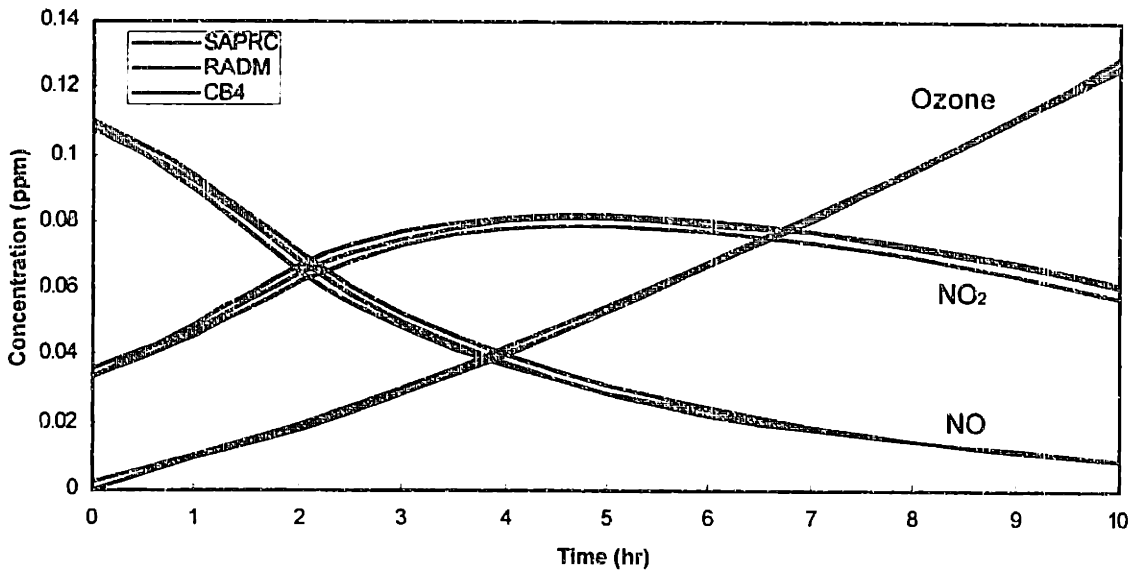


Figure 9-3. Comparison of Base Case Simulation Using the SAPRC, CB4, and RADM mechanisms.

9.4 Photolysis Rate Case Study

Photolysis reactions play a critical role in the chemistry of photochemical smog, as discussed in Chapter 2. Several photolysis rate constants were found to have dominating effects on the uncertainties in ozone predictions in the SAPRC mechanism in Chapter 8. Many common photolysis reactions are included in SAPRC, CB4, and RADM mechanisms. Therefore, this class of reactions was selected for investigation and for comparison between different mechanisms

9.4.1 Description of Uncertain Parameters

The photolysis reactions employed in the mechanisms are listed in Table 9-3. Many photolysis reactions are common between RADM and SAPRC. RADM includes more inorganic photolysis reactions and SAPRC emphasizes photolysis reactions of secondary organic compounds. CB4 contains fewer photolysis reactions than RADM or SAPRC. Relative rates are used in CB4 for O_3 (→ O^3P) , NO_3, HONO, H_2O_2, UAFP, and MGLY. These rates were taken to be directly proportional to another photolysis reaction rate, typically that of NO_2 or HCHO.

Shared Photolysis Rates	NO_2, O_3 (→ OSD), HCHO(→ radicals), HCHO (→ molecules), CCHO, ACET, ROOH, GLY (→ radicals), GLY (→ molecules)
Shared Rates, Relative Rates Used in CB4	O_3 (→ O^3P), HONO, NO_3, H_2O_2, Unknown Aromatic Fragmentation Product (UAFP), MGLY
RADM Only	HNO_3, HNO_4 (ignored in SAPRC) peroxy-acetic acid, organic nitrate higher organic peroxide (same rate as ROOH)
SAPRC Only	RCHO, MEK, BALD additional UAFP (rate constant correlated to the other UAFP) additional NO_3 photolysis pathway (RADM treats as product split)

Table 9-3. Photolysis Reaction Rates Employed in SAPRC and RADM.

Uncertainties in several photolysis rates not used in the SAPRC mechanism were not compiled previously by Stockwell (1994). For the study of the RADM mechanism, uncertainty factors were assigned to these photolysis rates based on Atkinson *et al.* (1992) and the references therein. These uncertainty factors were assigned to reflect the agreement / disagreement and availability of experimental data, and are listed in Table 9-4.

Photolysis Reaction	Uncertainty Factor of (QY * AC)
HNO3 —> HO + NO2	1.2 (1.3^)
HNO4 —> HO2 + NO2	1.6 (2.0^)
PAA —> MO2 + HO	2.1
ONIT —> 0.2 ALD + 0.8 KET + HO2 + NO2	1.3
OP1 / OP2	1.8

Table 9-4. Uncertainty Factor Assignment to Photolysis Reactions Used in RADM Only.
^ Stockwell, 1994 for SAPRC

The multiplicative factors used in CB4's relative photolysis rates were assigned the same uncertainty factors used for the independent photolysis rates for the identical species in SAPRC and RADM. Therefore, the resulting uncertainties in relative photolysis rate constants were higher than their independent counter parts, due to the correlation with jNO_2 or jHCHO.

9.4.2 Photolysis Case Study Results

Demmucom (Chapter 5) was used to propagate input photolysis rate uncertainties in the RADM, CB4, and SAPRC models. The results of a third order PCE representation of the uncertain outputs are shown in Figures 9-4 to 9-7. Figures 9-4a, 9-4b, and 9-4c compare the expected (mean) and median dynamics of O_3, NO, and NO_2 for SAPRC, RADM, and CB4 mechanisms respectively. The nominal results based on the best estimate photolysis rates were also included in these figures for comparison purposes. In the presence of uncertain photolysis rates, ozone formation was accelerated compared to the base case model predictions. By the end of the simulation, the mean concentrations were elevated roughly 10% relative to the nominal results for both the SAPRC and RADM mechanisms. The elevation of ozone level was even more significant for the CB4 mechanism. The median values of the predictions, on the other hand, were less elevated from nominal results in all three mechanisms compared to the mean values, indicating the skewed nature of the density functions describing the uncertain ozone concentrations.

For NO predictions, the mean predictions agreed better with the nominal the median predictions for all three mechanisms. CB4 seemed to be the only mechanism that showed a slight decrease in NO consumption rates relative to the nominal value. On the other hand, SAPRC and RADM mechanisms predicted slightly faster NO consumption rates in the presence of uncertain photolysis rate constants, especially during the first half of the simulation. All three mechanisms predicted NO_2 mean and nominal concentrations to be lower than the expected values, with CB4 predicting the biggest change.

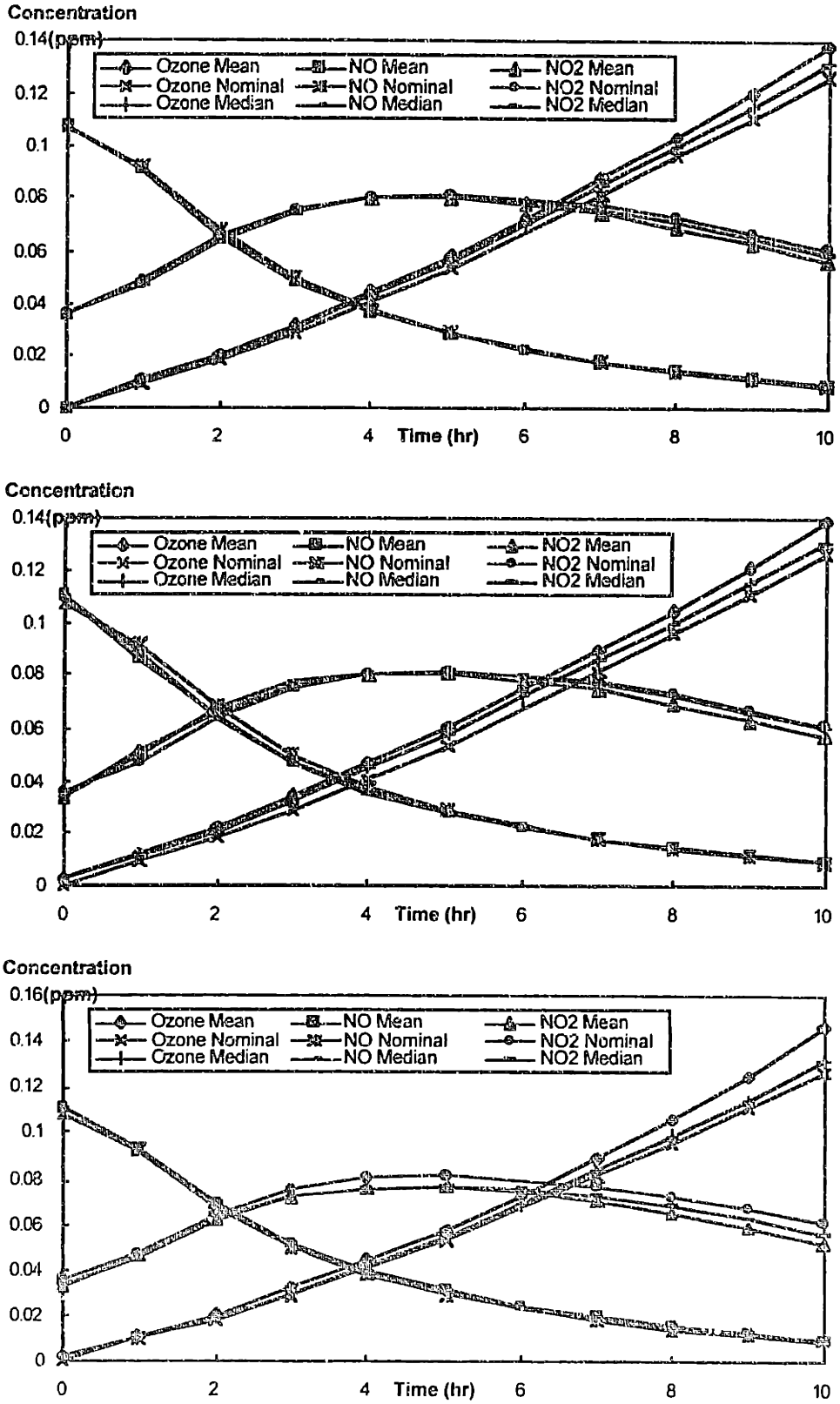


Figure 9-4. Dynamics of Ozone, NO, and NO_2 -- Response to Uncertain Photolysis Rates. (a) SAPRC, (b) RADM, (c) CB4.

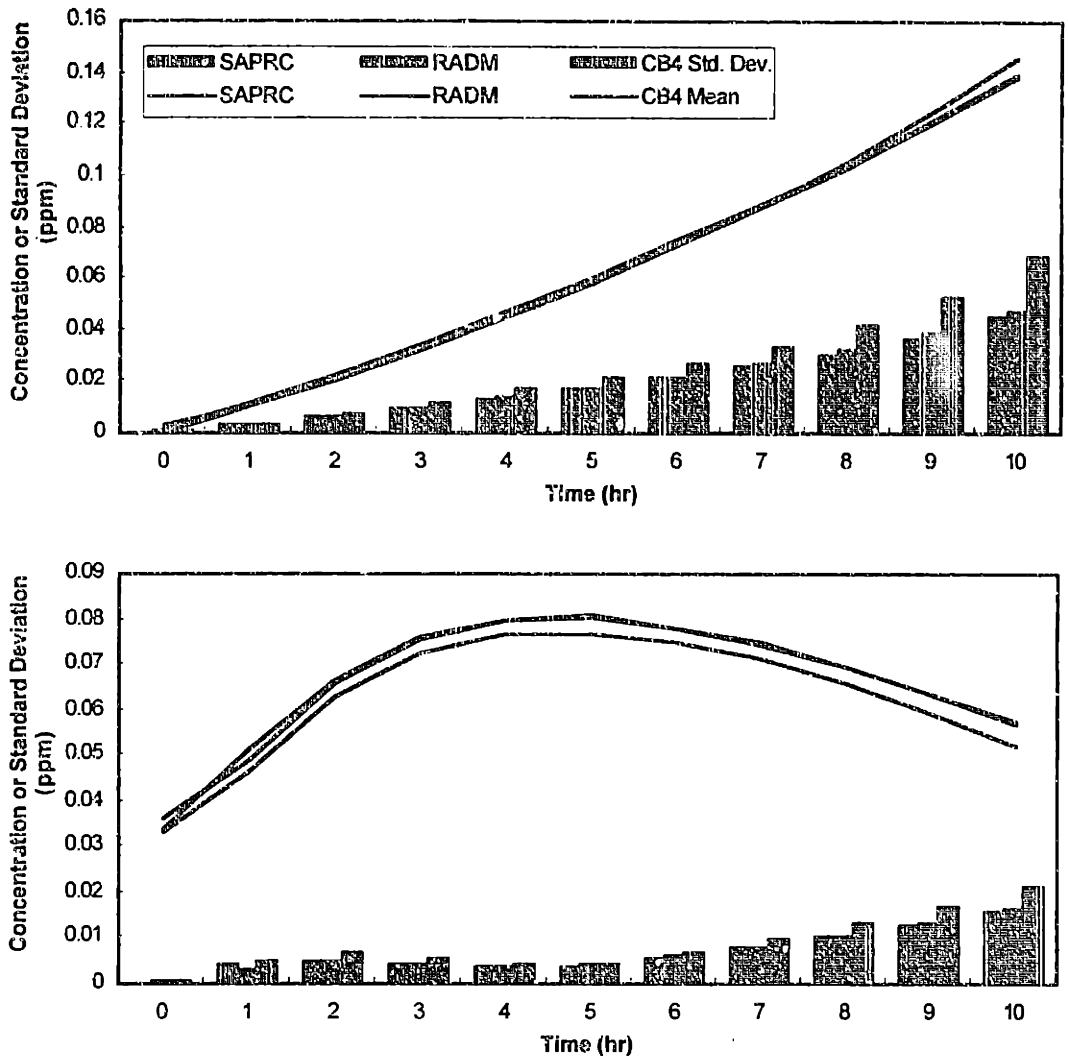


Figure 9-5. RADM, CB4, and SAPRC Dynamics in Response to Uncertain Photolysis Rates (a) ozone dynamics; (b) NO2 dynamics (Lines = expected concentration; Bars = standard deviation)

Figure 9-5 shows a comparison among the RADM, CB4, and SAPRC predictions for the case of uncertain photolysis rates. The maximum standard deviations were 33% to 40% for ozone and NO_2. Maximum standard deviations were predicted at the last time point of the simulation for both ozone and NO_2, and the uncertainties seemed to be growing at this time. For nitric oxide, the relative uncertainties was highest towards the end of the simulation when the remaining amount is small, even though the absolute uncertainty was decreasing at that point. It should be noted that the 67% confidence intervals were tighter for all species than the +/- standard deviation limits, indicating the non-Gaussian nature of the distributions. With fewer uncertain photolysis parameters, CB4 seemed to predict larger variances than RADM and SAPRC. Even though there were minor differences in the dynamics in the middle hours of the simulations, both the expected values and variances of the important compounds were remarkably similar for the RADM and SAPRC mechanisms.

The similarities in the uncertain predictions of the RADM and SAPRC mechanisms did not stop at the first and second moments of the uncertain output distribution. The estimated

probability density functions (PDFs) predicted at selected time points by both mechanisms are shown in Figure 9-6. Probability densities were found to be significantly skewed for ozone and NO_2. The distribution of ozone had a tail towards higher values and the NO_2 distribution had a corresponding tail towards lower values, due to the interconversion between the two compounds. The distributions for NO were quite symmetric. The PDFs for all the species investigated were essentially indistinguishable between the RADM and SAPRC predictions. This similarity in response of the two mechanisms to uncertainties in important photolysis parameters indicated that the structurally differences in SAPRC and RADM were masked, to a large extent, by parametric uncertainties.

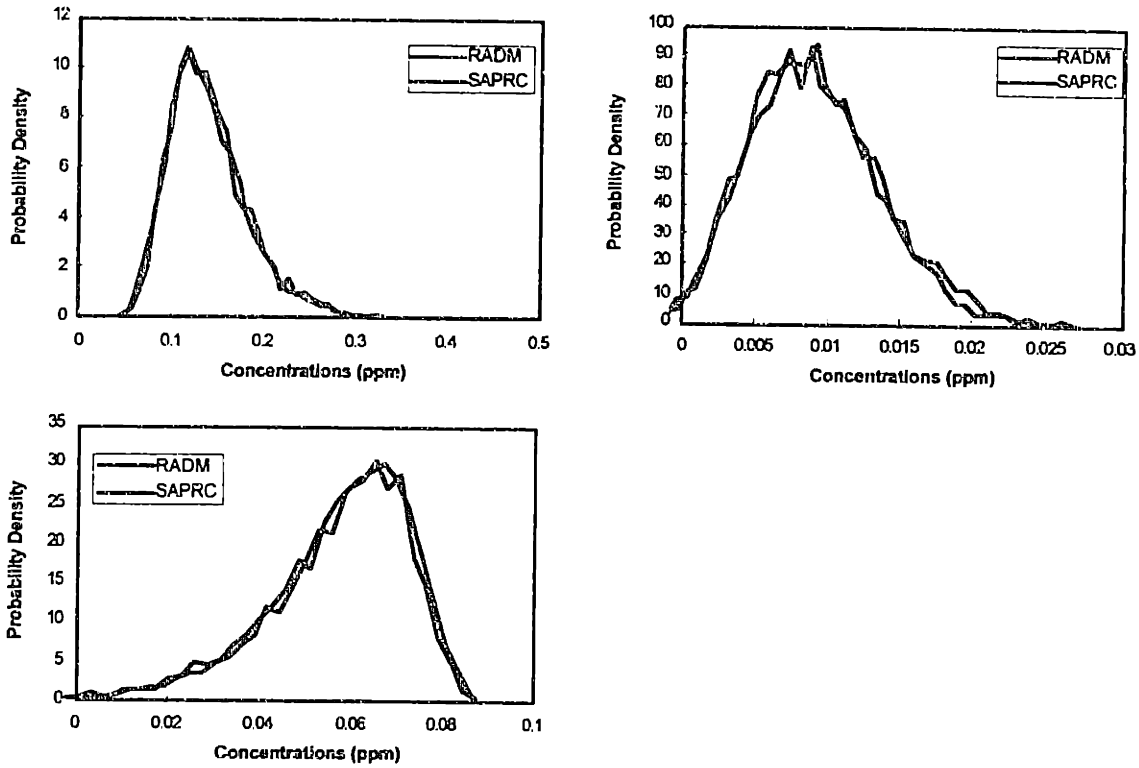


Figure 9-6. RADM and SAPRC Probability Density Functions at 10th Hour of Simulation – Response to Uncertain Photolysis Rates.
(a) Ozone, (b) NO, and (c) NO_2

An analysis of variance was conducted to identify the input parameters whose uncertainties contributed to the outcomes of-interest. The variance contributions to ozone from key uncertain photolysis rate constants are shown in Figure 9-7. Even though CB4 predicted a slightly higher variance than RADM and SAPRC, the variance breakdown looked similar for all three mechanisms. The photolysis rates of NO_2, HCHO, and O_3 (to $O(^1D)$) combined to account for over 90% of the variance to ozone at all times in the three mechanisms considered. During the first few hours of the simulation, the variance was dominated by the uncertainty in jNO_2, which then gave way to jHCHO (decomposing to radicals). The contributions of jO_3 and jHCHO (to molecular products) grew in importance towards the end of the simulation. The same parameters also account for over 90% of the variance in NO_2, with different time series behavior. The slightly higher variance contribution from jNO_2 and jHCHO in CB4 was probably a result of the correlation relations with other photolysis rates which were determined as a fraction of these photolysis rates. Since independent photolysis

rates measurements were available, the unrealistic correlation could be eliminated from the model. An alternative formulation of the model to use independent rates might reduce the contributions of uncertainties of jNO_2 and jHCHO.

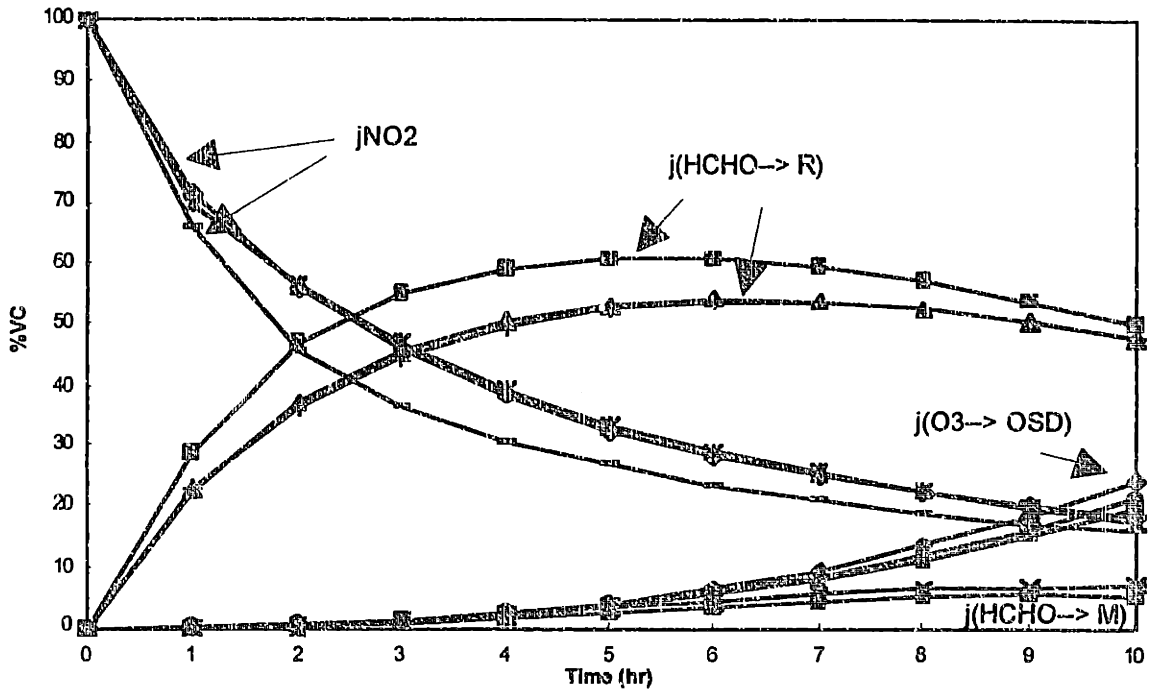


Figure 9-7. Photolysis Study. RADM, CB4, SAPRC VC Analysis of Ozone.

The photolysis uncertainty experiment showed that even though RADM and SAPRC were formulated differently, they behave similarly under the effects of these uncertain parameters. The effects of uncertain photolysis parameters were magnified in the CB4 mechanism due, most likely, to correlation. Although CB4 showed somewhat larger uncertainties, the parameters that drove the uncertainties were found to be the same as those in RADM and SAPRC. The photolysis rates of inorganic species and formaldehyde drove the uncertainties in the predictions. It is not surprising, therefore, that the mechanisms predicted similar mean, variance, etc., because a standard set of inorganic and formaldehyde reactions is used by all mechanisms. Reducing the uncertainties in these parameters will reduce the uncertainties of the predictions of both the mechanisms and three dimensional models based upon these mechanistic descriptions.

9.5 Organic Reaction Studies

The similarity between the uncertainty responses of the RADM, CB4, and SAPRC mechanisms to uncertain photolysis reaction rates raised the following question: Are theoretically dissimilar models truly different in the presence of parametric uncertainties? As discussed in Section 9.2, the major difference between the three mechanisms is in their treatment of organic reactions. Therefore, an uncertainty analysis of the primary organic reactions was performed as another case study to elucidate the key differences in the uncertain predictions of different mechanisms.

9.5.1 Description of Uncertain Organic Parameters

This experiment focused on aspects of the mechanism related to primary organic reactions: organic lumping, reaction rates of emitted species, and the formation of radicals and second generation products. In the SAPRC mechanism, the primary reactions were lumped so that the rates and products were determined from the reactions of individual species. All the uncertain rate and product coefficients were represented in the small number parameters in the AAR (emitted alkanes and aromatics) and OLE (emitted alkenes) reactions. The lumped initial conditions also depended on the uncertain parameters of the species. The uncertainty analysis of SAPRC's primary organic reaction and parameters can be found Section 8.4. Corresponding organic initial conditions, rates constants, and product coefficients were investigated in the RADM and CB4 mechanisms so as to compare the effects of primary organic reactions. (Refer to Figure 9-2.)

In the RADM mechanism, the parameters of-interest were scattered in a larger number of reactions. According to the original "reactivity lumping" approach, the initial conditions would depend on both the uncertain organic composition and uncertain reaction rate constants for individual and modeled species (see Equation 9-1). However, since the individual organics were first lumped into emission categories using the approach of Middleton *et al.* (1990), the assumed reaction rates of the lumped category were not considered to be uncertain. Therefore, in this case study, uncertainties in the organic initial conditions in RADM were only functions of the uncertain amounts of individual organic compounds but not the rate constants. Reactions that were assigned uncertain parameters are listed in Table 9-4. The left hand column in Table 9-4 contains a list of all primary reactions in the RADM mechanism. The right hand column is a list of all peroxyl radical reactions leading to secondary carbonyl products.

Reactions with Uncertain Rate and Product Coefficients	Reactions with Uncertain Coefficients Only
ETH + HO	ETHP + NO
HC3 + HO	HC3P + NO
HC5 + HO	HC5P + NO
HC8 + HO	HC8P + NO
OL2 + HO	OL2P + NO
OL2 + NO_3	
OL2 + O_3	
OLT + HO	OLTP + NO
OLT + NO_3	
OLT + O_3	
OLI + HO	OLIP + NO
OLI + NO_3	
OLI + O_3	
TOL + HO	TOLP + NO
XYL + HO	XYLP + NO

Table 9-4. Uncertain Parameters Investigated in the RADM – Organics Experiment.

The uncertainty description and assignments of the 64 uncertain variables used in the RADM case study can be found in the Demmucom input file provided in the appendix. ETH and OL2 were the only organic compounds modeled explicitly. The reaction rates of HC3, HC5, and HC8 were determined based on averaging the rate constants of species present in the National Acid Precipitation Assessment Program (NAPAP) inventory. The uncertainty factor of the reaction rate constant of a typical HC reaction was assigned to be equal to that of an individual compound, even though the uncertainty of an "averaged" reaction rate constant was expected to be lower on a percentage basis because of the central limit theorem. Since the rates of OLT reactions were assumed to be equal to those of propene, uncertainty factors for OLT were assumed to be the same as those for propene, although OLT represented a range of alkenes. OLI reactions were assumed to be the same as those of trans-2-butene; therefore, the reaction rate constant for OLI was assigned the uncertainty factor of trans-2-butene. Similarly, TOL reaction rates were assumed to have the same UF as toluene, the compound. The modeled species XYL was assumed to be more reactive than xylene, but the uncertainty factors of xylene reactions were assigned to XYL reactions. The product yields for alkanes were derived from "condensing" a detailed reaction mechanism. Alkene reaction products were based on estimates of US averaged emissions. Uncertainties in the product coefficients were represented using uniform distributions and the ranges were assigned using multiplicative factors in an analogous manner as those of an individual species, using the procedure described previously in Section 7.3.

The initial conditions used in CB4 were determined from the organic mix. The assignment of individual species to modeled bond types was not treated as uncertain, since it was based on the structure of the compound. Therefore, uncertain initial amounts of the compounds were the only factors contributing to uncertain initial conditions for PAR, OLE, etc. Despite having the fewest reactions, the CB4 mechanism contained the largest number of parameters describing the chain of events leading from breakdown of primary organics to the formation of stable secondary compounds. The large number of parameters was a result of the condensation techniques used to determine the reactions of the bond type. The condensed representation tended to be highly parameterized and therefore contained many uncertain product coefficients. The reactions involved in this study are listed in Table 9-5. The description of uncertain parameters can be found in the Demmucom input file included in the appendix.

95 uncertain parameters were identified for the CB4 organic case study. 14 were reaction rates, 8 were initial concentrations, and the remaining 73 were product coefficients of the reactions considered. The uncertain initial conditions were determined by sampling (Monte Carlo) the individual organic species and fitting the resulting distributions for the model species (such as PAR, OLE, FORM) using lognormal distributions. Reaction rates of each bond type were determined based on an assumed composition of the organic mixture. Since the reaction rates represented an averaged value of reactivity of the particular kind of bond, the parametric uncertainty would be smaller than that of the individual compounds' reaction rates due to the central limit theorem. However, the composition of the organic mixture used for determining these coefficients was fixed and was likely to be outdated. Instead of assigning uncertainty based strictly on the definition of the parameters, uncertainty in the rates and product coefficients of model species were assigned analogously to typical compounds in the group to investigate a "worse case scenario" in which these parameters were highly uncertain. For example, since most alkane + HO reaction rates were assigned uncertainty factors between 1.2 and 1.4, an uncertainty factor of 1.3 was assigned to PAR + HO. Product coefficients were assumed to be uniformly distributed. The range of the distributions were determined using multiplicative factors described in Section 7.3. The product coefficients of each model species were assumed to have similar uncertainty characteristics as a typical compound contributing to that bond type.

Reactions with Uncertain Rate and Product Coefficients	Reactions with Uncertain Coefficients Only
PAR + HO	ROR → products
OLE + O	
OLE + HO	
OLE + O3	
OLE + NO3	
ETH + O	
ETH + HO	
ETH + O3	
TOL + HO	TO2 + NO
XYL + HO	
(ISO + O)	
(ISO + OH)	
(ISO + O3)	
(ISO + NO3)	

Table 9-5. Uncertain Parameters Investigated in the CB4 – Organics Experiment.

9.5.2 Results of Organics Case Study

Ozone uncertainties resulting from uncertainties in input organic composition and reaction parameters are shown in Figure 9-8, which compares the uncertainty analysis of SAPRC, RADM, and CB4. The expected values of ozone concentrations predicted by CB4 were elevated over the nominal value by 40%. RADM's expected ozone predictions were about 20% higher than that predicted by SAPRC, which were quite close to nominal values. Furthermore, CB4 also predicted the highest variance, almost 30%, about the mean value of ozone concentrations. The RADM-predicted variance was about 20%. At any given time, the uncertainty in ozone predicted by the SAPRC mechanism was smaller than that predicted by the other two mechanisms.

As postulated, the structural differences in the three mechanisms led to significant differences in the mechanisms' responses to uncertain organic parameters. The difference in lumping strategies in the mechanisms could provide an explanation of the behavior. SAPRC's lumping diminished the effects of uncertainties of the input parameters because of the use of averaged quantities that had low relative uncertainties based on the central limit theorem. The parameters used in the primary organic reactions of CB4 and RADM were inflexible and were determined based on a fixed, nationally averaged, organic composition which was also likely to be out-dated. The uncertainty factors of the reaction rates and product coefficients of these reactions were assigned to values of a representative individual species due to the nature of these parameters. Both RADM and CB4 mechanisms predicted higher variances in ozone concentration than the SAPRC prediction. Of the two, the CB4 mechanism exhibited higher output uncertainties than RADM. This could be due to the larger number of uncertain parameters used to describe primary organic reactions in CB4 mechanism, or the sensitivity of ozone predictions to organic reaction rates and product parameters that is intrinsic in the formulation of the CB4 mechanism.

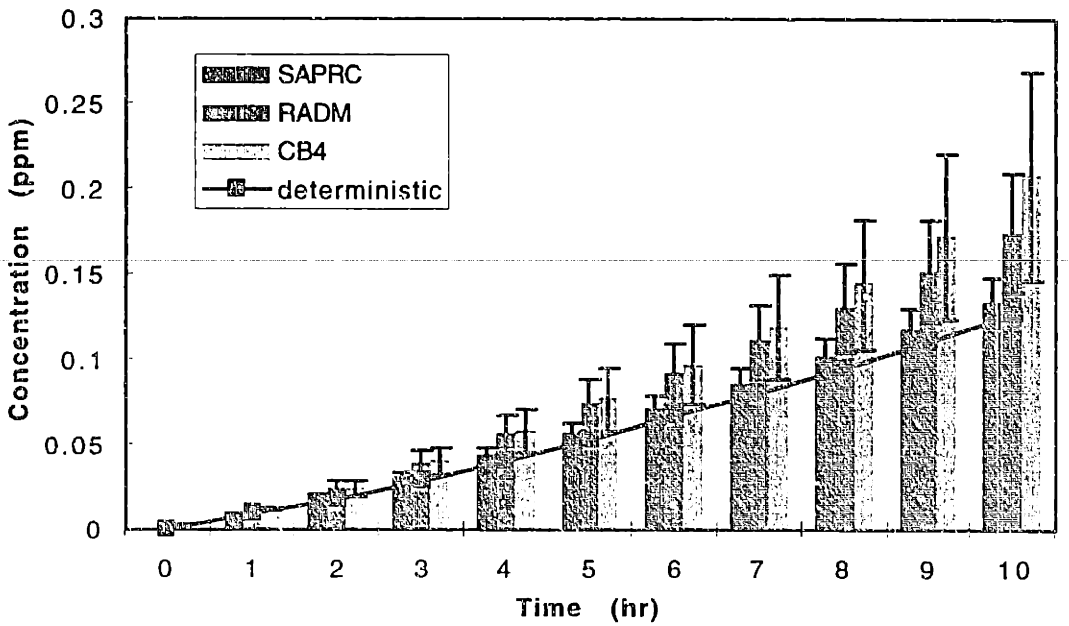


Figure 9-8. Effects of Uncertainty in Primary Organic Parameters on Ozone Predictions – SAPRC, RADM, and CB4.
(Columns = Mean Value, Errorbars = Standard Deviation)

The organic parameters that contributed to uncertainties in the predictions of ozone in the SAPRC mechanism were found to be initial conditions of aromatic compounds and the product coefficients of the unknown ring-opening product of substituted benzene compounds. These parameters also contributed to uncertainties in the ozone predictions of the RADM mechanism. RADM's prediction of ozone variance, shown in the next figure, were largely explained by uncertainties in parameters related to the reactions of poly-substituted aromatics, represented by XYL, especially in the early part of the simulation. The aromatic parameters of importance included the reaction rate of XYL + HO, the initial condition of XYL, and the product coefficients of aldehydes, formaldehyde, peroxyl radical, and (especially) the ring opening compounds from the XYL + HO reaction. The product coefficients of aldehydes from the OLTP + NO reaction contributed to the variance of ozone at later times, when the reactions of secondary products became an important source of ozone. The same parameters also appeared to account for the variances in the predictions of NO_x and HO.

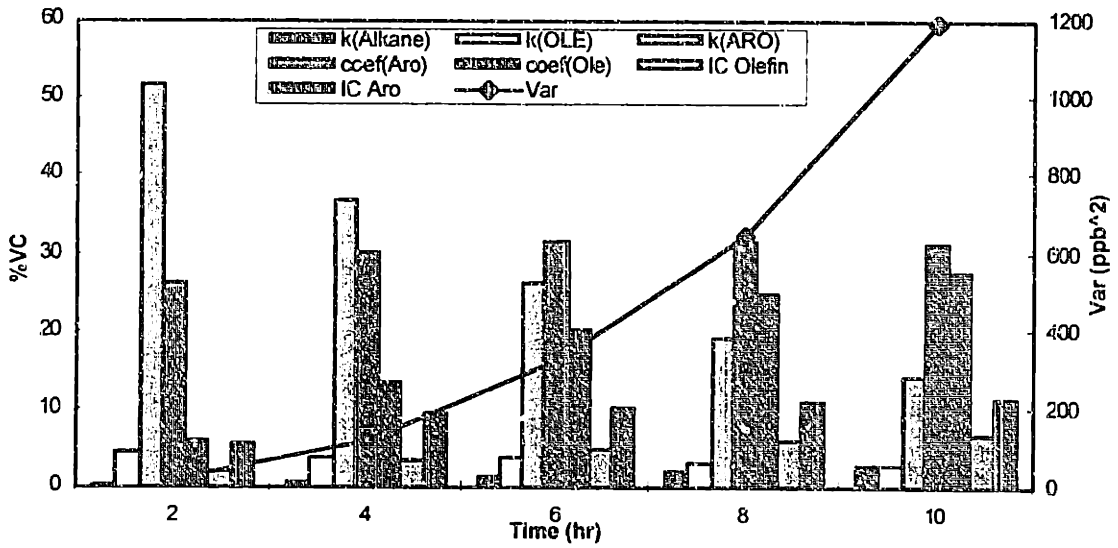


Figure 9-9. VC to Ozone Predictions (RADM Organic Case Study)

Figure 9-10 shows the variance contribution to ozone, as predicted by the CB4 mechanism. The reaction rates of HO + XYL and HO + PAR contributed to the ozone uncertainties. In contrast to the RADM and SAPRC mechanisms' predictions, product parameters combined to account for a large fraction (over 50%) of the uncertainties in ozone throughout the CB4 simulation. The product coefficients that had significant effects on the uncertainties of ozone predictions included parameters from the reactions PAR + HO (ROR yield), XYL + HO (MGLY yield), OLE + HO (beginning of simulation only), and ROR decomposition (HO_2 yield). On the other hand, the uncertain initial concentrations did not contribute significantly to uncertain predictions, probably because the uncertainties in the initial conditions were small to start with. The importance of the PAR reactions in explaining the uncertainties of the predictions was unique to the CB4 mechanism. This sensitivity indicated an aspect of the CB4 mechanism that was structurally very different from the other mechanisms investigated in the study. The dominance of the PAR parameters may simply be a result of the great abundance of PAR groups modeled than any other bond types.

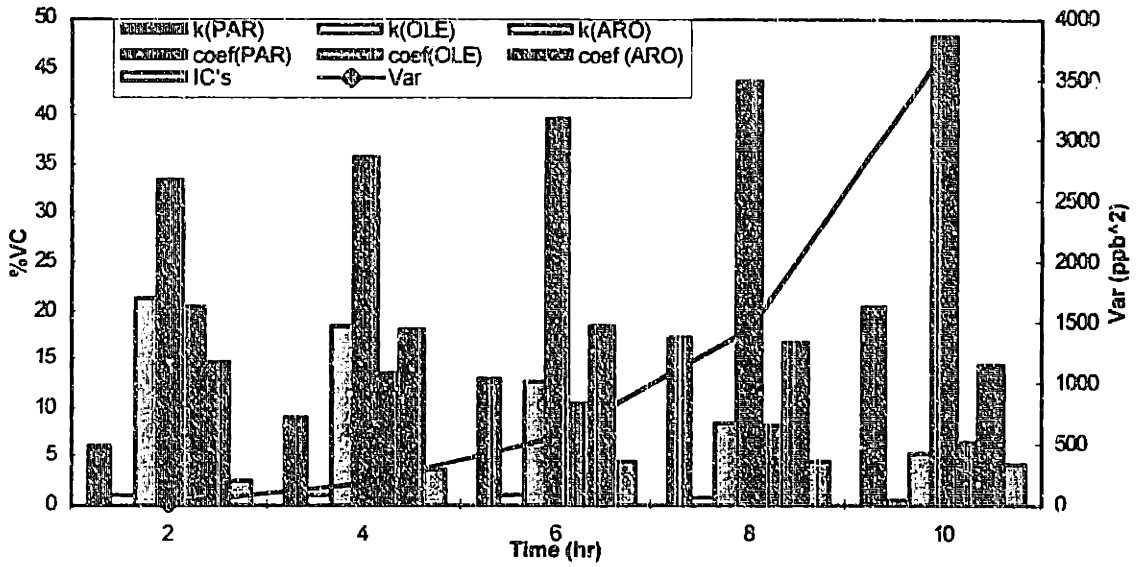


Figure 9-10. VC to Ozone Predictions (CB4 Organic Case Study)

The structural differences between the mechanisms considered in this study were highlighted in this uncertainty analysis case study of primary organic compounds. Both SAPRC and RADM mechanisms' predictions of ozone were sensitive to uncertainties in aromatic parameters. To a smaller extent, so were the predictions of CB4. However, the representational uncertainties in the products of the C-H bond (PAR) reactions in the CB4 mechanism contributed significantly to the uncertainties of its predictions. The lumped structure treatment of organics in the CB4 seemed to result in greater uncertainties in the output caused by uncertain organic parameters. This difference in the structure is a distinguishing feature of the CB4 mechanism.

9.6 Comprehensive Uncertainty Analyses

The parameters identified to affect the output uncertainties in the previous case studies were combined with other VCIs identified for the SAPRC mechanism in an uncertainty analysis. The lists of uncertainty parameters can be found in the appendix for both the RADM and CB4 mechanisms and included organic and inorganic reaction rates and initial conditions (20% uncertain) in addition to primary organic parameters identified as VCIs in Section 3 of this chapter. The 25 uncertain parameters identified led to a 25% elevation of the ozone predictions from the corresponding nominal values in the RADM simulation. In the CB4 study, 41 parameters were investigated, due to the large number of primary reaction parameters found to be important in explaining uncertain outcomes. In CB4, the elevation of the mean ozone concentration above the nominal value was even more significant. The predicted mean values are shown in Figure 9-11.

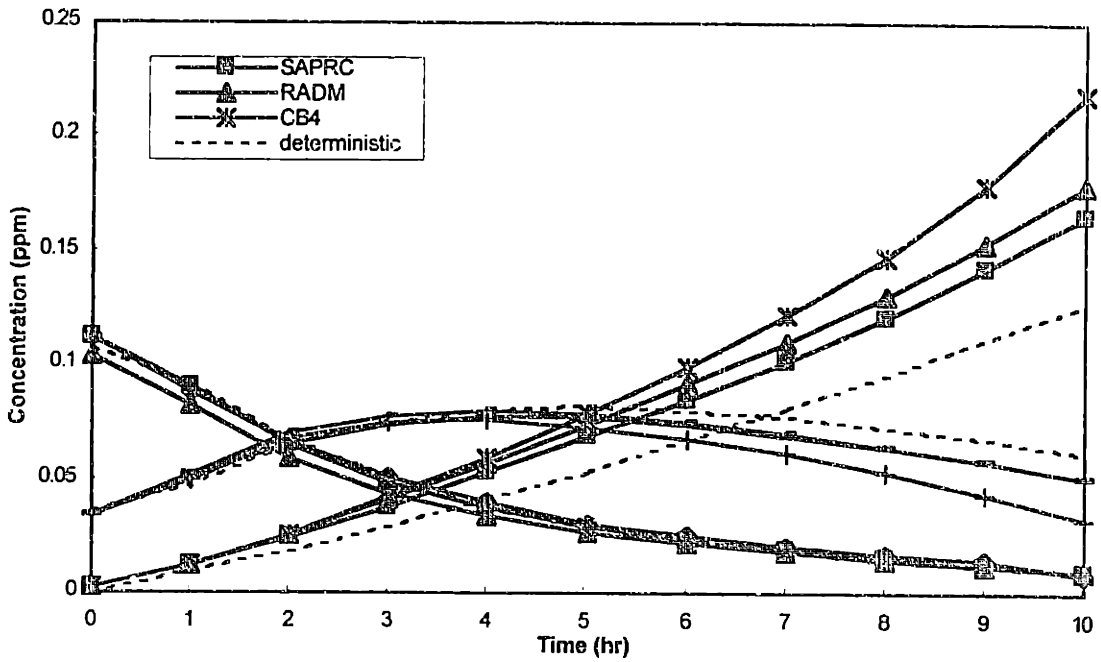


Figure 9-11. RADM, CB4, SAPRC Ozone Predictions, All City Case, All Parameters.

The higher mean values were also associated with higher absolute amount of uncertainties. The standard deviations of the uncertainties associated with the predictions of both RADM and CB4 were as much as 55%. The variances in HO and products such as HNO_3 and PAN were even higher on a relative basis. At a third order representation, the maximum relative error in the ozone response surface were 2% (RADM) to 3% (CB4).

The level of uncertainty has significant implications to the issue of model discrimination. Here, the distributions of uncertain predictions overlapped because relatively large ranges of uncertainties were associated the expected predictions of mechanisms which differed only slightly. Any measurements taken for ozone and HCHO could be reproduced by any of the three mechanisms using parameters within their ranges of uncertainty. Therefore, they could not be distinguished based on measurements. Figure 9-12 illustrates the small differences in

the concentration predictions from the three models compared to the large uncertainty bounds of these predictions.

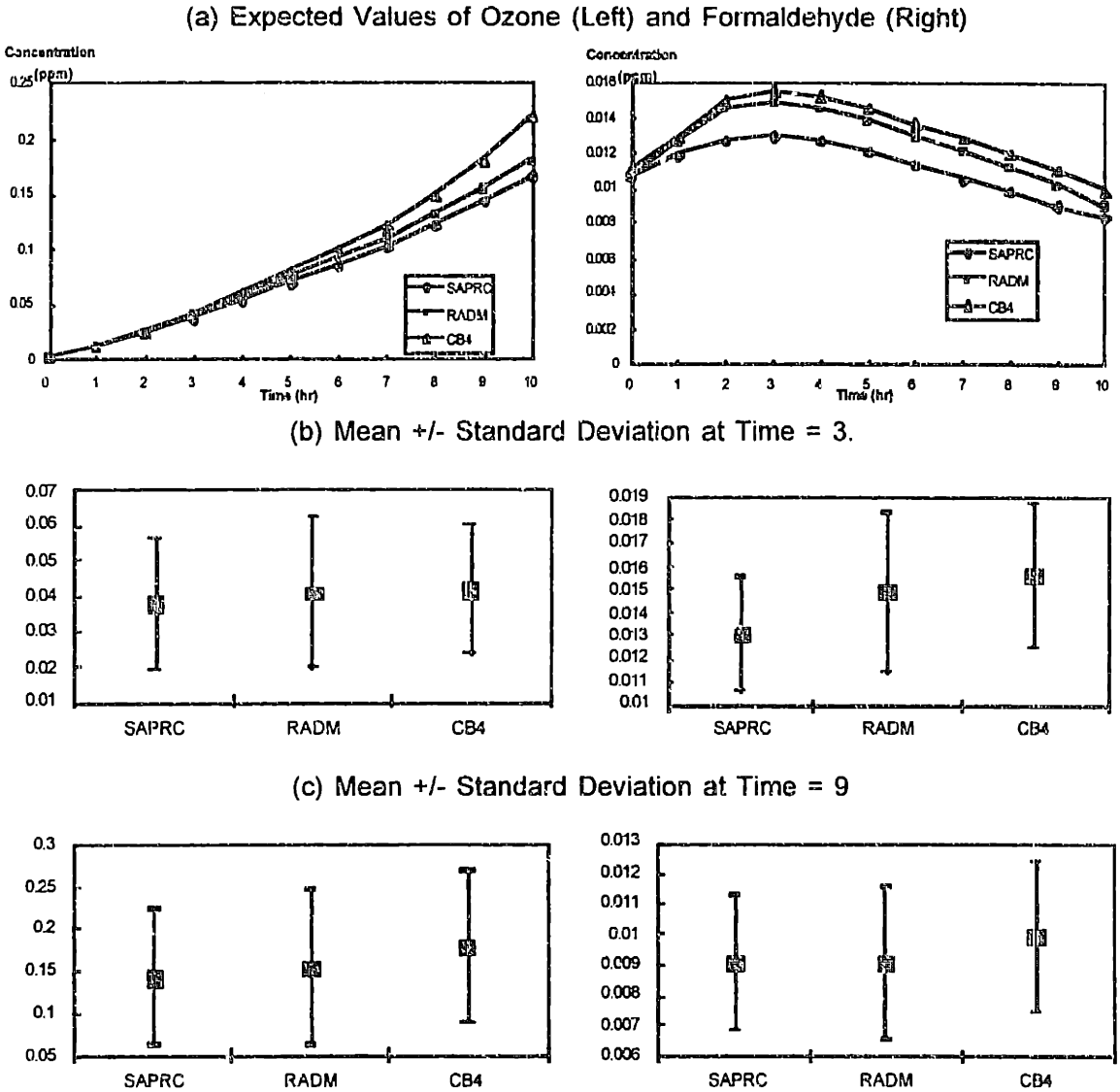


Figure 9-12. RADM, CB4, SAPRC Uncertain Predictions of Ozone and HCHO (Left Panels – Ozone; Right Panels _ HCHO)

Figures 9-13 and 9-14 show the variance contributions of various parameters to ozone as determined for the RADM and CB4 studies. As in the SAPRC study, the reaction rate of HO + NO_2 and the initial condition of NO_x were the main contributors to ozone uncertainty in the RADM predictions. In the very beginning of the simulation, the photolysis rate of NO_2 accounted for a majority of ozone variance. The photolysis rates of O_3 and formaldehyde were the other important parameters. Figure 9-13 shows that the predictions of RADM were sensitive to the same set of parameters as those of SAPRC (See Figure 8-8).

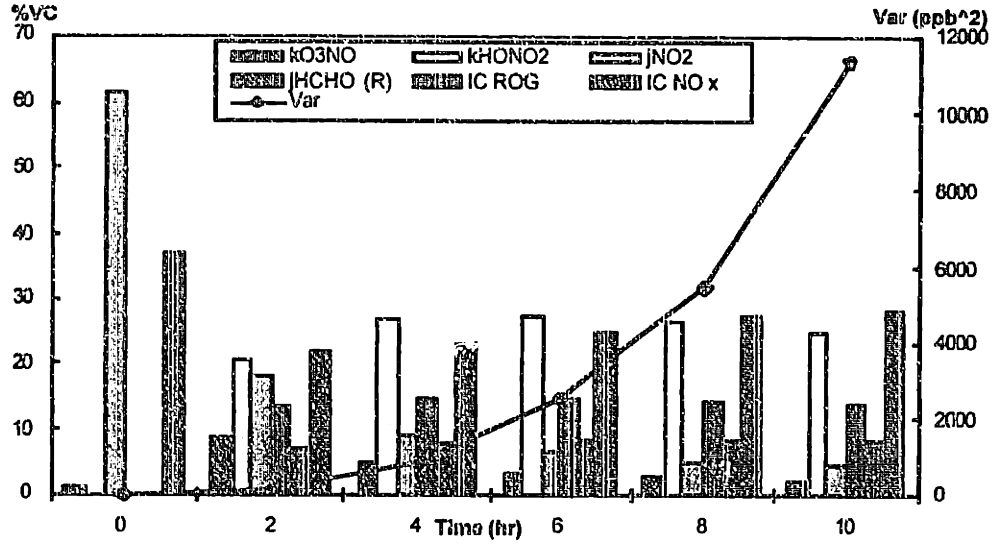


Figure 9-13. RADM Ozone VC, All City Case, All Parameters

As shown in Figure 9-14, the CB4 mechanism also predicted the same set of variance contributing inputs (VCIs) to the uncertainties in the concentrations of ozone. However, the uncertainty in the product coefficient of the alkoxyl radical species ROR also contributed to the uncertainties in the prediction of ozone. These results confirmed that the structural difference between CB4 and the lumped molecule mechanisms, SAPRC and RADM, were detectable in the VCIs. In SAPRC and RADM, the output responses were insensitive to uncertainties in organic parameters relative to the VCIs listed above (and in Figure 9-13). Since the product yield of ROR was not a measured quantity, but a condensed representation of several pathways, it was concluded that the parameterized representation in the CB4 mechanism introduced additional uncertainties in the predictions of CB4. This kind of parametric uncertainties was unique to the lumped structure mechanism and might be difficult to resolve by new measurements.

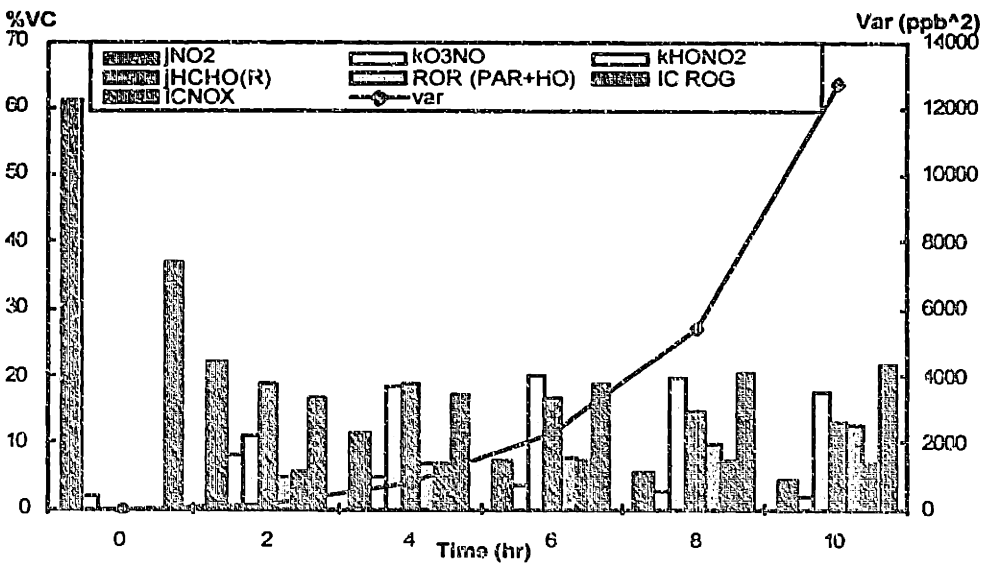


Figure 9-14. CB4 Ozone VC, All City Case, All Parameters

The predictions of the RADM and CB4 mechanisms seemed to respond strongly to many of the same uncertain parameters found to cause uncertainties in the predictions in the SAPRC mechanism (see Figures 8-8 and 9-15). Therefore, reducing the parametric uncertainties of initial conditions, photolysis rates, and the rate of HO + NO_2 was expected to improve the confidence of the predictions of all three mechanisms.

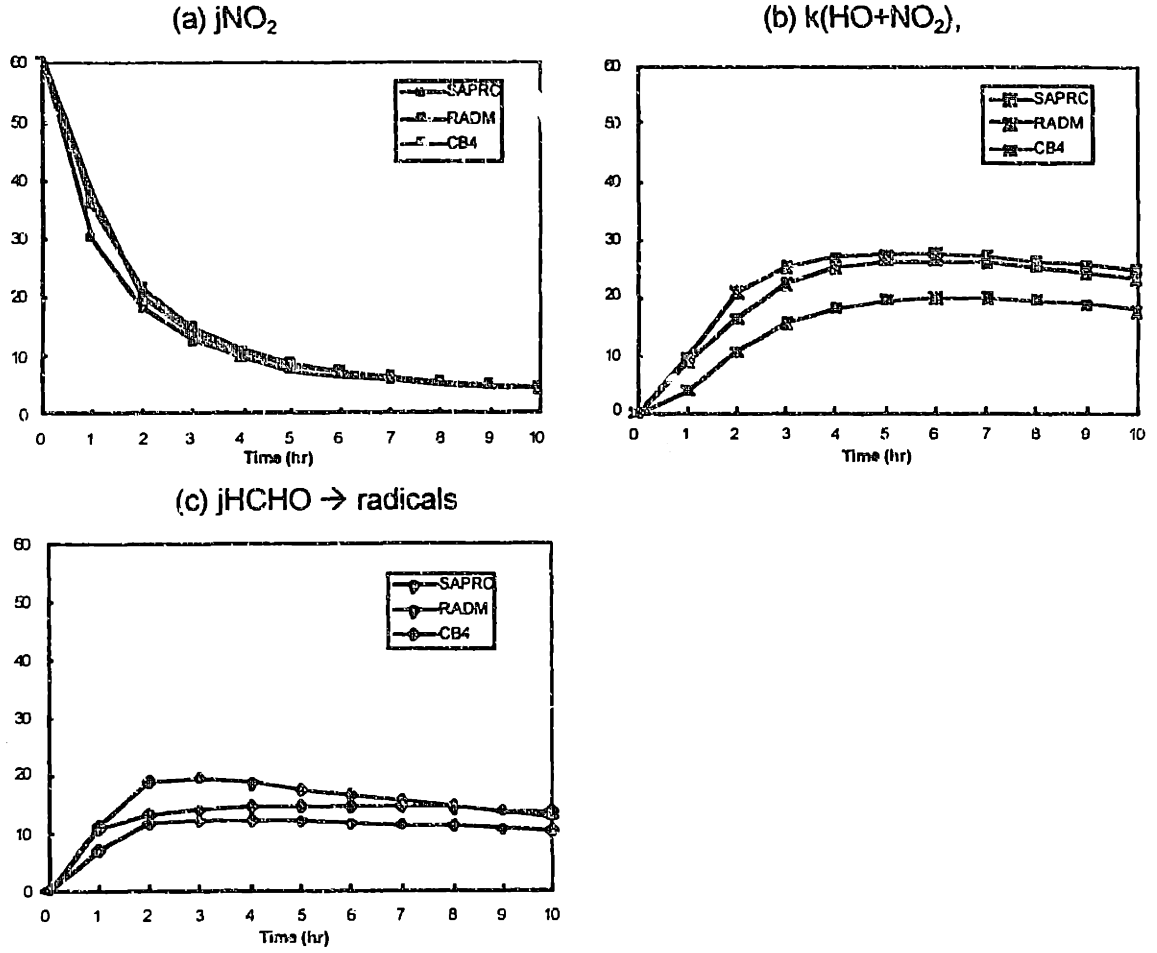


Figure 9-15. VC to Ozone, Comparison between Predictions by SAPRC, RADM, and CB4.

9.7 Conclusions

- The three mechanisms behaved similarly in many respects under the influence of uncertain inputs (e.g., responses to photolysis rates).
- Of the organic parameters, SAPRC and RADM were most sensitive to aromatic parameters, while the product coefficients of PAR (C-H bond) reactions also contributed to uncertainties in the predictions of CB4.
- CB4 appeared to respond more strongly to parametric uncertainties than SAPRC or RADM.
- CB4 predictions were affected by parametric uncertainties due to the condensed representation of the reactions of bond types. This type of parametric uncertainties did not exist in lumped molecule mechanisms such as SAPRC and RADM, and should be more carefully characterized and evaluated in the future.
- The three mechanisms share many of the same variance-contributing parameters, such as the reaction rate constant of HO + NO_2, the photolysis rates of NO_2 and formaldehyde, and initial amounts of NO_x.
- Given the high level of uncertainties associated with the mechanisms, CB4, RADM, and SAPRC would probably not be distinguishable using measurements commonly available such as ozone and formaldehyde.

Even though the uncertainties predicted by the CB4 mechanism seemed higher than those of the RADM and SAPRC mechanisms for key pollutants like ozone, the uncertainty analyses did not support any discussion of the relative merits of the mechanisms. Parametric uncertainty analyses revealed structural differences of the lumped structure vs. the lumped molecule approach in the behavior of uncertain responses. Structural uncertainties were not quantified in the study. Given the high level of parametric uncertainties, it was impossible to "distinguish" between the three mechanisms. Any set of measurements can be easily reproduced by each mechanism using parameters well within their uncertainty bounds. In fact, all these mechanisms were shown to simulate smog chamber results well. The uncertainty analysis performed here provided evidence that their predictions are indistinguishable from one another. Research focused on reducing input uncertainties would be needed before a meaningful mechanistic discrimination exercise can be undertaken. At present, research resources would be better invested on reducing uncertainties of key parameters through carefully designed experiments and field studies, than on the development of a new generation of chemical mechanisms / models based on the available body of data.

10. Airshed Uncertainty Analysis

10.1 Introduction

"...the (air quality) models contain important uncertainties about chemical mechanisms, ... (uncertainties) must be considered when using such models to project the effects of future emission controls." (NRC, 1991)

The predictions from large, complex atmospheric models are uncertain for a variety of reasons. A simplified view is that either the model is inadequate (wrong), or the data needed to run the model are inaccurate. One of the goals of a parametric uncertainty study is to provide error bounds to the predictions based on the current formulation of the model. A meaningful comparison between model predictions and observations requires the knowledge of the model prediction uncertainties as well as the measurement errors. As long as the observations are within the range covered by the uncertainty bounds, the model can reproduce the observed values with parameters whose values lie within their uncertainty estimates, and the model cannot be rejected as wrong even though the agreement between model and observations may be poor. The key is then to understand which parameters cause the uncertainties in the predictions and target these parameters in experiments and field measurements for the purpose of improving model predictions. This information can be gained through an analysis of variance, which is an integral part of an uncertainty analysis. However, if the differences between model predictions and observations are larger than the uncertainties of the models and the data, the conclusion should be drawn that either the measurements are inaccurate or the model equations are inadequate to describe the physical processes.

Air quality models solve the atmospheric diffusion equation (ADE), which is the turbulent mass transport equation for trace pollutants in the atmosphere:

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (u c_i) = \nabla \cdot (\mathbb{K} \nabla c_i) + R_i[c_1, \ldots, c_n; T, t]$$

advection diffusion reaction

$$\text{Boundary Condition : } \underset{\text{deposition flux}}{v_g^i c_i} - \mathbb{K} \nabla c_i = \underset{\text{emission}}{E_i[\mathrm{x}, t]} \qquad (10.1)$$

In Equations 10.1, c_i is the concentration of trace species i, u is the average horizontal wind speed, and v_g^i is the deposition velocity of species i. The bold $\mathbb{K}$ represents the turbulent diffusion coefficient, a vector quantity. In the ADE, all chemical reactions are aggregated into one single volume source term (R_i) for each chemical species.

Even though atmospheric chemistry is only one source of uncertainty in a three-dimensional air quality model, there are many parameters in the kinetic subroutine. An uncertainty

analysis that considers the effect of the uncertainties in the chemical mechanism is interesting because of the importance of the chemistry sub-modules. Three-dimensional air quality models typically spend 70-80% of the execution time solving the differential equations governing the chemical transformation of species.

10.2 Goals of Study

Understand the limit of predicitability of current air quality models by characterizing the uncertainties in the model outputs due to input parametric uncertainties in the chemical process representation and source amounts of nitrogen oxides and organic compounds.

Characterize the uncertainty contribution from various parameters to model predictions as a function of location and time.

Recommend field measurements and experiments to obtain better (less uncertain) air quality model predictions.

Determine if parametric uncertainties in chemical mechanisms can account for the observed discrepancies between measurement data and model predictions.

10.3 Case Study

The air quality model used in this study is the CIT airshed model (McRae, 1981), which has been validated against field data obtained during the 1987 Southern California Air Quality Study (SCAQS) (Blumenthal *et al.*, 1989; Seinfeld, 1989; Harley et al., 1993).

The modified LCC mechanism (Lurmann et al., 1987) used in the CIT airshed model can be considered a simplified version or a precursor of the SAPRC mechanism (Carter, 1990) described in Chapter 7. The LCC mechanism contains 35 modeled species including nine lumped organic classes. These nine species represent more than 100 emitted organic compounds using a lumped molecule approach. There are 95 reactions in the mechanism with 34 inorganic reactions. The photolysis reactions of NO_2, NO_3 (2 channels), O_3 (2 channels), HONO, HCHO (2 channels), ALD, MEK, MGLY, DIAL are included. The lumping of organics is typically performed on a per-mole basis. A few product coefficients are determined according to key characteristics of the lumped class, such as the fraction of terminal alkenes (ALKE + HO reaction) or the fraction of tri-substituted aromatics (AROM + HO reaction).

An uncertainty analysis was applied to the Los Angeles basin. This example used a scenario of August 27-29, 1987 during Southern California Air Quality Study (SCAQS). During this study period, a variety of special meteorological and air quality measurements were carried out to supplement the routine measurements made in the Los Angeles area. The CIT Airshed model was used to simulate the concentrations of pollutants observed at several monitoring stations including upwind, urban, and downwind locations with different pollution characteristics. As shown in Figure 10-1, simulations results did not match the observed measurements very well at several monitoring stations.

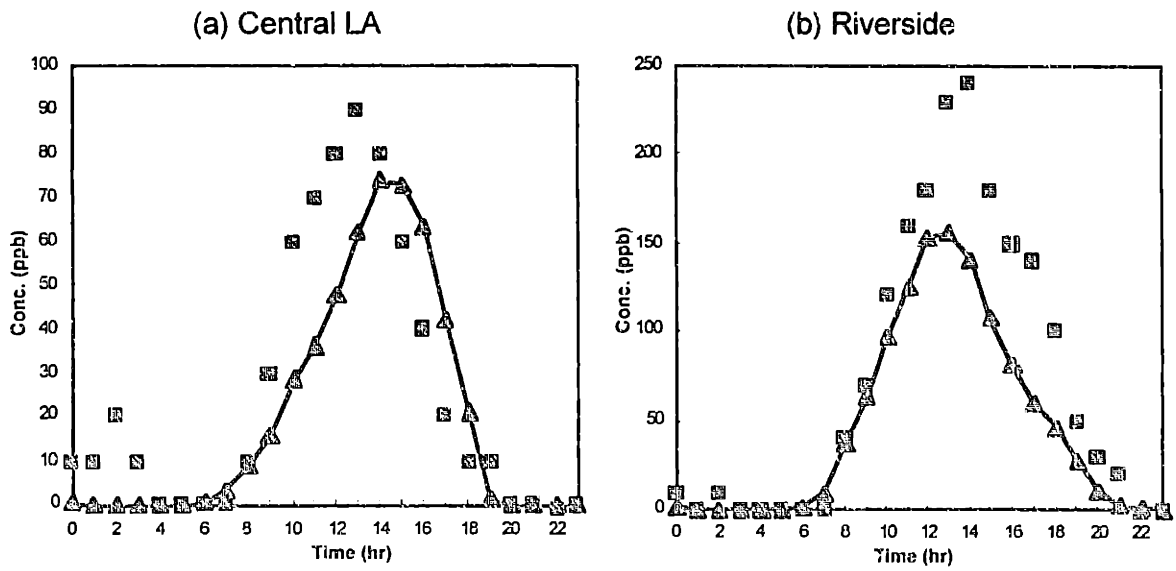


Figure 10-1. Comparison between CIT Airshed Predictions and Observation of Ozone at Two Monitoring Stations on August 27, 1987.

10.4 Uncertain Parameters

The computational requirement for an uncertainty analysis increases with the dimensionality of the problem and the accuracy requirement, as well as with the execution time of the model itself. The execution time for CIT Airshed model is more than two hours for a 24-hour simulation on a desk-side work station. Even with an efficient uncertainty analysis method such as DEMM, the answer quickly becomes unaffordable with increasing number of parameters considered. It is therefore imperative to keep the number of uncertain parameters small in the analysis. A full analysis of the 95 reactions in the LCC mechanism would be prohibitively expensive at this time. The chemical mechanism (one-dimensional) uncertainty analyses described in the previous chapters provided the ground work for analyzing the effects of uncertain chemistry parameters in a three dimensional air quality model. The variance-contributing inputs (VCIs) short-listed from the mechanism studies provided a starting point in the uncertainty analysis of the three dimensional air quality model.

The uncertain reaction parameters identified to contribute to uncertainties in the predictions of atmospheric chemical mechanisms are listed in Table 10-1. No product coefficients were found to contribute significantly to the predictions of the lumped molecule chemical mechanisms studied. In addition to the VCIs found in the comprehensive uncertainty analysis of Section 8.6, rate constants of the most important primary and secondary organic reactions, e.g., k(XYL + HO) and k(ACO3 + NO2), were also included in the list. Two uncertain factors applied to the sources and initial conditions of reactive organic gases (ROG) and NO_x were also included as uncertain inputs for the CIT Airshed model.

Rate Parameters	Uncertainty Factor used in Lognormal Distribution
$j(NO_2)$	1.3
$k(O_3 + NO)$	1.2
$j(O_3 \rightarrow OSD)$	1.4
$k(OSD + H_2O)$	1.26
k(OSD + M)	1.26
$k(HO+NO_2)$	1.28
j(HCHO –> R)	1.4
j(HCHO –> M)	1.4
k(ACO3 + NO)	2.0
$k(ACO3 + NO_2)$	2.0
k(PAN)	2.0
k(ALK+HO)	1.3
k(ETH+HO)	1.15
$k(ETH + O_3)$	1.25
k(OLE + HO)	1.2
$k(OLE + O_3)$	1.5
k(XYL + HO)	1.3
Factor (NO_x)	1.2
Factor (ROG)	1.2

Table 10-1. Uncertain Parameters Investigated in Airshed Chemistry Uncertainty Study.

Most rate constant uncertainty factors were assigned according to Stockwell's compilation for the SAPRC mechanism (Stockwell *et al.*, 1994). The 298K uncertainty factor was applied as a multiplicative factor for each rate constant. In other words, the uncertainty factors were assumed to be independent of temperature. The uncertainty factors for NO_x and ROG were assigned to all sources and initial condition terms. In the one dimensional simulations, uncertainty factors were applied to the initial conditions to simulate the effects of uncertainties in the amount of reactive organic gases (ROG) and nitrogen oxides (NO_x) present in the system. The effects of uncertainties in the initial conditions were found to be comparable to those of the rate parameters (see Chapters 8, 9). Therefore, for comparison purposes, uncertain factors of 1.2 were assigned to all initial conditions and source terms, even though the uncertainties in NO_x and ROG emissions are likely to be much more significant. For example, the emissions of ROG in the LA basin was postulated to be underestimated by as much as a factor of 2 (NRC, 1991).

10.5 Uncertainty Analysis of the CIT Airshed Model

Due to the high computational requirement of the CIT Airshed model, the iterative approach (Figure 8-1) used for the box models was not suitable for investigating the effects of uncertainties in the CIT Airshed model. A short list of parameters was chosen a priori from the mechanism studies, as shown in Table 10-1. For the 19 parameters listed in Table 10-1, an uncertainty analysis using DEMM requires the model solution for at least 20 points to obtain representations of the output parameters of interest. Table 10-2 lists the number of runs of the Airshed model required to solve the polynomial chaos expansions of various order.

Approximation	No. of Model Runs for PCE Coefficients	No. of Runs for Error Calculation	Total No. of Model Runs	Execution Time in hours (days)
Linear	20	39	59	148 (6)
Second Order	39	58	97	243 (10)
Second Order + Bilinear Terms	210	229	439	1098 (46)
Third Order (Complete)	1540	1559	3099	7748 (323)

Table 10-2. Computational Requirement for Uncertainty Analysis by DEMM (19 Parameters, execution time ~2.5 hours for each 24-hour simulation)

As shown in Table 10-2, even with an efficient method such as the DEMM, uncertainty analyses could be very time-consuming for large, complex models. There existed a tradeoff between the accuracy of the results and the amount of time spent on the analysis. For air quality models, which are plagued with many sources of uncertainty, an analysis of the uncertainties in chemical reactions and initial conditions would provide a lower bound of the amount of uncertainties in the predictions. Furthermore, it was found from the mechanistic study that the variance-contributing parameters could be identified even when the PCE representation was imperfect. Given the goals of the study, as listed in Section 10.2, it was determined that the results of the uncertainty analysis would be informative even if the approximation is not perfect. A second order approximation (without cross terms) was

chosen for this study. Interaction effects were not studied. The errors of the approximations were documented explicitly in the report.

A variety of output could be chosen as the response variables for the uncertainty analysis, such as the maximum concentration of ozone in the simulation domain or at particular monitoring stations, daily averaged ozone concentration, etc., it was determined that the time series outputs would provide additional information regarding the response of the ADE to uncertainties in chemistry parameters. In addition to ozone, species such as NO, NO_2, formaldehyde, PAN, HNO_3, H_2O_2, and radicals were analyzed as output variables of the uncertainty analysis. The hourly-averaged time series information were obtained at 5 locations: Hawthorn (HAWT), Central Los Angeles (CELA), Pasadena (PASA), Claremont (CLAR), and Riverside (RIVR), corresponding to the locations of upwind, urban, and downwind monitoring stations where observational data were recorded. Different locations were chosen to shed light on the effect of different ROG/NO_x concentrations on uncertain predictions.

10.5.1 Uncertain Predictions of AIRSHED

The uncertainty bounds (+/- standard deviation) of the model predictions are presented in the following figures. Each set of five figures are the predictions for CELA, CLAR, HAWT, PASA, and RIVR respectively. Figure 10-2 shows the mean and standard deviation time series of ozone. Figures 10-3 and 10-4 are the corresponding results for NO and NO_2. Formaldehyde predictions are shown in Figure 10-5. Nitric Acid and PAN results are depicted in Figures 10-6 and 10-7.

The concentrations of ozone in Figure 10-2 exhibited afternoon peaks at all 5 monitoring stations. The maximum values of ozone at the upwind and urban sites, HAWT, CELA, and PASA, were lower than those at CLAR and RIVR. The peak ozone concentrations occurred relatively early in the downwind sites at CLAR and RIVR. Unlike the uncertain predictions of the box models (Chapters 8 and 9), in which the expected dynamics were quite different from the nominal (deterministic) simulation, the mean values in the three dimensional model did not deviate very much from the nominal values from the deterministic simulation. Ozone uncertainties typically grew with the mean values. The maximum uncertainties for the predictions of ozone were higher at the downwind locations than the upwind sites. The standard deviations ranged from 25 ppb (35%) at upwind sites such as CELA and PASA to 40 ppb (25%) at the downwind site of RIVR. Therefore, the times at which peak ozone levels were registered were not changed by the consideration of uncertainty. The relative errors, defined in Section 5.5, for ozone predictions in the period of 9 a.m. to 5 p.m. are shown in Table 10-3. These relative error values show that the PCE representations and hence the ozone predictions were by no means perfect. However, based on previous experience, the uncertainty estimates should be acceptable and convergent at the accuracy level of a few percent. Moreover, the relative importance of the variance-contributing input parameters can be elucidated.

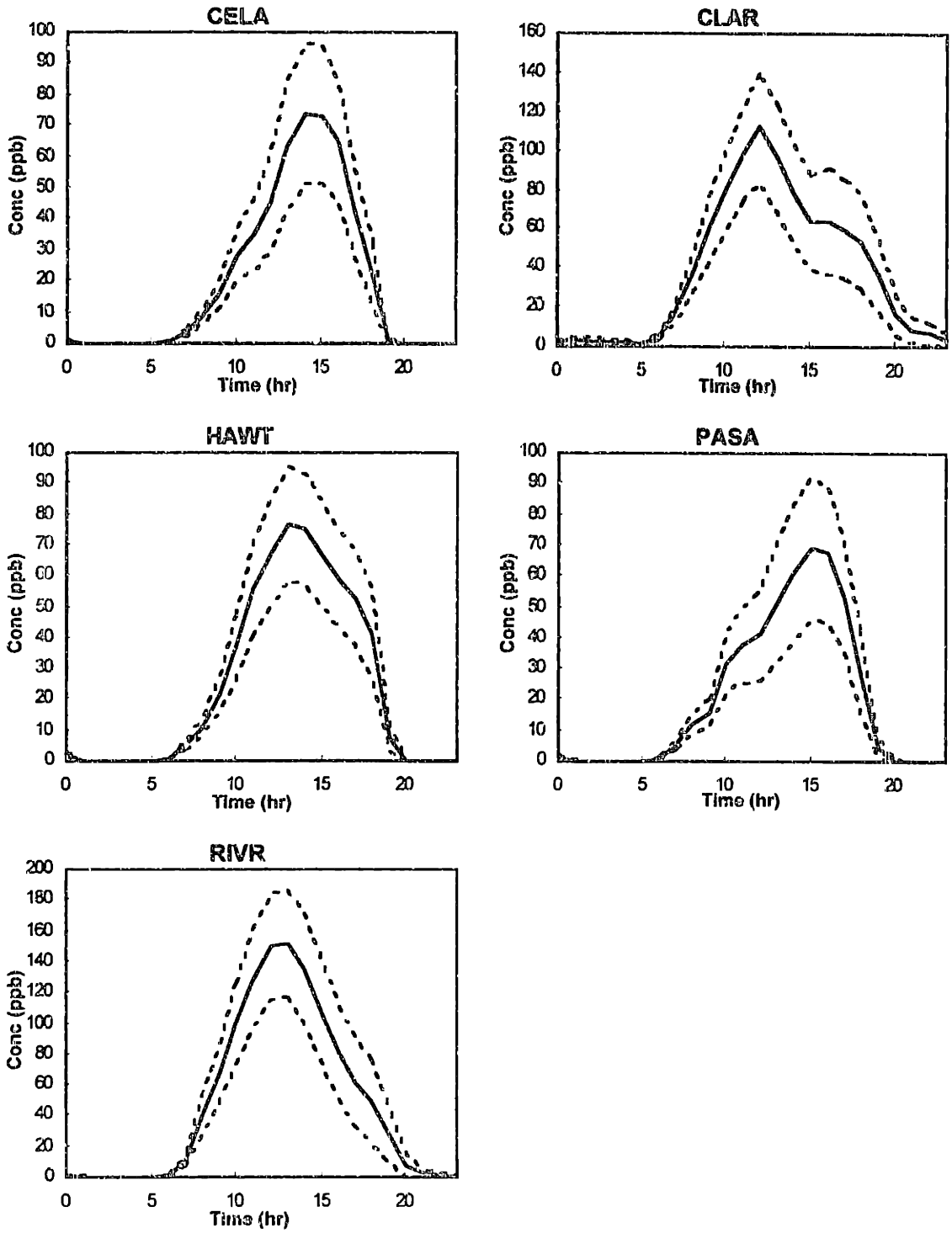


Figure 10-2. Predicted O_3 Mean and Standard Deviation Time Series (5 Monitoring Stations).

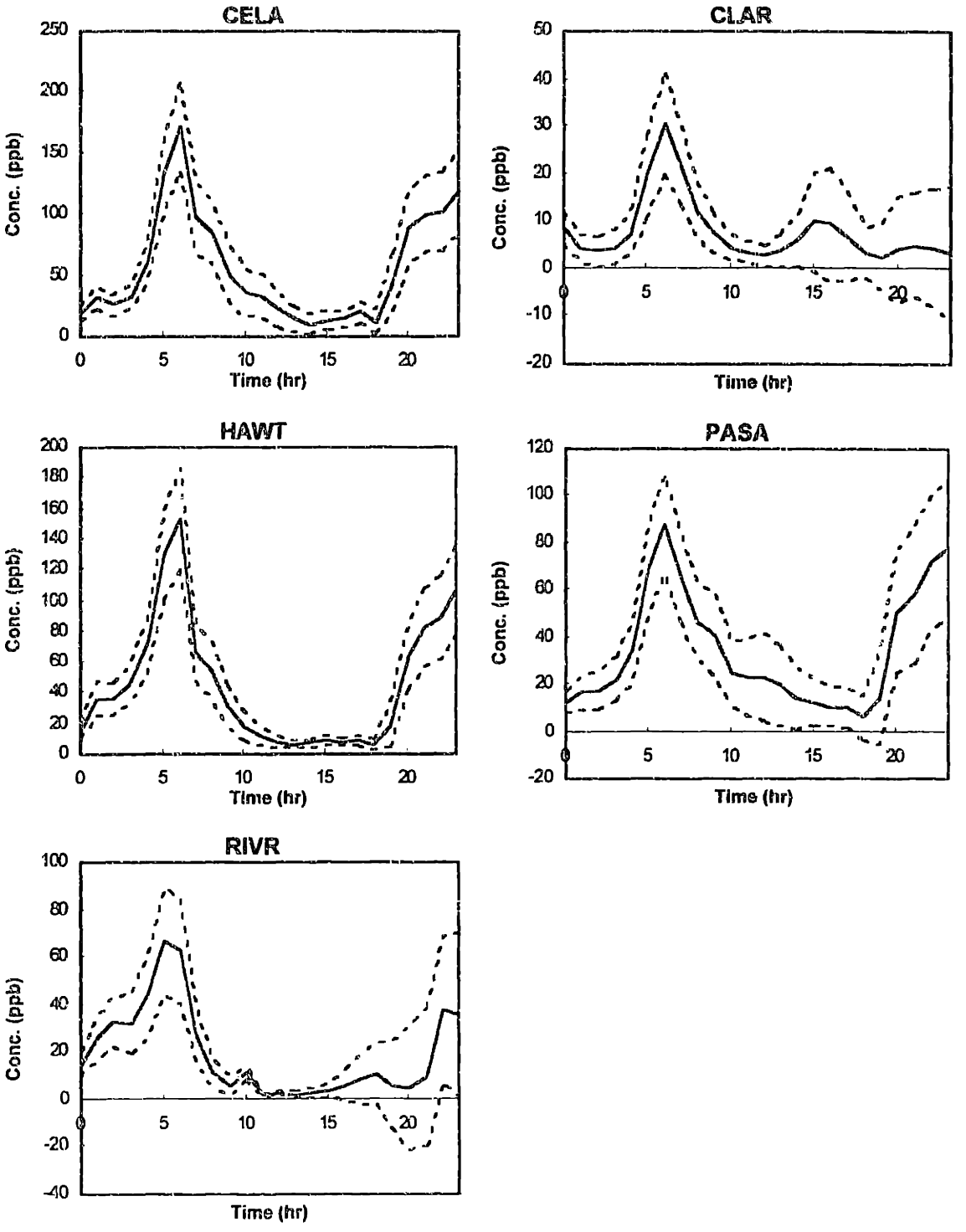


Figure 10-3. Predicted NO Mean and Standard Deviation Time Series at 5 Monitoring Stations.

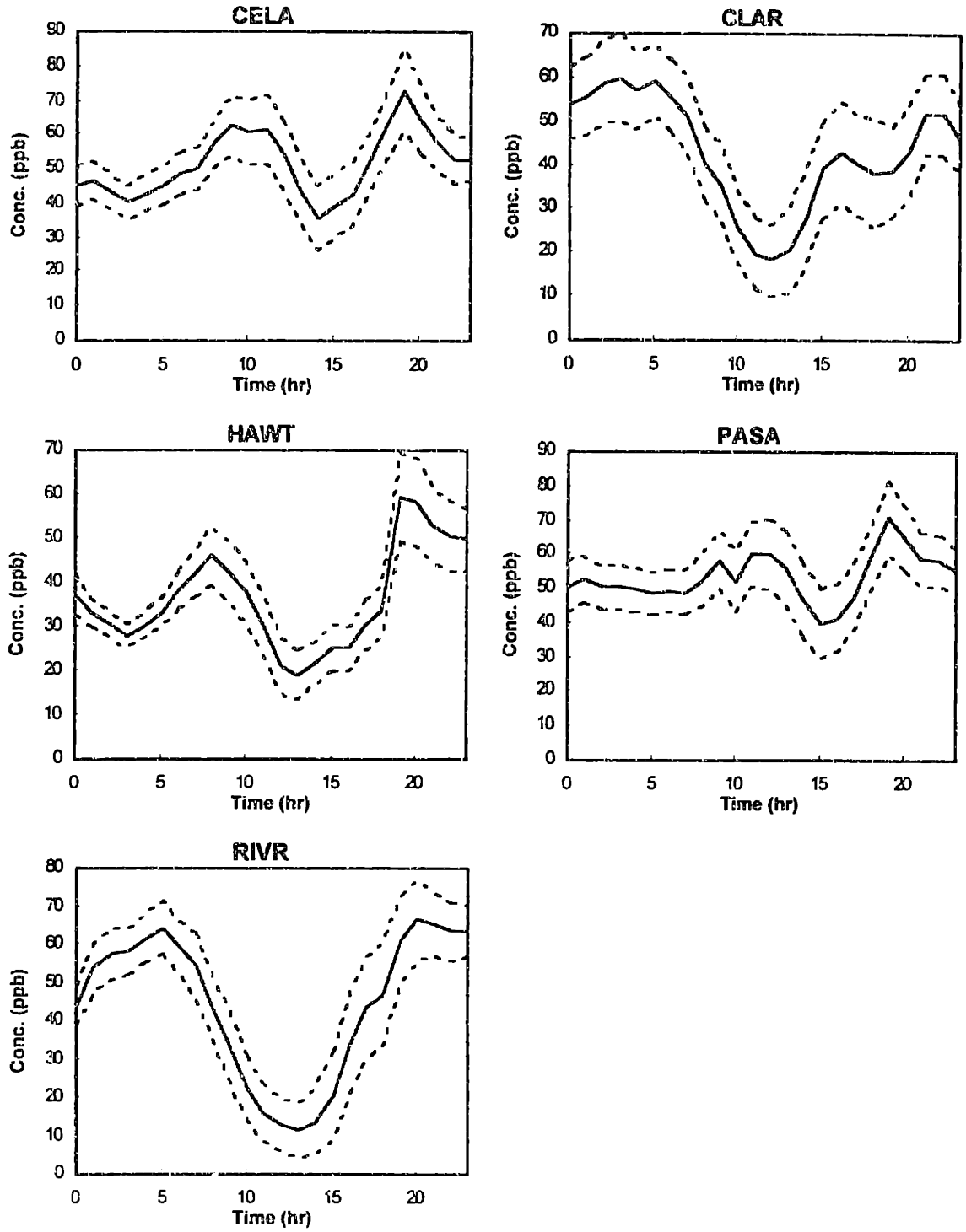


Figure 10-4. Predicted NO_2 Mean and Standard Deviation Time Series at 5 Monitoring Stations.

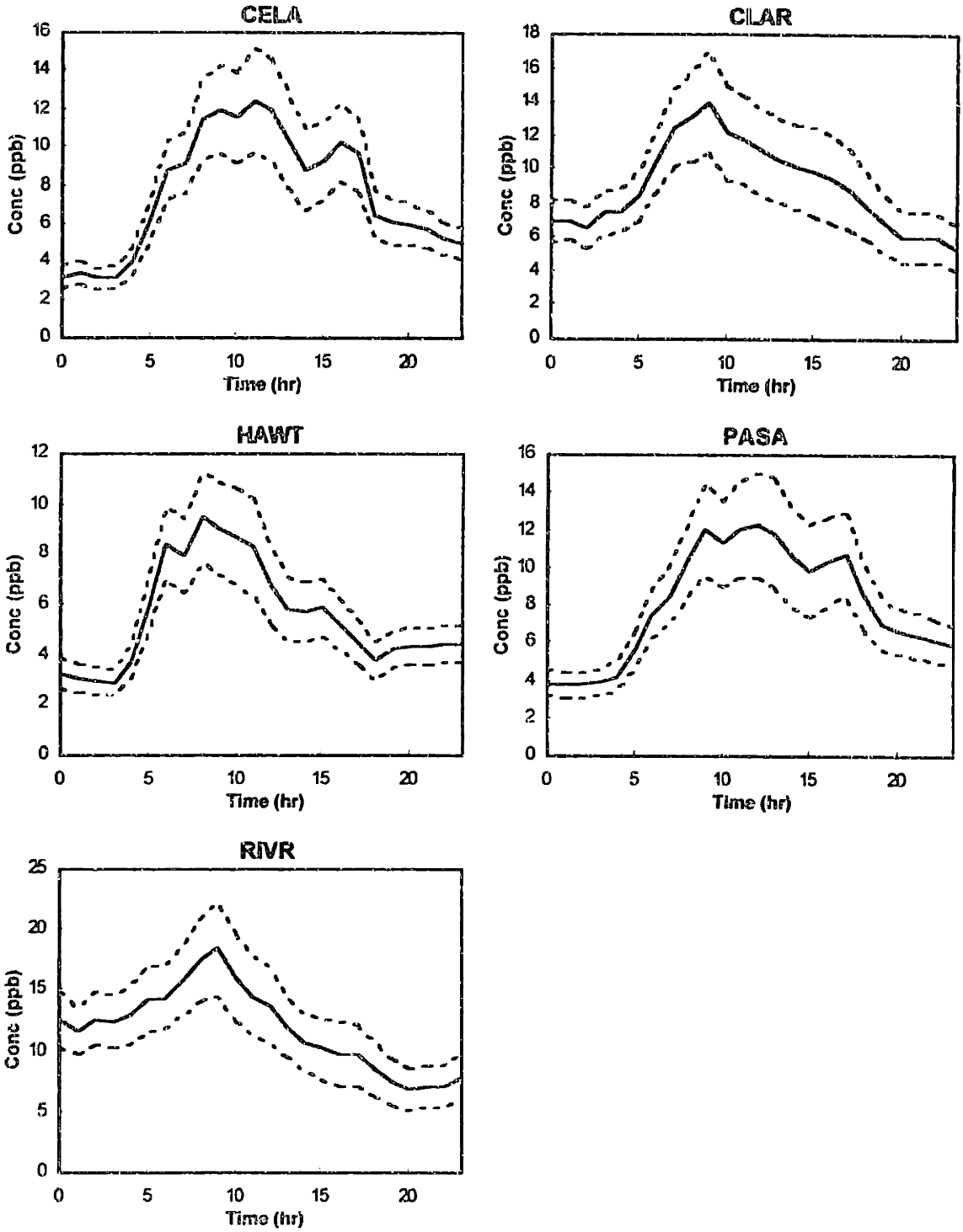


Figure 10-5. Predicted HCHO Mean and Standard Deviation at 5 Monitoring Stations.

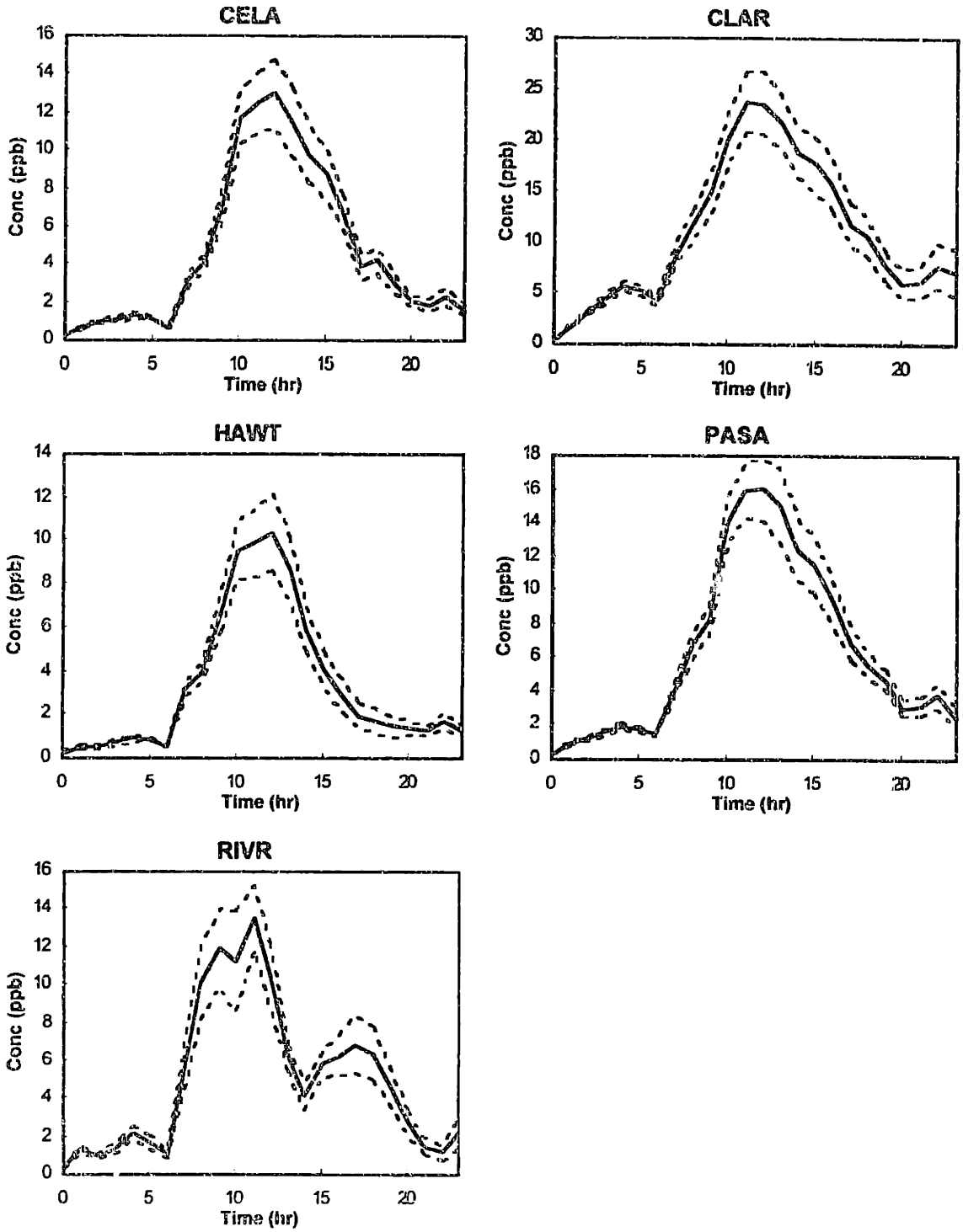


Figure 10-6. Predicted HNO_3 Mean and Standard Deviation at 5 Monitoring Stations.

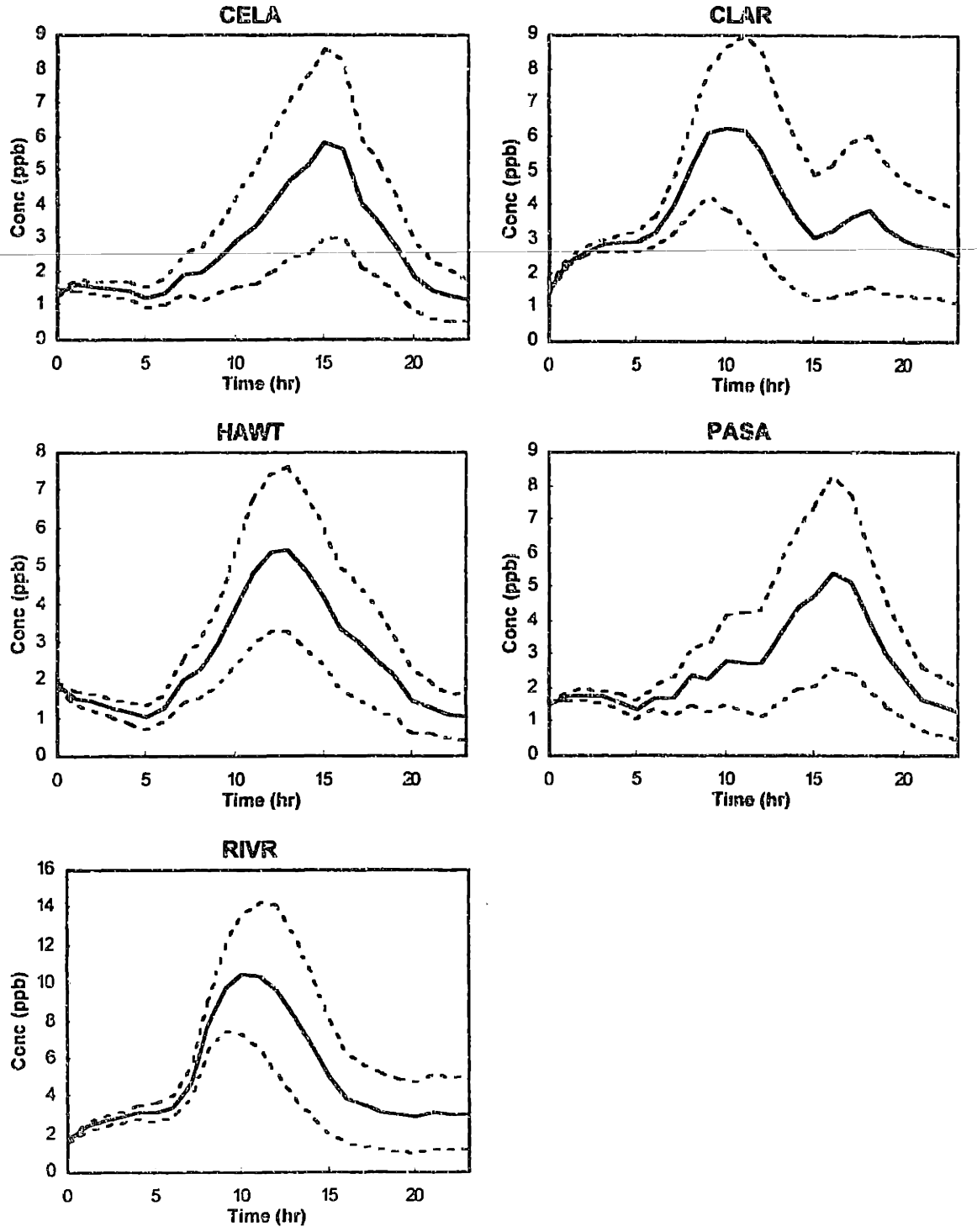


Figure 10-7. Predicted PAN Mean and Standard Deviation at 5 Monitoring Stations.

Time	CELA	CLAR	HAWT	PASA	RIVR
9	4.26E-02	1.37E-02	2.89E-02	6.02E-02	1.62E-02
10	4.02E-02	1.63E-02	3.66E-02	3.23E-02	1.23E-02
11	6.21E-02	1.85E-02	1.92E-02	3.93E-02	8.57E-03
12	6.17E-02	1.61E-02	1.92E-02	5.86E-02	9.07E-03
13	3.41E-02	2.22E-02	1.03E-02	5.49E-02	1.14E-02
14	2.37E-02	3.25E-02	1.13E-02	5.42E-02	1.94E-02
15	1.75E-02	5.25E-02	1.00E-02	4.44E-02	3.42E-02
16	1.41E-02	4.86E-02	2.73E-02	2.72E-02	3.13E-02
17	1.75E-02	4.57E-02	8.56E-03	2.15E-02	5.79E-02

Table 10-3. Relative Errors in Ozone Predictions at 5 Stations.

The limitations of a second order PCE representation were shown quite clearly in Figure 10-3 for NO predictions in the early evening. Even though negative concentrations were not consistent with the physical behavior of the system, and were not predicted by any of the Airshed model rums at the chosen collocation points, uncertain NO concentrations at several locations were predicted by DEMM to take on negative values with significant probabilities. Negative concentrations were predicted by the polynomial chaos representation frequently in the evenings at CLAR, PASA, and RIVR and were associated with extremely high error values (up to 80% at the RIVR site at 8 p.m.). These responses were rejected as erroneous and not analyzed further. During the day, the uncertainties of NO were highest when NO concentrations peaked, and generally decreased when the mean NO concentrations were low (except for CLAR and RIVR, where the standard deviation exceeded the mean in the evening hours, resulting in the occurrence negative concentrations with significant probabilities). At downtown locations such as CELA, the morning NO concentration were up to 40 ppb (25% of the mean value) uncertain due to uncertain NO_x sources.

NO_2 predictions exhibited a minimum at each station at the time corresponding to the peak ozone level. The magnitude of the uncertainty bounds for NO_2 were fairly constant with time at all monitoring stations. The lack of dependence with the predicted levels indicated that some kind of equilibrium process controlled the concentration of NO_2, making the concentration of NO_2 relatively insensitive to uncertainties in the sources of NO_x and in the rate constants. Compared to the errors in the representation of ozone, representations for the NO_2 response were more accurate, as shown in Table 10-4. Most of the errors during the daytime were of at the one to two percent level. Higher relative errors were associated with relatively low predicted values of NO_2 at RIVR.

Formaldehyde concentrations and the associated uncertainties are shown in Figure 10-5. The relative errors in the predictions of HCHO were generally less than 1%. Due to the uncertainties in the initial conditions, formaldehyde concentrations were uncertain from the start of the simulation. After dawn, the predictions of HCHO became even less certain due to the variance contributions from the chemical reactions parameters. Like in the case of ozone, the uncertainties associated with formaldehyde predictions grew with the absolute concentrations. The maximum relative uncertainty associated with HCHO predictions was 28% of the mean value, during the early afternoon at RIVR. The mean concentrations and the fractional uncertainties of HCHO decreased towards the end of the day at all locations.

Nitric acid and PAN, both products of radical termination reactions, exhibited very different behaviors. Nitric acid reached a maximum at all locations in the early afternoon, with RIVR

alone exhibiting a secondary peak in the late afternoon (Figure 10-5). At the time of maximum HNO_3, uncertainties in the concentrations of nitric acid were typically 10 - 20% of the mean. The absolute uncertainties decreased with the predicted concentrations towards the end of the day. At RIVR, the uncertainty reached 20% at 5 p.m., when a secondary peak in nitric acid concentration was observed, and eventually decreased as the concentration of HNO_3 decreased at the site. At CLAR, the other downwind site, the expected value of nitric acid concentration stayed above 5 ppb throughout the evening, and standard deviation associated with the HNO_3 concentrations remained above one ppb.

Time	CELA	CLAR	HAWT	PASA	RIVR
9	1.98E-02	4.27E-03	5.89E-03	8.76E-03	2.15E-02
10	3.55E-02	5.93E-03	1.48E-02	1.24E-02	1.41E-02
11	1.02E-02	9.69E-03	7.82E-03	2.69E-02	2.55E-02
12	1.10E-02	9.83E-03	9.17E-03	1.89E-02	4.20E-02
13	1.81E-02	2.15E-02	1.47E-02	1.72E-02	6.48E-02
14	1.10E-02	1.67E-02	1.11E-02	1.72E-02	5.95E-02
15	1.48E-02	2.03E-02	4.79E-03	1.64E-02	2.23E-02
16	9.80E-03	2.21E-02	8.28E-03	1.02E-02	1.66E-02
17	5.34E-03	2.17E-02	8.63E-03	7.61E-03	2.52E-02

Table 10-4. Relative Errors in NO_2 Predictions at 5 Stations.

The time series behavior of PAN varied from station to station, with peak concentrations at 10 a.m. at RIVR and not until 4 p.m. at CELA and PASA. The concentrations of PAN were lower than those of nitric acid by roughly a factor of two, except at RIVR, where the concentration of PAN was quite high. PAN also had the highest relative uncertainties (standard deviation / mean) of the compounds discussed so far. The ratio of the standard deviation to the mean value increased monotonically throughout the day, plateauing at 50-60% at the end of the 24 hour simulation period. Since the uncertainties were not directly proportional to the mean value, the time at which the concentration of PAN peaked may shift as a result of uncertainties in the chemical parameters. For example, at RIVR, it seemed that the peak of PAN could occur between 9 a.m. and 2 p.m. because of the uncertain predictions. The relative errors associated with PAN predictions are shown in Table 10-5. The highest relative errors observed were 8%, and the highest absolute error was 0.3 ppb.

Time	CELA	CLAR	HAWT	PASA	RIVR
9	2.46E-02	1.22E-02	1.23E-02	2.11E-02	1.04E-02
10	2.93E-02	1.70E-02	1.84E-02	2.43E-02	1.35E-02
11	3.88E-02	2.28E-02	1.82E-02	3.95E-02	1.92E-02
12	3.91E-02	3.13E-02	1.93E-02	6.40E-02	2.81E-02
13	3.54E-02	4.31E-02	1.88E-02	5.64E-02	2.88E-02
14	4.01E-02	6.58E-02	1.42E-02	5.11E-02	3.47E-02
15	2.78E-02	7.94E-02	8.55E-03	5.03E-02	5.78E-02
16	2.51E-02	8.08E-02	1.22E-02	4.05E-02	7.13E-02
17	1.86E-02	6.96E-02	7.69E-03	3.18E-02	7.66E-02

Table 10-5. Relative Errors in PAN Predictions at 5 Stations.

10.5.2 Variance Analyses of Key Species

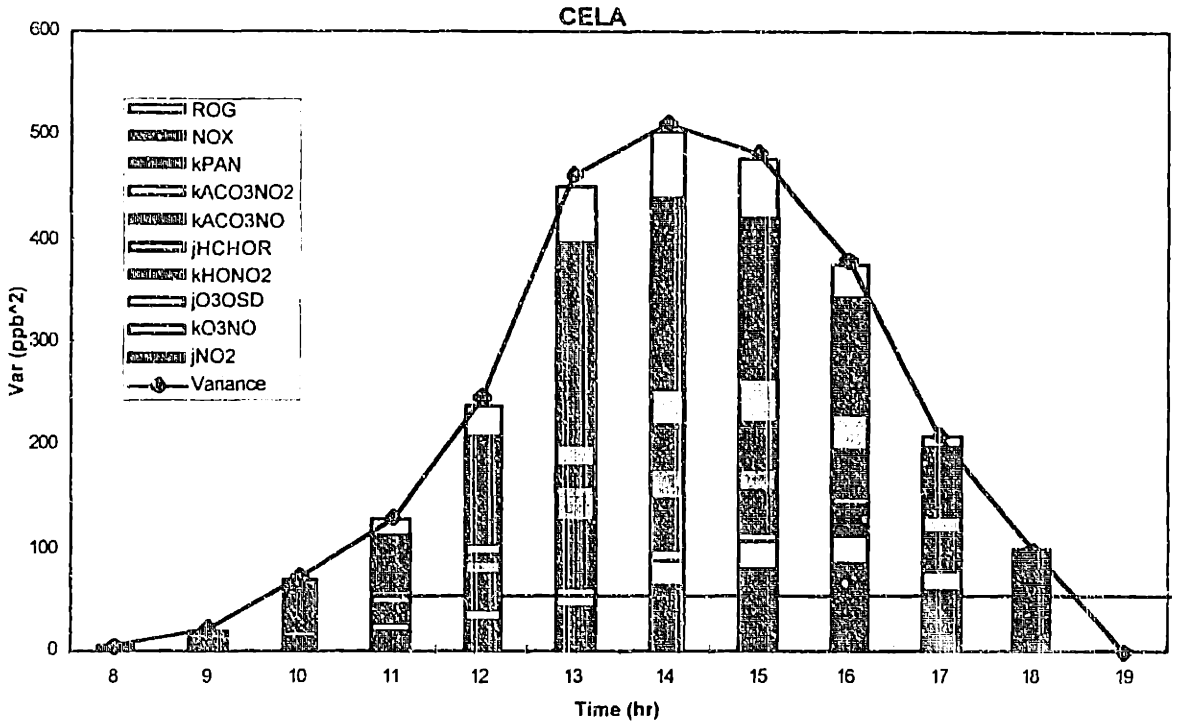
Ozone. Variance analyses were performed on the uncertain concentrations of several species. Figures 10-8 through 10-10 show the results for ozone, formaldehyde, and PAN for the five stations. Since the concentration of ozone was negligible before sunrise and after sunset, uncertainty analyses were only meaningful for the hours between 8 a.m. and 9 p.m. From the panels of Figure 10-8, it can be observed that the uncertainties in ozone were affected by the uncertainties of a range of parameters. Uncertainties of the 10 parameters included in the diagrams accounted for at least 95% of the uncertainties in ozone predictions. Most of these parameters were previously identified to be VCIs in the one dimensional mechanism studies, although the relative contributions were quite different. A few parameters, such as $j(O_3 \rightarrow OSD)$ and $k(O_3+NO)$, contributed to only a few percent of the total variance in ozone. The initial conditions and source strengths of the precursor species ROG and NO_x accounted for a significant fraction of the uncertainties in ozone in all sites except HAWT, a station upwind of most emissions. It should be noted that within a 20% uncertainty the effect of increasing NO_x was a reduction in ozone. At HAWT, the variance of the ozone time series was relatively low, partly due to the small role of the uncertain direct emissions. In fact, uncertainties in the source and initial conditions of NO_x and ROG contributed to ozone uncertainties only in the morning, accounting for about 30% of the variances in the predictions of ozone. At urban sites, most notably CELA and PASA, the variance contributions from the initial conditions and emissions of ROG and NO_x were highest in the morning and early afternoon, contributing up to half the total variance of ozone, and decreased slightly towards the end of the afternoon. CLAR and RIVR, the downwind sites, exhibited the opposite behavior, with growing variance contributions from the emissions as time elapsed. These two parameters can contribute up to 60% of the uncertainties in ozone downwind, even though emissions and initial conditions were only 20% uncertain. Of the two precursors, the effects of NO_x were almost always higher than that of ROG. The only counter example found among the responses here was at the downwind site of RIVR in the morning, where the VC of IC/source of ROG exceeded that of NO_x. However, it should also be noted that at this site at this time, the kinetic parameters had stronger effects on the uncertainties of ozone concentration.

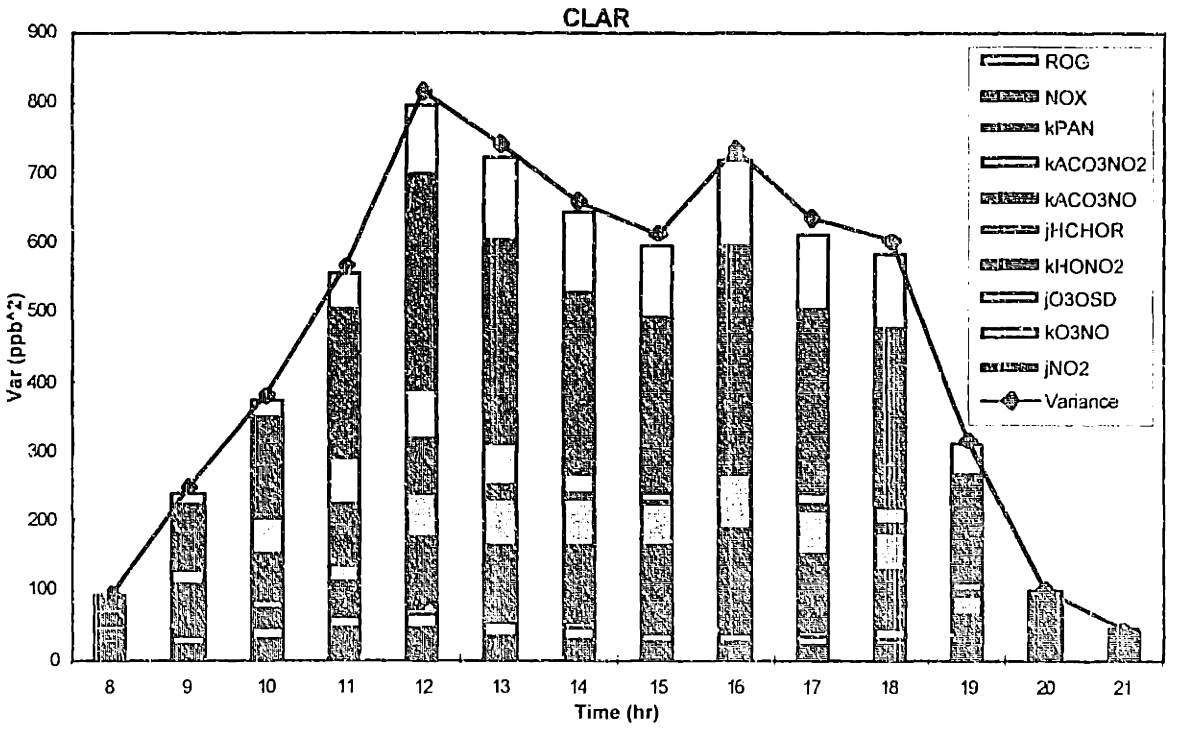
In Section 8.3, the reaction rates of peroxyacyl radical and PAN reactions were found to be VCIs for formaldehyde concentrations. The significant variance contributions of these uncertain reaction rate constants to the uncertainties in the predictions of ozone in a three dimensional model were not predicted by the mechanism studies performed in a box model. Acyl radicals, labeled ACO_3 in the chemical mechanism, could react with NO to form NO_2 and carbonyl products, or with NO_2 to form PAN.

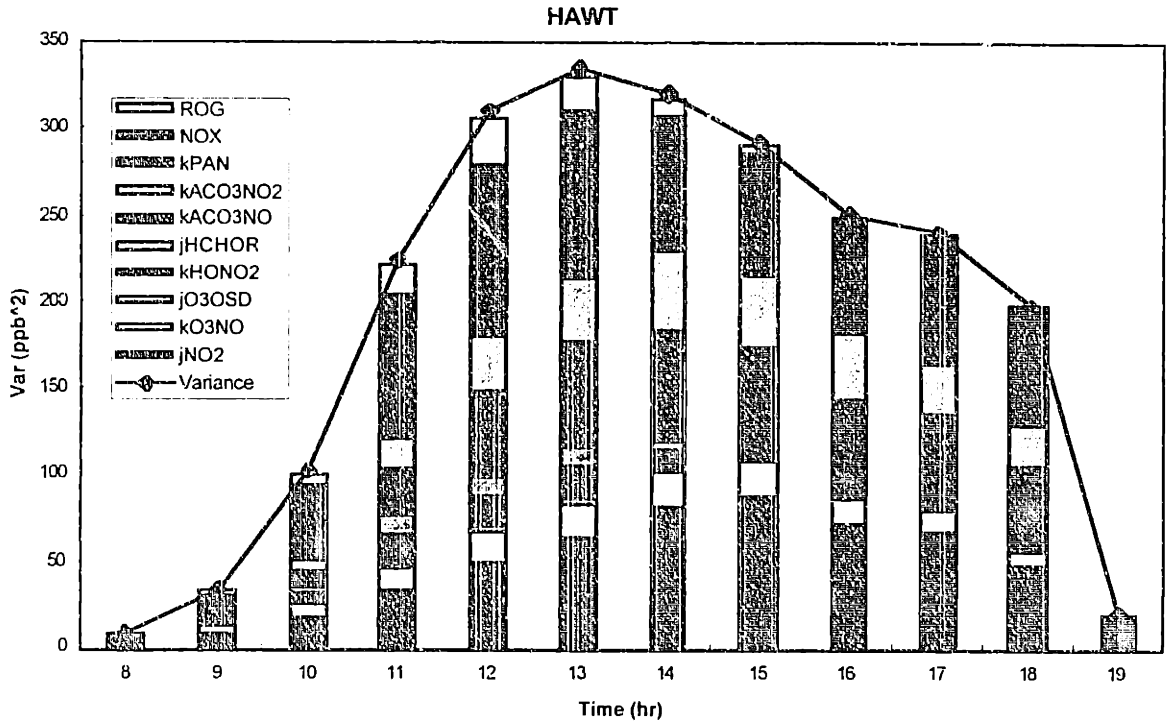
$$ACO_3 + NO \rightarrow HCHO + RO_2 + NO_2 \qquad (10.2)$$
$$ACO_3 + NO_2 \rightarrow PAN \qquad (10.3)$$

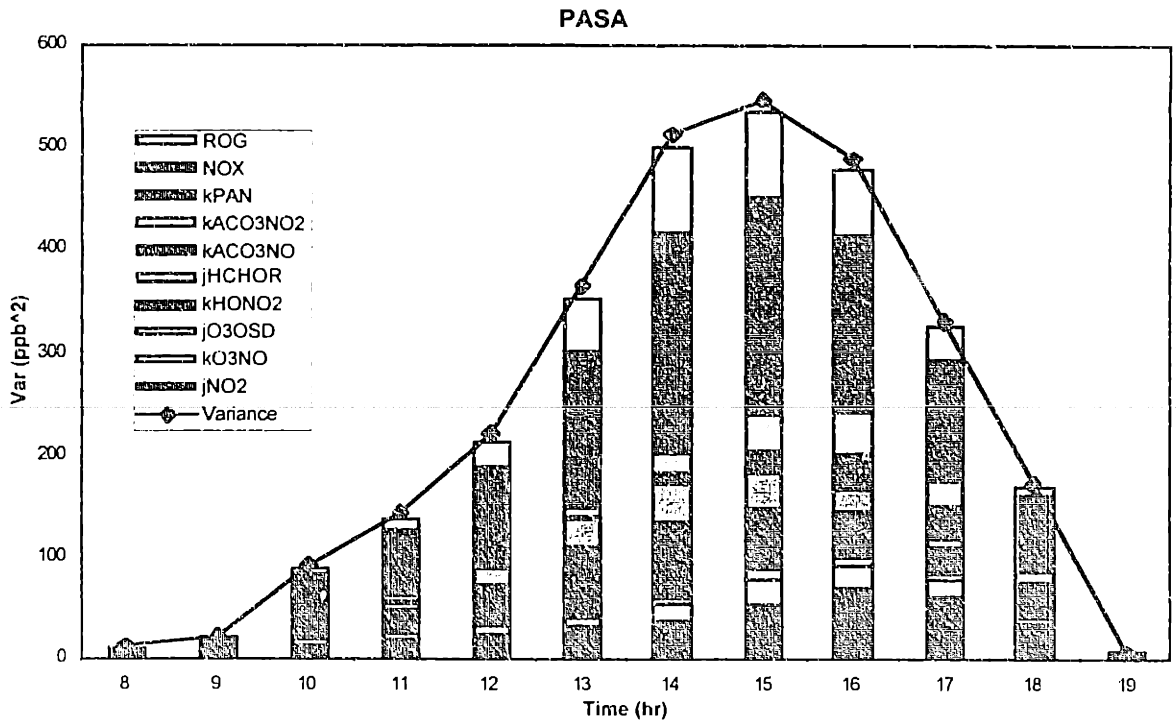
In Reaction 10.2, the NO_2 formed would photolyze and would eventually result in ozone formation. The formation of PAN in Reaction 10.3 temporary removed ACO_3 radicals and NO_x, and reduced the reactivity of the chemical system, thus the rate of ozone formation. The two reactions can then be considered competitive reactions for ACO_3 which have opposite effects in terms of ozone formation. The thermal dissociation of PAN back to its precursors freed up both NO_2 and organic radicals and promoted ozone formation:

$$PAN \rightarrow ACO3 + NO_2. \qquad (10.4)$$


CELA
ROG
NOX
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
kO3NO
jNO2
Variance
Var (ppb^2)
Time (hr)


CLAR
ROG
NOX
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
kO3NO
jNO2
Variance
Var (ppb^2)
Time (hr)


HAWT
ROG
NOX
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
kO3NO
jNO2
Variance
Var (ppb^2)
Time (hr)
0
50
100
150
200
250
300
350
8
9
10
11
12
13
14
15
16
17
18
19


PASA
ROG
NOX
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
kO3NO
jNO2
Variance
Var (ppb^2)
Time (hr)
0
100
200
300
400
500
600
8
9
10
11
12
13
14
15
16
17
18
19

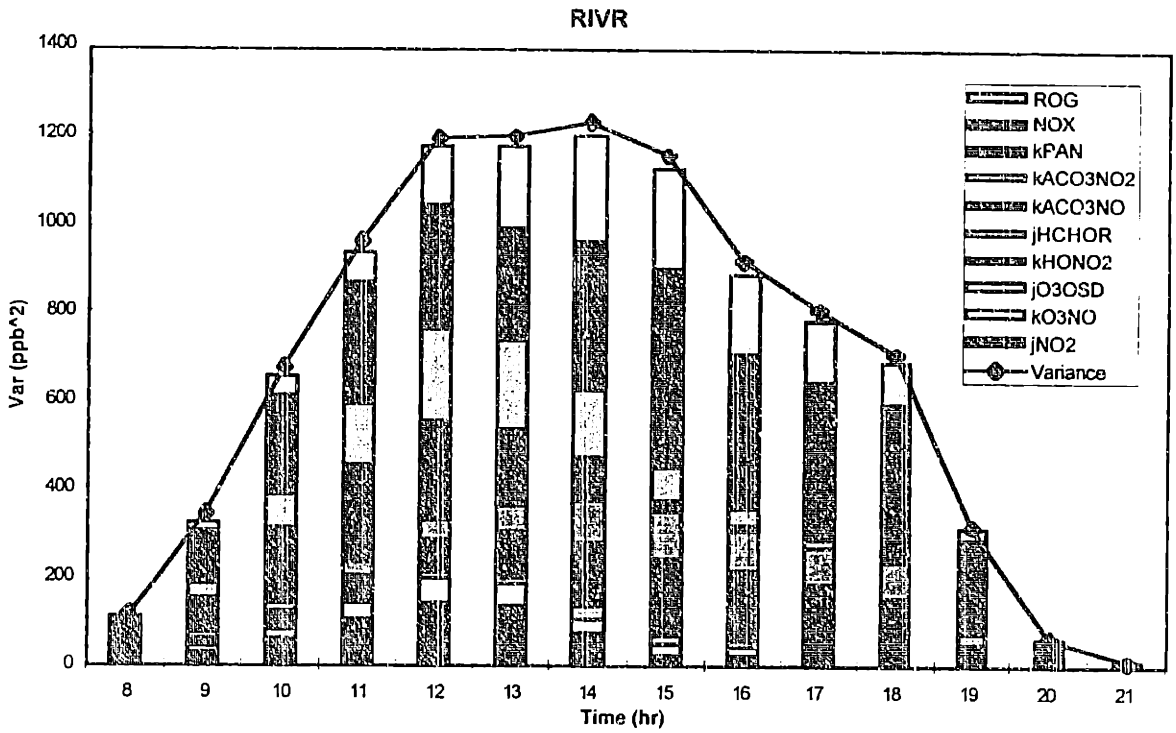


Figure 10-8. Variance Contributions from Input Parameters to Ozone Concentrations at 5 Monitoring Stations.

The uncertainties in the reaction rates of Reactions 10.2, 10.3, and 10.4 had strong effects on the uncertainties of ozone, especially in upwind and downwind areas where there were less direct emissions of ROG and NO_x, i.e., HAWT, CLAR, and RIVR. At HAWT, the contribution of uncertainties of the reactions rates of Reactions 10.2 to 10.4 to the variance of ozone was more than 50% throughout the day. The VC from the reaction rates of the NO_2-forming reactions, k(PAN→ACO_3+NO_2) and k(ACO_3+NO) were especially important starting late morning. This observation may be an indication that NO_2 is the limiting reagent in the chain of reactions leading to ozone formation at this time. At urban sites such as CELA and PASA, the relative contributions of the uncertainties of Reactions 10.2 to 10.4 to uncertainties in ozone predictions were smaller. The combined VC of these rate parameters exceeded 30% of the total variance at 3 or 4 p.m., when the VC of the NO_x source term was reduced. At the downwind sites of RIVR and CLAR, the variance contribution from the reaction rates of the ACO_3 and PAN reactions were high in the morning and shrank in the afternoon. Since the concentrations of PAN were quite high at these downwind sites in the morning (Figure 10-5), the VC of k(PAN→) and k(ACO_3+NO) might be an indication of PAN as a source of NO_x and radicals, and hence ozone, at downwind locations in the morning.

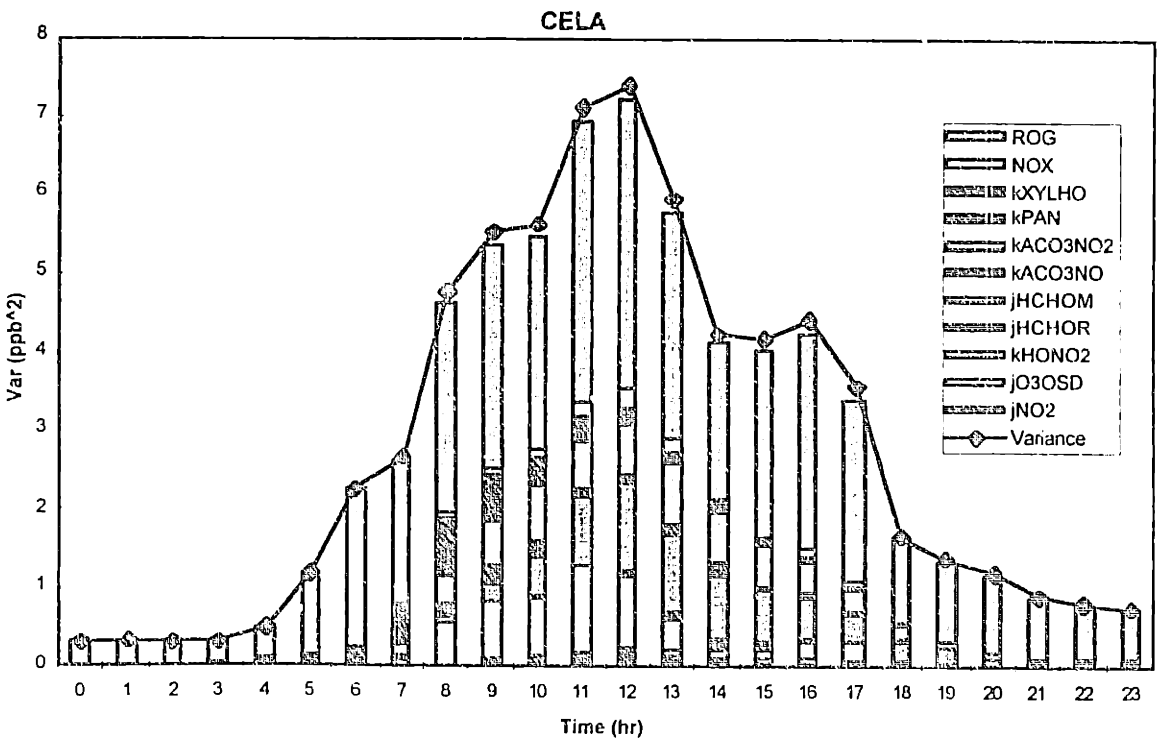
Another nitrogen / radical sink reaction is HO + NO_2 → HNO_3. At urban sites, the contribution of the uncertain in the reaction rate of HO + NO_2 was highest in the morning and early afternoons. On the other hand, at downwind locations, the VC of the nitric acid formation rate increased during late afternoon. Compared to the other sink-formation reaction, ACO_3 + NO_2, the influence of the uncertainty in k(HO+NO_2) was stronger than k(ACO_3+NO_2) until late afternoon in the urban locations. The opposite was true for the downwind locations. At HAWT, the variance contribution of k(ACO_3+NO_2) always exceeded that of k(HO+NO_2). It was possible that at HAWT, PAN was a more important sink for radicals and NO_x than nitric acid. At other locations, the relative importance of the two sink

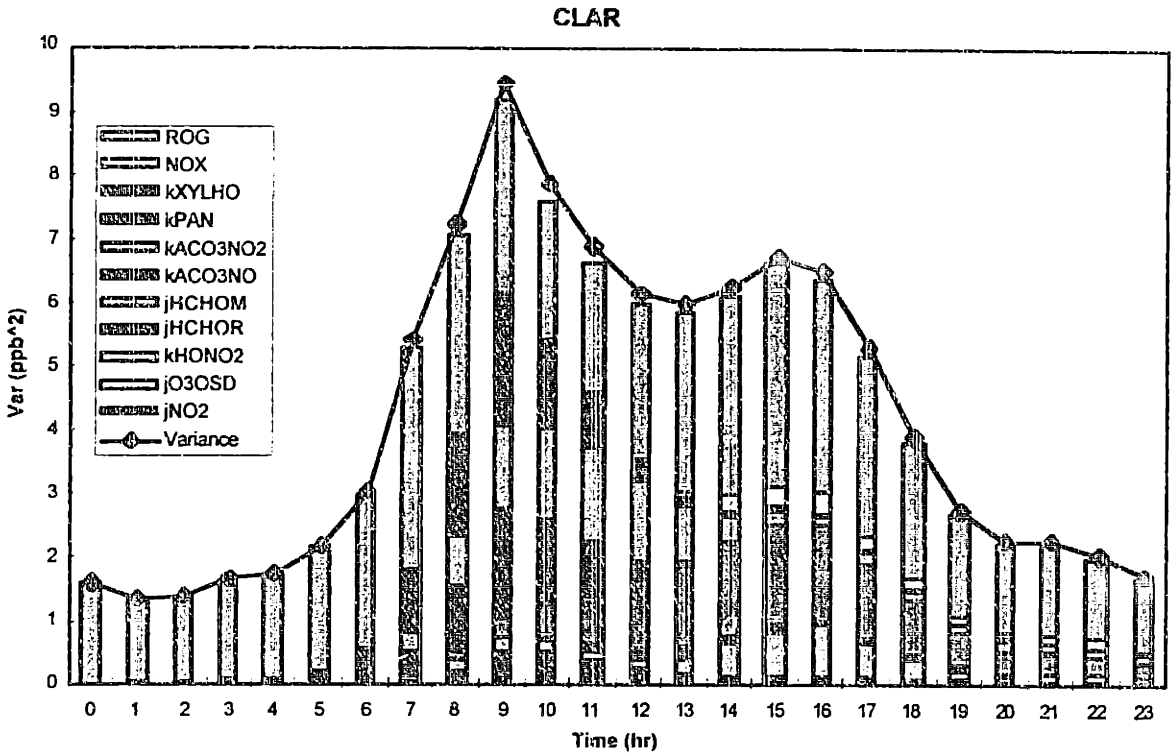
reactions varied with time. The formation of nitric acid upwind and earlier times may reduce the available raw materials, i.e., radicals and NO_x, for ozone formation downwind in the afternoon.

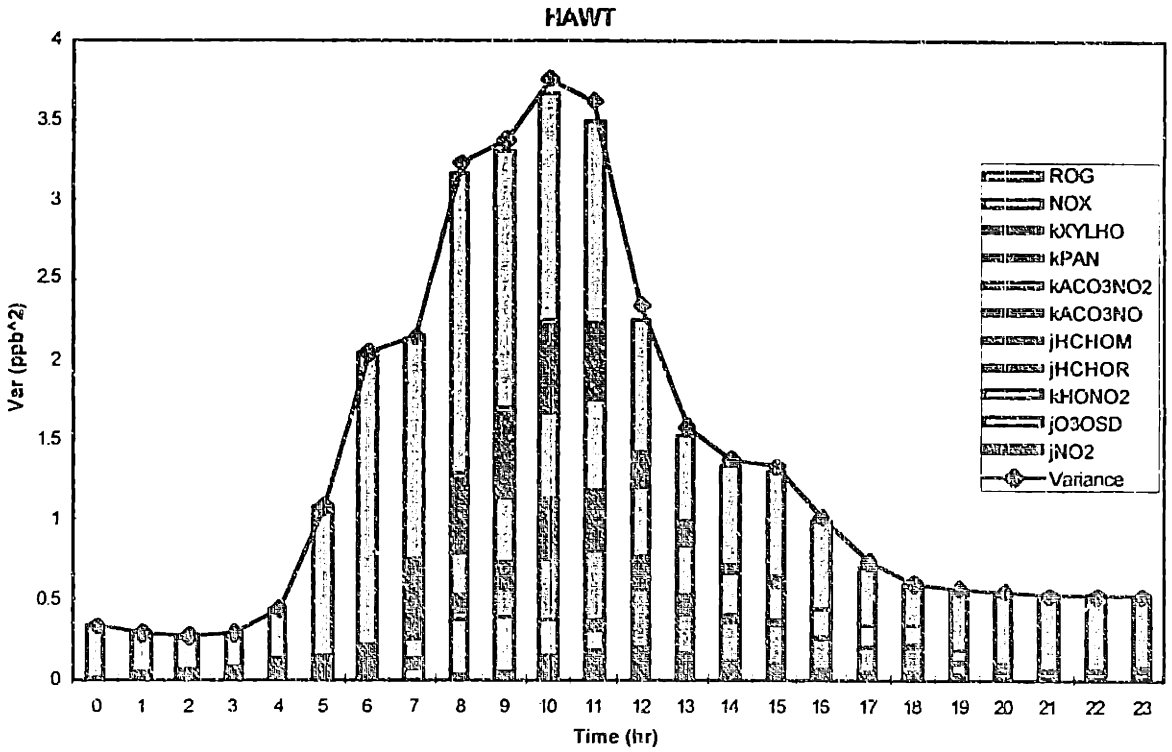
The photolysis rate of NO_2 also proved to be a variance contributor to ozone. This was especially true at HAWT, the upwind site, where the effects of the uncertainty in $j(NO_2)$ could be observed throughout the day. At the urban sites of PASA and CELA, the variance contributions of $j(NO_2)$ were highest in the late afternoon. The downwind ozone uncertainties seemed less sensitive to this parameter. On a fractional basis, the variance contribution of $j(NO_2)$ could be dominating in early morning, though this effect was not observable in Figure 10-8.

Two other VCIs should be noted: $k(O_3+NO)$ and $j(HCHO \rightarrow 2HO_2+CO)$. Compared to the parameters discussed so far, the effects of these two reaction rates were limited. The timing of maximum VC from $k(O_3+NO)$ coincided with the peak of the ozone concentration, because this was a sink reaction for ozone. High concentrations of ozone increased the effect of this rate parameter on the uncertain predictions of ozone. The interesting observation was that, comparing different geographical locations, the percentage of variance contributed by this reaction rate was highest at the upwind site of HAWT. This is consistent with the low ROG concentration at HAWT such that the fate of ozone would be to react with NO.

The radical route of formaldehyde photolysis played a relatively small role in the formation of ozone compared to the box model findings. This was especially true at HAWT, where the concentration of formaldehyde was low throughout the day. At the urban and downwind sites, the uncertainty of the formaldehyde photolysis rate contributed more to the uncertainty in ozone concentration. The effects were largest in the early afternoon at CELA and PASA, and continued to contribute to a few percent of the ozone variance until late in the afternoon at CLAR and RIVR.




CLAR
Var (ppb^2)
Time (hr)
ROG
NOX
kXYLHO
kPAN
kACO3NO2
kACO3NO
jHCHOM
jHCHOR
kHONO2
jO3OSD
jNO2
Variance


HAWT
Var (ppb^2)
Time (hr)
ROG
NOX
kXYLHO
kPAN
kACO3NO2
kACO3NO
jHCHOM
jHCHOR
kHONO2
jO3OSD
jNO2
Variance

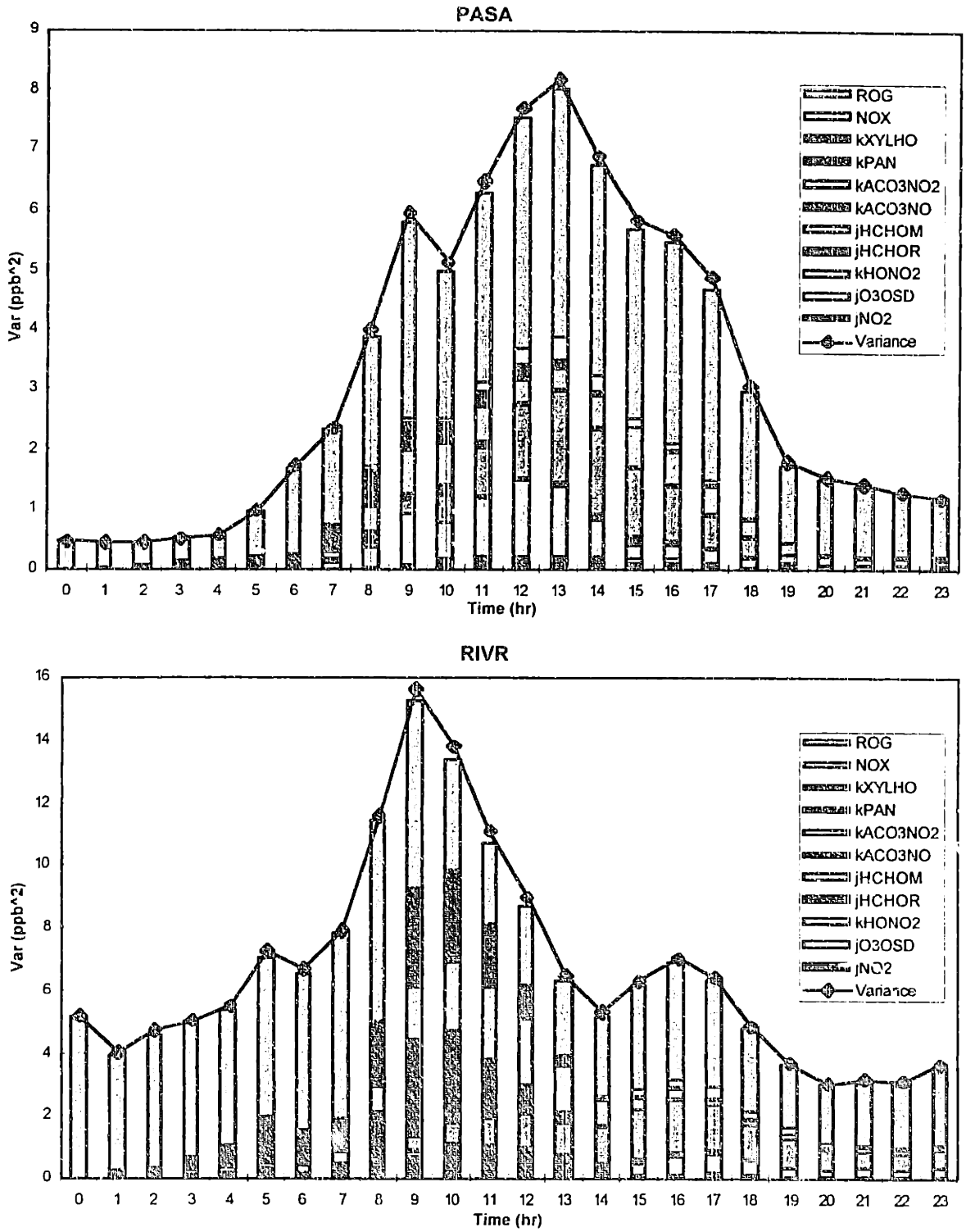


Figure 10-9. Variance Contributions from Input Parameters to Formaldehyde Concentrations at 5 Monitoring Stations.

Formaldehyde. Figure 10-9 shows the parameters that contributed to the uncertainties in formaldehyde predictions. Even though quite a few parameters were needed to explain 95% of the variance of formaldehyde, several contributed to only a few percent of the

uncertainties in formaldehyde predictions. Examples were the source strength of NO_x, k(XYL+HO), and $j(O_3 \rightarrow OSD)$.

The effects of the uncertainties in ROG sources and initial conditions were very significant on the uncertainties of formaldehyde predictions. Before sunrise, the uncertainties in formaldehyde concentrations were almost exclusively attributable to ROG initial conditions and sources. During the day, uncertainties related to chemical kinetics might contribute up to two-thirds of the total uncertainties in formaldehyde concentrations, but the ROG sources remained the single most important factor in explaining the uncertainties of formaldehyde concentrations throughout the day. The lack of variance contribution from NO_x sources and initial conditions probably indicated that NO_x was not a limiting reactant in the reactions preceding the formation of formaldehyde.

The reaction rate constants that contributed the most to uncertainties in AIRSHED's formaldehyde predictions were none other than those for the PAN-related reactions (Reactions 10.2, 10.3, and 10.4). During the morning, the variance contributions of these reaction rate constants accounted for roughly half of the total variance in formaldehyde in the downwind locations CLAR and RIVR. The contributions of these parameters to upwind and urban formaldehyde were comparatively small, maximizing at about 25% in the mid-morning. The thermal rate constant for the ACO_3 + NO and for dissociation of PAN, Reactions 10.2 and 10.4, seemed to contribute strongly to the uncertainties in the predictions of formaldehyde early in the morning. This indicated the importance of the dissociation of PAN followed by the reaction ACO_3 + NO as a direct source of formaldehyde in the system in the morning. Towards the late morning, as the PAN concentration dropped, the variance contribution from the reaction rate of Reaction 10.3, $ACO_3 + NO_2$, increased, indicating the sensitivity to the formation of PAN as a process that removed radical precursors of HCHO.

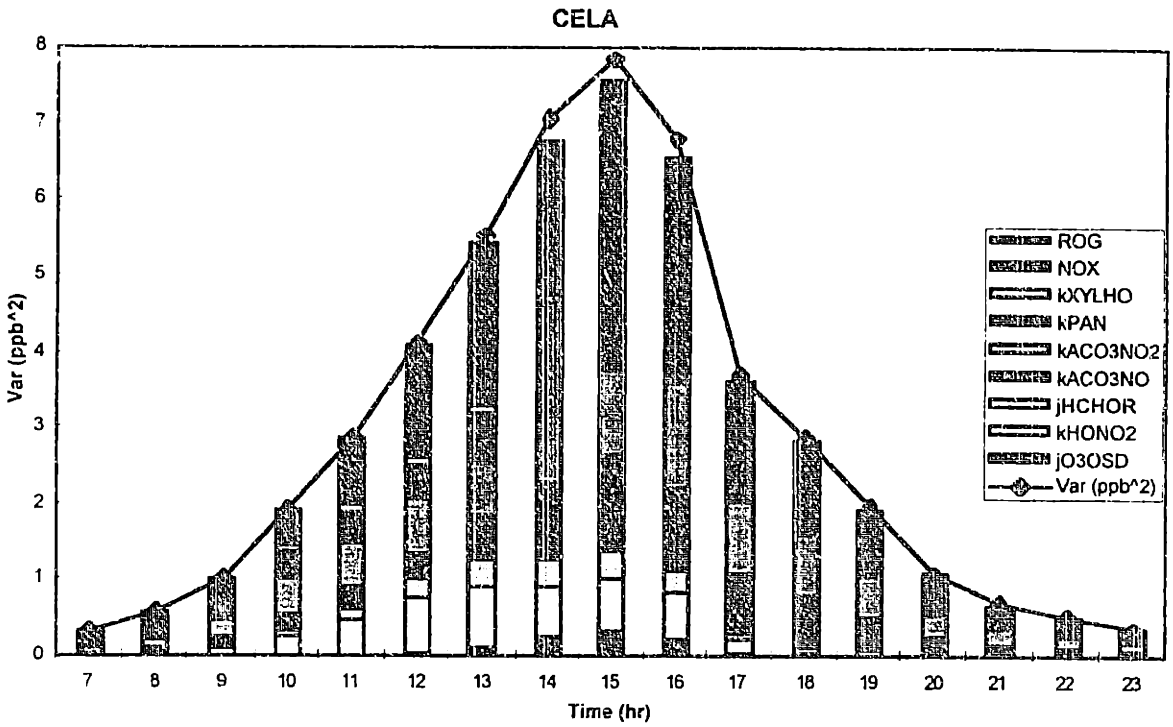
The influence of the above-mentioned parameters continued to drop in the afternoon, especially at the downwind sites. Other parameters gained influence on the uncertainties of formaldehyde. One important parameter turned out to be the photolysis rate of formaldehyde, a key removal process for formaldehyde. In particular, the molecular products route contributed more to the uncertainties of formaldehyde concentrations than any other kinetic parameters in the afternoon. Formaldehyde photolyzes via two pathways,

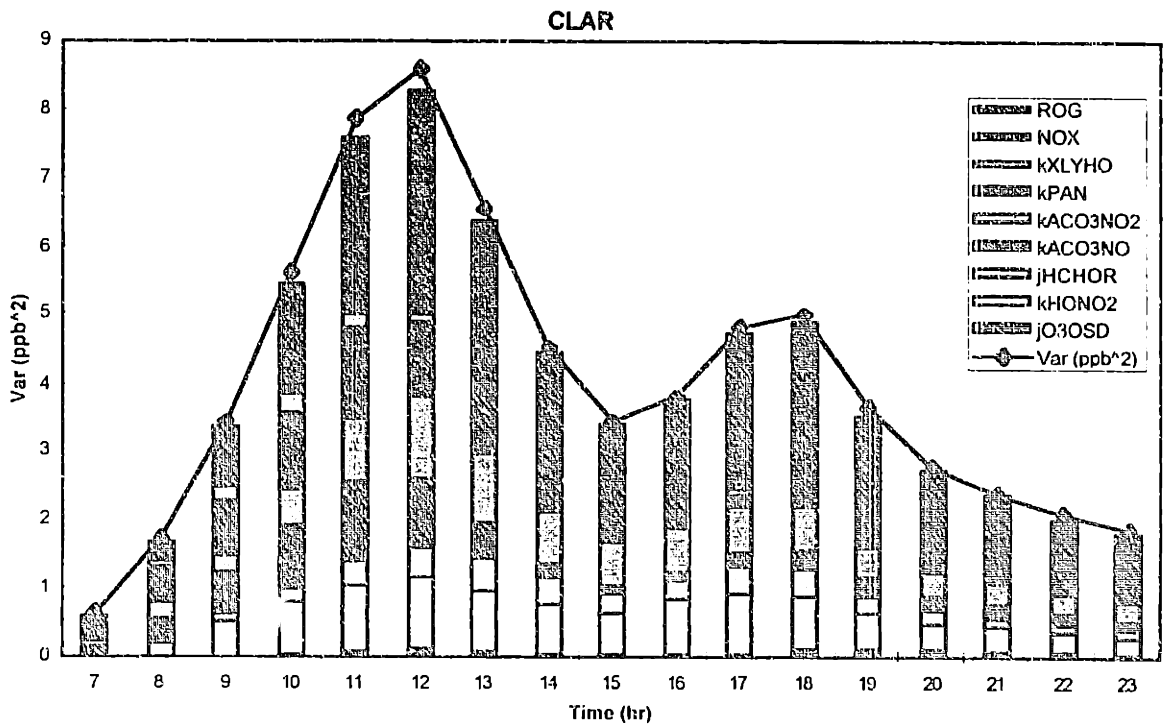
$$HCHO + h\nu \rightarrow H_2 + CO; \text{ and} \qquad (10.5)$$
$$HCHO + h\nu \rightarrow 2\,HO_2 + CO. \qquad (10.6)$$

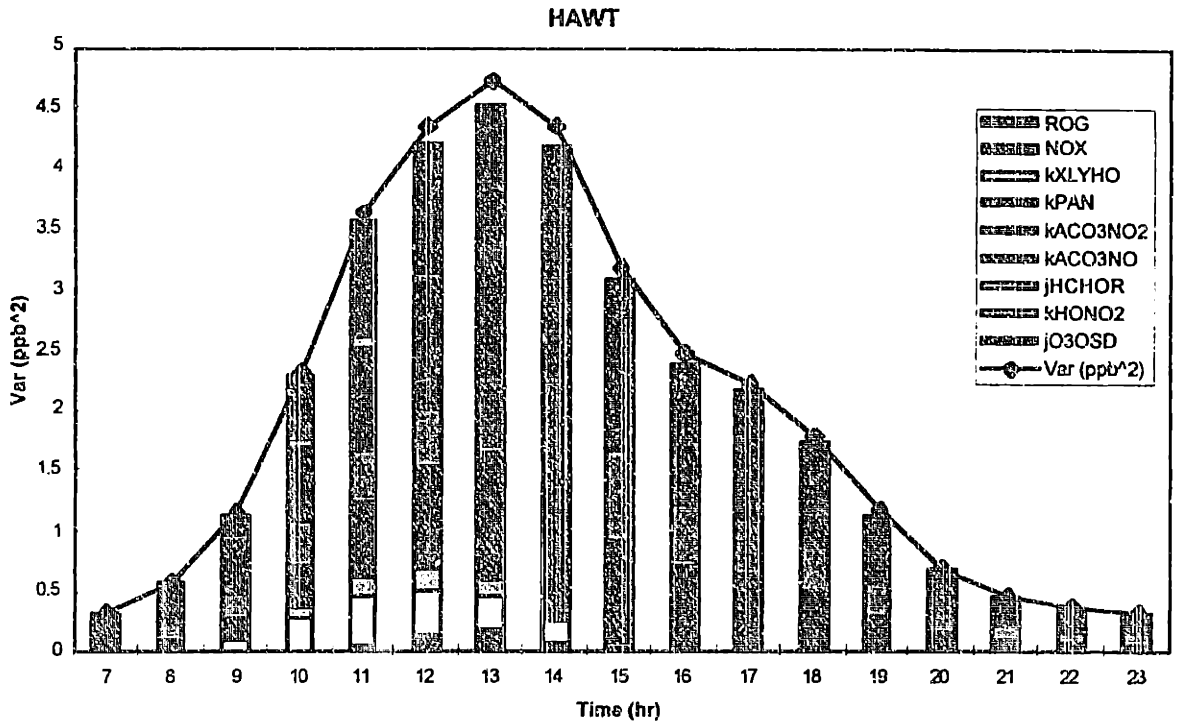
The radical route, Reaction 10.6, contributed to the uncertainties in the formation of ozone, as described previously. The uncertainties in formaldehyde, however, were influenced more significantly by the photolysis reaction rate of Reaction 10.5, which has a higher rate than Reaction 10.6, especially at high solar zenith angle (late afternoon). At CELA, HAWT, and PASA, the contribution of j(HCHO) to the variance of HCHO was highest early in the afternoon. At the downwind sites, the parametric VC was highest later in the afternoon.

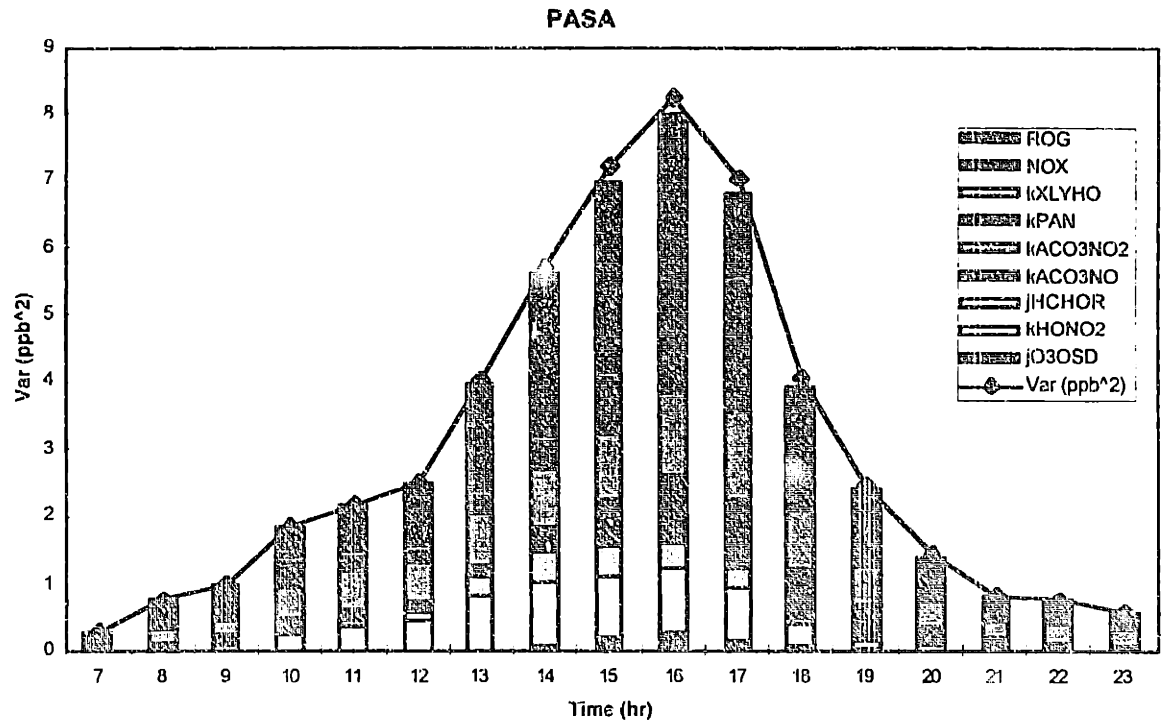
Although the concentration of formaldehyde was only weakly affected by reactions of NO_x, $k(HO+NO_2)$ was another parameter that contributed to variance of formaldehyde, especially in the upwind and urban sites. At the downwind sites of CLAR and RIVR, the VC of $k(HO+NO_2)$ were at the few percent level. Uncertainties in formation rates of sink species contributed generally to uncertainties in the radical concentrations, thereby having an indirect effect on formaldehyde. Other parameters that contributed to the uncertainties in formaldehyde predictions are parameters that dictated the chemical dynamics of the system, including $j(NO_2)$ and $j(O_3 \rightarrow OSD)$.

After sunset, the uncertain predictions of formaldehyde were controlled mostly by the uncertainties in ROG emissions at HAWT, CELA, and PASA. However, at the downwind sites, CLAR and RIVR, the kinetic parameters continued to contribute about 30 to 40% of the uncertainties of formaldehyde after 8 p.m., possibly an indication of the chemically-originated formaldehyde that persisted from the afternoon.






HAWT
Var (ppb^2)
Time (hr)
ROG
NOX
kXLYHO
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
Var (ppb^2)


PASA
Var (ppb^2)
Time (hr)
ROG
NOX
kXLYHO
kPAN
kACO3NO2
kACO3NO
jHCHOR
kHONO2
jO3OSD
Var (ppb^2)

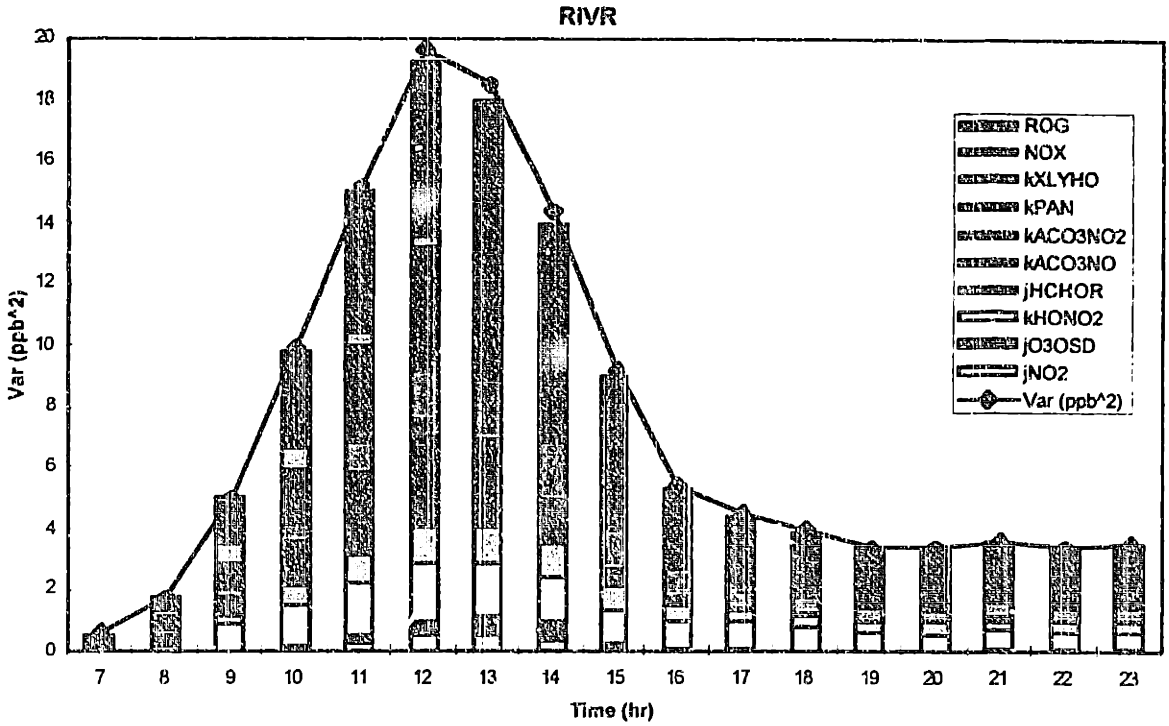


Figure 10-10. Variance Contributions from Input Parameters to Peroxyl Acetyl Nitrate (PAN) Concentrations at 5 Monitoring Stations.

PAN. PAN, a combination product of organic radicals and NO_x, seemed to be as sensitive to the uncertainties of the sources and initial conditions of the organic compounds as it was to the uncertainties of nitrogen oxides. Uncertainties of the source terms accounted for about a quarter of the uncertainties in PAN at HAWT, a third of the PAN uncertainties at CELA and PASA, and up to a half at CLAR and RIVR. The downwind sites were also the only ones where variance contribution from the source terms of ROG and NO_x were still observed after sunset. Not surprisingly, the chemical parameters whose uncertainties contributed the most to uncertainties in PAN are the formation and destruction rates of PAN, as well as the reaction rate of ACO_3 + NO, which could be viewed as a competitive reaction for the peroxyl radicals. The variance contributions of these uncertain rate constants were quite dominating at HAWT, the upwind site. Furthermore, the reaction rates of Reactions 10.2 to 10.4 accounted for most of the residual uncertainties in PAN after sunset. In addition to the ACO_3-NO_x-PAN reaction rates, those for the photolysis of formaldehyde and for the formation of nitric acid (HO + NO_2) also contributed to the uncertainties of PAN at upwind and downwind sites, probably via their effects on the radical pool. The variance contributions from these uncertain rates were smaller than the contributions of the sources and the ACO_3-NO_x-PAN reactions. At HAWT, the variance contributions died down after mid-afternoon. However, at downwind sites, the contributions from uncertain j(HCHO) and k(HO+NO_2) persisted into the night time PAN concentration variance. The lack of contribution from jNO_2 to PAN uncertainties at most sites was somewhat interesting. This observation will be discussed further in the next section.

10.6 Discussion and Conclusion

10.6.1 Ozone Uncertainties

A simple picture can be used to explain the variance contributions of the parameters to the uncertainties in the predictions of ozone. Figure 10-11 shows a simplified view of the transformation of atmospheric NO_x into ozone. The parameters identified to have significant effects on the uncertainties in ozone are related to this chain of events. Since uncertainty contributions are also related to model sensitivities to the parameters, variance contribution information may reflect the bottleneck or the rate-determining step in this chain of reactions at any given time and location. This view provides significant opportunities to improve model performances by focusing on a small number of uncertain input parameters.

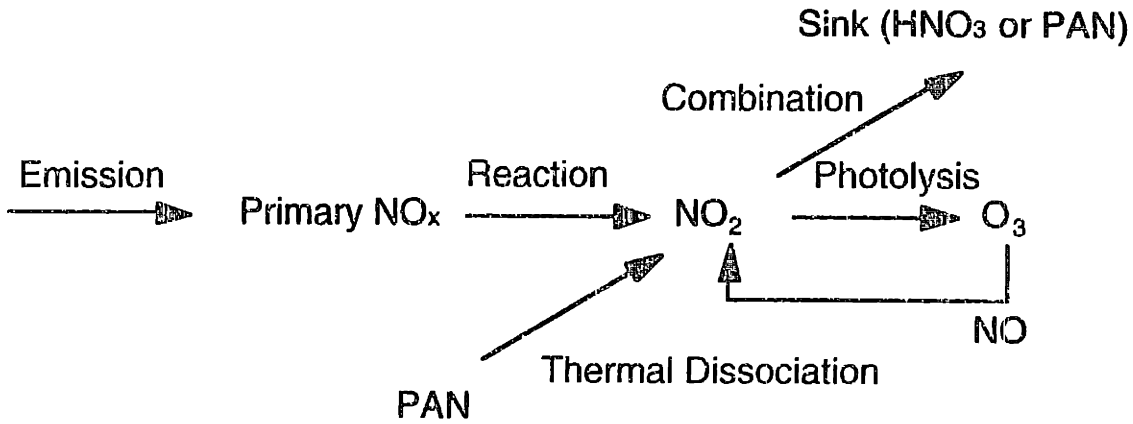


Figure 10-11. Steps for Ozone Formation.

At all five sites analyzed, the PCE coefficients of the NO_x term in the uncertain ozone representation were negative. Within a 20% uncertainty in the source strength of NO_x, increasing NO served to reduce ozone concentration at all times. Therefore, it was concluded that the amount of NO_x emitted into the system contributed to the uncertainties in ozone through the removal of ozone, either directly or indirectly. In the ranges of uncertainties considered, the production of ozone in this system seemed to be controlled by reaction kinetics rather than by the availability of NO_x.

At the upwind (clean) site of HAWT, the uncertainties in ozone formation seemed to be highly sensitive to the availability of organic radicals and the transformation of NO to NO_2 by these radicals. This interpretation was supported by the large variance contribution to ozone from the kinetic parameters of the formation and especially the dissociation of PAN, and the reaction rate of ACO_3 + NO throughout the day. These parameters controlled the availability of NO_2, which was probably the limiting reagent in the sequence of reactions. PAN concentration reached a peak in the late morning at HAWT, consistent with the postulate that it might act as a NO_2 reservoir. By mid afternoon, the variance contribution from jNO_2 increased, indicating that the photolysis reaction became the a bottleneck in the formation of ozone. HAWT was the only site where the uncertainty in the nitrogen dioxide

photolysis rate contributed up to 30% of the uncertainties in ozone predictions. Since there were not much emissions at the upwind site, the influence of uncertain source strengths of VOC and NO_x was comparatively small at HAWT than at the other sites.

At the urban sites, CELA and PASA, the concentrations of ozone was sensitive to the emissions of NO_x throughout the morning and early afternoon. During this time, the NO_2 concentrations are high and steady production of ozone was maintained at urban sites. Ozone concentration seemed to be sensitive to changes in the removal processes, most notably to uncertainties in the source strength of NO which titrated away available ozone. Uncertainties in the formation rates of nitric acid, which removed NO_x and hydroxyl radicals, also contributed to uncertainties in ozone in the early afternoon. The coincidental variance contributions of NO source and $k(HO+NO_2)$ indicated that the sensitivity of the urban ozone system was towards processes that reduced the general reactivity of the system during the day. Towards the late afternoon, photolysis became the limiting reaction and the uncertainty in the photolysis rate of NO_2 contributed more to the uncertainties in ozone formation. The reaction rates governing the formation ($ACO_3 + NO_2$ vs. $ACO_3 + NO$) and dissociation of PAN combined to contribute about 25% during the middle afternoon, coincidental with the maximum concentration of PAN and the minimum concentration of NO_2 at these locations.

The variance contributions of the parameters to ozone uncertainty behaved quite differently in the downwind sites of CLAR and RIVR. In the morning, the uncertainties in ozone formation were controlled by the uncertainties in the rate of dissociation of PAN and the subsequent reaction of ACO_3 with NO. This indicated the importance of PAN as a source of NO_2 early in the morning at these downwind sites, and was consistent with the high PAN concentrations observed at these sites in the morning. Late in the afternoon, the availability of NO_2 was also sensitive to the rate of formation of nitric acid. The relative variance contributions of jNO_2 to ozone concentrations at these sites were small compared to the upwind and urban sites, indicating that the photolysis was not a rate determining step to the formation of ozone compared to the availability of NO_2. Emissions of VOC and NO_x also combined to account for over 50% of the observed ozone uncertainty in the afternoon at downwind sites. Since local emissions were probably limited at these downwind sites, the sensitivity to VOC and NO_x emissions and to the rate of formation of HNO_3 may indicate the effects of the transport of primary pollutants to these locations in the late afternoon and early evenings.

The contribution of IC / source of NO_x (and ROG, to a lesser extent) to uncertain ozone concentrations at both urban and downwind sites illustrated the importance of good emission (source) data in improving the predictions of air quality models. Even though the uncertain NO_x sources always played a more important role than ROG in the variance contribution to ozone formation, uncertainties in ROG still contributed to 20% of the uncertainties in ozone at some sites. Furthermore, whenever the VC of ROG source was high, the variance contribution from j(HCHO → radicals) also become non-negligible. The combination of variance from these two parameters possibly indicated that ozone was indeed sensitive to radical productions at specific times and locations (afternoon at downwind sites or later afternoon at urban sites), even though the effects of the uncertain rates of the process depicted in Figure 10-11 were dominating most of the time.

PAN acted as a temporary sink for both NO_2 and radicals at the upwind sites and as a source for these compounds downwind. The reaction rates of $ACO_3 + NO$, and the PAN formation and dissociation reactions were found to have significant variance contributions to the uncertainties of ozone predictions. It is critical to characterize these reaction rates better so as to understand when PAN acts as a source of radicals and NO_2 and when it acts as a sink for these species. The uncertain reaction rate of $HO + NO_2$ was important for explaining both

the urban and downwind ozone uncertainties. At the urban sites, HO + NO_2 acted as a competitor for the photolysis of NO_2. At the downwind sites, the upwind formation of nitric acid had a direct effect on the NO_x available downwind in the afternoon, when NO_x was probably transported rather than emitted.

10.6.2 HCHO Uncertainties

Although many parameters contributed small amounts to the uncertainties of formaldehyde, the uncertain predictions of HCHO were dominated by a relatively small number of uncertain input parameters. Reducing uncertainties in sources and emissions of ROG would be particularly important for improving the model estimations of formaldehyde, particularly if secondary production of formaldehyde was to be estimated to elucidate the status of organic reactions. The dissociation of PAN followed by the reaction ACO_3 + NO was a very important source of formaldehyde in the system in the morning, especially at HAWT, CLAR, and RIVR. The high variance contributions of the PAN and ACO_3 + NO reaction rates also indicated the importance of peroxyl acyl radicals in jump-starting photochemistry at many locations. Towards the late morning, as the PAN concentration dropped, the variance contribution from the reaction rate of Reaction 10.3, ACO3 + NO_2, increased, indicating the sensitivity to the formation of PAN which was a competing reaction for the organic radicals, precursors of PAN.

The variance contributions from the uncertainty of the photolysis rate of HCHO ($\rightarrow H_2$ + CO) reached a maximum in the middle of the afternoon. The maximum contribution occurred slightly earlier in CELA, HAWT, and PASA (1 p.m.) than at RIVR and CLAR (4 p.m.). This difference in timing may reflect that formaldehyde was being transported downwind. It was rather interesting that of the formaldehyde removal processes, the molecular photolysis route contributed more to the uncertainties of formaldehyde than the radical route of photolysis and the HO reaction. The HO reaction rate, which was not considered in this study, should remove HCHO roughly three orders of magnitude faster than the photolysis routes. With an uncertainty factor only slightly smaller (1.26) than the photolysis reactions (UF = 1.4), the variance contribution of k(HCHO+HO) to formaldehyde concentration was small compared to j(HCHO→M) in the one dimensional study, and was roughly the same as the j(HCHO→R). (HCHO + hv → R is a factor of 2 or so slower than HCHO + hv → M.) Therefore, the reaction rates did not explain the variance contribution of these parameters to the concentration of formaldehyde, nor did the uncertainty factors. One possible explanation was related to the products from these reactions. Both the radical route of photolysis and HO + HCHO form HO_2, which would ultimately form HO and would activate more organic compounds, leading to further formation of HCHO. However, HCHO + hv → M removed HCHO from the reaction system without forming radicals, thereby reducing the reactivity of the system and reducing the opportunity to form HCHO. This may explain why j(HCHO→M) contributed more to the uncertain predictions of formaldehyde than the other HCHO removal processes.

The reaction rate of HO + NO_2 also contributed uncertainties to the concentration of formaldehyde, although there was no obvious relationship between formaldehyde and nitric acid. The maximum contribution of k(HO+NO_2) occurred at 9 a.m. at HAWT, 11 a.m. at CELA, 1 p.m. at PASA, 4 p.m. at CLAR, and 5 p.m. at RIVR, and were associated with VC from NO_x. Since the reaction HO + NO_2 irreversibly tied up HO and NO_2 in an air parcel, uncertainties in the rate constant reflected uncertainties of the reactivity of the air parcel as it traveled downwind. This was consistent with the observation for ozone uncertainties. The relative magnitude of its variance contribution indicated that at the urban sites, the organic

chemistry leading to the formation of formaldehyde (and ozone) was quite sensitive to the combination reaction of $HO+NO_2$.

The fact that VC from kinetic parameters continued to contribute to the uncertainties of formaldehyde concentrations in the evening at downwind sites seemed to confirm the persistence of formaldehyde that were formed photochemically during the day.

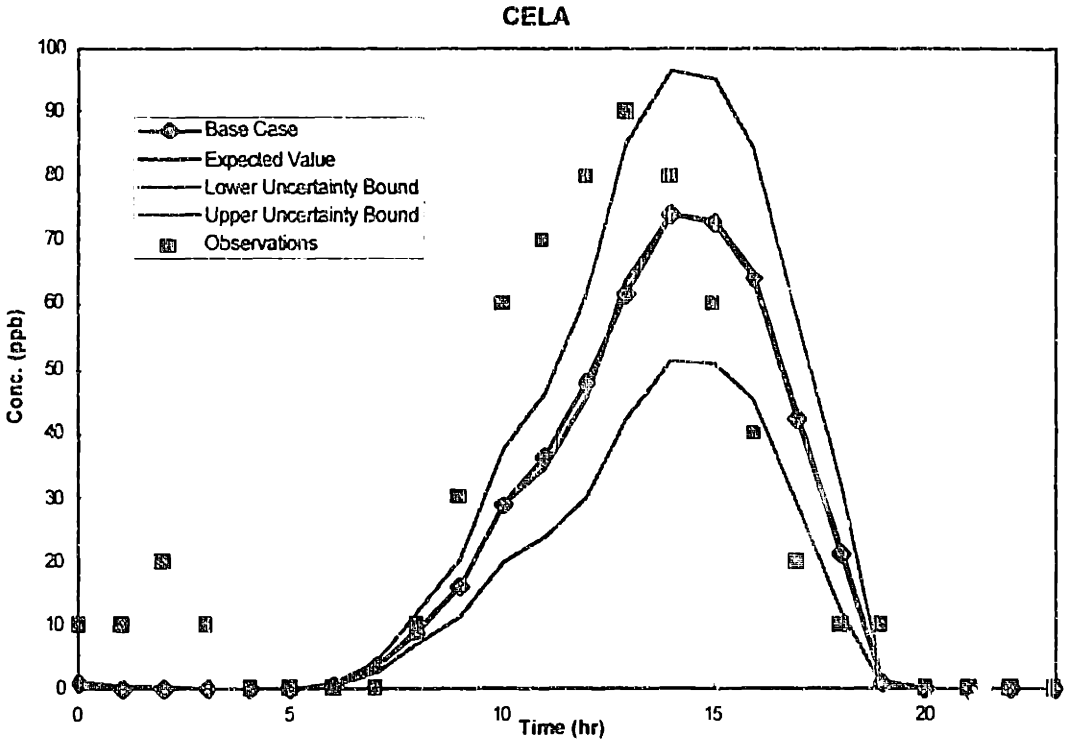
10.6.3 PAN Uncertainties

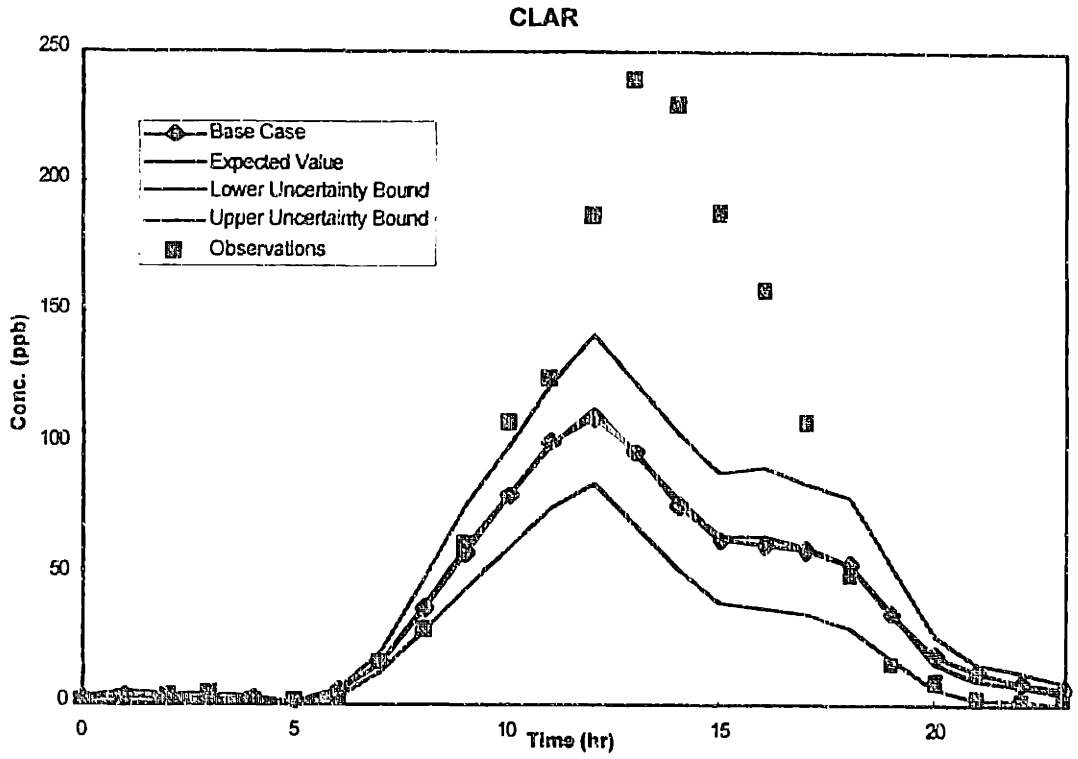
Uncertainties in the predictions of PAN were quite high compared to most of the other parameters. In the upwind site of HAWT, uncertainties in chemical parameters dominated the uncertainties of PAN. In particular, uncertainties in the reaction rates of $ACO_3 + NO_2 \rightarrow$ PAN, its competition, ACO_3 + NO, and PAN's thermal dissociation explained much of the uncertainties in PAN concentration throughout the day. From morning till early afternoon, a variety of other parameters contributed to up to 30% of the uncertainties of PAN. These parameters controlled the availability of PAN precursors, acyl radicals and NO_2, and included IC / source of ROG and NO_x, as well as $k(HO+NO_2)$ and $j(HCHO \rightarrow R)$. At the urban sites, CELA and PASA, the relative amounts of uncertainties accounted for by the PAN-related reactions (Reactions 10.2 to 10.4) were considerably smaller. Up to 60% of the uncertainties in PAN was attributed to the source factors and kinetic parameters that controlled the reactivity of the air parcel. At downwind sites, uncertainties in the sources could account for up to 40% of the uncertainties in PAN, while the uncertainties in $k(HO+NO_2)$ and $j(HCHO \rightarrow R)$ contributed to another 30% of PAN uncertainties in the afternoon.

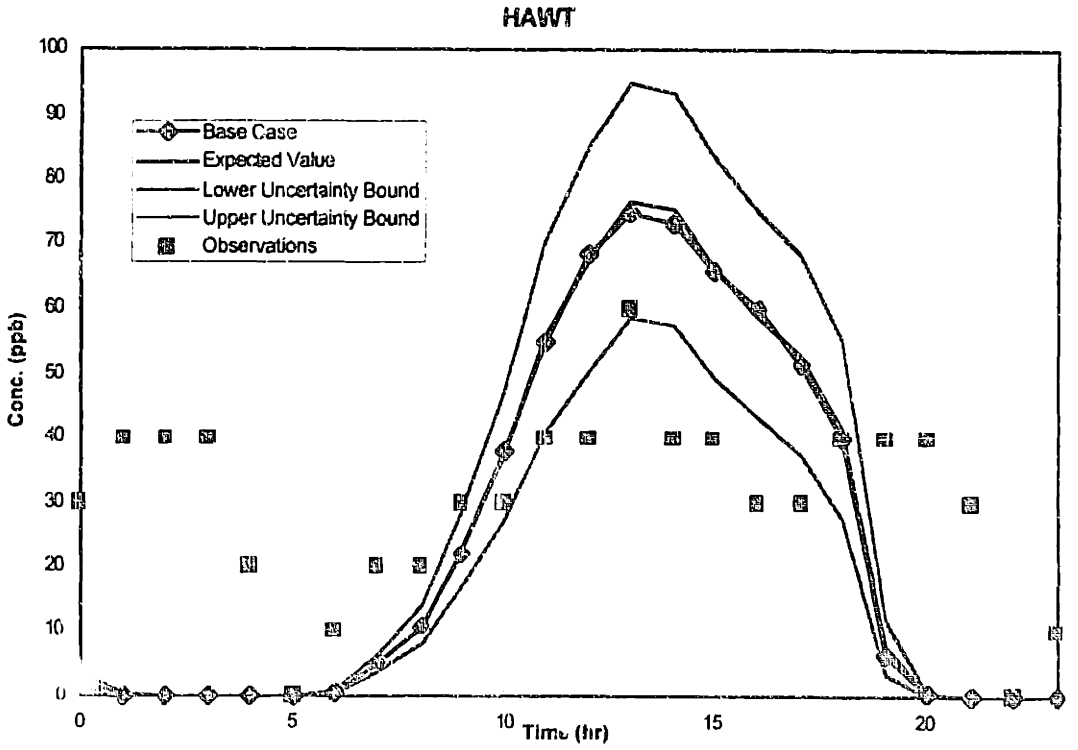
Uncertain parameters that contributed to uncertainties in the concentration of the PAN precursors NO_2 and ACO_3 were expected to contribute to the uncertainties of PAN. Since NO_2 was usually in photostationary steady state with NO, parameters that affect total NO_x contributed to the uncertainties of NO_2. The most important VCI for NO_2 was the IC / source of NO_x. The variance contributions of other parameters, including $k(HO + NO_2)$, $k(PAN \rightarrow ACO_3 + NO_2)$, and $k(ACO_3 + NO_2)$ were small, except early in the morning at RIVR. Note that NO_2 was not sensitive to jNO_2 during the day, but the uncertainties in both HCHO and ozone contained significant contributions from jNO_2 in the afternoon. On the other hand, the concentrations of the peroxyl acyl radical ACO_3 were quite uncertain. In addition to NO_x, the initial conditions and source of ROG, the reaction rates of ACO_3 + NO, $HO + NO_2$, and the photolysis rate $j(HCHO \rightarrow R)$ were also important variance-contributing inputs. The variance contributions of these parameters to PAN concentration uncertainties, as well as the coincidental nature of the peaks of PAN and ACO_3, seemed to provide circumstantial evidence that the uncertainties in PAN concentration were correlated to the uncertainties of ACO_3 rather than to uncertainties in NO_2 concentrations.

10.6.4 Do uncertainties in chemical kinetics explain the discrepancy between predictions and measurements?

The answer is no. At most stations, a few of the non-peak observations fell within the uncertainty bounds predicted by the airshed model taking into account input uncertainties in the kinetic parameters. However, it is quite obvious from the following figures that the model is still unable to predict some key behavior of the physical system.




CLAR
Base Case
Expected Value
Lower Uncertainty Bound
Upper Uncertainty Bound
Observations
Conc. (ppb)
Time (hr)
0
50
100
150
200
250
0
5
10
15
20


HAWT
Base Case
Expected Value
Lower Uncertainty Bound
Upper Uncertainty Bound
Observations
Conc. (ppb)
Time (hr)
0
10
20
30
40
50
60
70
80
90
100
0
5
10
15
20

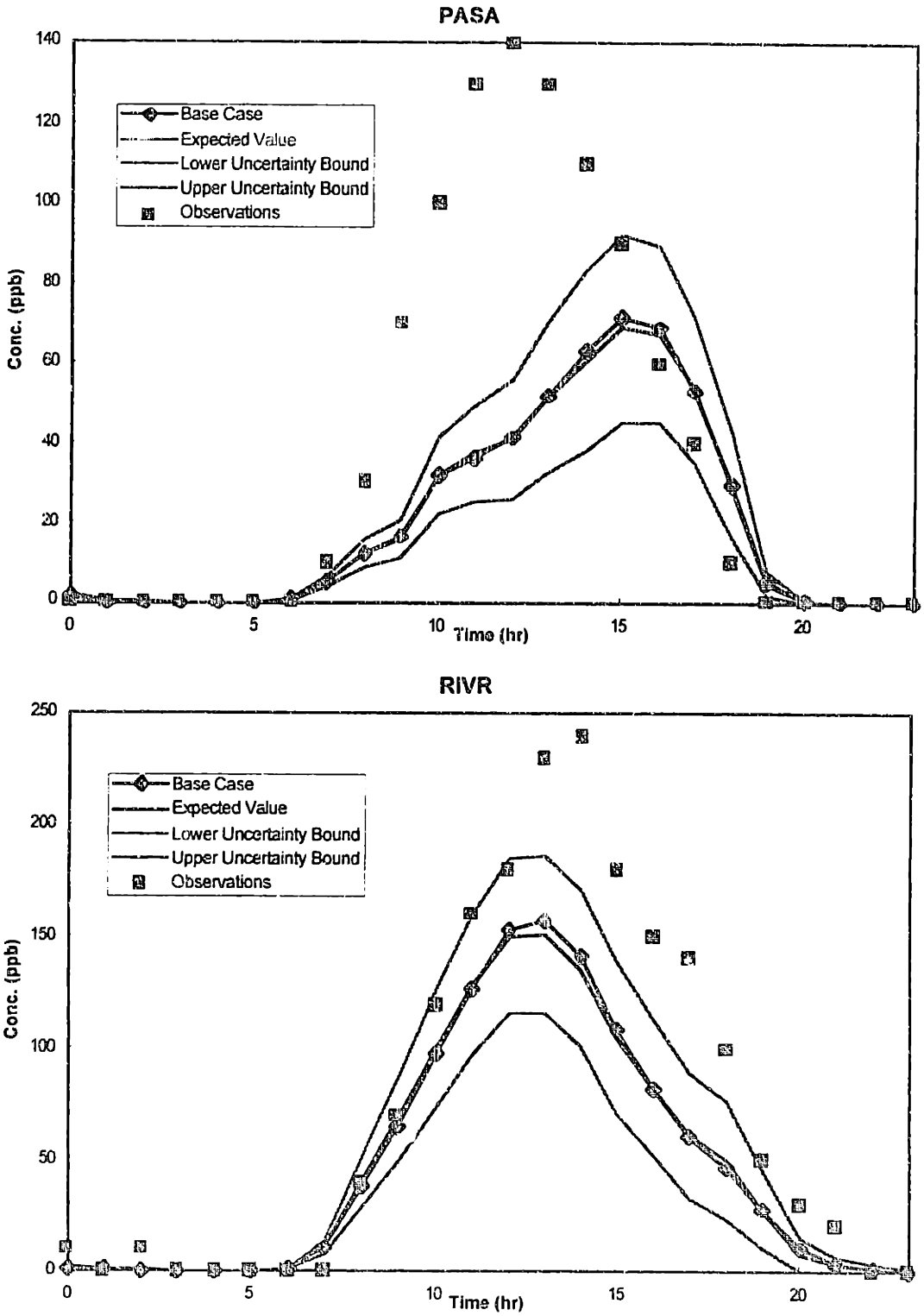


Figure 10-12. Comparison of Uncertain Ozone Predictions with Observations at 5 Stations.

At the upwind site HAWT, the data indicated less daily variation in ozone concentration than the model predicted. The uncertain ozone predictions followed more or less the same trend as the nominal simulation and did not predict the correct same time series behavior. At

both urban sites, the predicted ozone peak occurred later than the measurements indicated. The peak concentrations were also underpredicted. Uncertain parameters in the chemical mechanisms were propagated to provide a range of possible predictions. At PASA, only three hourly readings fell within that range during the day. Although the magnitude of the uncertain predictions agreed better with the experimental values at CELA, the time at which the peak level of ozone occurred was missed by the uncertain predictions. The same was true at the downwind sites of CLAR and RIVR. Uncertainties in the chemical kinetics parameters alone could not account for the lack of agreement between the measurements and the CIT / AIRSHED model predictions.

Other sources of uncertainty probably play a significant role in the uncertain predictions of the airshed model. Parametric uncertainties in other modules such as meteorology and emissions and in the input data should be investigated before the model can be rejected as wrong. One possible explanation for the discrepancy is that the time profiles of VOC or NO_x emissions used in the model are different from actual behavior. This was not considered since the entire emissions field was adjusted by an uncertain factor in the uncertainty studies. Other model assumptions should also be tested by comparing alternatively formulated models to find out if the model predictions are sensitive to structural uncertainties.

10.6.5 Conclusions

Ozone predictions from three dimensional models are typically uncertain by 25 - 30% due to uncertain parameters in the chemistry submodules and a 20% uncertain factor applied to all emissions and initial conditions. The predicted concentrations of other species may be more or less uncertain than ozone. For example, the uncertainties in the predictions of formaldehyde and NO_2 are relatively small, but those in PAN are very significant fractions of the expected value.

A 20% uncertainty in the source strength of NO_x may be responsible for up to 40% of the variance in the uncertain ozone predictions in urban and downwind locations. The contributions of uncertain emissions of organics to uncertainties in ozone predictions are usually smaller than those of nitrogen oxides. However, uncertain organics represent a non-neglible source of uncertainties in air quality model predictions, especially to important organic compounds like formaldehyde. It is critically important to obtain better estimates of the emissions source strength of VOC and NO_x to improve the model estimations of key photochemical pollutants.

Of the chemical parameters, the reaction rates of ACO_3 + NO, and the PAN formation and dissociation reactions have significant variance contributions to quite a few species. The uncertain reaction rate for the formation of nitric acid also affects predictions of the key species, especially at the downwind sites. Furthermore, the photolysis rate of nitrogen dioxide contributes to variance of the predictions of ozone and other species. Better measurements of the thermal kinetic parameters should help elucidate the roles of the sink species PAN and nitric acid in the urban atmosphere. Field measurements of actinic flux will help minimize the uncertainties in the NO_2 photolysis rate.

DEMM was applied successfully to characterize the uncertain predictions in a three dimensional air quality model due to parametric uncertainties in reaction rate constants and source strengths. The incorporation of uncertainty in the model estimates did not improve the agreement between model predictions and field measurements, indicating other short comings in the model or input data.

11. Isoprene Study Using the RADM Mechanism

11.1 Introduction

The importance of isoprene as a precursor of photochemical smog in areas with significant biogenic emissions has been generally accepted since the publication of the "Rethinking" report (NRC, 1991). Though many studies have been undertaken since that time to elucidate the reactions of isoprene in urban air (Aschmann and Atkinson, 1994; Grosjean *et al.*, 1993; Paulson *et al.*, 1992a, b; Tuazon and Atkinson, 1989, 1990a, b), the chemical pathways are still not fully characterized and stoichiometric coefficients of the identified products remain uncertain. With incomplete information, parameterized mechanisms have been proposed to represent these reactions. Most the major mechanisms currently used in air quality models include some treatment of isoprene, usually modeling their reactions analogously to another alkene. (Isoprene is modeled after propene in RADM, and is modeled after butene in SAPRC.) In recent years, a number of researchers have attempted to model isoprene reactions in a more realistic manner by supplementing existing atmospheric mechanisms (Paulson and Seinfeld, 1992; Zimmerman and Poppe, 1996) with reactions of isoprene (ISO) and the carbonyl products of its decomposition. These supplements usually include additional model species representing isoprene's decomposition products methyl vinyl ketone (MVK) and methacrolein (MAC) as well as special peroxyl radicals derived from them. Supplementary reactions include the breakdown of isoprene and carbonyl products by radicals (HO, NO_3, $O(^3P)$) and ozone, and the reactions of the biogenic peroxyl radicals with NO_x and with other radicals. Due to the reactivity of isoprene and the uncertain nature of the new reactions, it is of great interest to estimate the uncertainties in the atmospheric model predictions caused by including these explicit yet uncertain isoprene reactions.

11.2 Goal of the Study

In this study, the isoprene mechanism was probed using the RADM mechanism and Zimmerman's supplement that described isoprene reactions (Zimmerman and Poppe, 1996). A series of uncertainty analyses was carried out to:

- characterize the uncertainties of the mechanism predictions due to uncertainties in the isoprene reactions,
- test the effect of uncertainties in the isoprene inventory,
- determine the uncertain parameters of isoprene reactions that drive uncertainties in the model predictions, and
- compare the effects of uncertainties from isoprene reactions against those from reactions in the parent mechanism.

11.3 Method of Approach

A description of Stockwell's RADM mechanism can be found in Chapter 9. The original mechanism was compared to a version that contained Zimmerman's isoprene supplement (Zimmerman and Poppe, 1996). Zimmerman's supplement contained 16 additional reactions to describe the transformations of ISO, MAC, MVK and their radicals. This set of reactions was condensed from a list of 153 explicit isoprene reactions of 39 isoprene-related species.

Base Case. The effects of uncertainties in isoprene composition and isoprene reaction parameters obviously depend on the amount of isoprene present in the inventory and the ROG/NO_x ratio. The all city averaged base case scenario (see Appendix) had only a minute amount of isoprene (0.0034 ppbC/ppbC organics, or 0.00068 ppb / 0.24 ppb organics), and so isoprene played a tiny role in the overall dynamics of the reaction system. A revised base case, with a higher mole fraction of isoprene, was used so that the effects of uncertainties in the isoprene mechanism could be detected in the model output. The initial conditions of the revised base case are listed in Table 11-1.

The initial conditions used in this study differed from previous uncertainty analyses only in the isoprene fraction of organics, which was raised to 10%. The total amount of organic compounds and nitrogen oxides were kept at the same level, which simulated the maximum incremental reactivity (MIR) condition (Carter, 1991). Figure 11-1 shows the base case dynamics of ozone, NO, and NO_2 in a box model simulation. The organic mix with 10% isoprene resulted in a 30% increase in ozone production in a 10-hour simulation relative to the all-city organic composition (see SAPRC mechanism base case in the Appendix). If isoprene was represented explicitly using the reactions of propene instead of being lumped into internal alkenes, slightly more ozone was produced per NO_x used in 10 hours. Zimmerman's explicit representation of isoprene reactions slowered the dynamics of the reaction in the first hour of simulation, but predicted the highest ozone production in 10 hours. The maximum difference in ozone predictions was about 5% between the different representations of isoprene reactions.

Species	Initial Concentration (molec/cm^3)	(ppb)
ETH	2.42e11	9.84
HC3	4.06e11	16.5
HC5	3.68e11	15.0
HC8	2.26e11	9.19
OL2	2.24e11	9.11
OLT	1.72e11	6.99
OLI	7.91e10	3.22
TOL	1.47e11	5.98
XYL	1.78e11	7.24
HCHO	2.58e11	10.5
ALD	1.93e11	7.85
!SO	2.36e11	9.59

Table 11-1. Base Case Initial Conditions for Isoprene Reaction Uncertainty Analysis.

Using this base case, uncertainty analyses were carried out for the cases with uncertainties in the isoprene amount and in the mechanistic representation of isoprene reactions. The effects of input uncertainties were compared between Zimmerman's explicit representation and Stockwell's original model.

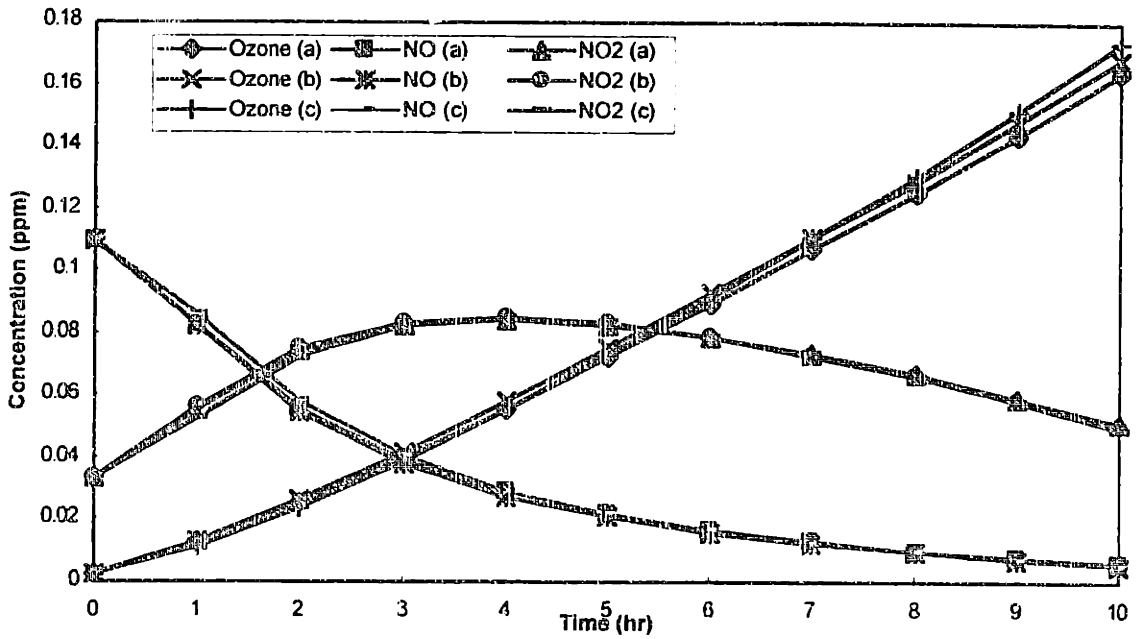


Figure 11-1. 10% Isoprene Base Case Dynamics of Ozone, NO, and NO_2.
(a) Isoprene lumped into internal alkenes, (b) isoprene represented as propene as in the original RADM mech. (c) Zimmerman's explicit isoprene reactions.

11.4 Uncertainty in Isoprene Amount

The estimation of isoprene emissions remains a difficult task even though much progress has been made in the field (e.g., Piers and Waldruf, 1991 and newer versions of the BEIS computer model). To simulate the resulting uncertainties in the predictions of ozone and other photochemical oxidant species, the initial condition of isoprene was treated as uncertain. Expert judgment was sought but not obtained as to the uncertainties of isoprene emissions, although some claimed that the estimation of isoprene was rather more like an art than a science. An arbitrary uncertainty factor of 1.2 was assumed for the lognormal distribution used to describe the uncertain initial condition of isoprene. As shown in Figure 11-2, the uncertainties in ozone concentrations derived from DEMM (collocation) was about 8-9% of the mean predictions due to a 20% uncertainty in the initial condition of isoprene for the original mechanism and for Zimmerman's representation. The uncertainties in the NO_x species were relatively small. The hydroxyl radical production dynamics were statistically different at the beginning of the simulation, considering the uncertainties associated with these predictions. At the 10th hour of the simulation, the uncertainty in the hydroxyl radical concentrations were higher on a percentage basis than the stable species such as ozone, NO, and NO_2.

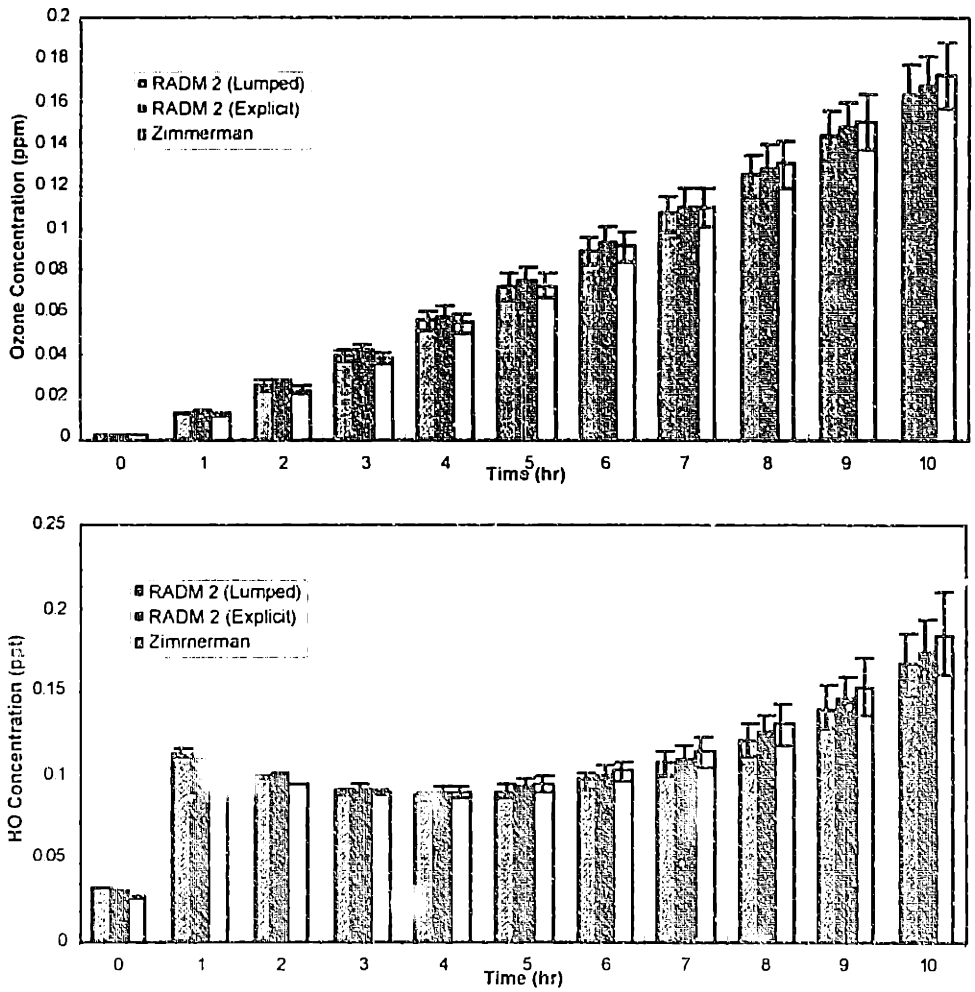


Figure 11-2. Uncertain Dynamics due to uncertain IC Isoprene (O_3, HO)

11.5 Uncertainty Analysis of Isoprene Reaction Parameters

This section describes the characterization of uncertain parameters involved in the reactions of isoprene and its intermediate products. Relevant data were compiled from recent studies listed in the introductory paragraph (Section 11.1). Uncertainties of the rate and product parameters were described by the range of reported values or a subjective measure of uncertainty factors when few measurements were taken. DEMM — Collocation was then used to propagate the isoprene-related uncertainties in the RADM mechanism and the Zimmerman supplement and their results were compared.

11.5.1 Characterizing Uncertainties in the Reaction Rates

Sixteen (16) reactions were added to the Regional Acid Deposition Model (RADM) mechanism by Zimmerman's condensed representation of isoprene reactions. In addition, three isoprene reactions in the original RADM mechanism were modified to produce special peroxyl radicals. Table 11-2 lists these reactions and the uncertainty factors of the reaction rate constants, which were described by lognormal distributions.

Rxn (in RADM)	UF		
ISO + HO	1.2		
ISO + NO3	2.5 (^2.0)		
ISO + O3	2.0 (^1.4)		
Rxn (Zim. sup.)	**UF**	**Rxn (Zim. sup.)**	**UF**
MAC + HO	1.4	MVK + NO3	2.0
MVK + HO	1.4	MAC + NO3	2.0
MAC + O3	2.0	ISOP + HO2	2.0
MVK + O3	2.0	MACP + HO2	N/A (k = ISOP + HO2)
ISOP + NO	2.0	MVKP + HO2	N/A (k = ISOP + HO2)
MACP + NO	N/A (k = ISOP + NO)	ISOP + ISOP	2.0
MVKP + NO	N/A (k = ISOP + NO)	MVKP + ISOP	2.0
ISON + NO	N/A (k = ISOP + NO)	MACP + ISOP	2.0

Table 11-2. Rate Constant Uncertainty Factors for Isoprene Reactions.
(^ Stockwell's estimates from the SAPRC rate constant uncertainty characterization)

The rate constant for the ISO + HO reaction had been published in several studies before the product studies and reaction mechanisms were investigated. The errors in three measurements by Atkinson (Atkinson 1989) were small (5%). An uncertainty factor of 1.2 was assigned due to the lack of measurements by other groups. For ISO + NO_3, the rate constant uncertainty factor was assumed to be 2.5. Stockwell (1994) argued when assigning uncertainty factors to SAPRC reactions that the reaction rate constants for NO_3 reactions should have uncertainty factors of at least 2.0 due to the experimental procedure. Here, an UF of 2.5 was assigned because one out of three reported values was significantly different from the other two. (The three reported values were (1.3±0.14)e-12, (5.79± 0.68)e-13, and

(6.52±0.78)e-13) cm^3/molec/s. An uncertainty factor of 2.0 was assigned to ISO + O_3 because of disagreement of reported measurements. Even though the error reported with measured values were usually less than 20%, five literature values range from (5.8±0.8)e-18 to 1.65e-17, with a mean value of 1.3e-17 (5 values) and a model nominal value of 1.5e-17.

The reaction rates of hydroxyl radical with MAC and MVK were assigned UF's of 1.4, which is the highest UF assigned to hydroxyl reactions in the SAPRC mechanism. For the ozone reaction rate constants with these two compounds, the uncertainty factor was 2.0, assigned analogously to the ozone-isoprene reactions. All ozone reactions seemed to have similar experimental error, and there were only a few reported measurements for MAC and MVK. The rate constants of MAC + NO_3 and MVK + NO_3 were assigned an UF of 2.0, according to Stockwell's estimate of uncertainty in the experimental procedure.

All peroxyl radical reactions with NO, HO_2, and other peroxyl radicals had reaction rates that were consistent with the analogous reactions in RADM. Therefore UF's were assigned to be equal to the peroxyl reactions in RADM. All peroxyl reaction rate constants in RADM had uncertainties factors of 2.0. The permutation reactions, ISOP + ISOP, ISOP + MVKP, and ISOP + MACP, were assumed to be independent.

11.5.2 Characterizing Uncertainties in the Product Coefficients

Many product coefficients were used in the description of isoprene reactions. Most peroxyl species were surrogate representations of several structurally-different radicals. On the other hand, a lot of the intermediate stable product yields were based on experimental observations. Uniform distributions were used to represent uncertainty product coefficients. As much as possible, the uncertainty range was based on the range of experimental values reported in the papers listed in the introductory section in this chapter. Several reactions were assigned no uncertain product coefficients because they represent model decisions or assumptions made in "condensing" a large number of reactions. These are listed in Table 11-3.

Reaction	Notes
ISO + HO —> ISOP	lumped representation of radicals
ISO + NO3 —> ISON	ISON represents the adduct
MVK + HO —> MVKP	lumped representation of radicals
ISOP + HO2 —> OP2	dominant pathway only
MACP + HO2 —> OP2	dominant pathway only
MVKP + HO2 —> OP2	dominant pathway only

Table 11-3. Isoprene reactions without uncertain product coefficients.

An example of uncertain product coefficients for the reaction ISO + O_3 is given in Table 11-4. The ranges were compiled from Paulson *et al.* (1992b), Grosjean *et al.* (1993), and Aschmann and Atkinson (1994). Both Paulson *et al.* and Grosjean *et al.* listed other products that were not included in the list of reaction products. No attempt was made to characterize the uncertainty due to product omission.

Product with Uncertain Coefficients	Range of Uniform Distribution (Nominal Value)
MAC	0.39 - 0.67 (0.55)
MVK	0.15 - 0.26 (0.21)
O(3P)	0.25 - 0.65 (0.45)
HO	0.18 - 0.83 (0.5)
HCHO	0.8 - 0.96 (0.89)
HO2	0.06 - 0.16 (0.11)
CO	0.17 - 0.19 (0.18)
OLT	0.0 - 0.07 (0.07)
ORA1	0.0 - 0.12 (0.07)
ORA2	0.0 - 0.15 (0.15)
MO2	0.07 - 0.11 (0.09)
GLY	0.0 - 0.03 (0.01)

Table 11-4. Uncertain product coefficients in ISO + O_3 reaction.

The other uncertain coefficients can be found in the RADM appendix where the Demmucom input script for the Zimmerman reactions were listed.

11.5.3 Uncertainty Propagation of Isoprene Parameters

There were 14 uncertain rate constants and 61 uncertain product coefficients in the reactions of isoprene, methy-vinyl-ketone, methacrolein, and radicals derived from these compounds. In addition, a 20% uncertainty was assumed (normal distribution) for the initial condition of isoprene. An uncertainty analysis was carried using DEMM – collocation with the iterative procedure described previously (in Chapter 8). The number of parameters was successively reduced as the order of polynomial chaos expansion approximation of the response surface was increased. A third order approximation complete with all cross product terms was used for the approximation of the output parameters. The maximum relative error of the response was less than 1%. Ultimately, 21 parameters were identified to contribute significantly to uncertainties in the predictions of ozone and NO_x.

Figure 11-3 depicts the expected dynamics of the RADM-Zimmerman mechanism under the influence of input uncertainties. The mean value of ozone was elevated over the nominal predictions by about 10%. The standard deviations associated with the predictions of ozone and NO_2 were highest at the end of the simulation (see Figure 11-3, errorbars = standard deviation). The maximum standard deviation for the ozone concentration was about 17%, or 30 ppb. NO and NO_2 exhibited much smaller uncertainties as a result of the uncertainties in isoprene parameters. Hydroxyl radical concentrations, on the other hand, were strongly affected by isoprene reactions, with standard deviations of as much as 30% associated with the predictions. These results were consistent with the postulation that uncertainties in isoprene parameters contributed to uncertainties in ozone through their effects on radical concentrations.

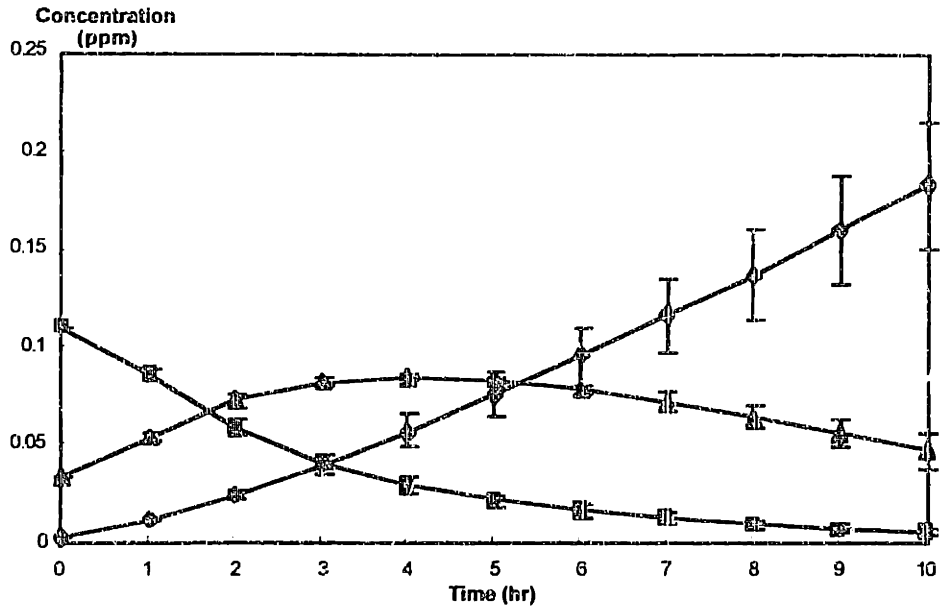


Figure 11-3. Dynamics of RADM-Zimmerman Mechanism with Uncertain Isoprene Parameters

Of the uncertain parameters considered in this experiment, the initial amount of isoprene and the product coefficients of the ISOP + NO reaction had the largest effect on the variance in ozone. Figure 11-4 shows that the variance contribution of the initial condition of isoprene is 20 - 35% of the ozone variance throughout the simulation. At the start of the simulation, the uncertainty the yield of NO_2 from ISOP + NO dominated the uncertainty in ozone. Since the yield of NO_2 was assumed to correlate to the yield of HO_2 on a one-to-one basis, the high variance contribution of the product coefficient could be attributable to either NO_2 or radical production. Towards the latter part of the simulation, ozone became sensitive to the stoichiometric coefficients of stable products such as MAC, MVK, and HCHO. The same set of parameters contributed to uncertainties in the concentrations of NO, NO_2, and HO radical.

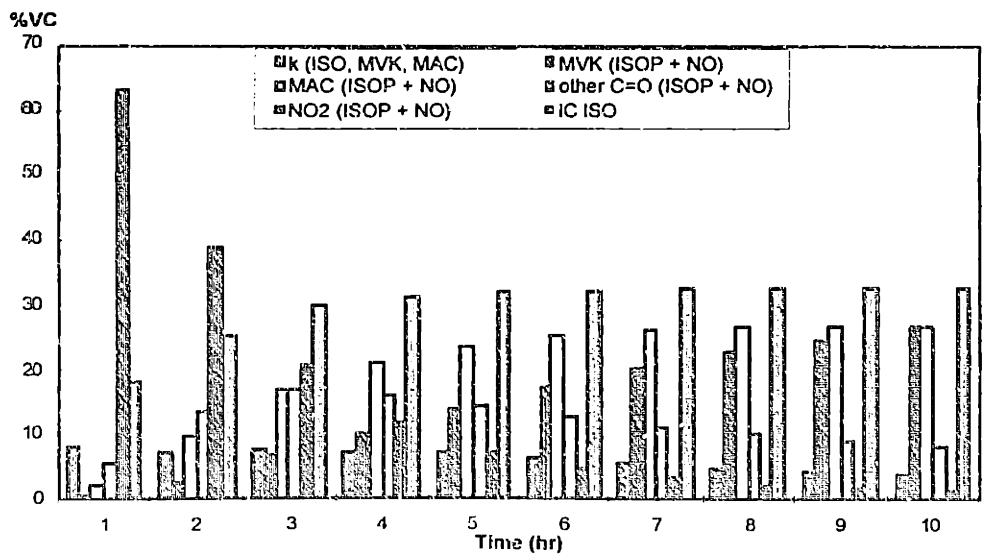


Figure 11-4. Variance Contribution to O_3 from Uncertain Isoprene Parameters.

11.5.4 Comparison between Uncertainty Results of RADM (Original) and Zimmerman Supplement

In the original RADM mechanism, isoprene was represented using the reaction of propene. This assumption was not tested in the uncertainty study, even though the different number of carbons in the two compounds made this representation less satisfactory. Only the reaction rates and product parameters were treated as uncertain to characterize the resulting uncertainties in the concentration predictions. The rate constant uncertainties for isoprene reactions were as determined in Section 11.5.1. Since ISO is modeled after propene, a terminal alkene, the reactions ISO + HO —> OLTP and ISO + NO_3 —> OLN did not contain any uncertain product coefficients. Therefore, only a few product coefficients in the ISO + O_3 reaction were uncertain. One way to characterize the uncertainties of the product coefficient uncertainties was to assign uncertainties that were analogous to the propene (or OLT) reaction, which were determined based on US averaged emissions. In this study, another approach was taken. Uncertainty in the ISO + O_3 reaction was related to the that determined for Zimmerman's with isoprene-specific products. The uncertainty assignments are illustrated in Table 11-5.

Product (Zim)	Scaled* Uncertainty Range (nominal)	Product (RADM)	Uncertainty Range (Nominal)	Note
HCHO	0.61 - 0.73 (0.68)	HCHO	0.47 - 0.59 (0.53)	same amt.
MAC + MVK	0.30 - 0.68 (0.53)	ALD2	0.30 - 0.70 (0.5)	same amt.
CO	0.129 - 0.145 (0.14)	CO	0.31 - 0.35 (0.33)	same %
ORA1	0.00 - 0.06 (0.06)	ORA1	0.10 - 0.30 (0.2)	same %
ORA2	0.00 - 0.11 (0.11)	ORA2	0.10 - 0.30 (0.2)	same %
HO_2 + $O(^3P)$	0.24 - 0.62 (0.31)	HO_2	0.18 - 0.46 (0.23)	same %
MO_2	0.05 - 0.07 (0.07)	MO_2	0.16 - 0.28 (0.22)	same %
HO	0.14 - 0.63 (0.34)	HO	0.04 - 0.19 (0.1)	same %
OLT	(0.05)	(none)	–	–

Table 11-5. Uncertainty Assignment for ISO + O_3 Reaction.
(* A scaled range was used because the OLT species has only 3.8 carbons, ISO has 5)

An parametric uncertainty analysis of the isoprene reactions in the RADM mechanism involved three rate constants, 8 product coefficients, and the initial condition of isoprene. The expected dynamics are shown in Figure 11-5. The expected concentrations of key species did not deviate significantly from the nominal predictions. Compare to the Zimmerman case described in the previous section, the uncertainties associated with the concentrations were small.

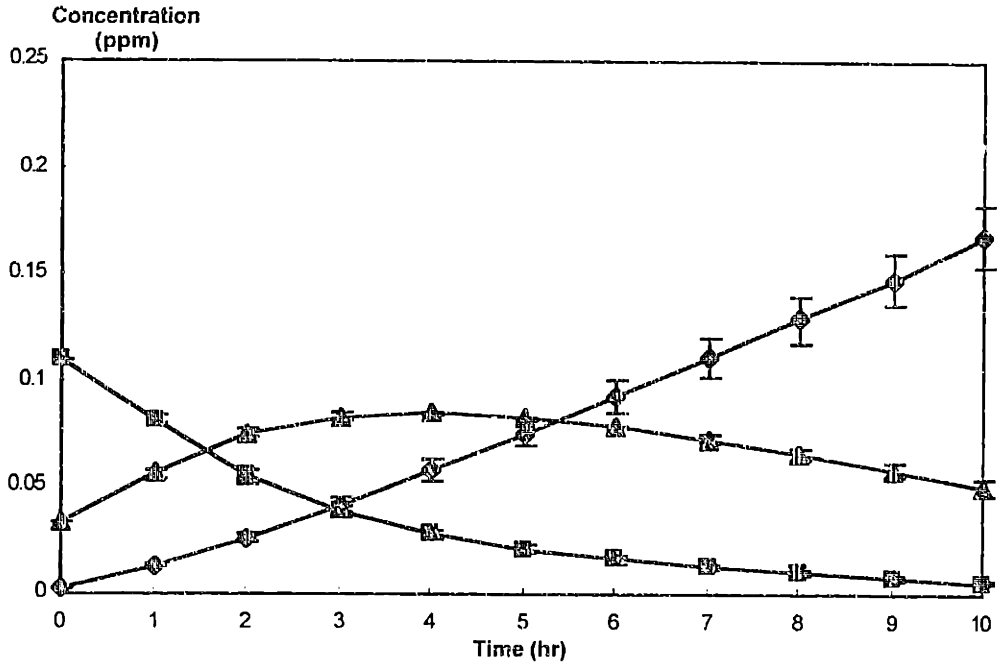


Figure 11-5. Dynamics of Stockwell's RADM Mechanism under Isoprene Uncertainties.

A variance analysis shows the uncertainties in all the key species in the mechanism were dominated by the uncertain initial condition of isoprene. The reaction rates of ISO with HO and with O_3 contributed to uncertainties during the first two hours of the simulation. None of the uncertain product parameters accounted for more than a few percent of the uncertainties in concentrations. This result was consistent with the uncertainty behavior in the Zimmerman study, which showed that the uncertainties in the product coefficients of MAC, MVK, and other secondary carbonyl compounds contributed to the uncertainties in ozone predictions. These products were not represented in the original RADM mechanism.

11.6 Combination of Isoprene Uncertainties with Other Parameters

The 22 reaction parameters identified to contribute to the variance in concentrations (from Section 11.5.3) were studied together with other variance-contributing parameters identified in a previous RADM study (see Chapter 9). The additional uncertainty parameters included photolysis rates of NO_2, HCHO, O_3, reaction rates of HO + NO_2, O_3 + NO, XYL + HO, parameters from the OLTP + NO reaction, and initial conditions of NO_x and ROG, etc. For a complete list, please refer to the Demmucom input file included in the appendix. The increase in the amount of isoprene in the organic mix resulted in significant increase in the sensitivity of the system to uncertainties in all parameters. The uncertainties in 46 input parameters resulted in a significant increase of ozone predictions (e.g., at the 10th hour, mean ozone = 227 ppb, nominal ozone = 122 ppb, a 37.5% elevation) over the nominal values. The maximum uncertainty associated with the ozone predictions was more than 50%, while that associated with the hydroxyl concentration was as much as 60%. Due to the increased variance of the responses, a 5th order PCE was required in the DEMM for an acceptable level of error.

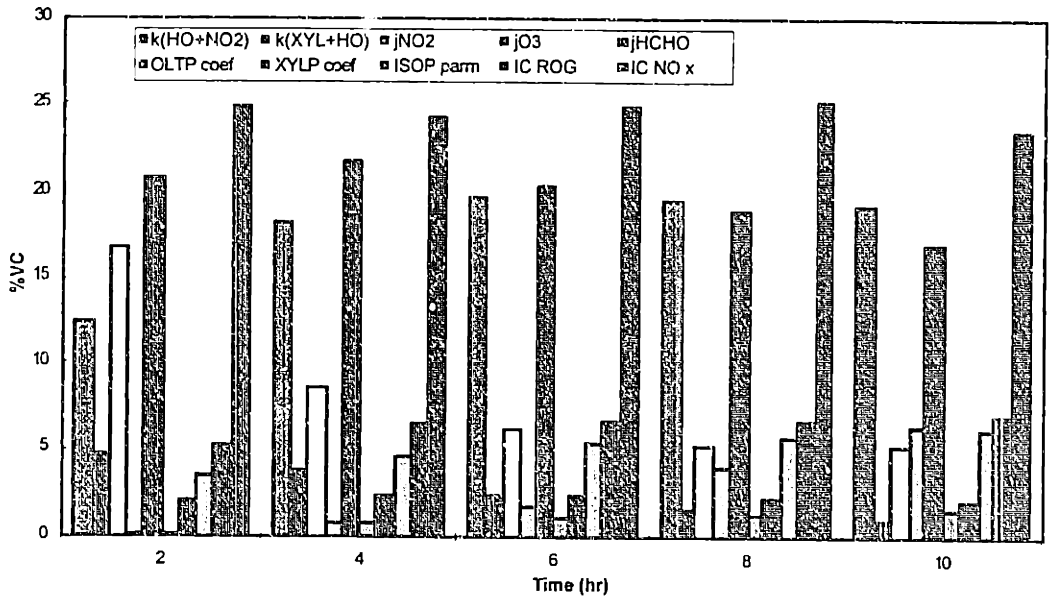


Figure 11-6. VC of Ozone (Isoprene + Other parameters).

Variance contributions to ozone concentration are plotted in Figure 11-6 as a function of time. Compared to the uncertainties in predictions caused by some of the inorganic parameters, the uncertainty contributions from isoprene reactions and parameters were small for ozone concentrations. Uncertainties in ozone were dominated by uncertainties in IC NO_x, jHCHO, and k(HO + NO_2). As found in previous studies, the uncertainties in jNO_2 also affected the predictions in the early part of the simulation. Isoprene parameters and the initial concentrations of ROG each contributed to only about 5-7% of the variance in ozone. In absolute terms, the variance attributed to isoprene parameters was smaller compared to the findings in Section 11.5.3 where that uncertainties in the isoprene initial conditions and parameters caused 30 ppb uncertainty in ozone concentrations when the other parameters were treated as deterministic.

11.7 Discussion

The importance of isoprene reaction rates and parameters as uncertainty contributors to the predictions of atmospheric chemical mechanism depends on the amount of isoprene present in the organic mix. In the all-city averaged mixture, isoprene comprised a very small fraction of the organics. As expected, isoprene-related effect could only be detected in uncertainty analyses when the amount of isoprene in the organic mixture was raised in the all city average scenario. Some observations drawn in this study would be uniquely applicable for this particular set of initial conditions, others might be more universal.

In the uncertainty analysis of the Zimmerman supplement to the RADM mechanism, the initial amount of isoprene, and the product coefficients of intermediate products of the ISOP + NO reaction were found to contribute to the variance of RADM's predictions. Isoprene rate parameters only contributed (to a small extent) to the uncertainties of the predictions in the first hour of the simulation. This is because isoprene was very reactive under the simulation conditions, for example, the isoprene-HO reaction has a time scale of about 1.3 hours. In the early part of the simulation, the shared product coefficient of NO_2 and HO_2 was found to be the most important of the product parameters. Isoprene contributed to a large number of NO-to-NO_2 conversions in the short time scales because of its high reactivity. The production of radicals from isoprene reactions at the beginning of the simulation was probably an important factor in the chemical dynamics of the system. Towards the end of the simulation, the stoichiometric coefficients of stable intermediate products of ISO, such as MVK and MAC, contributed significantly to the uncertainties of the predicted concentrations. These results indicated the importance of the intermediate products from isoprene as contributors to atmospheric ozone production. Therefore, there is a need to understand better the reactions of secondary products from isoprene decomposition.

The comparison between Zimmerman's isoprene mechanism and Stockwell's representation by propene reactions showed that it is important to incorporate newly available knowledge into atmospheric mechanisms so that parametric uncertainties can be elucidated and meaningful conclusions can be drawn regarding the parameters that require further studying. The lack of parametric uncertainties in the original RADM mechanisms isoprene reactions was due to the lack of details in the description of isoprene reactions. In other words, small parametric uncertainties did not imply good predictions, because of uncertainties in the model structure. In this case, the uncertainties due to the assumptions about modeling isoprene were probably a significant limitation of the RADM mechanism and the Zimmerman supplement reduced the "structural uncertainties" by allowing more information about the process to be incorporated.

The absolute amount of uncertainties in the predictions of atmospheric models associated with uncertain isoprene parameters depends on the amount of isoprene present in the organic mix as well as the relative amounts of ROG and NO_x. This study shows that at least in some urban locations, isoprene reaction uncertainties may be unimportant when compared to uncertainties in other parameters in the chemical mechanism. However, isoprene can comprise up to 50% of the organics present in the atmosphere. In such cases, the effects of isoprene emissions and reactions uncertainties on model predictions may be significant.

12. Need for New Instrumentation

12.1 Gaps of Information

Urban smog has been studied for over 20 years, and is much better understood than many problems of emerging importance such as aerosol haze, aerosol-greenhouse gas-climate interactions, and regional photochemical air pollution. Models have been built to integrate existing theories and data, and to predict the effects of source controls. Despite the research efforts invested, there are still gaps in the knowledge, as reflected in the inability to control effectively the levels of urban ozone. Chapter 3 summarized the data requirements for studying urban atmospheric chemistry. Examples of research areas within the field where fundamental understanding is still shaky are: the ground level actinic fluxes (hence, photolysis of key organic and inorganic compounds) under a variety of cloud cover conditions and the reactions of complex organics, especially aromatics and biogenic alkenes. In terms of field experiments, accurate methods are direly needed to sample secondary products such as formaldehyde, PAN, hydrogen peroxide, and organic peroxides. Another urgent need is artifact-free nitrogen dioxide measurements, which would help unravel the important processes in atmospheric chemistry. Active research is underway to detect radicals in the atmosphere. It is generally agreed that current field techniques badly undersample the atmospheric system (Kolb, 1995). Mobile, real-time measurements should replace or supplement the current method of sample collection followed by laboratory analysis. Kolb outlined the requirements of innovative field measurement designed, reproduced here in Table 12-1.

Goal:
Exploit capabilities of advanced instruments and platforms to form effective and responsive field measurement systems tailored to the issues being addressed.
Field Measurement System Characteristics:
Simultaneous real-time measurements of many trace species and meteorological parameters
Each measurement technique calibrated and quality assured
Mobile, fast response instruments replace multiple, slow fix-site monitors
Advanced platforms cover spatial scales necessary to bound problem
Careful blending of static and mobile point and remote sensing measurements to assure intercomparability and synergy
Immediate model analysis of field data to assess measurement success and scope ongoing measurement efforts

Table 12-1. Innovative Atmospheric Chemistry Field Measurement Design (Kolb, 1995)

Specific field data needed for the accurate modeling, as identified in Part I of this work, include UV radiation and source strengths of reactive organics and NO_x. In addition to ozone, measurements of nitrogen species such as NO_x, PAN, and nitric acid, and intermediate species

such as formaldehyde are expected to aid the validation of atmospheric chemical models for urban applications.

12.2 Examples of Emerging Technologies

One of the challenges facing the measurement community is to incorporate new technologies into existing measurement plans, so as to take advantage of measurements with improved sensitivity, increased resolution of species, and different time and length scales. Many new technologies are built to be mobile to supplement the traditional point measurements. A variety of new and emerging technologies were discussed in the EPA Center on Airborne Organics Symposium that was held in Dedham, MA, on July 10th and 11th, 1996 (EPA CAO, 1996). Spectroscopic techniques have gained attention in the recent years because of their accuracy and specificity. Table 12-2 lists a variety of new measurement techniques and their applications (additional references include Nriagu, 1992 and Newman, 1993).

Measurement Techniques	Species Measured
Chemical Ionization Mass Spectroscopy (CIMS)	$ClNO_2$, OH, HONO, nitric acid (HNO_3), NH3, HOCl, SO_2, SO_3, DMS, pyridine, acetone, NO_2, CH_6H_7N, and other larger organic molecules
Light detection and ranging (LIDAR) e.g., elastic scattering, Raman scattering, fluorescence, Differential Absorption LIDAR (DIAL)	water, ozone (with UV-DIAL), aerosols (with IR-DIAL), CO, methane maybe SO_2, ammonia, NO_2, NO
Tunable Infrared Differential Absorption Spectroscopy (TILDAS) or Tunable Diode Laser Absorption Spectroscopy (TDLAS)	methane, N_2O, CO, CO_2, ozone, NO, NO_2, SO_2, H_2O_2, HCHO, NH_3, HOCl, HO_2, HNO_3
Differential Optical Absorption Spectroscopy (DOAS)	NO_2, HONO, NO_3
Fourier Transform Infrared Spectroscopy (FTIR)	organic nitrates, nitric acid, NH_3
Single Particle Mass Spectrometer	size and composition of individual aerosol particles

Table 12-2. Measurement Techniques and Applications

Many of the light-based spectroscopic techniques listed in Table 12-2 can measure pollutant concentrations *in-situ* and over a long spatial range. *In-situ*, real-time methods are gaining popularity over the traditional method of sampling followed by laboratory analysis, especially for reactive compounds such as formaldehyde and compounds which cannot be effectively sampled because of interactions with the sampling devices (e.g., nitric acid and hydrogen peroxide). By eliminating the time lag between sampling and analysis and the need for canisters, *in-situ* measurements also reduce the possibility of contamination.

With well-designed sampling cells, mirrors can be used to "fold" the light to increase the optical path for a point measurements using the TDLAS technology. LIDAR requires high powered lasers and is generally very expensive. Of the trace species in the atmosphere, DOAS will not detect NO, CO, and most VOCs because these species do not absorb in the visible / near UV wavelengths.

Although FTIR can detect complex organic compounds and nitrogen compounds, its sensitivity is expected to be so low that the application will be limited to the most polluted urban atmospheres. Gas chromatography remains the most widely used method for the detection and measurement for airborne organic compounds. A cold trap may be used at the head of the column to focus the sample. In addition to the commonly-used flame ionization detector (FID), photoionization, mass spectrometry, atomic emissions, and ion mobility can be used to detect various organic species after the separation column. Oxygenated compounds (especially ones with low molecular weight) do not behave very well in the GC. Effective methods are still being developed for small aldehydes and ketone.

12.3 Focus of this Work

Better field measurements would provide critical data for improving the understanding of urban organic-NO_x chemistry. In the design of field measurements, particular attention should be paid to nitrogen species, which were shown to have significant effects on the uncertainties of model predictions in Part I of this thesis. There is an especially dire need for artifact-free measurements of nitrogen dioxide (NO_2).

An effective strategy should measure several species simultaneously to elucidate information concerning their correlations. Taking advantage of recent technologies with real-time sampling and analysis, information concerning turbulent fluctuations can also be obtained with high frequency data. A TDLAS instrument was designed to measure ozone, NO, and NO_2 simultaneously to determine their fluctuation and correlation characteristics in the seconds time-scale.

Meteorological information should be taken during field measurement campaigns to provide accurate inputs for subsequent modeling. For example, solar radiation measurements are necessary to provide good estimates of photolysis rates to reduce the uncertainties of model simulations. Although the data collected in this project is not used to support a modeling effort, meteorological information was taken at the same site to facilitate deposition fluxes analysis using simultaneous wind measurements.

13. TDL Description

13.1 Introduction

Tunable diode laser absorption spectroscopy (TDLAS) is emerging as a useful tool in trace gas detection and monitoring. Tunable diode lasers operate in the middle infrared (MIR) region between 2 and 15 μm. Within this region, there are so-called "atmospheric windows" (9 - 11 μm and 3 - 4 μm) which are free from interference absorption of water vapor and carbon dioxide. Almost all trace species of interest have strong spectroscopic absorption in this region. These include green house gases (e.g.methane and nitrous oxide), urban smog (ozone, nitric oxide, nitrogen dioxide, and formaldehyde) and acid rain precursors (sulfur dioxide and nitric acid), and chlorinated species (hydrochloric acid and hypochloric acid HOCl). Some species that are detected in this region and their absorption regions are listed in Table 13-1. In addition, many radicals of interest in the atmosphere, such as hydroxyl (HO) and peroxyl (HO_2), also have absorption features in the MIR region.

Trace Species	Absorption Regions (cm^{-1})
Methane (CH_4)	1300, 3000
Carbon Monoxide (CO)	2100, 4200
Ozone (O_3)	1020, 1800, 2100
Ammonia (NH_3)	930, 1600
Nitrous Oxide (N_2O)	1200, 2200, 2550
Nitric Oxide (NO)	1900
Nitrogen Dioxide (NO_2)	1600
Nitric Acid (HNO_3)	890, 1300, 1700

Table 13-1. Absorption of Some Atmospheric Trace Gases (Park *et al.*, 1987)

The advantages of TDLAS include specificity, sensitivity, fast response, and absolute measurement capability. Due to the high spectral resolution of diode lasers, individual rotational-vibrational (non-overlapping) features of molecules can be isolated and measurements can be taken free from interference from other MIR absorbers, particularly water and carbon dioxide. This allows unequivocal identification and measurement of the species of interest. The sensitivity of the measurement is further enhanced by increasing the absorption signal. Since the amount of absorption increases with optical pathlength, special designs of mirror arrangements are used in multipass cells to "fold" the optical path in a small volume. The advantages of keeping the sampling volume small are reduced pumping requirements and improved time response of the measurement. Samples are frequently collected at reduced pressure to increase the sensitivity of the measurement: pressure broadening effects are minimized and interference from other absorption features can be avoided. The detection limit for most species are in the low ppb to ppt range, suitable for monitoring stable species in even the clean atmosphere. Unlike chemiluminescence and

chromatography techniques, detection is instantaneous and the time response in a TDL instrument is limited only by the sampling and data processing procedures. The sampling time is not a concern in open path operations. Faster pumps and smaller sampling volumes can be used in the closed path configuration to reduce the time response to below 0.1 second, when low pressure sampling is desired. TDLAS has been used successfully in eddy correlation applications for methane and nitrous oxide. The data processing time requirement is strongly dependent on the speed of the control computer and is usually a fixed time lag. TDL measurements are absolute, in the sense that the concentration can be determined from the absorption based on the spectral line parameters, which can be determined independently. No calibration is needed. TDLAS has been used extensively in field intercomparison campaigns as a standard against which other less definitive measurements are compared.

The next sections will describe the theory of infrared absorption, and the general designs of TDLAS systems. In this study, two TDLAS systems were coupled together to monitor ozone, nitric oxide, and nitrogen dioxide in a single sampling cell. The details pertaining the optical and electronic modules of the design will be addressed, followed by a discussion of the operating procedures.

13.2 Rotational Vibrational Spectra of Simple Molecules

The rotational and vibrational transitions of most molecules occur in the infrared region. For each vibrational transition, simultaneous rotational energy change can take place, as illustrated in Figure 13-1. Each vibrational transition, observed as a "branch" or a "band", contains discrete lines that correspond to different rotational transitions. The shape and width of these lines are completely determined by the pressure and temperature, based on compound-specific pressure broadening parameters. Therefore, concentration of the sample can be calculated from the height of the absorption once the specific transition can be identified.

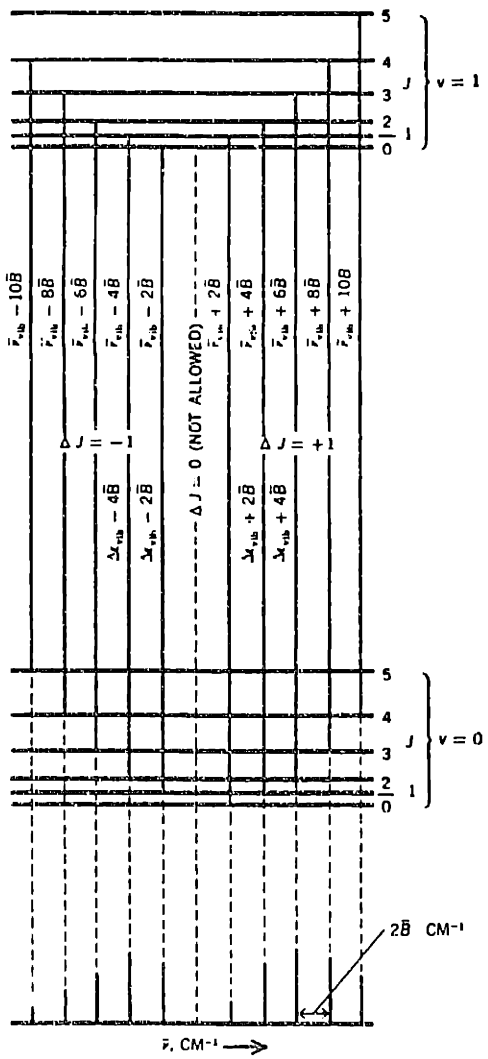


Figure 13-1. Rotational-Vibrational Transitions.
Source: Barrow, Gordon M. Physical Chemistry, 4th Ed. (Figure 14-12)

13.3 Absorption Relationships and Broadening

Beer's Law. The Beer-Lambert Law relates the amount of transmitted power to the concentration of the absorbing species and the pathlength of the measurement.

$$\frac{I(\lambda)}{I_0(\lambda)} = Exp(-\sigma(\lambda)\cdot c\cdot L) \quad , \tag{13.1}$$

where $I(\lambda)$ and $I_o(\lambda)$ are the transmitted and incident power, $\sigma(\lambda)$ is the absorption cross section, c is the concentration of the species and L is the absorption pathlength. The optical depth at a wavelengh λ is defined as the product of σ, c, and L. At low concentration, the following approximation can be used to determine the concentration of the species.

$$a = 1 - \frac{I(\lambda_0)}{I_0(\lambda_0)} = \sigma_0 \cdot c \cdot L \quad , \tag{13.2}$$

a is the absorbance at the line center (λ_0) and σ_o is the absorption coefficient at the line center.

The absorption line strength of a rotational-vibrational feature,

$$S = \int_{-\infty}^{\infty} \sigma(\lambda) d\lambda \quad , \tag{13.3}$$

is the integral of the absorption cross section over the entire line. S is a function of temperature only, and is independent of pressure. This parameter is tabulated in databases such as HITRAN (Rothman, 1992), together with two pressure broadening coefficients, one in the pure gas and the other in air. Pressure broadening coefficients describe the line width as functions of pressure. Together with the line strength parameter, they are used to provide parameters for the absolute calibration for measurements.

Doppler Broadening. The width of the absorption line decreases with pressure until it reaches the doppler limit, usually below 10 torr. The width of the absorption feature at the Doppler limit is characterized as follows (Zahniser *et. al.*, 1995):

$$\alpha_D = 3.58\times10^{-7}\sqrt{\frac{T}{M}}\cdot \upsilon_0 \quad . \tag{13.4}$$

α_D is the Doppler half width at half maximum, T is the temperature in Kelvin, M is the molecular weight in atomic mass unit (amu), and υ_0 is the line center frequency in wavenumbers (cm^{-1}). Doppler line broadening is a result of the random thermal motion of the gas molecules and the Doppler width is independent of pressure. Therefore, changes in pressure in the Doppler regime result in proportional changes in the height of the absorption with the number of absorbing molecules. Since the absorption shape in the Doppler regime is a Gaussian function, the line center absorbance is

$$A_{\upsilon}(\text{Doppler}) = \sqrt{\frac{\ln 2}{\pi}}\frac{SLmP}{\alpha_D} \quad , \tag{13.5}$$

where S (cm^{-1} atm^{-1} cm^{-1}) is the absorption line strength, L (cm) is the pathlength, m (dimensionless) is the mixing ratio, P (atm) is the pressure, and α_D is the Doppler width.

Pressure Broadening. At high pressures, the absorption shape is a Lorentzian profile. At this limit, which usually occurs above 100 torr, the dominant broadening effect is collision-induced. The width of the line is directly proportional to the pressure:

$$\alpha_L = bP\,, \tag{13.6}$$

where b, the pressure broadening coefficient (cm^{-1} atm^{-1}). At low mixing ratios, b is determined by the broadening effect of air, as opposed to the self-broadening effect which is important at high mixing ratios. The center absorption for the Lorentzian profile is given in Equation 13.7.

$$A_0(\text{Lorentz}) = \frac{SLm}{\pi b}\,. \tag{13.7}$$

Notice that the Lorentzian peak height is also the maximum center peak height with respect to pressure for any given mixing ratio. However, due to excessive broadening at the wings of the absorption feature, it is more difficult to distinguish the signal from the baseline. Moreover, the broader the feature, the higher the chances are of interference.

Voigt Regime. The Voigt regime is the intermediate between the Lorentz and Doppler regimes. The Voigt profile represents a convolution of the Gaussian and Lorentzian functions. The line center absorbance can be related to the complementary error function (erfc) as follows:

$$A_0(\text{Voigt}) = \sqrt{\frac{\ln 2}{\pi}}\frac{SLmP}{\alpha_D}\exp(a^2)\cdot\text{erfc}(a)\,, \tag{13.8}$$

where

$$a = \sqrt{\ln 2}\,\frac{\alpha_L}{\alpha_D}\,. \tag{13.9}$$

The Voigt regime usually represents the best tradeoff between absorption height and contrast between the absorption feature and the baseline. Figure 13-2 illustrates the effect of pressure on line center height and the optimum sampling range.

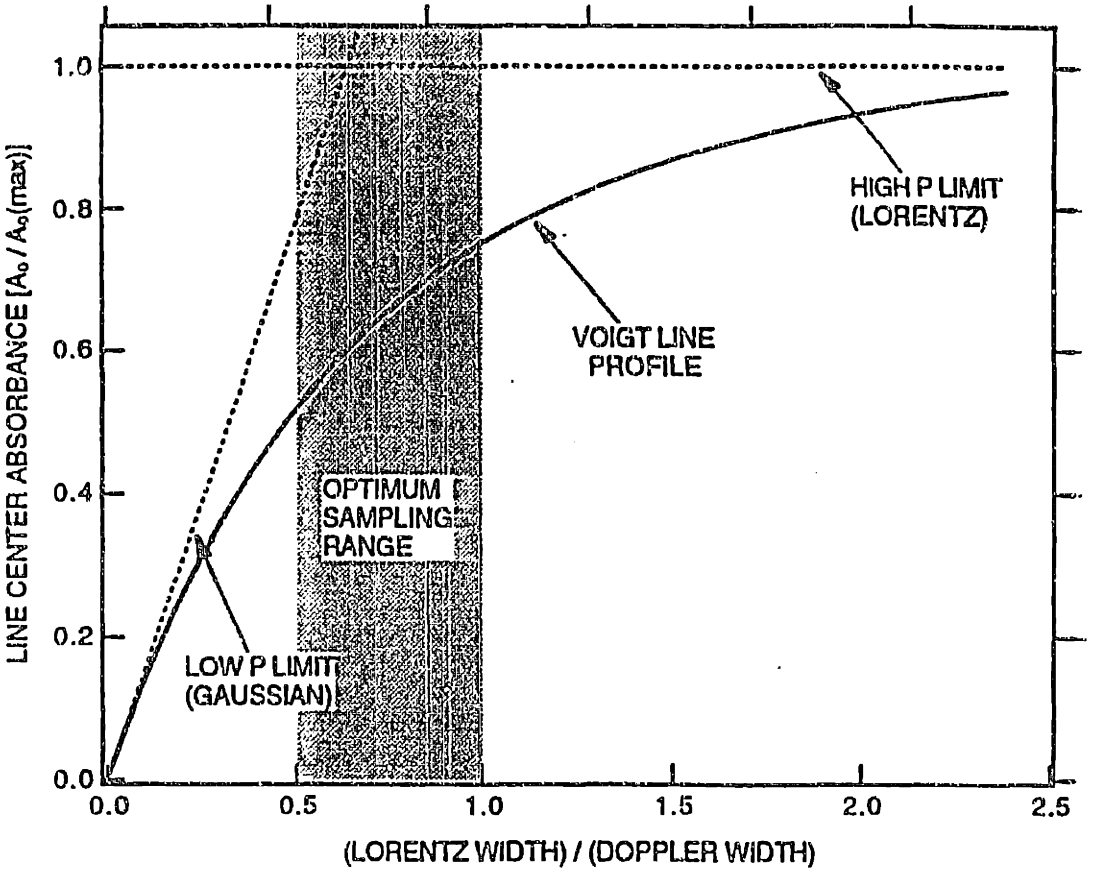


Figure 13-2. Effect of Pressure Broadening.

13.4 Optical Components of the Tunable Diode Spectrometer

A tunable diode laser absorption spectrometer contains two main modules: the optical module and the electronic module. A simplified block diagram is shown in Figure 13-3. The optical module contains the dewar that houses the diode laser(s), the transfer optics which direct the light into and out of the multipass absorption cell, the infrared detectors, and the associated flow control devices. The electronic module contains the laser controller, and the signal acquisition and processing computer. The key components of the optical module will be presented in the following paragraphs, followed by a description of the laser control and signal processing procedures and the software used.

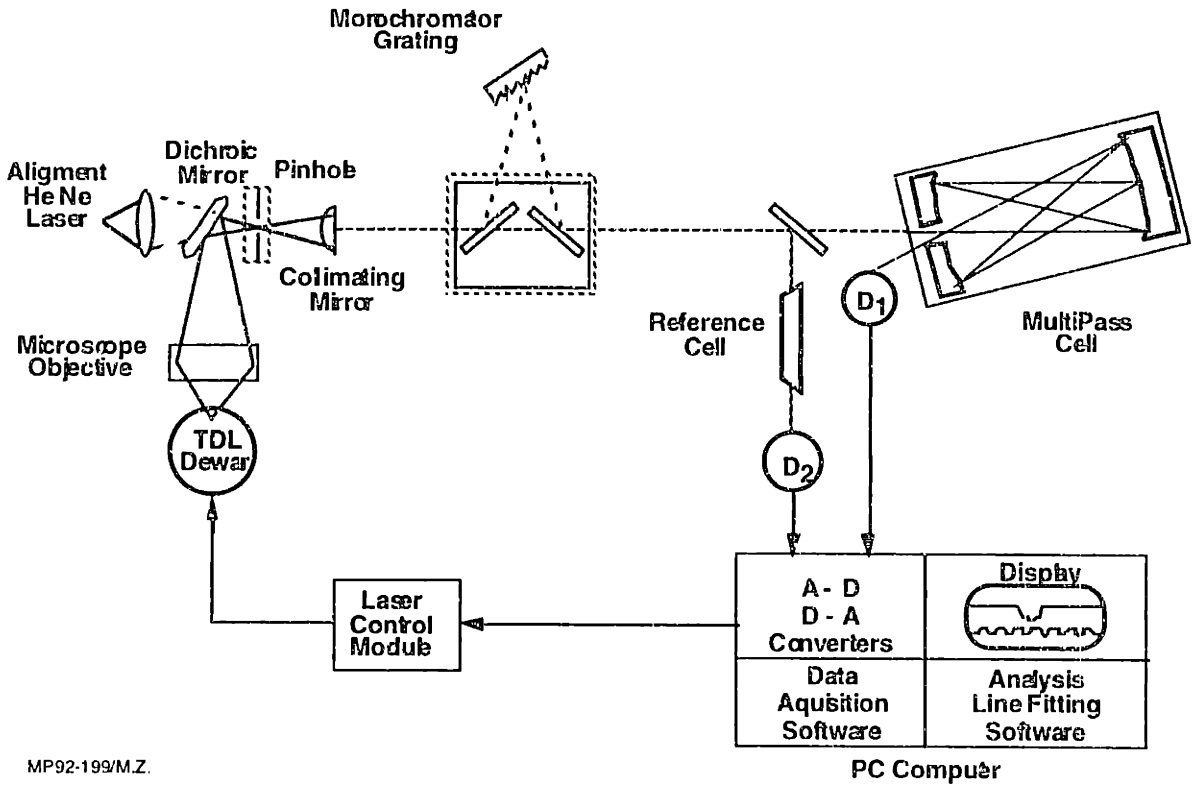


Figure 13-3 Block Diagram of the TDL System.

Tunable Diode Lasers. Tunable diode lasers are lead salt semi conductor chips that form p-n junctions with small energy band gaps. When a forward biased current for 0.1 to 1 A is applied across the junction, a population inversion is produced. Electrons are pushed up to the conduction band from the valence band. The electron-hole recombination causes the emission of light associated with the band gap energy, typically of wavelengths of 3 - 30 μm. This process is illustrated in Figure 13-4.

Diode lasers usually operate at below 120 K. Therefore, they are generally housed in liquid nitrogen or liquid helium dewars. Closed-loop cryostats are available that achieve temperatures down to 12K. Broad laser frequency tunability is achieved by varying the temperature of the diode. In practice, the diode current is varied for fine tuning. As the current increases, the diode tunes through a series of wavenumber segments, called "modes." One of the biggest drawbacks of tunable diode lasers is the instability of modes. The

temperature/current combination at which infrared light of a certain frequency is emitted drifts with time. If the laser is warmed up to room temperature (in a process referred to as "temperature cycling"), mode structures may change significantly when it is cooled down again to lasing temperature.

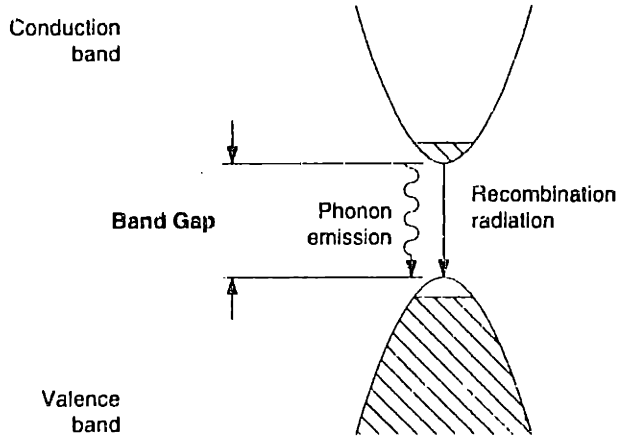


Figure 13-4 Electron-Hole Combination Process Resulting in Laser Emission (Adapted from W.T. Silfvast, 1996, Figure 5-19)

Multipass Absorption Cell. As shown in Equation 13.1, the amount of absorption for a given concentration increases with line strength and absorption pathlength. To achieve high sensitivity, multipass cells are used to provide long pathlengths at reduced pressure. Special mirror arrangements are used to maximize the number of passes of the light (pathlength) in a minimum volume to maintain high sampling rate. Two traditional designs are the White cell and the Herriott cell. The reflections on the White cell mirrors form two horizontal line. The Herriott cell uses 2 spherical concave mirrors to generate a series of reflection spot in an elliptical pattern. The spot patterns formed by the passes of light in these two mirror arrangements are shown in Figure 13-5 (a, b).

A patented astigmatic design (Figure 13-5c) from Aerodyne Research Inc. was used in this study. This design is a variation of the traditional Herriott cell. Due to the difference in horizontal and vertical radii of the curvature of the mirrors, the spots on the mirrors appear in a Lissajous pattern. The optical path fills the absorption volume better than the traditional designs. The volume of the absorption cell can be minimized for a given pathlength. The cell being used has 182 passes and a base length of 60 cm. The total pathlength is 100 m and the volume is 3 L (McManus *et al.*, 1995). This design represents a tradeoff between an increase of the pathlength and a loss of laser intensity due to imperfect reflectivity of the astigmatic mirrors.

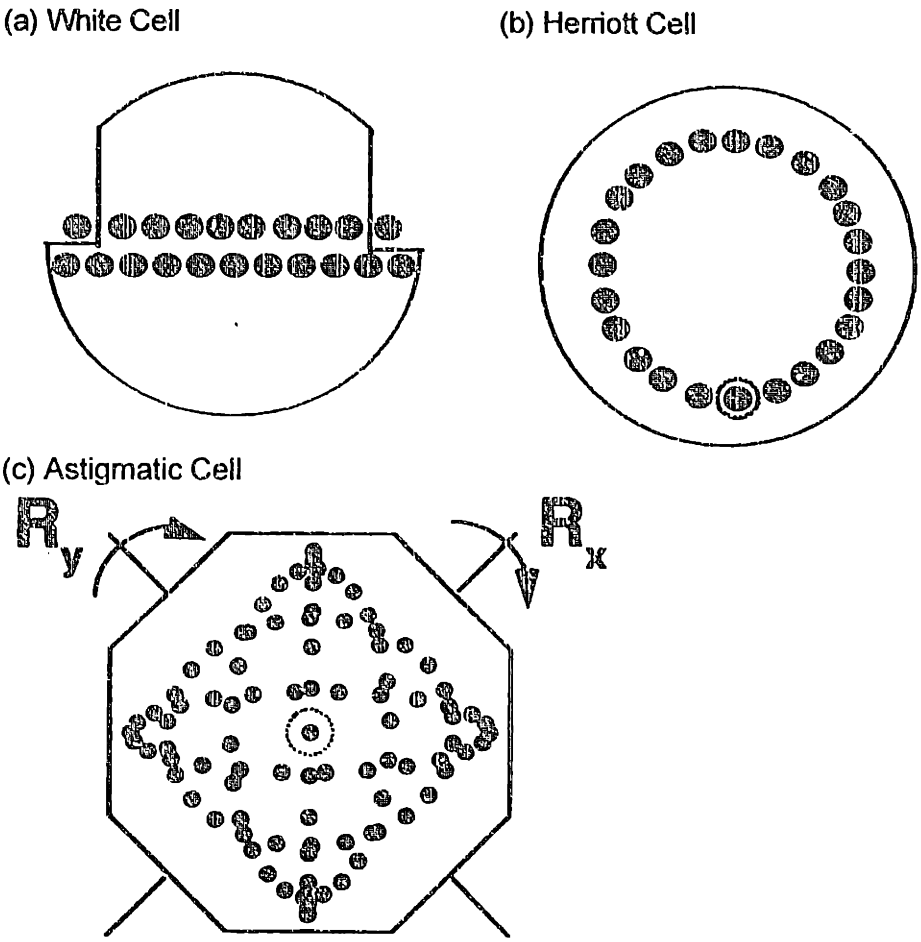


Figure 13-5 Mirror Configuration in Sampling Cell.

Detectors and Preamplifiers. Semiconductor photon detectors, in which incident photons interact with electrons within the detector material to give an electric signal, are used to detect infrared radiation in the TDL system. Two types of detectors are available in the wavelength regions of interest (Moore, 1983). In photoconductive (PC) detectors, the incident photons increase the number of charge carriers in a material or change their mobility. This results in a change in the electrical resistance in the detector directly proportional to the amount of light falling on the detector. A bias current is applied to the detector material and the change in detector resistance produces a change in the voltage across the detector. A volt-to-volt preamplifier is used to generate a signal detectable by the data acquisition circuitry in the computer. In photovoltaic (PV) detectors, the absorption of photons leads to the generation of a potential difference across a junction between two materials. An external resister turns this potential difference into a current through the detector. A voltage signal is generated by the current-to-voltage preamplifier.

Detector response is a function of the wavelength of light. The responsivity usually exhibits a cut-off wavelength above which the detector does not respond to the light. This cut-off wavelength and the wavelength of peak response are dependent on the composition of the detector material. As an example, the responsivity of the photovoltaic detector for the NO and NO_2 lasers is shown if Figure 13-6. The peak detectivity of this detector is at 8.69 μm and the cut off wavelength is 9.43 μm. At the wavelengths of NO and NO_2 lasers, which are

approximately 5.3 μm and 6.3 μm, the response is about 60% of maximum. Due to the dependence of the detectors' responsivity on wavelength, detectors have to be matched to the light being detected. This photovotaic detector would be a poor choice for the 9.8 μm light used for ozone detection because the responsivity is down to about 10% of the maximum at that wavelength.

Infrared detectors operate at liquid nitrogen temperatures and have to be housed in cold dewars. The optical tables are designed such that it is possible to house detectors and lasers in the same dewars. This is discussed in the next section.

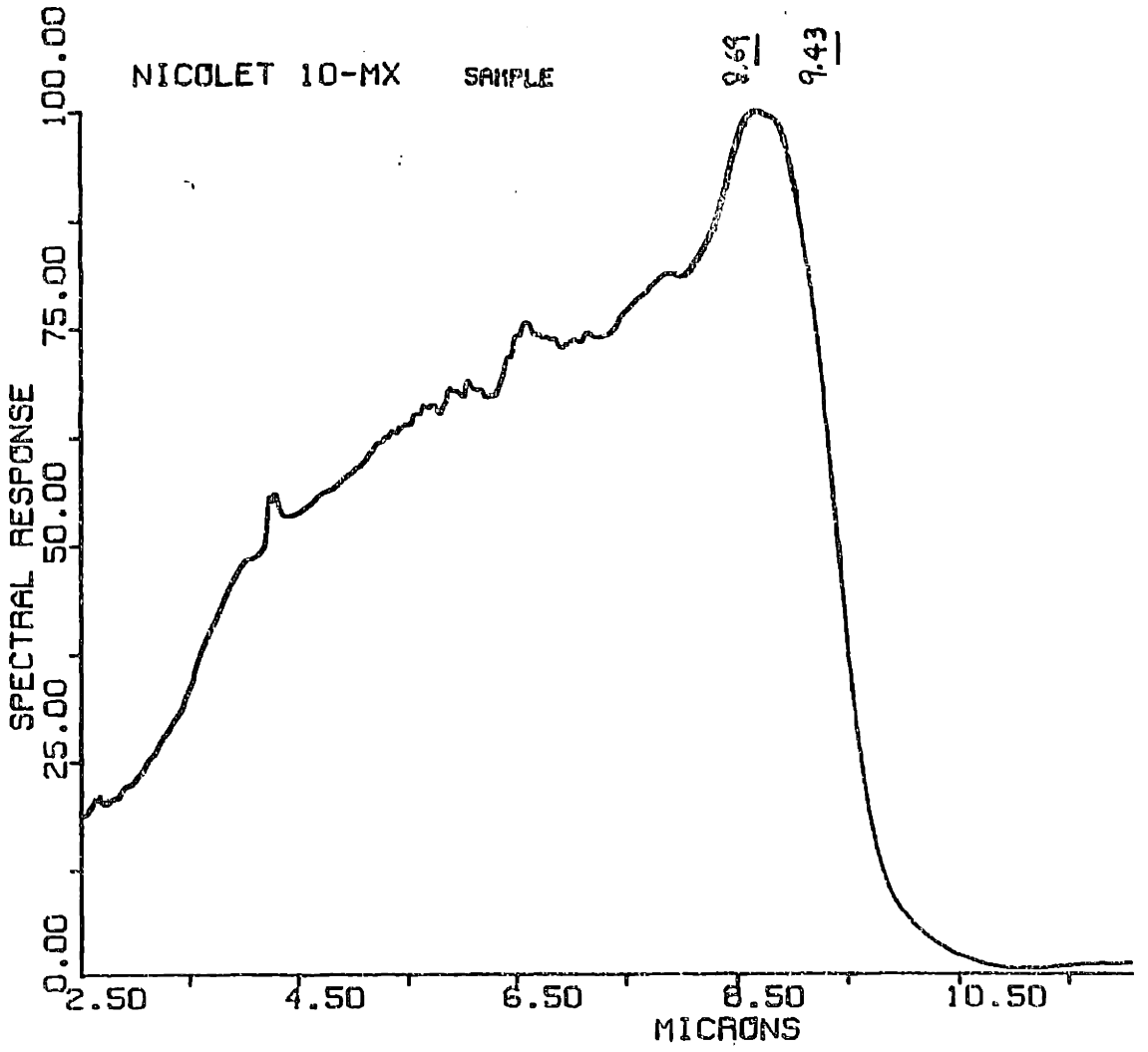


Figure 13-6 Responsivity of a Dectector as a Function of Wavelength

Typical Optical Layout. The optical train of a single-laser TDL module is depicted in Figure 13-7. Infrared light from the laser diode in the dewar is collected by a reflective microscope objective (15X) and focused onto a 200 μm pinhole which defines the input aperture. There is a visible optical system which employs a red helium-neon (HeNe) trace laser to aid alignment and setup. This tracer laser is co-aligned with the infrared beam through the pinhole. The pinhole is used only during alignment, and it is mounted on a removable kinematically indexed base. The microscope objective is mounted on an X-Y-Z translator to

allow accurate focusing of the laser bean into the fixed aperture. Past the input aperture, the infrared beam is directed past a dichroic beam splitter, where a small fraction of the light (typically 5%) is reflected to form the reference beam. The remainder of the light is then focused into the optical absorption cell, in a narrow beam at approximately F#200. The exit beams have approximately the same focal parameters as the entering beams, but at an angle of 0.1 radians (6 degrees) from the input beam. The output beams are focused onto the detector located close to the absorption cell.

The reference beam from the dichroic beam splitter is focused onto a by-passable diffraction grating used as a monochromator. By comparing the diffraction angle of the laser to that of a co-aligned helium-neon laser of known wavelength, the exact wavelength of the infrared light can be determined to within ± 1 cm^{-1}. A short reference cell can also be inserted into the reference path to identify spectral line positions, as well as to provide the reference for peak locking when the signal is small. Peak locking procedures will be discussed in detail in the next chapter.

An optical table has been designed for simultaneous measurements using two lasers and a single detector, as described in Wormhoudt *et al.* (1995) and shown in Figure 13-8. Light is collected from two lasers with alternative and out-of-phase laser on (frequency scanning) and laser-off periods using the same objective and pinhole configuration described above. The beams are then collimated using curved mirrors, and combined at the main beamsplitter, where one beam is transmitted and the other reflected. The resulting co-aligned beams are passed to a pair of flat mirrors and onto a large gimbal-mounted steering mirror, where it is sent off the optical table to the absorption cell and detector.

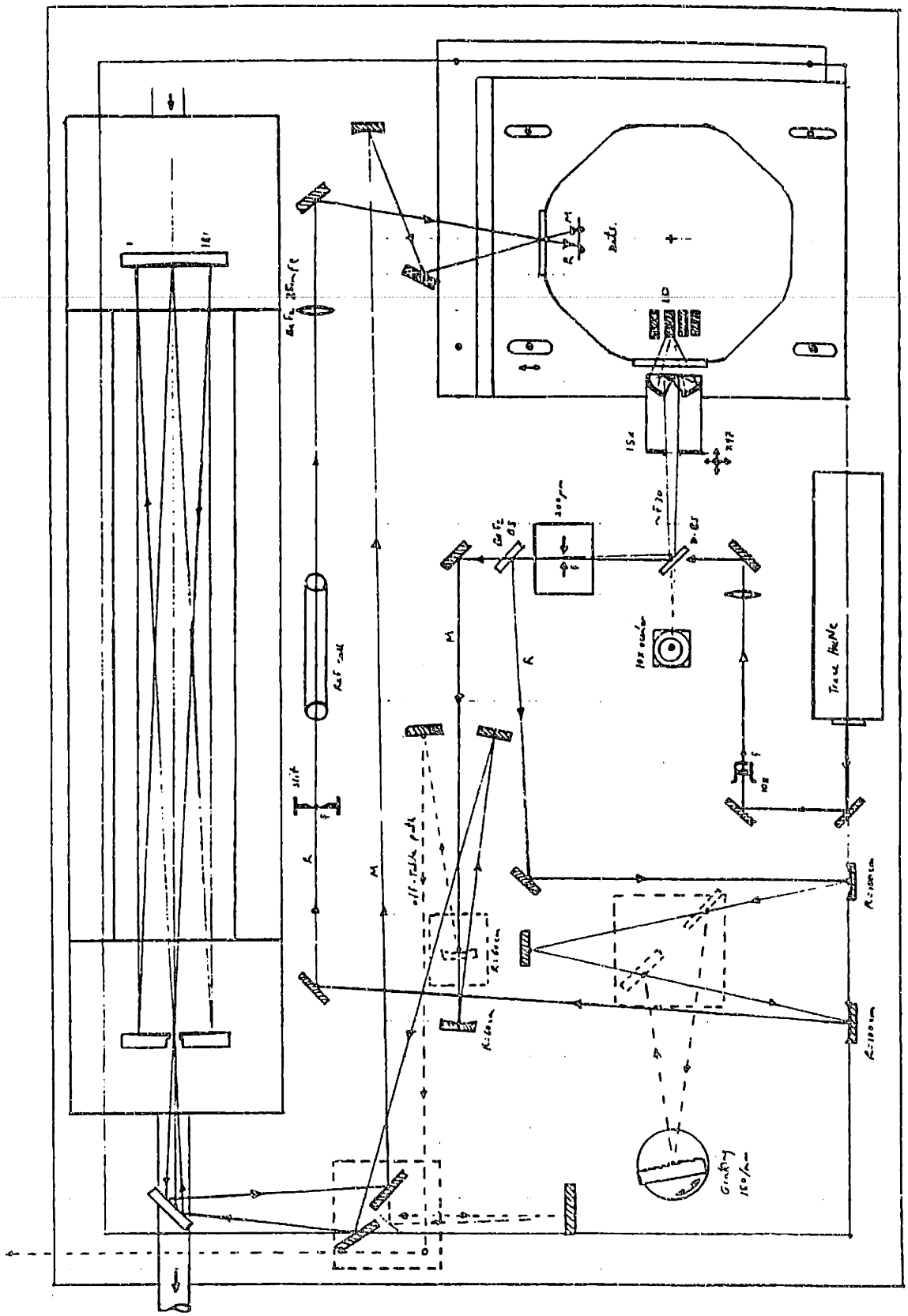


Figure 13-7 Basic TDL Optical Table Design

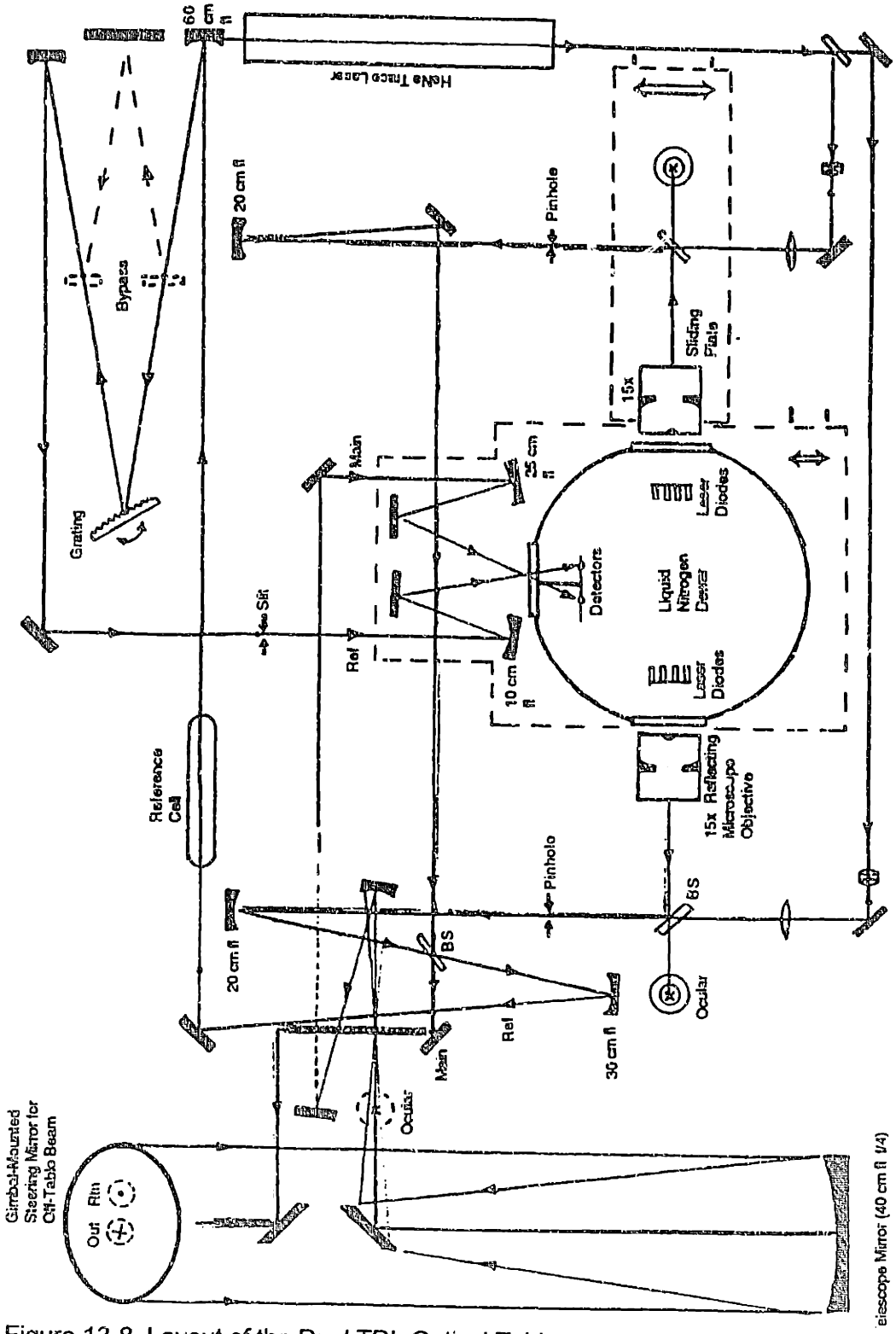


Figure 13-8 Layout of the Dual TDL Optical Table

13.5. Laser Control and Signal Processing Software

The data acquisition method is based on rapid sweep integration over the infrared transition line. The laser frequency is scanned under computer control and the transmitted infrared light intensity is measured synchronously. A software program serves as the user interface to the laser controller. It also reads the pressure of the system from a pressure transducer (baratron) and analyzes the spectral information in real time to output the concentration of the trace species being investigated. A block diagram showing the interactions between the computer program and the experimental hardware is shown in Figure 13-9.

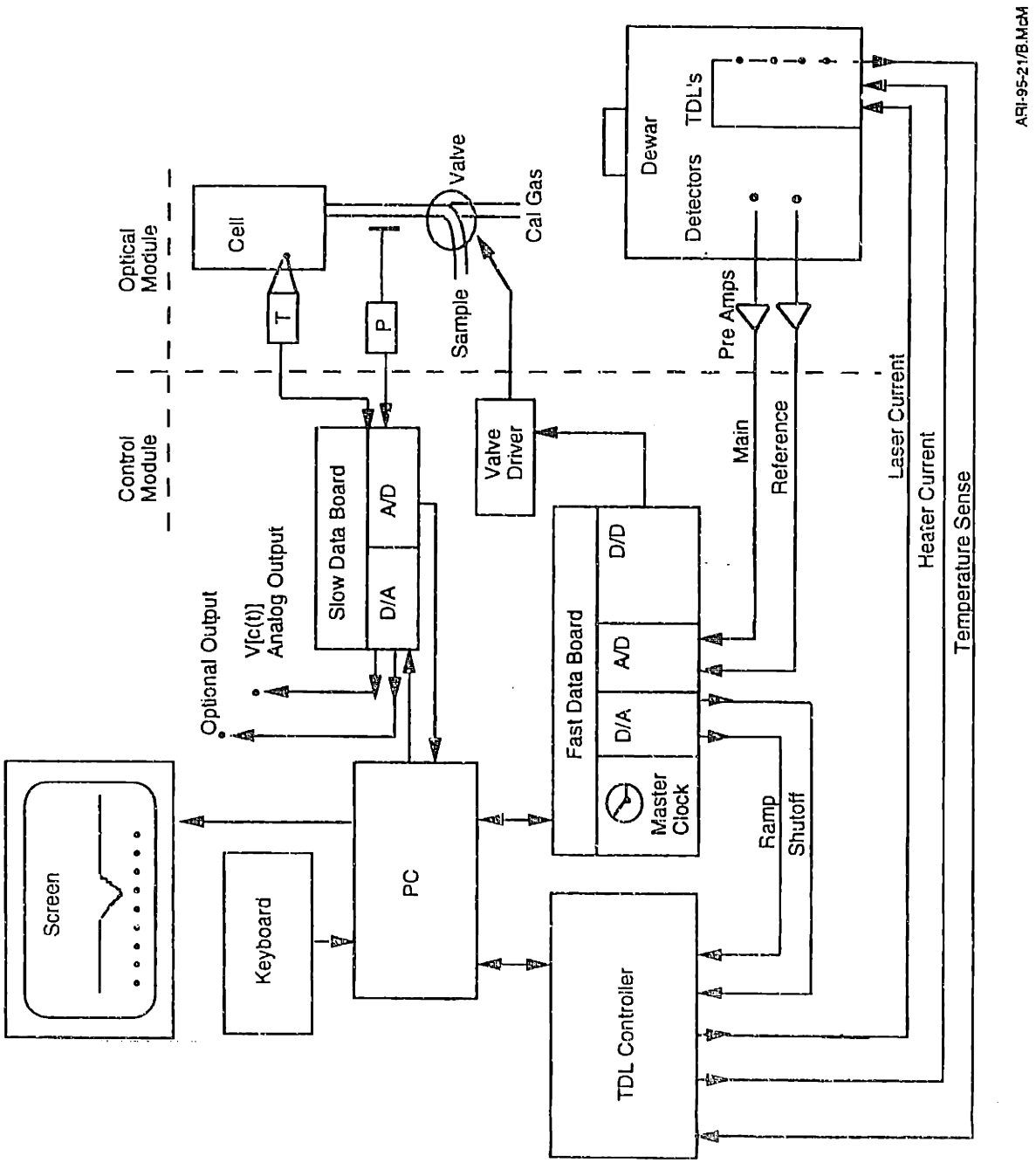


Figure 13-9 Computer - Hardware Interaction for the TDLAS System

Laser Control for Sweep Integration. The laser temperature and current are controlled by Laser Photonics controllers. Typically, the laser temperature is held constant and the current is modulated with a computer generated sawtooth waveform to sweep the laser frequency across the transition feature of interest. The sawtooth is generated in 50 to 200 discrete steps by the computer, and the absorption at each point is displayed on the computer screen real-time (usually at about 1 Herz, depending on the user's preference). The height of the sawtooth can be varied as integral multiples of the number of display points up to a total of 4096, the resolution of the 12 bit digital to analog converter board that is used to convert the computer digital signal into an analog voltage. The analog voltage is fed into the laser controller at a rate of 6 kHz to 1.5 kHz to modulate the current of the diode. At the end of each sweep, the ramp is shut off so that the laser current drops below the lasing threshold to provide a precise measurement of zero light intensity.

Nonlinear Least Square Fitting for Direct Absorption. The analog voltage signal from the detector and pre-amplifier is fed into a analog to digital converter board. In one second, results from 6000 to 15000 scans can be averaged and displayed as a power spectrum on the computer screen. Since this process takes relatively little CPU time, data fitting and analysis and data collection can be done simultaneously in real time. A sample spectrum shown in Figure 13-10.

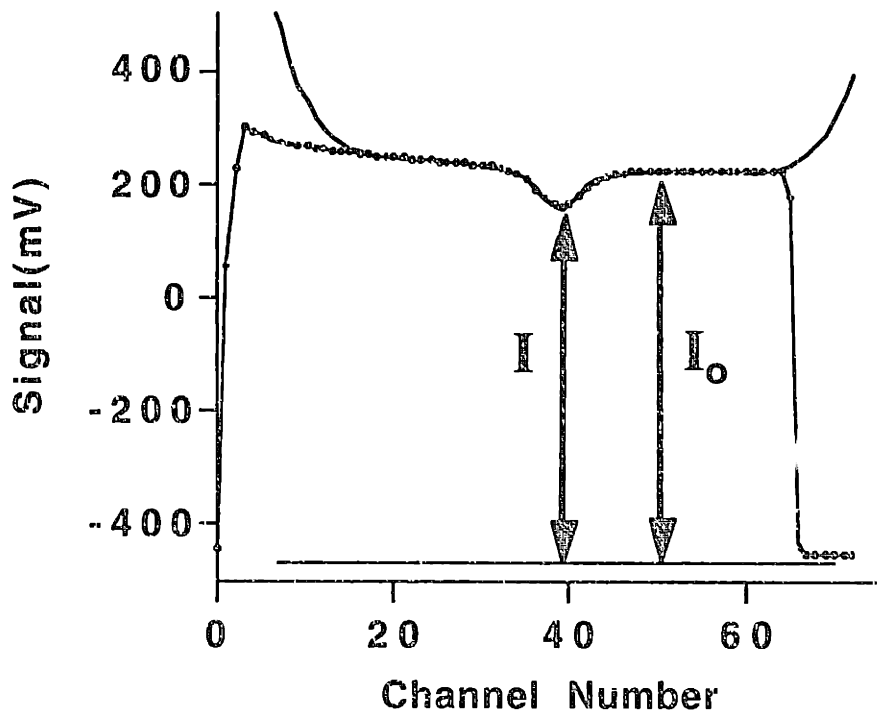


Figure 13-10 A sample Spectra of NO at 1900 cm^{-1}

Since the spectral line shape is completely determined by the temperature, the pressure, and the line parameters that are documented in the HITRAN database, the spectra can be fitted to the known shape using a least square procedure to estimate the concentration. There are two parts to the fitting routine. First, the baseline is fitted to a polynomial representing the power variations of the laser as the current is varied, providing an estimate of the available

power from the laser, I_0 at each frequency. Polynomials of up to 7 terms (6th order) may be used to represent highly non-linear shapes of the laser power spectra. Second, the absorption line profile is fitted. At a fundamental level, three line types can be expected depending on the pressure: gaussian, voigt or lorentzian. The position, width and height of the line are simultaneously fit together with the diode laser polynomial baseline. The fit is displayed in real time, as shown in Figure 13-10. The absorption under the base line (I) is determined based on the "zero level" light intensity at the end of the scan, when the laser is turned off. The absorption profile is converted to the concentration of the trace gas using the pathlength and the known line strength parameters (functions of pressure and pressure) by Equation 13.1. No external standard is needed for calibration. Sources of error of the calibration include: uncertainty in the zero light level (multiple mode operation or non-shut-off conditions), pressure reading errors, non-linear baselines that are not well-approximated by polynomials, and uncertainties in the archived spectral parameters.

Real time observation of the concentration as a function of time is also a feature of the program. The same information is also be recorded in a data file and can be sent through a RS-232 data output channel to be logged separately.

13.6 Noise Characterization

No measurement is exact. Noise is all unwanted output features that mask the true signal in the detection system. This section is devoted to a discussion of the various kinds of noise that contribute to uncertainties in TDL absorption measurements, and ways to minimize them.

The following definitions are adopted in this work for a few terms that are used more or less interchangeably to describe the quality of measurements. *Detection Limit* is the minimum detectable signal that is distinguishable from the noise. *Precision* is used here to describe the measurement root mean square (rms) noise on a constant (steady) concentration signal. *Accuracy* measures how close the measurement value is to the "true value."

Detector Noise. The most common source from the detector noise is thermal noise. Thermal noise is due to fluctuations of the electronic charges in conducting components of the detector. It is inherently random but is sensitive to the temperature of the operating environment. Each detector is characterized by a D* value, its detectivity.

$$D^* = \frac{\sqrt{A}}{\mathrm{NEP}}, \tag{13.12}$$

where NEP, the noise-equivalent power is the rms value of a sinusoidally modulated radiant power falling upon a detector that gives an rms signal voltage equal to the rms random noise voltage from the detector. To obtain the noise level for a particular setup, the wavelength dependent responsivity and amplification factors are needed. Responsivity is the current/voltage per watt of energy incident on the detector at the given wavelength. The current/voltage due to noise in the detector is amplified along with the signal. Typically, detector noise less than a few μV is desired, so that the detector noise will only be dominant at signal fractional absorption levels of $<10^{-5}$. This can be achieved by choosing a detector with high detectivity in the wavelength region of interest, or increasing the signal averaging time. Since thermal noise is random in nature, it deceases linearly with the square root of averaging time (Werle *et al.*, 1993). Furthermore, since detector noise is independent of the incident power of the laser, the detection limit can be improved if a more powerful laser is available to increase the signal voltage and thus the S/N ratio when the detector noise is dominant over other sources.

Laser Excess Noise. The sensitivity of TDLAS measurements is frequently limited by fluctuations in laser power. Some of these fluctuations are intrinsic to the behavior of the diode. For example, mode competition can distort laser power spectrum even when 99% of the laser power is in one mode. There are also external effects, such as noise in the injection current, temperature instabilities, and optical feedback. Injection current noise may be stable enough to be subtracted away using a background subtraction technique. Temperature instabilities can be controlled by a good temperature/heater feedback circuit with a well-located temperature sensor. When the temperature sensor is too far from the diode, the diode may sense and respond to temperature instabilities from the ambient environment, such as room temperature changes or radiative cooling from the dewar itself when it is being filled. Optical feedback is produced by spurious reflections and scattering in the external optical system that feed back into the laser cavity, resulting in the deterioration of the laser spectral and noise characteristics. One manifestation of optical feedback is a distortion of line shape of absorption features. Optical feedback cannot be completely eliminated, although some diodes are more sensitive to it than others. Proper alignment usually helps in reducing

spurious reflections. Apertures and scrapers can be used to prevent light from being reflected back to the laser diode. Decreasing the laser intensity generally decreases feedback light intensity in a square relationship. Screen attenuators and apertures can be used for this purpose.

Fringes. Interference fringes are caused by light following different paths to the detector. The path difference causes light to interact constructively and destructively at the detector. Fringes appear as sinusoidal waves on the signal. If the fringe frequency is comparable to the width of the absorption feature, the accuracy of the measurement will be compromised. The fringe spacing (F) in wavenumbers is related to the path difference in the following manner:

$$F = \frac{1}{2Ln}, \tag{13.12}$$

where L is the pathlength difference (in cm) and n is the refractivity index (1 in air). This permits the identification of optical elements causing the fringes. Fringes are typically caused by multiple reflections on an optical element, e.g., a beam splitter, or by light scattering on dirty surfaces or misalignment, the latter two of which can be minimized with relatively little effort. If the fringe field is stable, signal-background subtraction can be used to obtain a fringe-free signal. Other ways to minimize fringe noise include vibrating the optical elements involved in producing the fringe, or applying high frequency modulation to the laser driving current the laser so that fringes "average out". Another technique is to us pressure modulation. Fringes are a dominant source of noise in the current design of the TDL systems.

Methods to reduce detector noise, laser excess noise, and fringes are unfortunately not usually compatible. Increasing the laser power reduces the fractional contribution of the detector noise but may result in higher laser excess noise. Increasing the averaging time reduces the variance of the random detector noise, but also reduces the efficacy of signal subtraction that might be used to remove fringes. Achieving the best sensitivity with the TDL system is frequently a skillful trade off between the various sources of noise.

Fit Noise. The absolute accuracy of the area under the peak returned by the fit is a few percent. The short term (~1 hour) precision can be better than one part in one thousand. However, numerical instabilities can cause the interpreted signal to fluctuate. When high order polynomials are used to fit non-linear baselines, the optimization procedure may find multiple minima for the least square fit, this would cause the reported concentration to fluctuate between 2 or more values on a constant signal. It is therefore recommended that a low order polynomial be used for fitting the base line.

In order to increase the stability and decrease the time taken to do the non-linear least square fit, the number of parameters to be fitted should be minimized. The position of the peak on the scan is frequently known quite well. However, when doing long term monitoring, long term drift may occur. The drift in light wavelength at the injected current or laser emission characteristics may cause the peak position to change with time. As a result, the accuracy of the concentration suffers as the peak moves away from the position at which the fit is fixed. This problem is partially corrected by using a peak lock procedure that would check the position of the peak at regular intervals and "follow" the peak with the fit. Limitations do exist for this procedure. These will be discussed in the procedure section.

14. Experimental

14.1 Site Information

Field measurements were conducted at Aerodyne Research, Inc. 45 Manning Road, Billerica, MA, 01821. Billerica is in suburban Boston, MA. The site is located about 20 miles due NW of metropolitan Boston, about 500 m due East of Route 3, an interstate highway that is a New Hampshire - Massachusetts thoroughfare. Sampling took place on the roof of a two-story building, with the sampling port attached to a 10-foot mast.

14.2 TDL Experimental Setup

This section is devoted to the discussion of the necessary preparations and some of the caveats of continuous monitoring of trace gases using TDLAS.

14.2.1 Absorption Feature Identification, Laser Detection, and Calibration.

The first step in the preparation is to identify the absorption region of interest and extract the relevant information from the HITRAN database (Rothman, 92). A simulation is run based on these parameters, the pressure, and the pathlength to generate a "map" for the absorption features in the region. Candidate features used for monitoring are identified based on the availability of interference-free absorption and the absorption strength of the individual feature. Since tunable diodes do not emit light continuously, but in modes of about 1 - 3cm^{-1}, finding the ideal features for monitoring is somewhat a matter of chance. Once the absorption feature was positively identified in the observed spectrum, by comparison with the "map" and by wavelength verification using the monochromator, TDL measurements are calibrated by the spectral parameters for the specific lines that are archived in databases such as HITRAN. Figure 14-1 shows sections of the HITRAN maps chosen for monitoring O_3, NO, and NO_2 corresponding to the spectra shown in Figure 14-2. The top trace of each panal is the absorption feature of-interest. The bottom trace displays the spectrum as it should appear in the TDL scan, with the background absorption of species present in the atmosphere (e.g., CO_2 and H_2O).

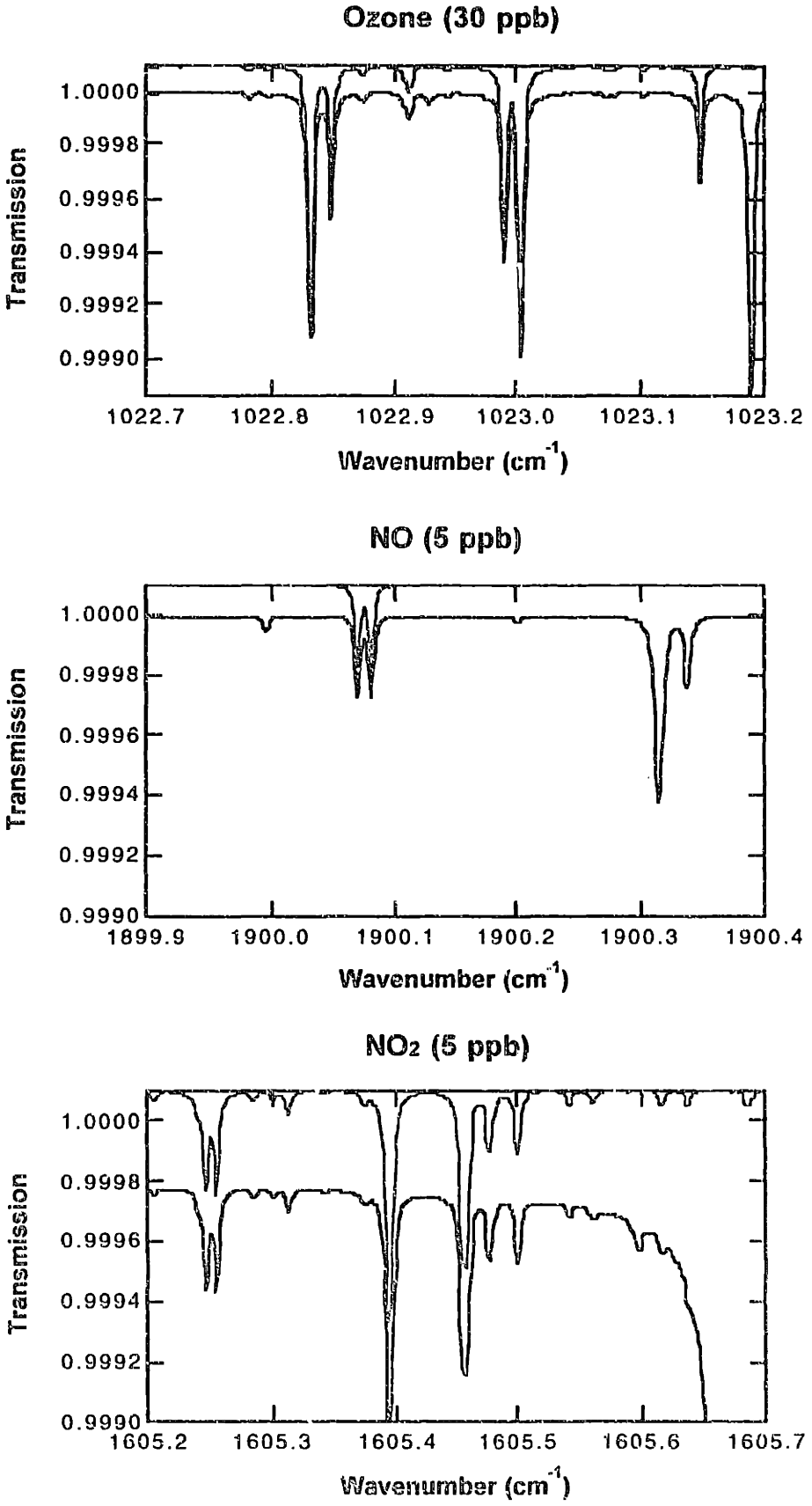


Figure 14-1. Hitran Simulations of Ozone, NO, and NO_2. (20 Torr and 100 m Pathlength)

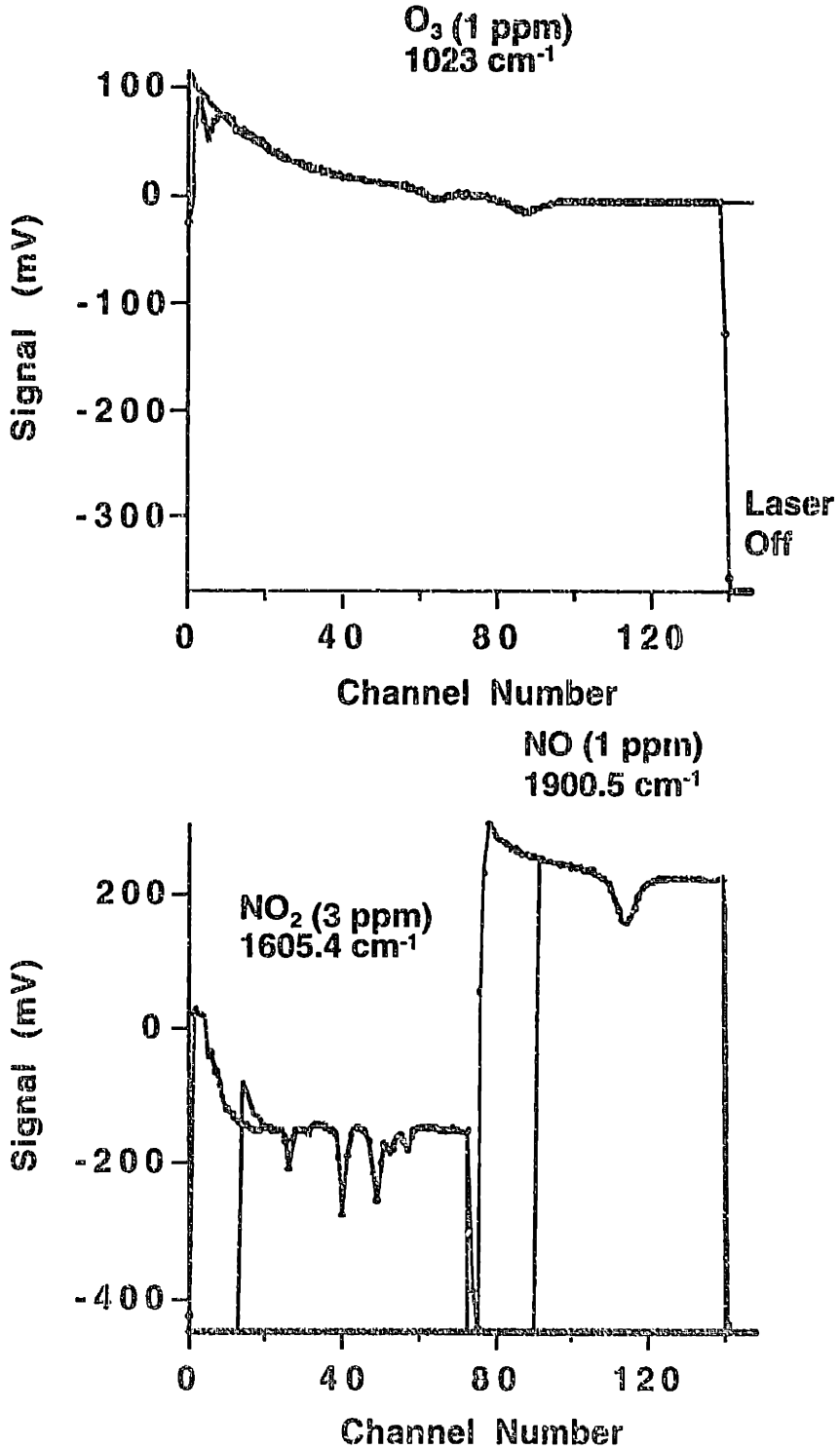


Figure 14-2 Observed Spectra for Ozone, NO_2 and NO.

When a series of absorption features occur close enough together that they can be observed in one laser scan, it is advantageous to fit the entire "fingerprint" region of the scan (Zahniser *et al.*, 1995b). The reference spectrum is obtained from the model calculations at

the same pressure and temperature as the sample spectrum. The line shapes and relative positions of the peaks on the scan are precisely determined based on the frequency tuning rate of the laser. The only unknowns are the exact position of the peaks, and the amplitude of the pattern. Confidence can be gained on the concentration determined from the fitted area under the peak when more features are utilized effectively. Fingerprint fitting was applied for the NO_2 feature at 1600 cm^{-1} in this study (the left hand side scan of Figure 14.2b). This approach was significantly more effective than fitting a single NO_2 line in separating the signal from the noise. The detection limit was reduced from 5 ppb to less than one ppb.

Three tunable diode lasers were used in this study for detection of ozone, nitric oxide, and nitrogen dioxide. These lasers were manufactured by Laser Photonics, Inc. The characteristics of the lasers are summarized in Table 14-2. Since the operating temperatures are above the boiling point of liquid nitrogen (77K), they are housed in liquid nitrogen dewars. Table 14-3 lists the corresponding detectors used in this study.

Laser Serial Number	Species	Wave number (cm^{-1})	Operating Conditions
4256-06	NO_2	1600	87.32K 300 mA
6218-06	NO	1900	91.5K 270 mA
509 -09	O_3	1020	87.32K 310 mA

Table 14-2. Tunable Diode Lasers Used in this Study.

Make/Model (Serial Number)	Light Detected	Peak Responsivity Wavelength	Cut Off Wavelength
Kolmar KV104-1-BM (3129-5PV)	NO, NO_2 signal 1900 cm^{-1} (5.3 μm) 1600 cm^{-1} (6.3 μm)	8.7 μm	9.4 μm
Judson (serial number not available)	NO, NO_2 reference 1900 cm^{-1} (5.3 μm) 1600 cm^{-1} (6.3 μm)	N/A	N/A
Judson J15D12-M204-S01M-30 (1547)	O_3 signal 1020 cm^{-1} (9.8 μm)	9.8 μm	12 μm
Kolmar KMPC11-1-BM (3129-3PC)	O_3 reference 1020 cm^{-1} (9.8 μm)	11.1 μm	11.8 μm

Table 14-3. Detectors Employed in the Study

14.2.2 Pump, Response Time, Operating Pressure.

A pump was connected to one end of the sampling cell, while the sampling line (teflon tubing) was connected to the other. The response time of the cell was the pumping rate divided by the volume of the cell. The sampling cell used in this study was connected to a Busch RB0021 pump with a nominal pumping speed of 14 cubic feet per hour (cfm) at standard pressure. At lower pressure, the pump performance decreased to 12.2 cfm at 23 Torr. The response time of the multipass cell was calculated to be 0.7 s. This was confirmed experimentally by fitting the tail of the response to a pulse ozone addition, as shown in Figure 14-3.

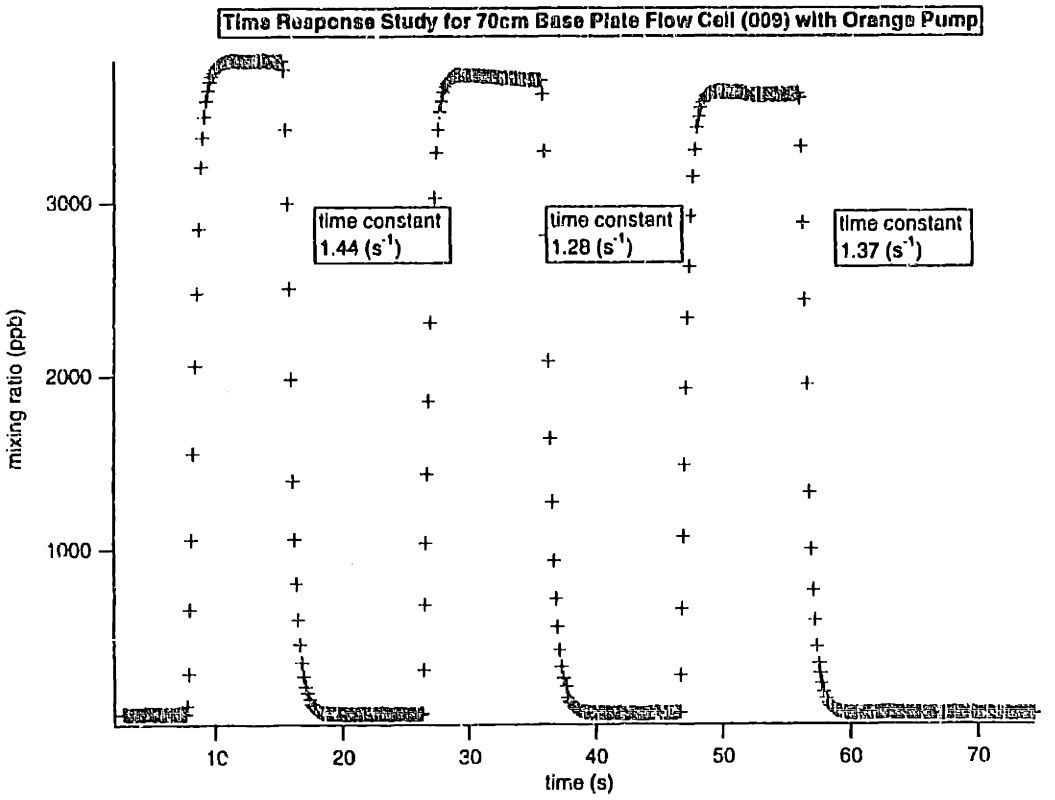


Figure 14-3. Pump Response Time

The operating pressure of the system was roughly 20 torr due to the drag on the flow within the 100 m teflon tubing from the roof of the Aerodyne building. The pressure is chosen as a compromise of height of signal vs. peak-baseline contrast for the three species that are measured in the same volume (as explained in Section 13.3). The sampling line also introduced a time delay of about two-seconds from the sampling port to the multipass cell. No significant chemical reactions was expected in this time.

14.2.3 Optical Layout and Alignment.

Two optical tables were used in this experiment. The experimental set up contained two separate TDL laser modules and a detection module. The first module was a single-TDL module referred to as the ozone module. It contained a dewar that housed the ozone diode laser and all the transfer optics used to send the ozone laser (1023 cm^{-1})laser into the multipass cell. The second module, the NO_x module, was a dual channel designed to combine the light from two lasers (NO and NO_2 lasers in this case) onto one single optical path, time-shared between the two lasers. The light was sent off the table and collected at the entrance to the multipass cell. The third module is the absorption and detection module and is located on the same table that houses the ozone module. This module contained the multipass cell, the detectors, and the transfer optics in between.

The current experimental design required two separate and independent optical paths in one absorption cell so that the pollutants can be measured in the same volume of air. The ozone laser was directed into the multipass cell in the horizontal entrance/exit plane. The NO/NO_2 laser beams, time-sharing the same path, were received from a separate optical table and manipulated to enter the absorption cell in a 90 degree rotated orientation (vertical entrance / exit plane). The resultant spot pattern on the astigmatic mirrors is shown in Figure 14-4. The beams follow different trajectories, each completing 182 passes and return to the same coupling hole as the entrance beam. The exit beams have approximately the same focal parameters as the entering beams, but at an angle of 0.1 radians (6 degrees) from the input beam. The output beams are focused onto their respective detectors located close to the absorption cell.

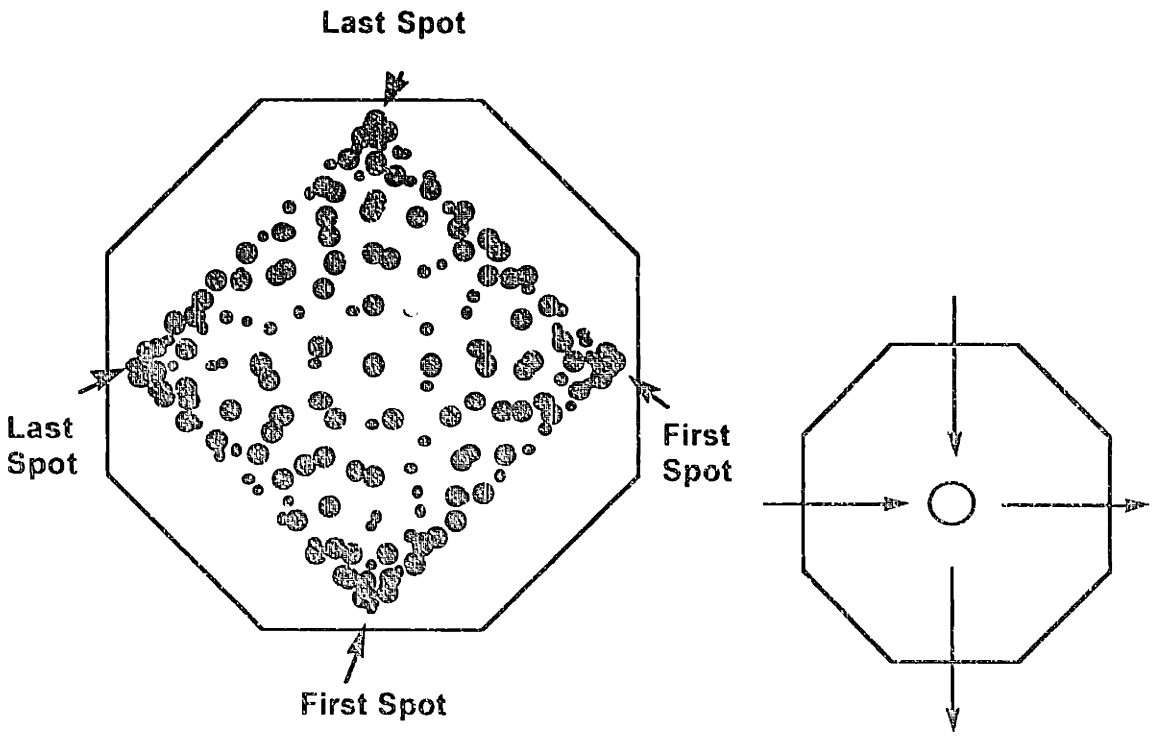


Figure 14-4. Spot Patterns (RHS figure: back of coupling hole)

Good alignment of the optical system was extremely important for maximizing the signal to noise ratio. The signal was increased by better focusing the light through the system so that more IR reached the detector. Laser-related sources of noise included optical feedback into the laser and interference from another laser beam leaking from the perpendicular optical path, rotated 90 degrees in the multipass cell. Since the internal alignment in the multipass cell is fixed, the optical elements that needed regular adjustments included the microscope objective that focuses the laser light, the aiming into the multipass cell, and the final focus onto the detector. Regular realignment was needed to assure that the laser goes through the pinhole and follows the designed optical path. Changes in alignment can be due to spatial changes in the diode emission, "creeping" of the positions of focusing optics, and changes in the optical path due to bumps, movements, and temperature-expansion changes of the transfer optics.

14.2.4 Signal Locking.

The laser's wavelength was generally fairly consistent for any given temperature and current. However, laser temperature control was not perfect, and a number of factors can cause drifts in the light frequency of the emitted laser. Since most laser scans were centered on the absorption feature of-interest, which is magnified to obtain a clearer signal (by using small current steps to scan the region), there were usually little room for the laser to drift before the absorption peak was not observed on the scan at all. It was necessary, therefore, to "lock" the scanning frequencies onto the signal for automatic and long term operation. Two methods were used in this study. In each of the two methods, a high concentration pulse signal was provided periodically. The computer software searched for location of the peak in the scan, and used that information in the subsequent spectral fits. If the drift was larger than some threshold, the computer would use a feedback control scheme to correct the drift by changing the current control parameters according to the difference between the peak's current and expected locations.

Reference Locking for NO and NO_2. For stable gases, the a reference cell can be used to provide the strong signal required for the computer to "find" the absorption peak. The NO / NO_2 TDL system used a 5 cm-long reference cell which contained high concentrations (ppm levels) of NO and NO_2 to generate a clear absorption signal. The reference spectra were recorded on a separate detector which received the laser light diverted by a beam splitter onto the reference leg (see Figure 13-8). At given intervals, usually every two to three minutes, the computer was instructed to analyze the signal from the reference detector, which it would used to locate the absorption peak. Then the computer would return to the sample channel, fixing the location of the absorption peak in fitting routine according to the peak location found in the reference channel. The program also made any necessary adjustments in the current parameters so as to keep the peak in the middle of the scanned wavelengths. When an adjustment was made, the reference locking procedure would record the new position of the peak before the computer returned to measuring ambient samples.

Flow cell reference locking for O_3. Ozone is a prime example of reactive gases that degrade in a closed cell because of chemical reaction on surfaces. In this experiement, Teflon sampling lines were used because of the inertness of ozone on Teflon surface. These compounds need to be produced each time they were required to provide a reference spectrum. In the case of ozone, a mercury lamp provided a source of short wave UV light at the wavelength range that dissociates oxygen. The resulting oxygen atoms quickly add to oxygen molecules in air, producing ozone. Using this idea, an ORION mercury lamp (model number 6035), emitting UV at 189 nm was kept in a semi-enclosed container to provide a

convenient source of ozone at relatively high concentration. A flow system, rather than a static one was used to obtain reference spectra for ozone. At specific intervals of two to three minutes, a valve was toggled open to introduce a burst of ozone from the reservoir container into the multipass cell. The computer program was instructed to go into "peak checking" mode after the valve opened, where it would stop recording ambient data. Using the peak location as a floating parameter in the least square fit for the spectrum, the computer would find the current location of the ozone absorption feature and fix the location parameter in the measurement to follow. After these tasks were completed, the valve to the ozone reservoir would close, and data-taking would resume after an extra data drop-out time (two seconds or so) to allow for the burst of ozone to clear the sampling cell.

14.2.5 Laser State (tuning rate, mode state, LN2 filling)

The success of continuous monitoring using the TDLAS hinges on the stable operation of the mid IR tunable diode laser. This section documents some of the more challenging aspects of the operation.

Tuning Rate. The tunable diode laser responds to both temperature and electric current changes. If the temperature of the diode is constant, the assumption can be made without large errors since the tuning rate (cm^{-1}/mA) is constant. In the current application, the tunable diode laser was turned off after every sweep through the absorption region. Sometimes, the wavelength region needed corresponded to high current modes. Modes that occur more than 100 mA above the threshold current were typically associated with higher laser power. This current produced a heating effect to the laser diode, so that when the current was dropped below the lasing threshold after each scan, the diode cooled. When the current was reapplied again, thermal effects dominated the tuning rate at the beginning of the next scan, causing the laser to tune more rapidly. After the laser relaxed to the operating temperature, the tuning rate became linear with current. This non-linear effect may dominate the beginning part of the scan or the entire scan. If the thermal effect was limited to the first part of the scan, the linear region was be used for line absorption determination. However, when a series of peaks were used, it might not be possible to avoid recording spectral peaks in the non-linear region. In this case, the tuning rate needed to be specified before any fitting routines could be applied successfully to the observed spectra. Based on the current understanding of the thermal and electrical tuning characteristics (Nelson, 1996), the tuning rate (in wavenumbers per display point) of the scan was fitted to the following functional form:

$$\text{Tuning rate} = a + b\exp(-cx) \qquad (14.1)$$

where x (channels) is the independent variable. In Equation (14.1), a (cm^{-1}/channel) is the linear tuning rate with current, c is the exponential decay from the thermal effect. The tuning rates at three points on the scan were used to re-calculate the coefficients regularly during continuous monitoring.

Mode state. Single mode operation was required for the accurate determination of the fractional absorption information calculated based on the "laser off" intensity. The simplest way to confirm single mode operation is to observe the extinction of light by a strong absorption feature, a "black" line, and compare the signal level to that when the laser was turned off. A monochrometer can also be used to separate light of different frequency from different laser modes, although the resolution of a monochrometer is limited to about 1

wavenumber, and competing modes may not be distinguishable. Multiple laser modes were more frequently observed at high currents than low. None of the lasers used in this study was operated in multiple mode conditions.

LN2 filling. One of the operational challenges to continuous monitoring using TDLAS is the need of liquid nitrogen (LN2) to keep the laser diode at low enough temperatures that allow it to operate. Typical liquid nitrogen dewar can hold about 12 hours' worth of liquid nitrogen, while a helium pump system would keep the laser cold for longer. Automatic liquid nitrogen filling systems were used. These systems included a tank of LN2 and a liquid level sensor, and will automatically fill the dewar system every time the level drops below some threshold level.

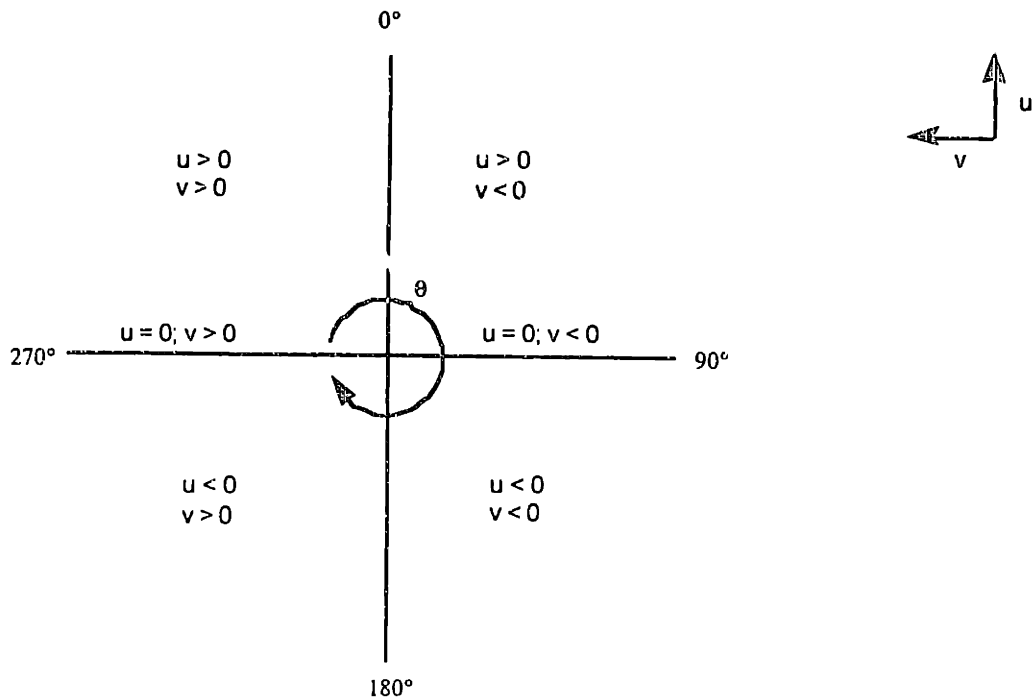
One of the dewars experienced difficulty in temperature control during LN2 auto filling. Even though the LN2 reservoir and the diodes were separated by vacuum, a rapid filling of LN2 by the auto fill system lowered the temperature of diode by radiative cooling. This temperature change was not regulated well by the temperature control circuit in the laser controller, because the temperature sensor was placed too far from the diode and protected against radiative cooling in the old design of the dewar. Therefore, temperature drifted after each auto fill, causing the laser to emit light in lower frequencies. The peak fit, being fixed at the previous peak location on the scan, captured a continually reducing portion of the peak area until the next reference lock procedure corrected the peak location and the injection current to compensate the temperature effect. Sometimes the peak drafted out of the scan, causing the program to crash because a fit was not found. The problem was communicated to the dewar manufacturer who improved the temperature control circuit design in newer dewars. No satisfactory scheme could be designed to correct these systematic drifts. They can be observed in the plots of data in the Appendix.

14.3 Micro-Meteorology Data

Micro-meteorology measurements were taken simultaneously with the TDL measurements of pollutant concentrations. These include wind, temperature, and UV radiation.

Sonic Anemometer. A sonic anemometer uses three pairs of sound transducers to measure wind speed in three directions. The wind speeds in the line of each pair of transducers can be deduced from the difference in the sound velocity in the forward and reverse directions. The transducers are arranged so that orthogonal wind speeds (u, v, w) can be either measured directly, or calculated from measurements. The output of the sonic anemometer also includes the instantaneous temperature, which can be deduced from the velocity of sound.

The output of the sonic anemometer is continuous. In this study, a RS232 signal was used to transmit the signal from the roof to the data logging computer in the laboratory. The data rate was 10 Hz. Each data stream contained three wind measurements in u, v, and w, and one temperature reading. The anemometer was aligned such that the w direction is exactly vertical, with zero mean wind speed. The horizontal winds, u and v, are perpendicular. If positive u is the forward direction, positive v is from right to left. Wind direction (in whole circle bearing) can be deduced from a knowledge of u, v, and the boom asmuth angle, which was measured to be 159 degrees. Figure 14-5 illustrates the calculation of wind direction from the horizontal wind speeds.



Boom Asmuth Angle = 159°

Wind Direction = θ+ 159°

Figure 14.5. Calculation of wind direction.

In addition to wind velocities, the local temperature was automatically derived by the sonic anemometer. The absolute calibration was only precise to about one degree, but fluctuations of 0.01 °C could be observed.

UV Pyranometer. The amount of UV radiation was measured using a radiometer (Model TUVR, Epply Laboratory, Inc.) The radiometer measures total UV radiation from 295 to 385 nm with a roughly-cosine response curve. This wavelength range corresponds to the absorption of photochemically active species such as NO_2 in the troposphere. An analog signal was transmitted from the pyranometer. An analog-to-digital board was used to convert the signal at a rate of 10 Hz. This data rate was chosen to match the sonic anemometer rate. The response time of the pyranometer was much faster than 0.1 seconds, and most changes in UV radiation in the local atmosphere takes place at a slower time scale. The TUVR instrument had a response of 2.01mV per mW/cm^2. The observed maximum analog voltage is about 10 mV. With a 16 bit A-to-D board whose full range is 0 - 1 V, the resolution of the signal was about 15 μV.

14.4 Data Storage

Three PC computers were used to take and store the data. A Quick Basic program was written to convert the analogue signal from the pyranometer and record the RS-232 signal from the sonic anemometer at the same time. This computer also received digital information from one of the TDL control computers and log the TDL output time and concentration as a duplicate for data synchronization purposes. Each TDL controlling computer also logged the time, concentration, and other relevant spectral information. Backups of all the measurements were made on a zip disk daily.

14.5 Quality Assurance

Before the raw data can be used in analysis, procedures were applied to correct or remove any erroneous readings that resulted from electronic glitches and peak locking procedures necessary for continuous operation of the TDL instrument. These procedures assured that the data made available to analysis were the best representation of the ambient concentrations being sampled. In addition to glitch removal, treatment of unevenly spaced data and time synchronization between the different computers used to take data are also discussed in this section. Re-sampling the raw data into a constant data rate stream enable the use of traditional routines for calculating autocorrelation and power spectrum information. Synchronization between the data streams was important for the analysis of covariance between different concentration streams, as well as between concentration and meteorological data.

14.5.1 Glitch removal

A computer program was written to analyze the raw data and to identify several patterns of the imperfect data generally termed "glitches." The cause of the glitches, especially in the ozone TDL system, and the corrective action taken by the computer program are outlined in this section.

Flow peak lock spike. Ozone was injected periodically into the multipass cell for the computer to locate the spectral peaks using an ozone pulse as a signal. During the reference peak lock, the signal was not recorded. In addition, there was an extra data drop out time corresponding to about three e-folding times of the multipass cell before data taking resumed, so that the pulse of ozone cleared the sampling cell. Usually, the three e-folding time was sufficient for the burst of ozone to decay to negligible levels. Occasionally, the concentration of ozone in the pulse source interfered with the first ambient reading after data taking resumed, particularly when ambient ozone concentration was small . In this case, the first data point after a peak lock was deleted.

Ozone burst. A burst of concentrated ozone was occasionally added to the flow cell by the operator to check the location of the peak as used in the fitting routine. It was assumed that any concentration above 300 ppb ozone was a result of ozone addition. These readings were also discarded.

One point glitch. Single point glitches were usually the result of erroneous fitting of the spectral scan by the computer, resulting in single points of negative readings amidst measurable concentrations. One point glitches were identified by the large difference (>35 ppb) between adjacent data points. These readings were also removed from the data stream.

Step glitch correction. A random step glitch occurred occasionally during the peak lock procedure, when the computer reset the current ramp that fed into the laser controller. An unknown electronic problem caused the current fed into the controller during the peak correction phase to be offset from that during the subsequent data taking phase. As a result, the laser frequency was inconsistent between the two periods and the peak position found during the peak locking procedure was different from the peak position of the spectra obtained during the actual data taking phase. The peak fitting routine, using the peak location found in the peak lock phase, fixed the spectral fit at the wrong position in the scan, and captured only a fraction of the peak area. This step glitch was characterized by its

coincidental occurrence with peak lock periods. A step decrease in concentration was observed during the entire data segment after the electronic glitch occurred. Then, the next peak lock corrected the position and caused a sharp increase in concentration, back to roughly the same level of ozone as before the glitch. It was assumed that a constant fraction of the area under the peak was captured by the best least square fit fixed at a different location. The fraction could be determined by the observed ratios of the concentrations observed during such a glitch to that observed before the peak lock and after the next peak lock, when the computer generally found the correct peak position again. All the readings obtained by misplaced spectral fits were increased by the multiplicative factor so determined.

Ramp glitch correction. A more annoying kind of glitch occurred during dewar filling in the ozone system. As discussed previously, one design flaw was the imperfect temperature control with respect to radiative cooling of the laser diode the there was a quick temperature change in the LN2 dewar environment. During such an event, the TDL computer may crash because it failed to fit the observed signal. Sometimes, a slow decrease of the concentration was observed, when the laser frequency drifted and the fit position was fixed. At present, there is no good way to correct for such a drift, as the dynamics seemed to vary according to the diode characteristics and the timing of the peak locking procedure at the time of the fill. Other factors such as the concentration of ozone observed, the rigor of the fill, the as well as the ambient temperature also seemed to affect the behavior of the system or the drift dynamics during the event. One way would be to determine a corrective factor much like in the case of a step glitch. Very little success was observed with such a correction procedure, and these ramp glitch or peak drifts were simply noted as they were found by the computer program.

14.5.2 Interpolating Non-uniform Data onto a Fixed-Interval Time Grid

Traditional methods for calculating correlation and spectral information (described in the next chapter) assume a regular interval, Δt, between data points. The TDL concentration data are not exactly uniform, because:

- the duty cycle of the measurement depended on the size of the signal (small fluctuations in the sampling rate)
- single point glitches that were removed
- data acquisition was interrupted by regular reference lock checks
- data acquisition was interrupted by comments.

This section outlines the procedures to deal with these non-uniformities.

Since most of the non-uniformities in the concentration data appear as data gaps, one viable solution would be to interpolate the data onto a fixed time grid that corresponds to the dominant sampling rate. Such a procedure is outlined below and is implemented in Matlab.

Find the dominant sampling rate. The sampling frequency of the TDL instrument could be specified in the laser control and data processing computer program. This rate was not always the same as the observed data rate, because data processing introduced a delay between detection and file output depending on the speed of the computer. The first step to generate uniform data from a data stream was to determine the dominant sampling rate. Two parameters were used to define the dominant sampling rate: fraction and tolerance. Due to the reference peak lock procedure, the TDL instrument sampled ambient air only 90 to 93% of the time. When the TDL instrument sampled ambient air, the data rate varied due to

variations in the duty cycle. The fraction parameter specified the minimum fraction of measurements with the same sampling rate before that rate could qualify as the dominant rate. Since the sampling rate fluctuated, the tolerance parameter defined the range of fluctuation allowed within the sampling rate, and was generally set to 10%.

Snap existing data onto uniform grid. After the dominant rate was determined, the existing data was re-sampled onto the new grid. Re-sampling could be achieved by interpolation of the data onto the time grid, or by simply snapping the closest data onto the time points on the grid. The latter method was chosen for this study, because the concentration data, though unevenly spaced, resembled more-or-less evenly spaced data with gaps.

Fill in single misses by interpolation. Single "misses" refer to one missing data point in the uniformly spaced data. There are at least two reasons for single misses. They could be due to single point glitches from the TDL data taking program, like a misidentification of peaks that led to unreasonable concentrations (e.g., negative concentrations that were deleted). Another possibility was that actual concentrations "snapped" to adjacent time points if the interval between samples happened to be larger than the sampling rate determined for the uniform grid. Single misses were filled in by interpolation, using an average of the two closest data points.

Use a predictor to fill in data gaps. An autoregressive approach was used to interpolate in longer gaps like those due to reference peak lock check. This approach was chosen because simple interpolation seldom result in satisfactory behavior in the power spectrum (Press *et al.*, Numerical Recipes in C, Chapter 13, 1992). The gap was filled in a manner consistent with the autocorrelative behavior of the data stream before it. A two sided predictor was used in which each point in the gap is determined by linear combinations of three points on either side of the gap. For example, if the gap was of length *ngap* starting at position *i*, as shown in Figure 14-6,

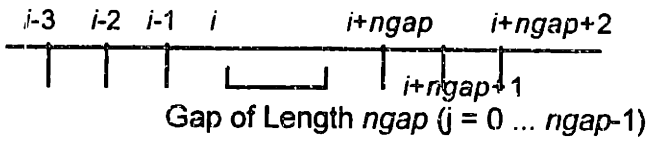


Figure 14-6. Illustration of Gap Fixing Technique.

The predicted concentration (c) of each point j within the gap was determined by Equation 14.2, where x_1 through x_6 are the weights assigned to $c(i\text{-}3)$, $c(i\text{-}2)$, $c(i\text{-}1)$, $c(i+ngap)$, $c(i+ngap+1)$ and $c(i+ngap+2)$.

$$c(i+j) = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{bmatrix} \cdot \begin{bmatrix} c(i-3) \\ c(i-2) \\ c(i-1) \\ c(i+ngap) \\ c(i+ngap+1) \\ c(i+ngap+2) \end{bmatrix} \qquad (14.2)$$

The weights ($x = [x_1\ x_2\ x_3\ x_4\ x_5\ x_6]$) of the points adjacent to the gap were determined by the autocorrelation coefficients at different time periods, with a constraint that the weights add to one, as shown in Equation 14.3.

$$Ax = b;$$
$$\text{with } x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 1 \qquad (14.3)$$

where A is a matrix of autocorrelation coefficients a(k) of k intervals, and ng is short for ngap, the length of the data gap.

$$A = \begin{bmatrix} a(1) & a(2) & a(3) & a(3+ng+1) & a(3+ng+2) & a(3+ng+3) \\ a(2) & a(1) & a(2) & a(2+ng+1) & a(2+ng+2) & a(2+ng+3) \\ a(3) & a(2) & a(1) & a(1+ng+1) & a(1+ng+2) & a(1+ng+3) \\ a(1+ng+3) & a(1+ng+2) & a(1+ng+1) & a(1) & a(2) & a(3) \\ a(2+ng+3) & a(2+ng+2) & a(2+ng+1) & a(2) & a(1) & a(2) \\ a(3+ng+3) & a(3+ng+2) & a(3+ng+1) & a(3) & a(2) & a(1) \end{bmatrix}$$

$$b = \begin{bmatrix} a(j+1+3) \\ a(j+1+2) \\ a(j+1+1) \\ a(ng-j+1) \\ a(ng-1+2) \\ a(ng-1+3) \end{bmatrix}$$

(14.4)

Numerical methods available for spectral estimates of unevenly sampled data are discussed in Chapter 16.

14.5.3 Synchronizing data from different computers

As discussed before, synchronization was an issue because data were taken on several computers and correlation between the different concentrations, and the concentrations and meteorological data were of interest in this study. Synchronization at the hardware level was preferable but difficult to implement. Therefore, each data stream was marked with the time on the respective computers. There were three different clocks: the ozone TDL computer, the NO_x computer, and the computer logging meteorological data. The concentration data were annotated periodically with a "time marker" input manually on both computers simultaneously, thus providing a correspondence between the clocks on the two TDL computers. The time-stamped ozone data was also transmitted from the ozone computer to be duplicated by the sonic anemometer logger computer with its own time stamp, correlating the clocks on the computers which recorded concentrations and meteorology. The time lag between the concentration data and the meteorological data due to the sampling procedure was determined by maximizing the correlation between the two. Using these information, the recorded data were synchronized to one another, including the sonic anemometer and UV pyranometer data.

15. Aerodyne Data Analysis

15.1 Field Deployment

The TDL instruments as described in the Chapter 14 were deployed in August, 1996. A four-day sample of data was analyzed in this study as an illustration of the utility of this type of data. The data analyzed were from August 20 through 23, 1996. Throughout that period, the weather was mostly sunny and clear. High temperatures in the afternoon were less than 90°F, with low relatively humidities. These were considered typical summer days in the New England area, and the ozone National Ambient Air Quality Standard was not exceeded during this period.

The TDL instruments took continuous measurements during the four-day period. During this period, the ozone base line noise was less than 0.0003 fractional absorption, which was experimentally determined to correspond to +/- 2 ppb for a steady ozone signal. The NO signal was estimated to be accurate to within 1 ppb. The NO_2 measurements were sensitive to 0.5 ppb. However, the 1600 cm^{-1} diode was very sensitive to temperature changes. Its high thermal tuning rate resulted in drifts in the laser frequency that were not effectively corrected by the peak locking procedure, which based on current control. NO_2 measurements were therefore most susceptable to errors caused by poor spectral fits by the data processing software. This was particularly problematic when the laser frequency drifted upwards so that the absorption feature was observed at the beginning of the scan where thermal turning dominated. The more rapid the tuning rate, the fewer the number of points in the absorption feature that were useful to the fitting procedure. During the day, operators could apply manual correction to the current and temperature control. When operating unattended, the main source of uncertainty for NO_2 concentrations was poor spectral fitting, which could miss up to 20% of the absorption features' area.

Species	Sensitivity	Ambient Concentration Range
Ozone	2 ppb	0 - 100 ppb
NO	1 ppb	0 - 150 ppb
NO_2	0.5 ppb	0 - 80 ppb

Table 15-1. Sensitivities of TDL Measurements

15.2 General Observations

Figure 15-1 shows the hourly concentration data for ozone, NO, and NO_2 during the period described. Plots of the raw data are included in the Appendix.

Ozone concentration displayed a cyclical behavior. It built up quickly during the morning after sunrise, reaching peak concentrations in the early afternoon. Although the peak ozone

concentrations occurred at different times during the afternoon on different days, ozone concentrations usually remained high throughout the afternoon. Because of the suburban site location, ozone was quickly removed by a titration reaction with nitric oxide (NO) from local traffic during the late afternoon/early evening rush hour.

Due to the proximity of the monitoring site to a local commuter route, daily NO peaks were observed in the morning at 6-7 a.m., corresponding to the morning traffic rush hour. High NO concentrations were also observed at night on August 21 and August 22, but not August 20. The low levels of pollutants in the evening of the 20th was associated with relatively strong winds. In evenings with relatively still conditions, such as the second and third nights of the data set, the accumulation of NO could reach instantaneous values of 100 ppb or more. This high night-time concentration may indicate the importance of highway traffic as a local source of nitrogen oxides in suburban areas.

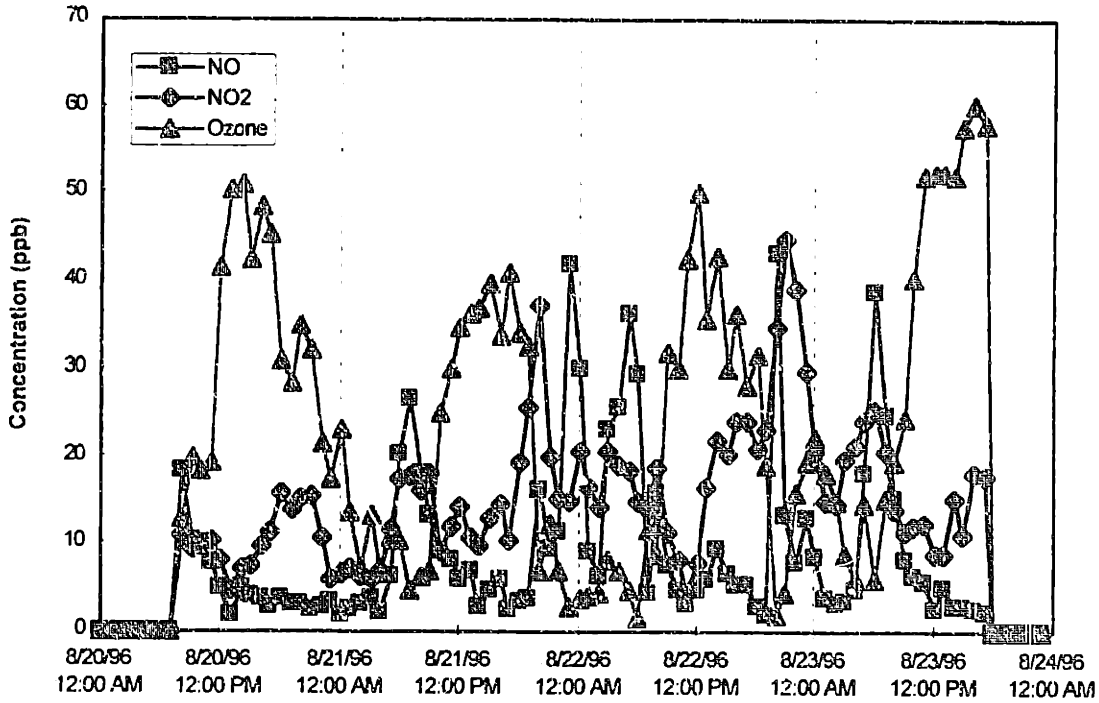


Figure 15-1. Hourly Concentration Data for Ozone, NO, and NO_2, August 20 - 22, 1996

Of the three pollutants, NO_2 exhibited the least amount of daily variability. A late afternoon peak was frequently observed. NO_2 was formed when NO from the afternoon traffic titrated ozone, but the destruction of nitrogen dioxide by the photolysis route was suppressed due to decreased sunlight in the late afternoon.

Even at this coarse time scale, anti-correlation between ozone and nitric oxide can be observed. The concentration of NO was usually low during the day when ozone was high. In the evening, and early morning, high NO concentration could be observed, when ozone concentrations were frequently low.

16. Time Series Analysis of Concentration Data: Ozone and NO_x

The analysis methods introduced in this chapter are selected from literature of signal processing, random data analysis, and statistical time series analysis.

16.1 Introduction

Before we set out to discuss the analysis of the 1 Herz concentration data and 10 Herz wind and solar data, some basic concepts need to be discussed. Figure 16-1 shows a general classification of data. Any observed data representing a physical phenomenon can be broadly classified as deterministic or non-deterministic (random). Deterministic data can be described by an explicit mathematical relationship with reasonable accuracy. Examples of periodic data are the vibrational motion of a slightly stretched spring and the voltage output of an electrical alternator. Non-periodic deterministic data can be close to periodic or transient in nature. Radioactive decays and step changes are examples of non-periodic data.

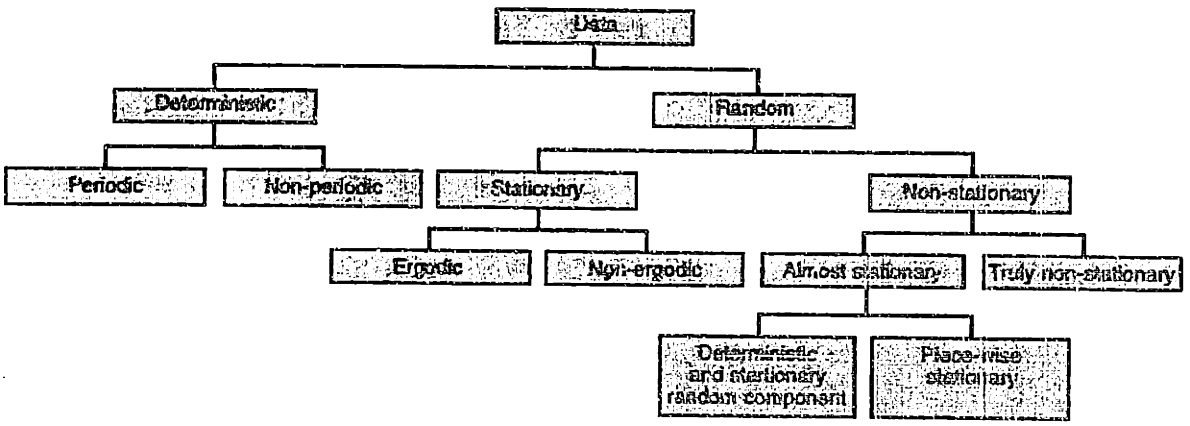


Figure 16-1. Classification of Data (adapted from Bendat and Piersol, 1986).

Random data are described in terms of probability statements and statistical averages rather than explicit equations. Experiments cannot be designed to produce identical results when repeated. Each time history representing a random phenomenon is a *sample record*, one physical realization of the *random process* or *stochastic process*, the collection of all possible sample records. The properties of a random process, such as the mean, variance, probability density functions, are described by the *ensemble average* calculated from a collection of sample functions. A special class of random process where all the moments and joint moments are invariant with time is *stationary random processes*. It is generally assumed that a process is stationary if its mean and autocorrelation function (defined later) do not vary with time. A stationary process whose ensemble properties (e.g., mean and autocorrelation) equal that of the time-averaged values is an *ergodic process*. Most random

data representing stationary physical phenomena are ergodic; therefore, meaningful properties can be determined from a single observed time history record.

The characterization of non-stationary processes require the analysis of ensemble averages unless further restrictions are applied. In many cases, the nonstationary random data produced by actual physical phenomena can be classified into special categories of non-stationarity that simplify the measurement and analysis. Examples are non-stationary processes {x(t)} described as follows:

$$x(t) = a(t)\, u(t), \tag{16.1}$$

where x(t) is a product of a deterministic time trend a(t) and a random stationary process {u(t)}. With piecewise stationary processes, properties can be estimated by stationary procedures as long as the data frequency is much faster than the change in the underlying trend.

16.2 Descriptive Properties of Random Data

Consider a discrete time series with N data points:

$$\{u_n\} \qquad n = 1, 2, \ldots, N \tag{16.2}$$

taken at intervals of Δt, such that

$$t_n = t_0 + n\,\Delta t \tag{16.3}$$

$$u_n = u\,(t_0 + n\,\Delta t) \qquad n = 1, 2, \ldots, N. \tag{16.4}$$

The length of the record is $T = N\,\Delta t$. The Nyquist frequency, which is the maximum frequency of a periodic signal discernible by the data, is

$$f_c = (2\,\Delta t)^{-1}. \tag{16.5}$$

16.2.1 Mean and Variance

The mean and variance, the first and second moments, are the most fundamental characteristics of a distribution of random data. The mean value of the sample data $\{u_n\}$, $n = 1, 2, \ldots, N$, is given by

$$\bar{u} = \frac{1}{N}\sum_{n=1}^{N} u_n. \tag{16.6}$$

The variance, or the second central moment, of the sample data is given by

$$s^2 = \frac{1}{N-1}\sum_{n=1}^{N} (u_n - \bar{u})^2. \tag{16.7}$$

Defining a transformed data set $\{x_n\}$ with zero mean, the variance is given by

$$s^2 = \frac{1}{N-1}\sum_{n=1}^{N} x_n^2, \tag{16.8}$$

where

$$x_n = x(t_0 + n\Delta t) = u_n - \bar{u}. \tag{16.9}$$

It should be noted that the mean value $\bar{u}$ and the variance s^2 are unbiased estimates of the ensemble mean μ and the ensemble variance σ^2 for ergodic data.

16.2.2 Probability Density Functions

This distribution of a sample of random data is represented by its probability density function (PDF). Consider a large enough sample of N data values, $\{x_n\}$, n = 1, 2, . . ., N, the PDF can be estimated by

$$\hat{p}(x) = \frac{N_x}{NW}, \tag{16.10}$$

where W is a narrow interval centered at x and N_x is the number of data values that fall within the range (x ± W/2).

16.2.3 Autocorrelation Functions

The autocorrelation function measures the time-related properties in the random data that are separated by fixed time delays. The autocorrelation function for a set of discrete data can be estimated by

$$\hat{R}_{xx}(r\Delta t) = \frac{1}{N-r}\sum_{n=1}^{N-r} x_n x_{n+r}, \tag{16.11}$$

where r = 1, 2, . . . , m, is the lag number. The maximum lag number, m, is assumed to be much smaller than N, the total number of records. The factor of (N - r) is used to provide an unbiased estimate.

An alternative method exists to calculate the autocorrelation function via fast fourier transform of the data.

16.2.4 Autospectral Density Functions

Spectral analysis seeks to describe the frequency content of the random process, based on a finite sample of data. An autospectrum is also called a power spectrum because the energy content between two frequencies is equal to the area under the spectrum between the frequencies. Consider a record of length T_r, divided into n_d segments of length T, a "two-sided" (symmetric) autospectral density function is estimated by

$$\hat{S}_{xx}(f) = \frac{1}{n_d T} \sum_{i=1}^{n_d} |X_i(f,T)|^2. \qquad (16.12)$$

$X_i(f, T)$ is the fourier transform of x_i, defined as

$$X(f,T) = \int_0^T x(t) e^{-j2\pi ft} dt. \qquad (16.13)$$

In the discrete case, $x_n = x\ (n\Delta t)$, n = 0, 1, 2, . . ., N-1, and the fast fourier transform can only be calculated at discrete frequencies as follows

$$X(f,T) = \Delta t \sum_{n=0}^{N-1} x_n \exp[-j2\pi fn\Delta t]. \qquad (16.14)$$

It can also be proven that

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) \exp[-j2\pi f\tau] d\tau, \qquad (16.15)$$

or

$$S_{xx}(f) = \sum_{r=-\infty}^{\infty} R_{xx}(r\Delta t) \exp[-j2\pi fr\Delta t] \qquad (16.16)$$

for the discrete case. The spectrum estimate as defined by one term (the fourier transform of just one segment) in the summation in Equation 16.12 is called a periodogram. The problem with periodogram estimates is that the variance of the estimate is large. Therefore, averaging results from non-overlapping segments is desired. Overlapping segments can also be used, although the correlation of the data segments reduces the variance benefits gained from having more segments.

Another way to improve the periodogram estimate is by applying a non-rectangular data window to sections prior to computing the periodogram. This procedure, called time series tapering, reduces the "spectral leakage" and increases the width of the spectral peaks. A cosine squared window, called Hanning window, can be used in conjunction with overlapping of data segments to produce a low-variance estimate of the spectrum.

16.2.5 Autospectral Functions of Unevenly Distributed Data

Numerical methods also exist for the analysis of unevenly sampled data, such as missing data or irregular intervals between samples. In a previous chapter, a technique was discussed to "fill the gaps." Filling data gaps using autocorrelation coefficients has an advantage over simple interpolation techniques, which create artificial correlation at short time scales. This section draws from Numerical Recipes in C (Press *et al.*, 1992, Chapter 13) which discusses various routines for spectral analysis. Using Lomb's normalized periodogram (Section 13.8 in Numerical Recipes), spectral power as a function of angular frequency ($\omega = 2\pi f$) is defined by:

$$P_N(\omega) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_j (h_j - \bar{h})\cos\omega(t_j - \tau)\right]^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{\left[\sum_j (h_j - \bar{h})\sin\omega(t_j - \tau)\right]^2}{\sum_j \sin^2\omega(t_j - \tau)}\right\} \tag{16.17}$$

where

$$\bar{h} = \frac{1}{N}\sum_{i=1}^{N} h_i,$$

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(h_i - \bar{h})^2, \tag{16.18}$$

$$\tan(2\omega\tau) = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}.$$

Equation 16.17 is identical to the equation that would be obtained by estimating the harmonic content of a data set, at a given frequency ω, using linear least square fitting to the model:

$$h(t) = A\cos\omega t + B\sin\omega t. \tag{16.19}$$

This method weighs the data on a per point basis rather than a per time interval basis, which is a distinct advantage when treating unevenly sampled data.

16.2.6 Cross Correlation Functions / Cross Spectral Functions

Cross correlation functions measure the association between data points from two different data streams that are r time intervals apart. Its definition is similar to the autocorrelation function defined in Equation 16.11:

$$\hat{R}_{xy}(r\Delta t) = \frac{1}{N-r}\sum_{n=1}^{N-r} x_n y_{n+r}. \tag{16.20}$$

The sample cross correlation function $\hat{R}_{xy}(r\Delta t)$ may be normalized to have values between -1 and 1 by division by $\sqrt{\hat{R}_{xx}(0)}\sqrt{\hat{R}_{yy}(0)}$. This defines a sample cross correlation coefficient function:

$$\hat{\rho}_{xy}(r\Delta t) = \frac{\hat{R}_{xy}(r\Delta t)}{\sqrt{\hat{R}_{xx}(0)}\sqrt{\hat{R}_{yy}(0)}}, \qquad r = 0,1,2,\ldots,m. \tag{16.21}$$

A high absolute value of $\hat{\rho}_{xy}(r\Delta t)$ indicate high correlation. For data streams of zero mean, $\hat{\rho}_{xy}(0)$ is equivalent to the widely used linear correlation coefficient, r, for pairs of quantities (x_i, y_i):

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}. \qquad (16.22)$$

As in the case of the auto-correlation function, an FFT procedure can also be applied to calculate the cross-correlation function.

A cross spectral density function is defined analogously to the power spectral function shown in Equation 16.12, and is related to the cross correlation function in the same manner as shown in Equation 16.16. For further details, please refer to Bendat and Piersol, "Random Data Analysis and Measurement Procedures, Second Edition, 1986.

16.3 Statistical Analysis of the Concentration Data

Ambient data of pollutant concentrations are by no means stationary. Except for a few highly stable species, atmospheric concentrations of trace species are strong functions of time, as shown in Figure 15-1. This section focuses on the concentration data and various statistical descriptions of the concentration data. The moment statistics and distribution information of the non-stationary data are interesting in light of widely used sampling procedures and air quality standard descriptions, both of which are in the form of hourly averages. No assumption was made that the data is piece-wise stationary or ergodic in nature. This issue is addressed further in the discussion of autocorrelation and spectra in Section 16.3.3.

16.3.1 Mean and Variance

The properties of a collection of measurements are often quantified by the first and second moments, which are the mean and variance of the distribution. The following three figures present the mean and standard deviation (square root of variance) information of NO, ozone, and NO_2, measured during the third week of August, 1996. In each of the graphs, the mean values are indicated by the marker, and the standard deviation of the hourly data are represented by the bars.

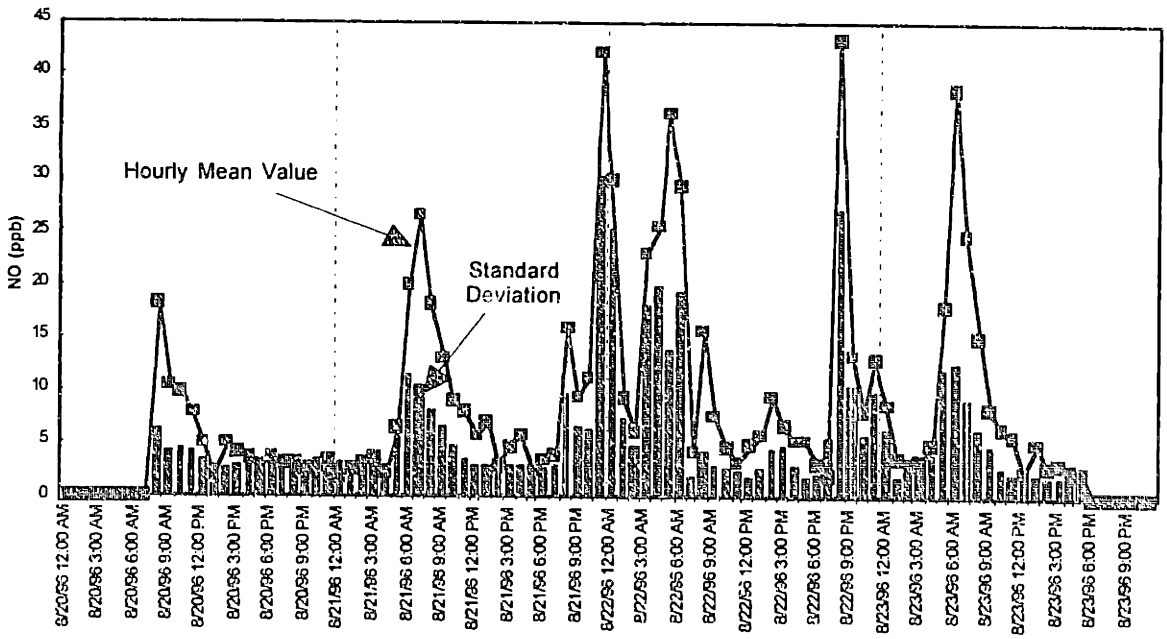


Figure 16-2. Mean and Standard Deviations of Hourly Concentration Data for Nitric Oxide.

Of the three trace species measured, the concentration of nitric oxide was the most variable during the course of a day. Morning peaks in NO concentrations corresponding to traffic emissions were observed everyday of the observation period. The afternoon rush hour was not registered as clearly in the hourly averaged NO observations. During the evenings of the second and third day of the monitoring period, spikes of very high nitric oxides

concentrations were observed in the NO data in Figure 16-2. In fact, sudden spikes of high nitric oxide were also evident in the smaller time scale of individual measurements on those two nights. The difference in the NO behavior seemed to be a result of the mixing conditions. The evening of August 20 was relatively windy compared to the next two nights. The average wind speed was 0.8 ms^{-1}, with gusts of up to 3 ms^{-1}. The next two nights were practically still, with average wind speeds of no more than 0.4 ms^{-1}. Still conditions probably enhanced the observed source effects from trucks on the nearby highway, because pollutants accumulated locally when the mixing condition was poor.

Higher concentrations were frequently associated with larger standard deviations. Standard deviations were typically 70 - 80% of the observed mean, with a minimum value of 2 ppb or so. Due to the low mean data observed throughout the day, the daytime standard deviations probably contained a significant contribution from the intrinsic error of the TDL measurement (1 ppb). One noticeable exception to the tracking of mean and standard deviation was during the morning rush hours, when the high averaged concentration was accompanied by a proportionately smaller increase in the standard deviation. This might be an indication that high nitric oxide was quite uniformly distributed in the volume of air sampled in the morning. Late evening peaks of NO were associated with higher variances compared to morning traffic peaks, indicating a high degree of variability of NO concentration from one sample to the next. High variability of NO relative to the mean was consistent with strong source effects and limited mixing of air parcels during the night.

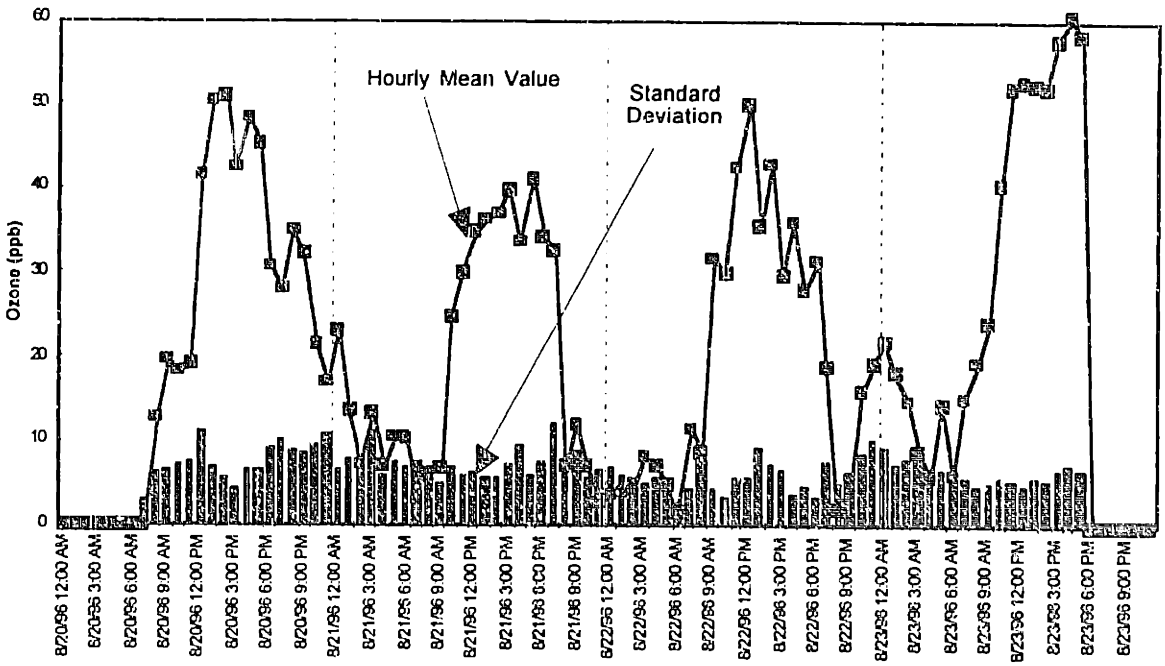


Figure 16-3. Mean and Standard Deviations of Hourly Concentration Data for Ozone.

Ozone exhibited a strong diurnal cycle, with peak concentrations occurring in the afternoon. However, the time of peak ozone concentration varied from noon to 4 p.m. during the monitoring period. Unlike the observations for NO, the standard deviation of ozone, a secondary pollutant, was not observed to have a strong dependence on the observed mean value. The standard deviation was about 7 ppb for ozone, with a minimum observed standard deviation of 2.7 ppb, and a maximum of 12.4 ppb. The magnitude of ozone fluctuations seemed slightly cyclical, with relatively high variability observed in the late afternoon or

early evening. It was postulated that high variability in ozone measurements were associated with relatively still conditions that stifled mixing after the sun went down, as well as the rapid change of ozone concentration due to chemical reaction with NO from the local traffic. When air parcels were not well mixed, a point measurement sampled a large range of concentrations. Figure 16-4 illustrates this point further. A comparison was made for the ozone measurements on September 15, 1994 and September 16, 1994. A different experimental setup was used at that time (background subtraction to reduce fringe noise) which enhanced the sensitivity of the ozone TDL instrument to 0.5 ppb.

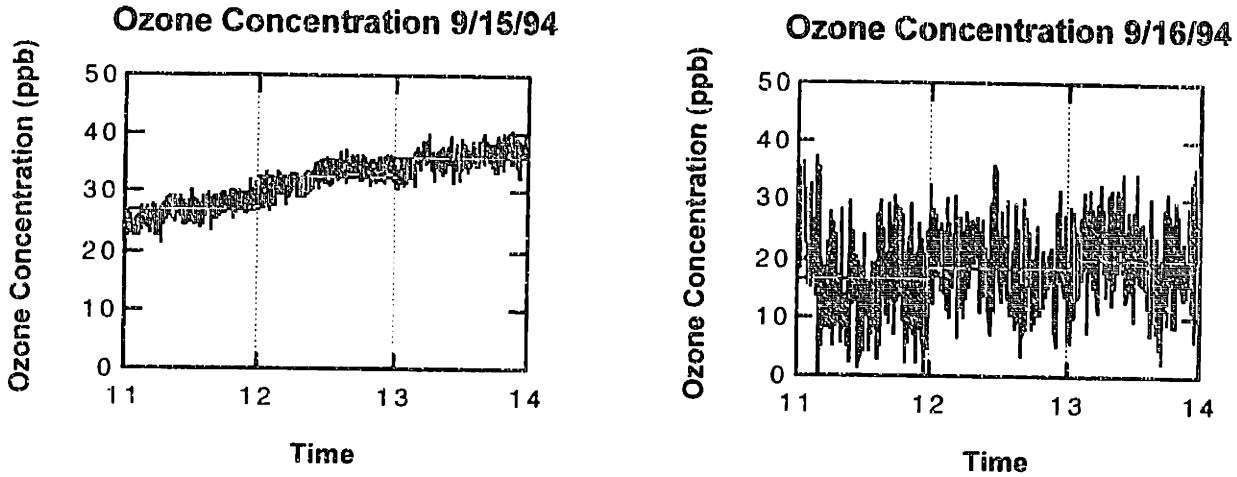


Figure 16-4. Ozone Fluctuation Levels on Two Consecutive Days in September, 1994.

September 15, 1994 was sunny and warm (19°C) with moderate winds (1.4 ms^{-1}) out of the NE, while the second day was overcast and cooler (17.3°C) with lighter winds (1.0 ms^{-1}) from the SW direction. With conditions more conducive to mixing, the ozone measurements taken during the morning and early afternoon on September 15 exhibited little fluctuation (< 2ppb standard deviation in one hour, very close to the effect of the rising trend alone). On the second day, the fluctuations were higher (with a standard deviation of 7 ppb, or 30% of the mean) due to meteorological effects.

The mean and standard deviation associated with NO_2 measurements are shown in Figure 16-5. Of the three trace gases, NO_2 exhibited the least amount of daily variability in terms of absolute amounts. Peak concentrations of NO_2 tended to occur in the early evening, while minimum concentrations were observed in the afternoon.

The variability of NO_2 seemed to combine characteristics of NO, a primary pollutant, and ozone, a secondary pollutant. While there was a tendency for the standard deviation to vary with the hourly mean value, the variability of NO_2 measurements was small compared to that observed for nitric oxide, both in absolute terms and in relative terms to the mean values. The fluctuations associated with NO_2 measurements exhibited a diurnal cycle, much like the one for ozone measurements. High variability were observed at night and relatively low standard deviations were observed during the day. As in the case of ozone, the fluctuations at night were probably due to meteorology unfavorable for mixing.

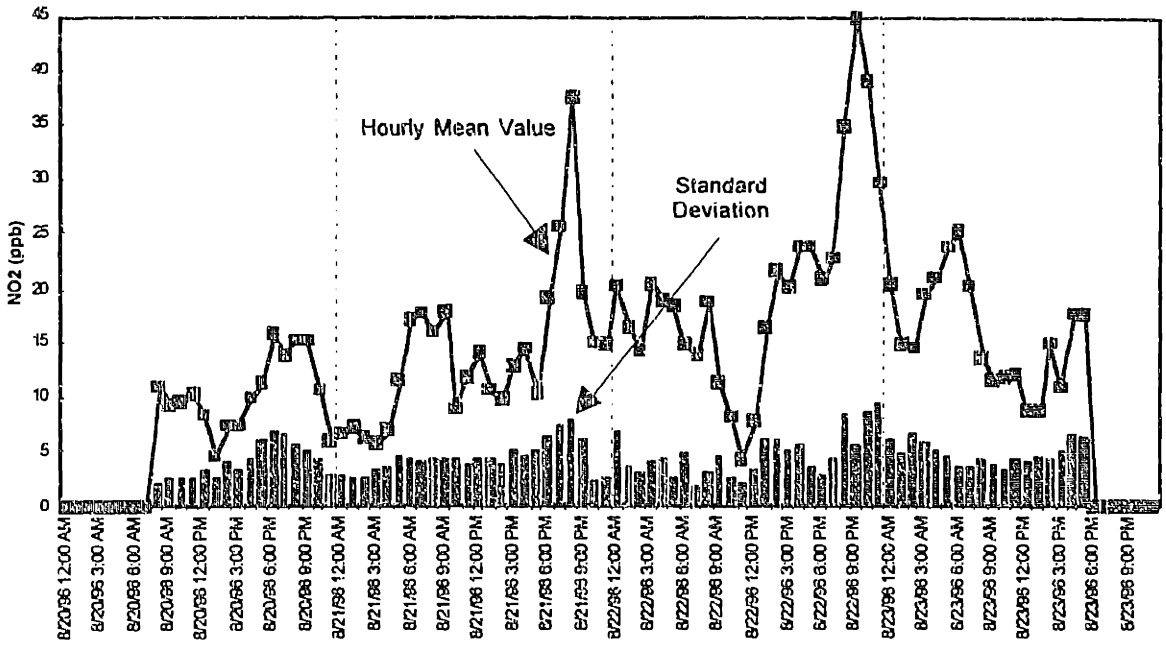


Figure 16-5. Mean and Standard Deviations of Hourly Concentration Data for NO_2 (ppb).

16.3.2 Probability Density (Histogram)

The frequency distributions of short time-scale measurements may give additional insights to the variability of field measurements and the suitability of an hourly average as the sole reporting format. To make the histogram, the range of observed measurements was divided into 50 bins, and the number of samples recorded in each bin determined the histogram. Therefore, the width of the bins was related to the range of the observations. Negative readings were observed when the ambient concentration fell below the detection limit, or they could be a result of an offset due to shifts in the shape of the baseline. Though physically impossible, negative concentrations were not removed by the histogram procedure, doing so would bias the result.

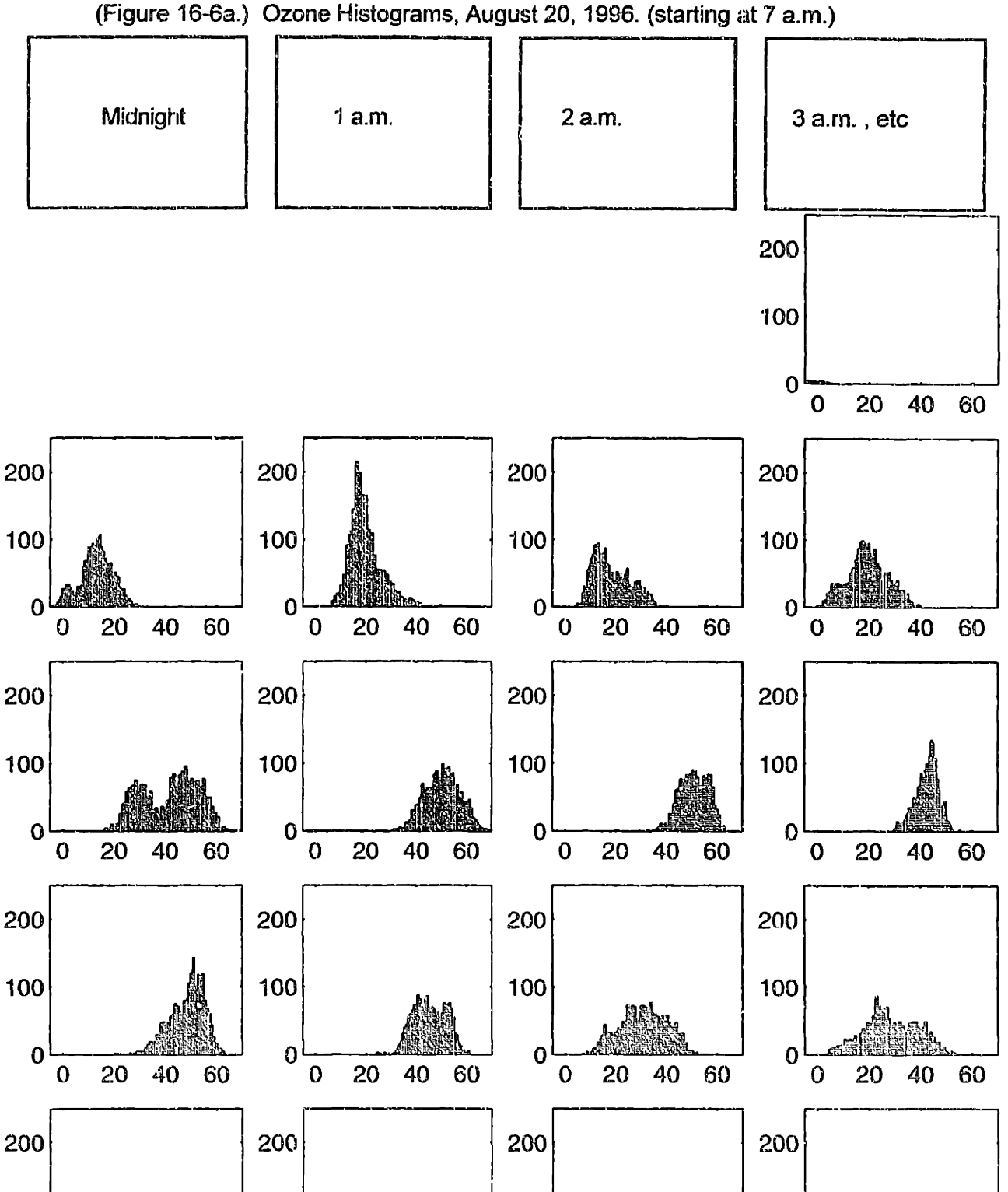
Plotted in Figures 16-6 (a to d) are frequency distributions of ozone concentrations associated with each hour of the day on August 20 through August 23, 1996. A few ozone histograms showed negative concentrations, particularly those during the evening.

The key observation made was that the variability of atmospheric pollutant observed at a stationary monitor was not always represented well by a Normal distribution. Some of the distributions were symmetric, others were skewed one way or another, still others were bimodal. No general trends were found regarding the shape of the histograms, with the time of the day or with the concentration of ozone. The histograms associated with hourly measurements seemed to vary significantly from day to day, and from hour to hour. A bell-shaped distribution could be followed by a skewed one during the next hour, and many bimodal distributions were observed one day and none the next.

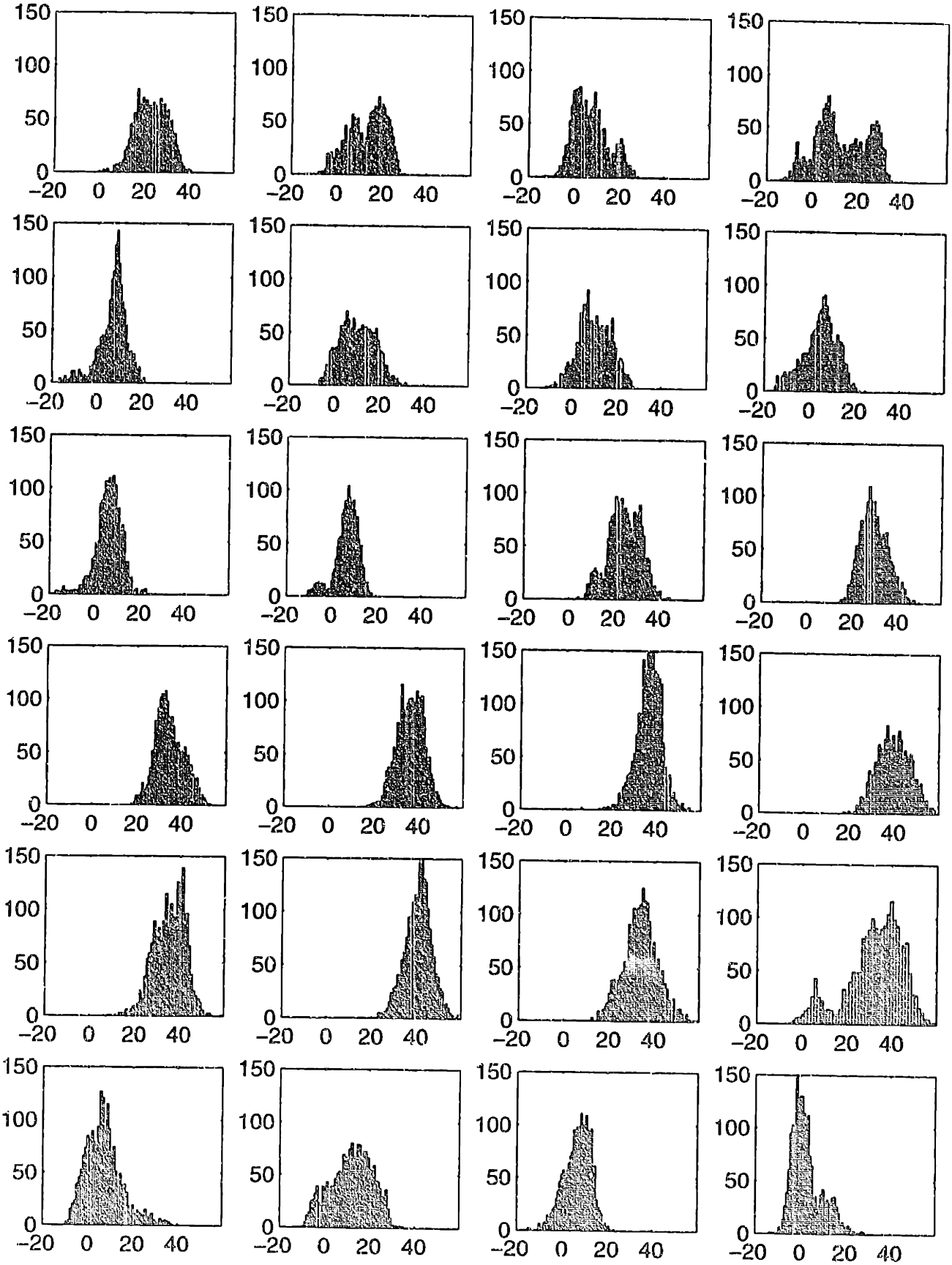
About one quarter of all recorded histograms of hourly frequency distribution of ozone concentration were bi-modal or multi-modal. Bi-modal and multi-modal distributions occurred more frequently on some days than others, and were postulated to be related to the meteorological conditions in the general vicinity of the site. An analysis was conducted to investigate if any relationship could be found between the dominant modes of ozone concentration and meteorological parameters measured at the site (described in the next chapter). None was found. Multi-modal measurements were observed at different times on different days (with the exception of August 23, 1996, when none was observed). In most instances, the observed modes were significantly far apart, with differences of 20 - 30 ppb. There were a few relatively narrow distributions observed on August 22: the differences in the concentration modes at 9 am and 5 p.m. were less than 10 ppb.

Several families of distributions were compared to the observed histograms with a single mode. These distributions included Gaussian, gamma, lognormal, and Weibull distributions. The parameters of the distributions were determined by moment matching or other means described in standard text books (e.g., Morgan and Henrion, 1992). The Matlab™ scripts written for this analysis were included in the Appendix. The distribution fits are not shown here. None of the four distributions named proved satisfactory in representing many of the distributions observed. Qualitatively, or visually, most single mode distributions could be roughly approximated by Normal distributions, although statistical chi square tests (Press *et al.*, 1992, Chapter 14) suggested that few frequency distributions were in fact Gaussian with the same first and second moments. None of the other distributions tested fitted the observed histograms any better than the Gaussian or gave better chi squared values. Therefore, the mean value and standard deviation were probably still the best choice of parameters to characterize the fluctuations of measurements in a period of an hour provided the sample distribution had only one mode.

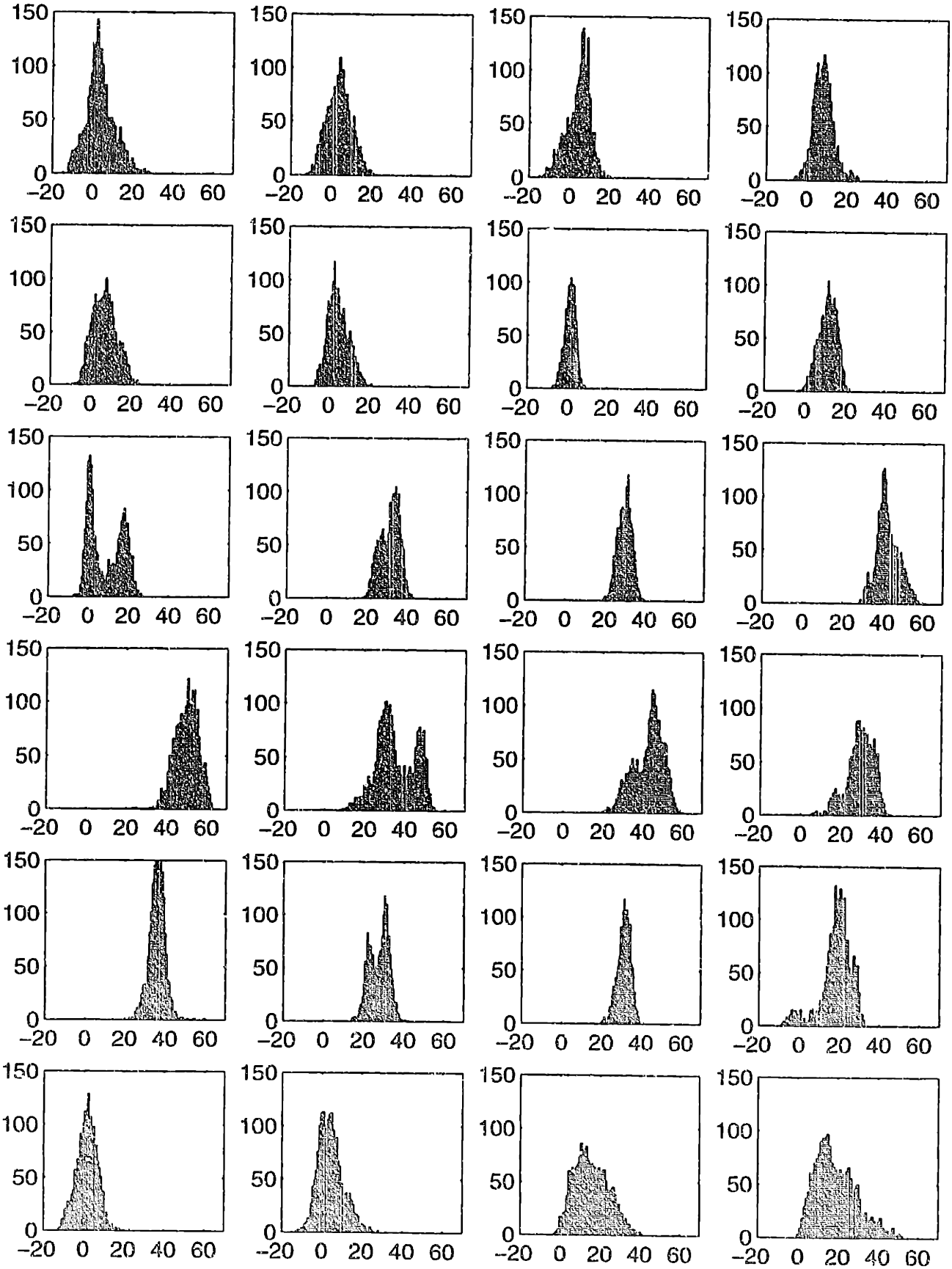
(Figure 16-6a.) Ozone Histograms, August 20, 1996. (starting at 7 a.m.)



(Figure 16-6b.) Ozone Histograms, August 21, 1996



(Figure 16-6c.) Ozone Histograms, August 22, 1996



(d.) August 23, 1996.

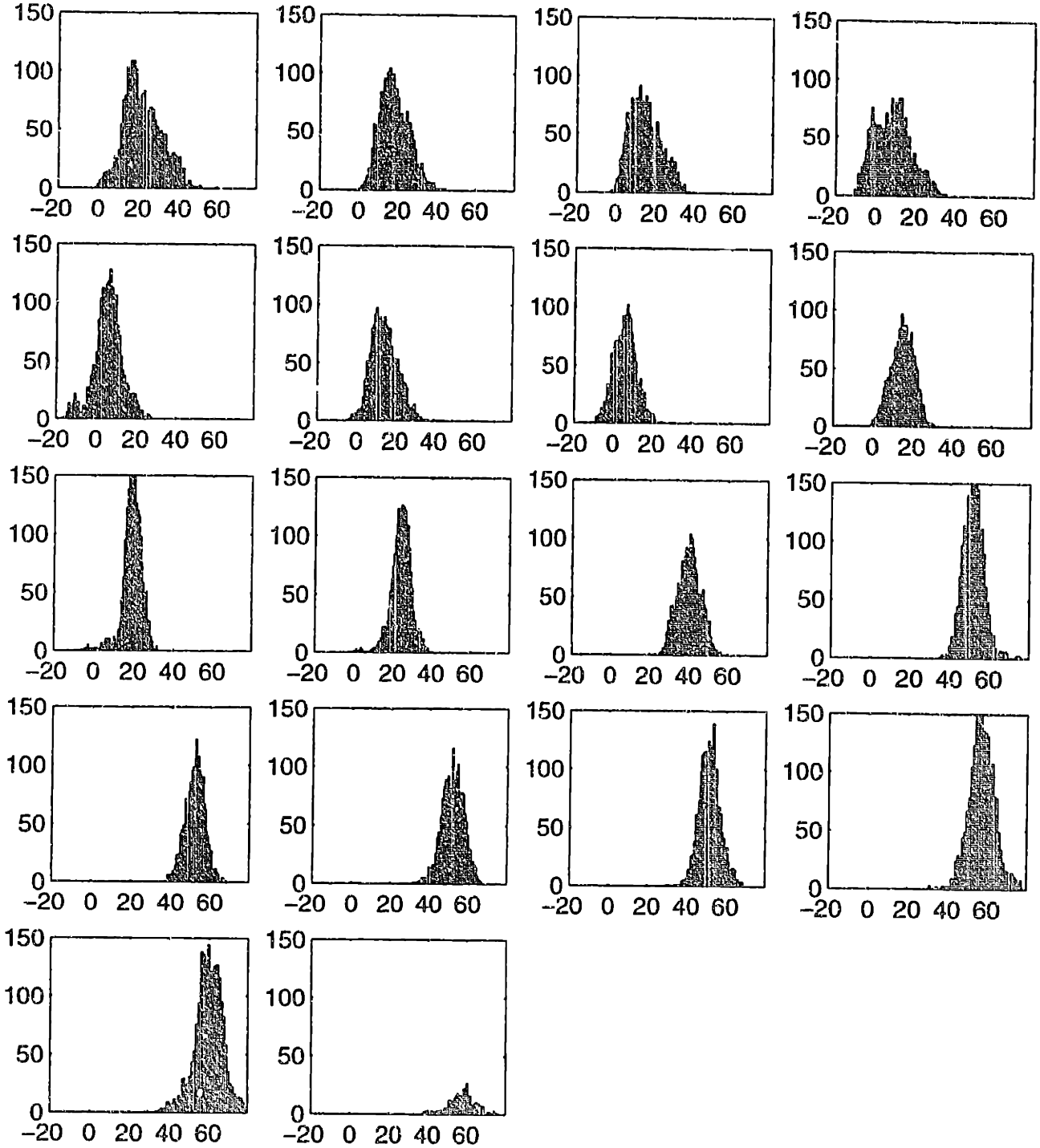


Figure 16-6. Hourly Ozone Histograms, August 20 to 23, 1996.
(Top row, from left to right: midnight, 1 a.m. ,2 a.m. ,3 a.m.
Second row, from left to right: 4 a.m., 5 a.m., 6 a.m., 7 a.m.
etc.)

Figure 16-7 shows the histograms of ozone measurements on August 20 to 23 at the time of the peak hourly concentration. As discussed previously, the maximum hourly average occurred on at different times on different days. All four histograms appeared to have only one mode. However, the shapes of the histograms were very different. None of the four observed distributions could be reasonably approximated by any standard forms of distributions tested. The two distributions from early afternoons were wider that the Normal and the two from later in the afternoon were narrower.

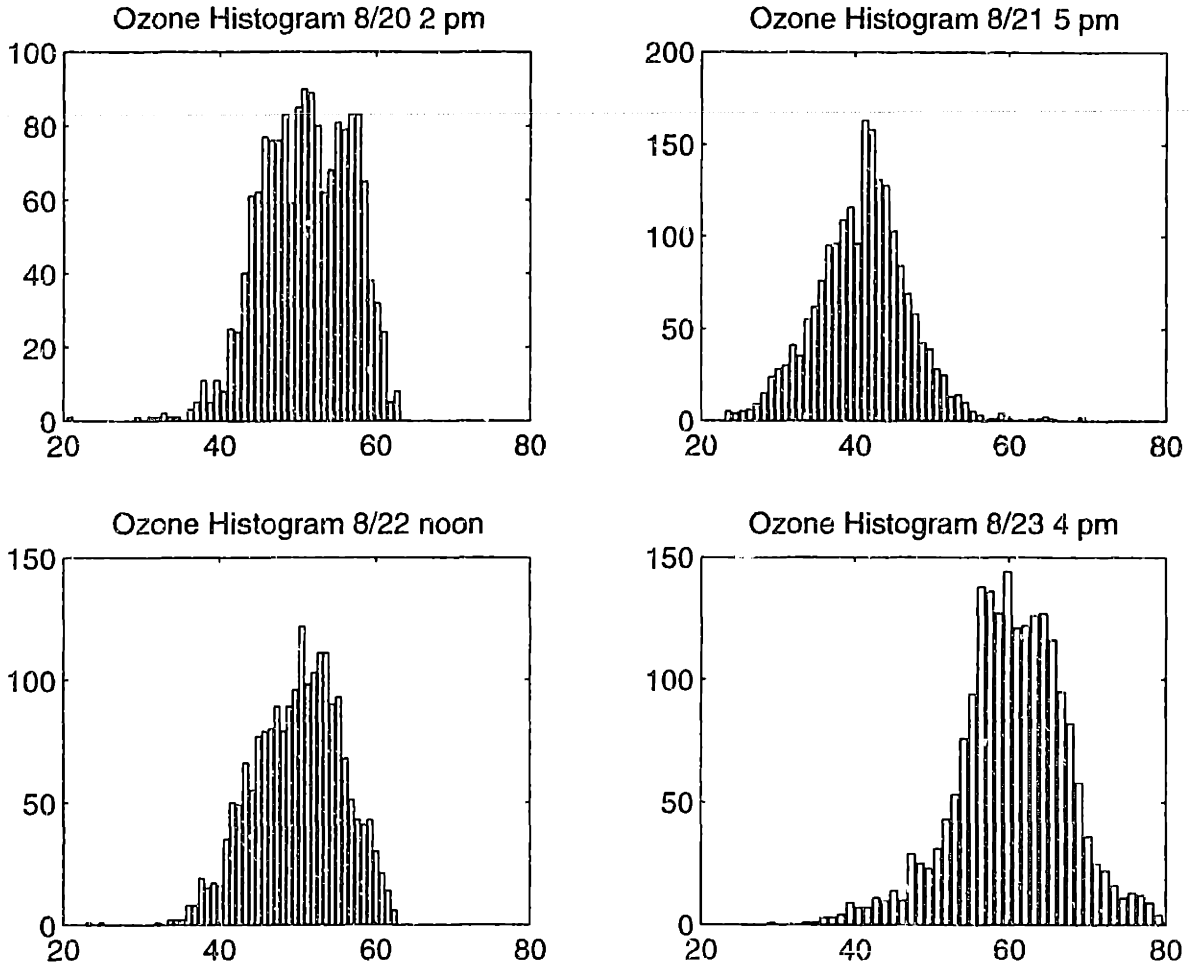


Figure 16-7. Ozone Histograms at Peak Ozone Hours.

Bi-modal distributions with two distinct modes that separated by more than 20 ppb could be reasonably represented by two normal distributions, with mean values at the modes. One example is shown in Figure 16-8. The hourly mean in this case was 41.7 ppb. and the two fitted gaussian had modes of 28.4 and 48.5 ppb respectively. When the modes are closer than 10 ppb together, or when multiple modes were present, this approach failed to provide a satisfactory representation.

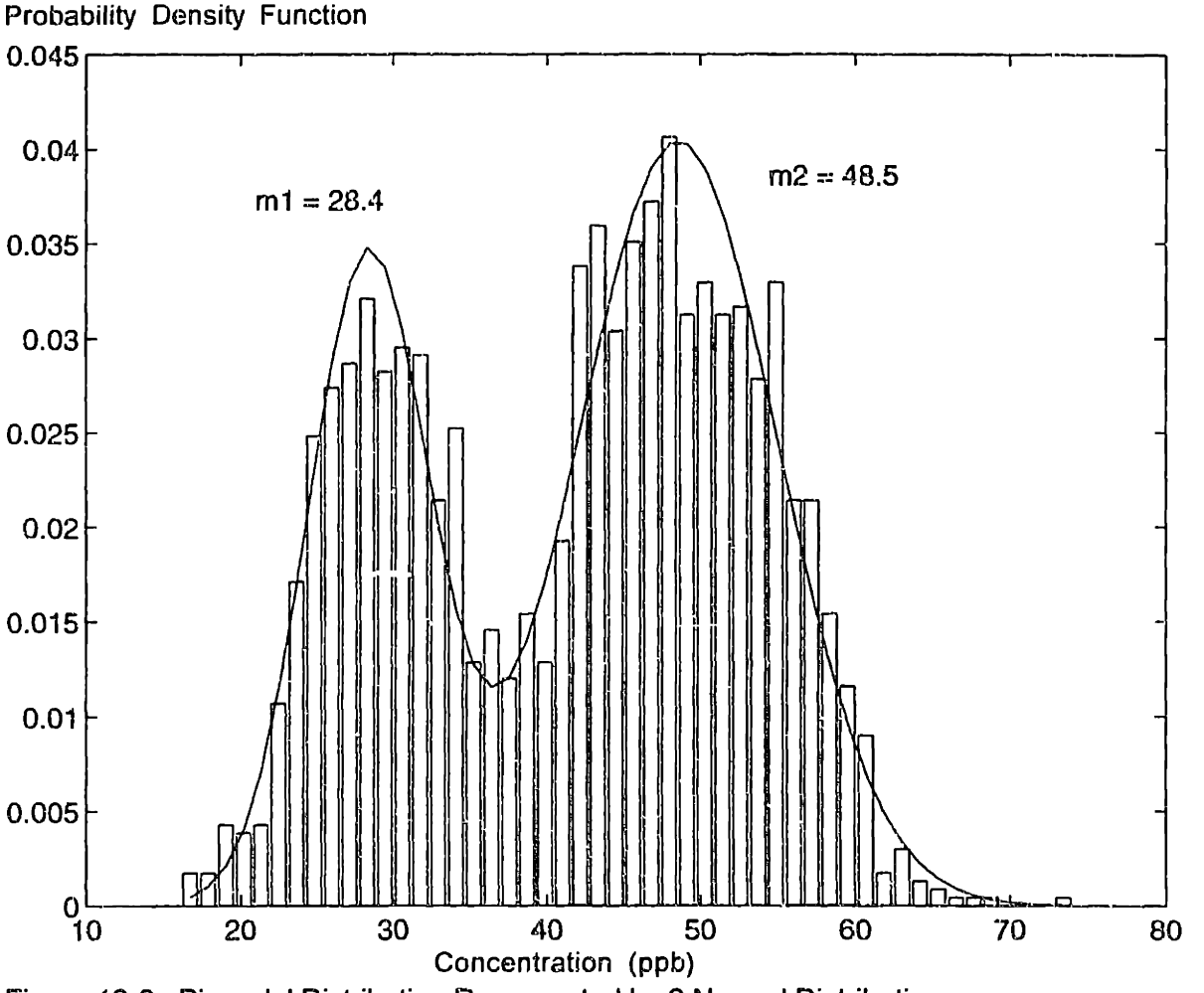


Figure 16-8. Bi-modal Distribution Represented by 2 Normal Distributions. (8/20/96 12 p.m.).

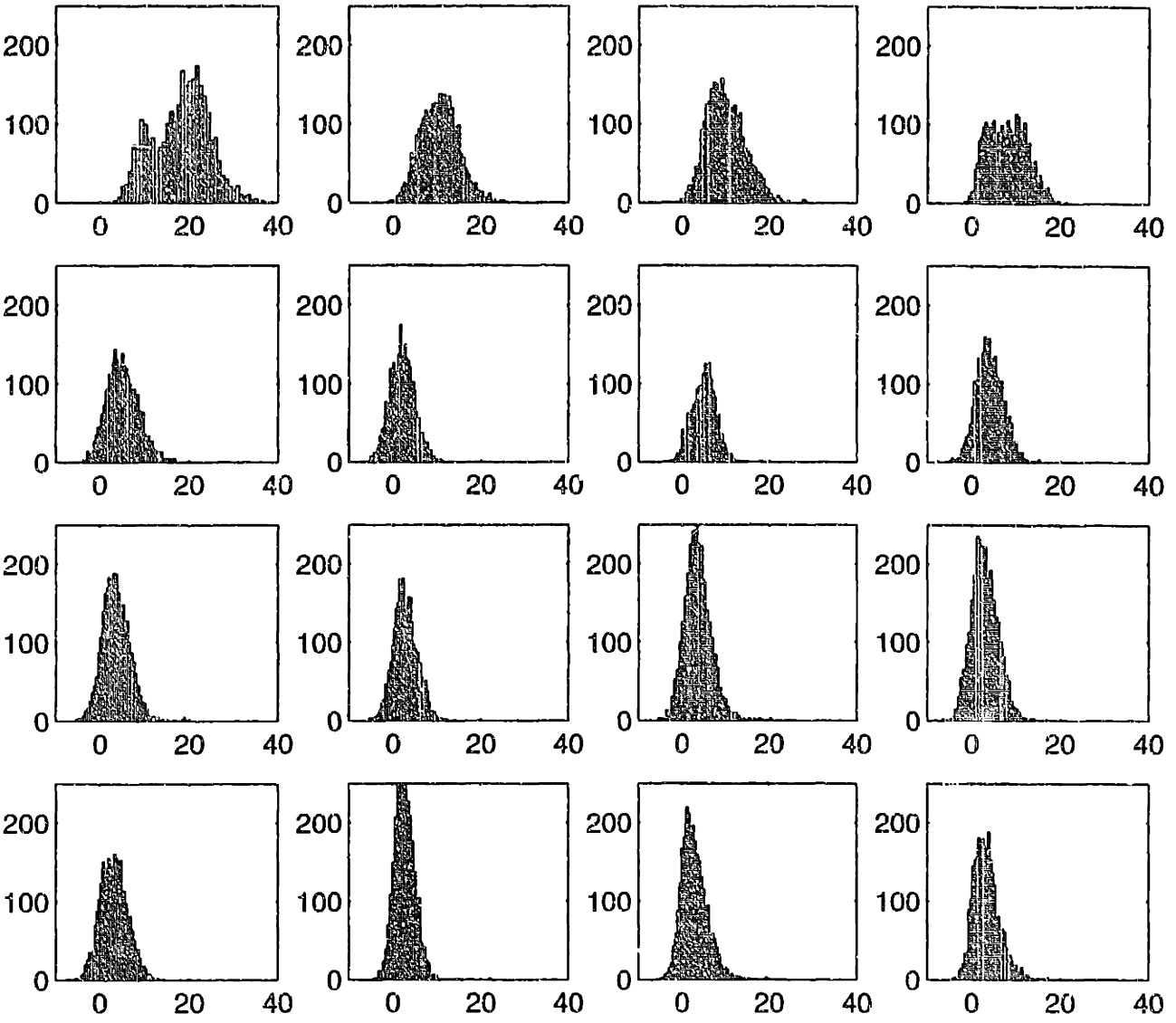
The histogram distributions for NO are shown in Figure 16-9 (a to d). The range of NO concentrations associated with hourly distributions was much more variable. Some distributions were very tight, while other have very long tails, all in the course of one day. For example, on August 2, 1996, distributions with long tails up to 100 ppb were observed before sunrise. Then the range of measurements dropped to 20 ppb at noon. The late afternoon distributions (3-5 pm) on the same day contain a small number of reading of very high concentrations. Generally, small narrow histograms were observed during the midday hours. During the high traffic hours, or at reduced mixing conditions, the observed distributions were skewed with very long tails, consistent with the observation of "spikes" of very high NO that occurred at a small time scale. A few bimodal histograms were observed. These do not necessarily occur at the same time as the ozone bimodal distributions.

The Normal distribution successfully fitted only a few mid-day distributions that were tight. Many of the afternoon and early evening distributions for NO deviate significantly from normality. Their long tails could sometimes be represented satisfactorily by the gamma distribution, with an offset applied to data to make readings positive. The lognormal distribution, another form with a long tail, was too "tail heavy" and did not predict sufficient height at the mode of the distribution to provide a good fit for NO data. The multi-modal distributions observed for nitric oxide were more variant in shape than those observed for ozone, and twin normal distributions did not provide reasonable fits for these NO frequency distributions.

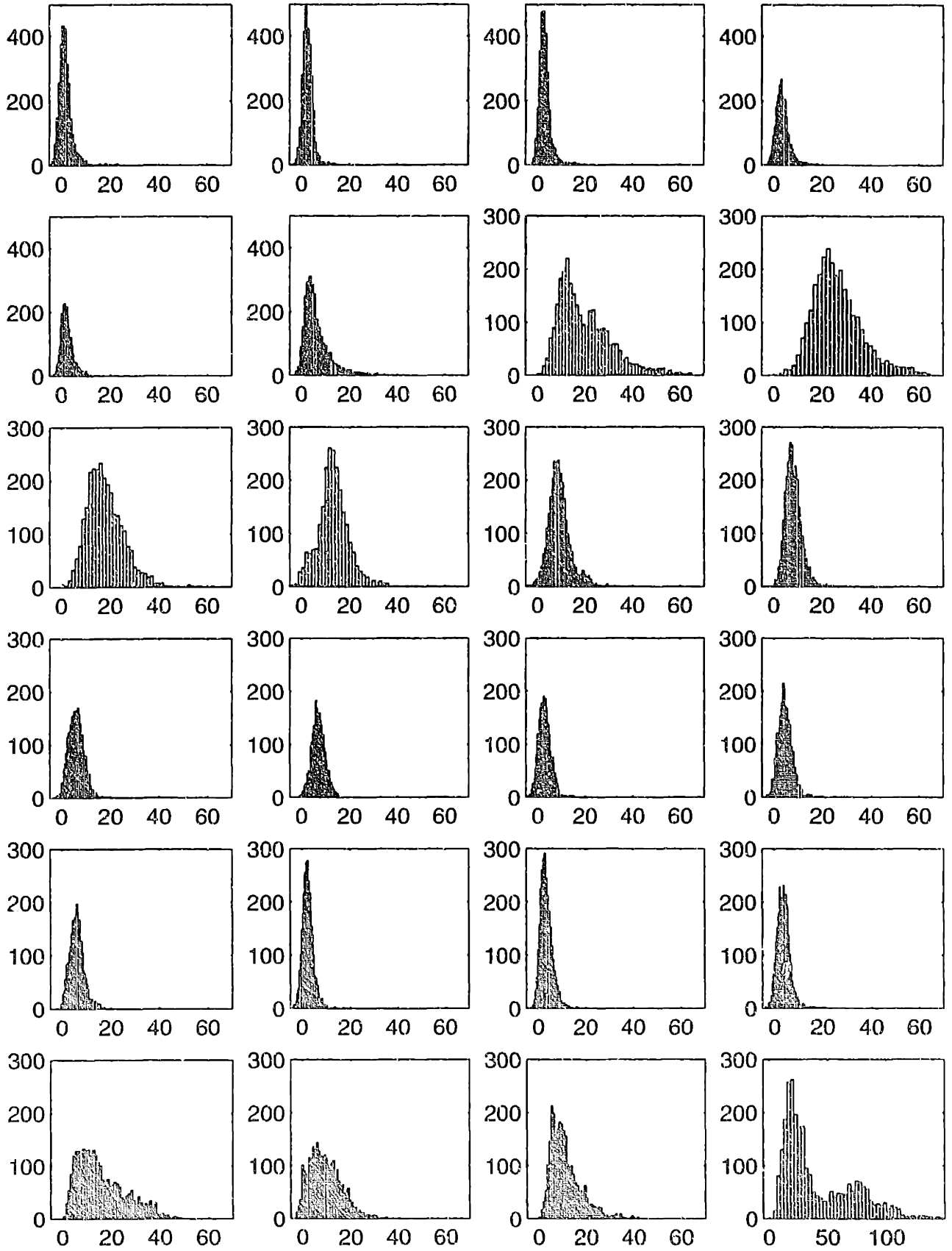
The histograms for nitrogen dioxide are shown in Figures 16-10 (a to d). Qualitatively, they looked different from both ozone and NO data. The majority of the distributions had one mode, even when the corresponding ozone or NO data exhibited multiple modes. The distributions were relatively sharp compared to both ozone and NO histograms. The range of data associated with each hourly interval did not vary greatly. Therefore, the widths of the histogram bins were fairly consistent from one hour to the next and from one day to next. Of the few bi-modal distributions for NO_2, some seemed to correspond to bi-modal ozone distributions. Examples could be found at 9, 10 a.m., 7, 11 p.m. on August 21 and 7 p.m. on August 22, 1996 (anti-correlated), as well as 8 and 9 a.m. on August 23 (same shape).

In contrast to nitric oxide distributions, which have long positive tails, NO_2 distributions frequently exhibited the opposite skewedness, having tails towards the lower concentration side. As in the case of ozone, the normal distribution could fit most single modal hourly data well, indicating that the fluctuations could be described well by a standard deviation around an hourly mean value. A few of the distributions were noticeably narrower than the Normal fit with the same mean and standard deviation. One example was the distribution on August 22 at 10 am. Because of the a tail towards the negative side, the Weibull distribution was a better choice than the gamma distribution for NO_2 data.

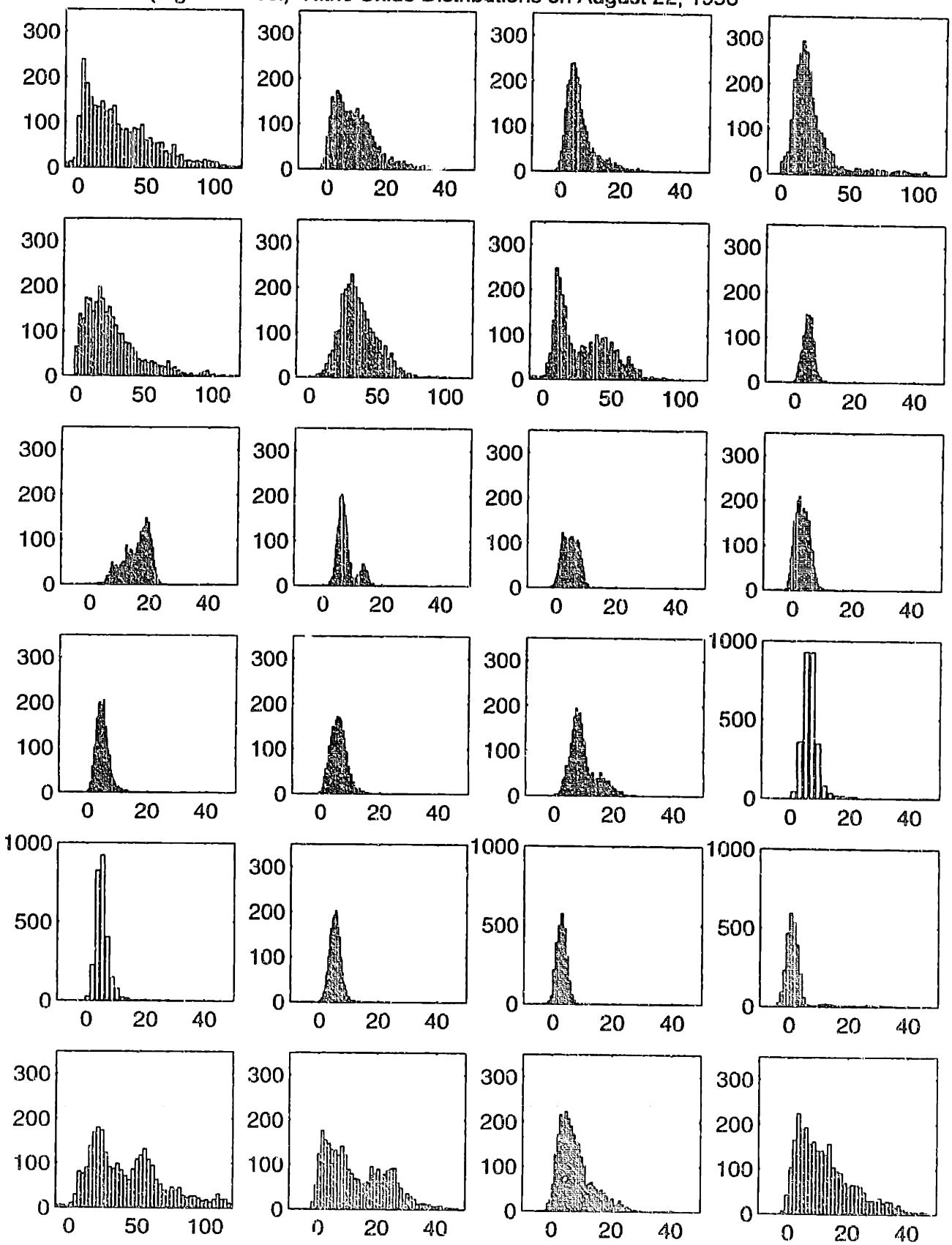
(Figure 16-9a.) Nitric Oxide Distributions on August 20, 1996 (starting at 8 a.m.)



(Figure 16-9b.) Nitric Oxide Distributions on August 21, 1996



(Figure 16-9c.) Nitric Oxide Distributions on August 22, 1996



(d.) August 23, 1996

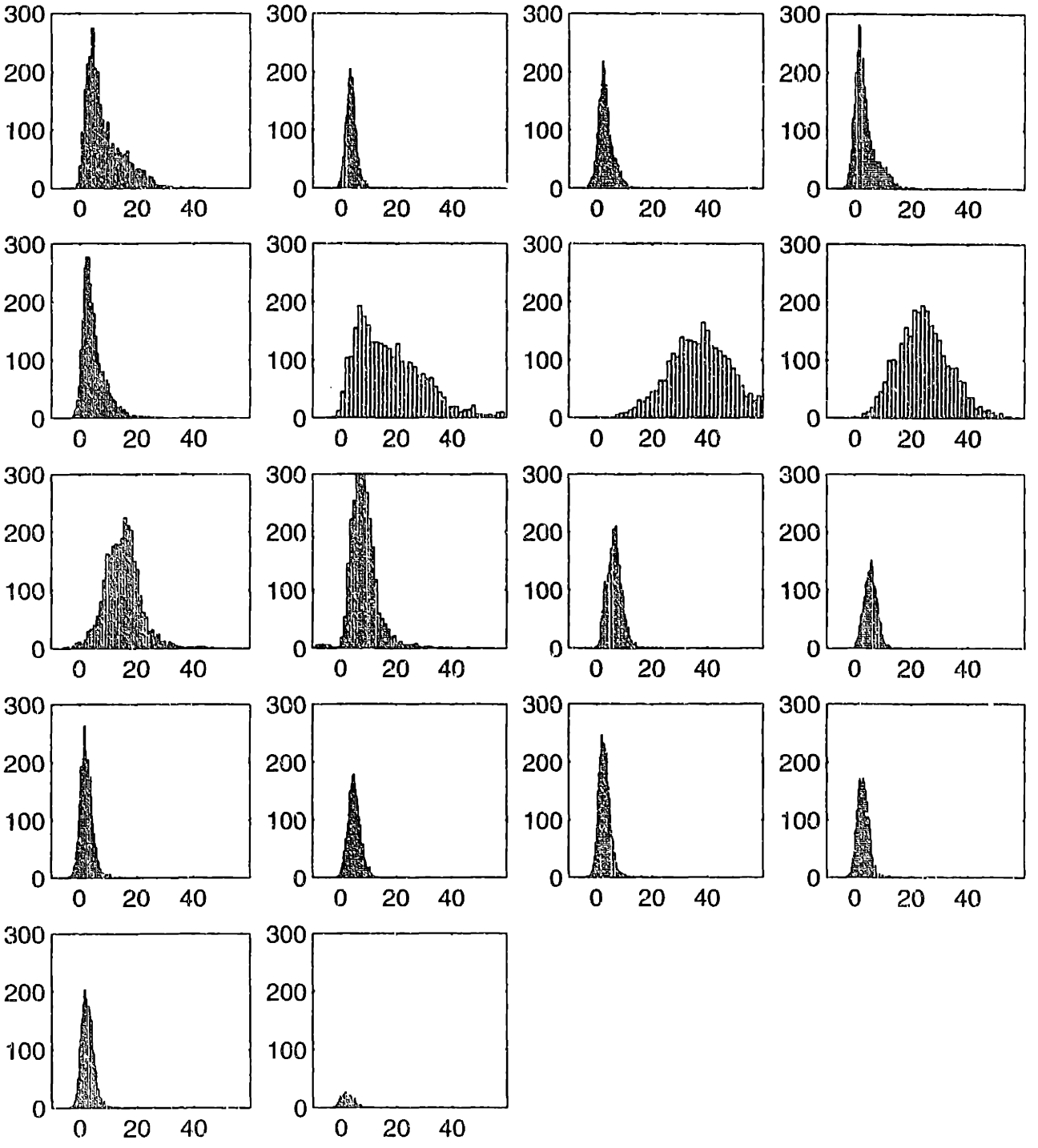
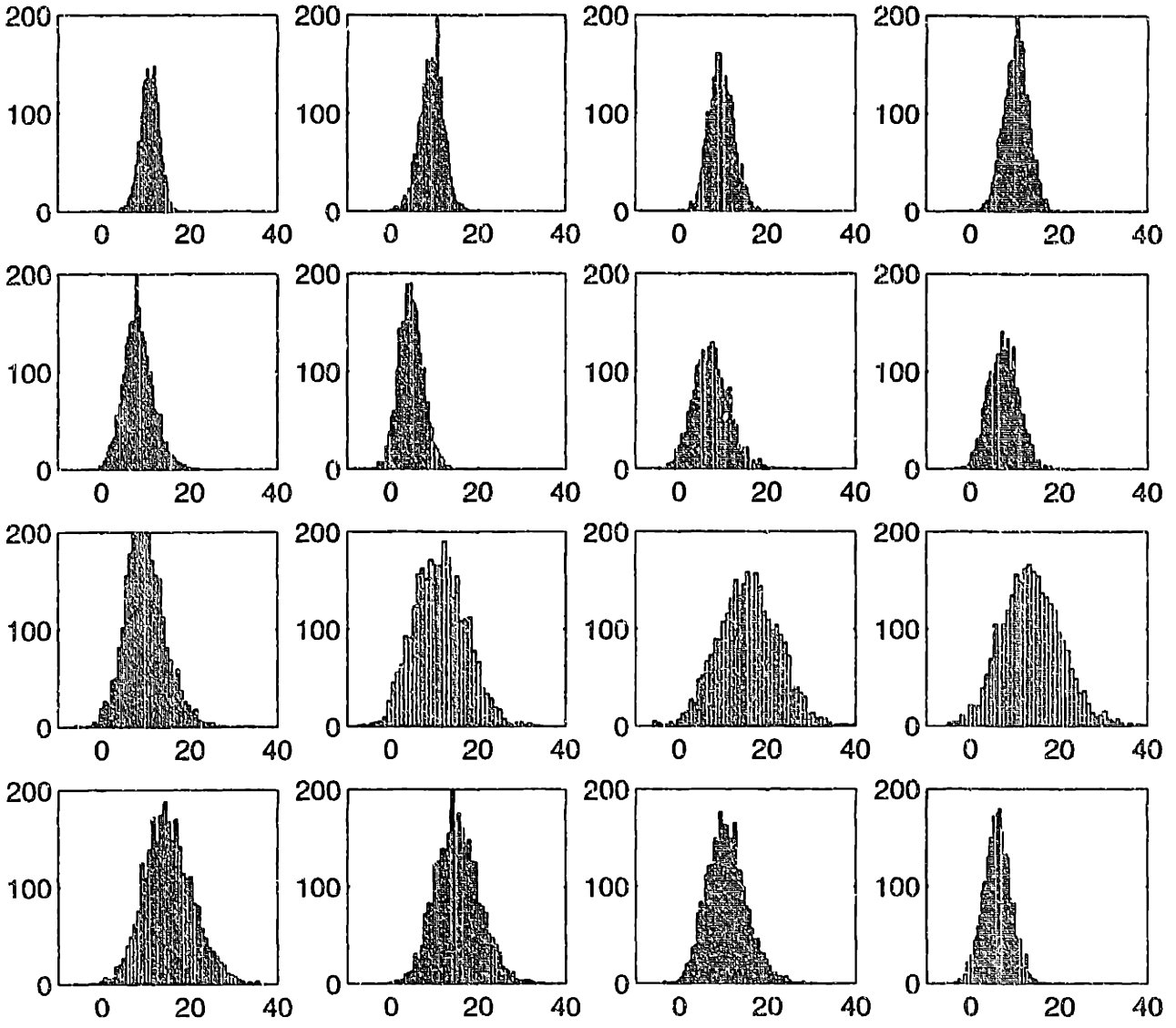
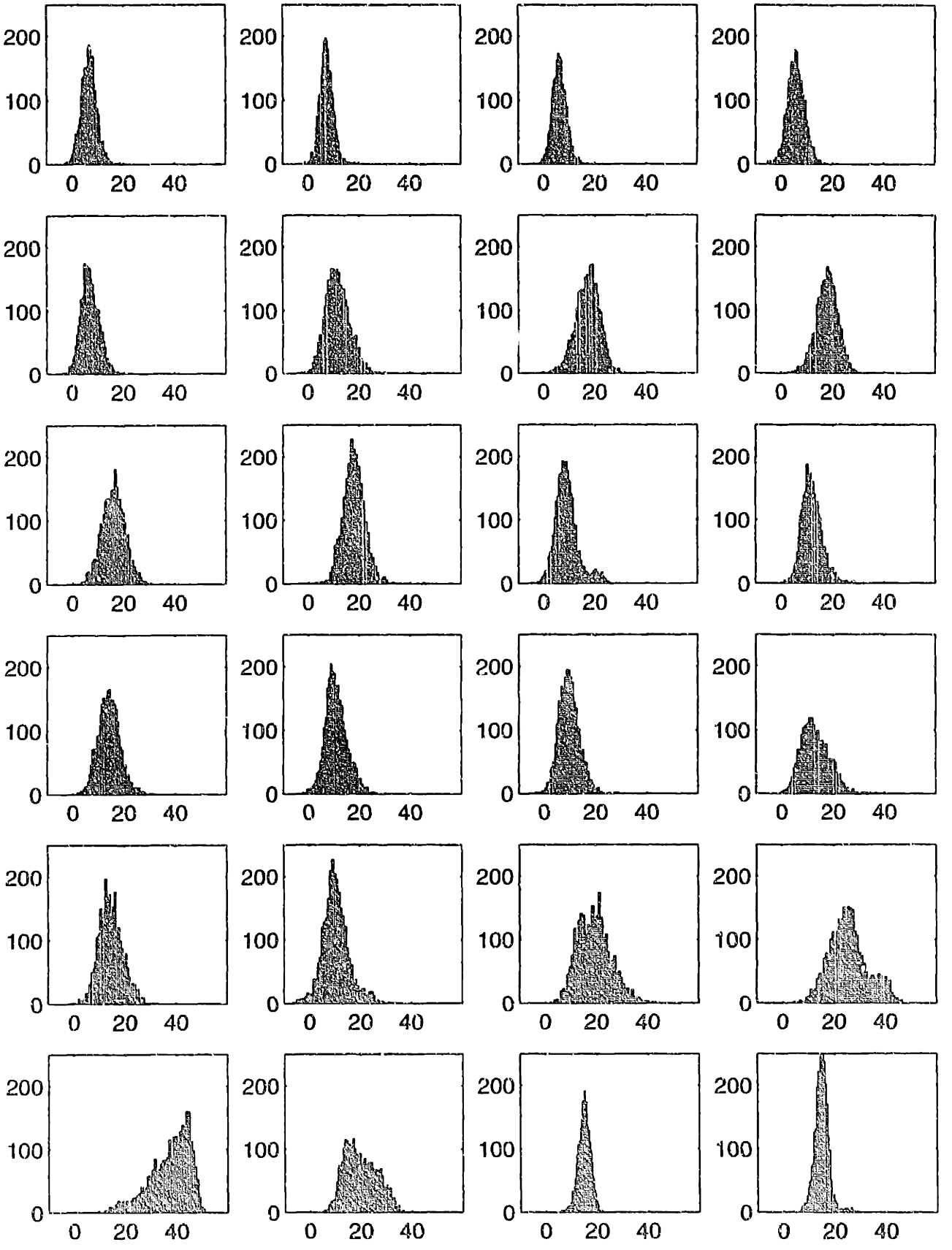


Figure 16-9. Hourly Nitric Oxide Histograms, August 20 to 23, 1996

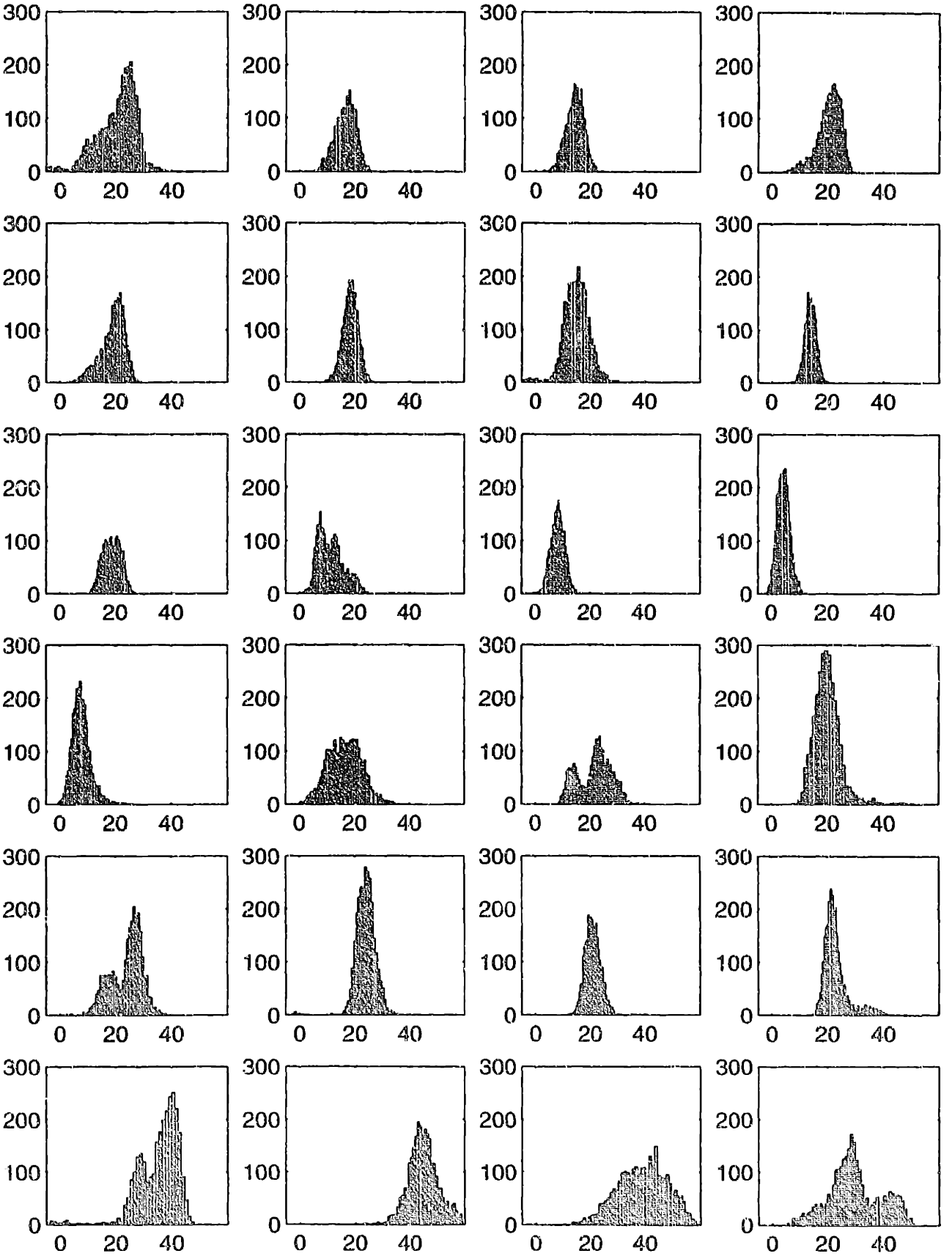
(Figure 16-10a.) Nitrogen Dioxide Distributions on August 20, 1996 (starting at 8 a.m.)



(Figure 16-10b.) Nitrogen Dioxide Distributions on August 21, 1996



(Figure 16-10c.) Nitrogen Dioxide Distributions on August 22, 1996



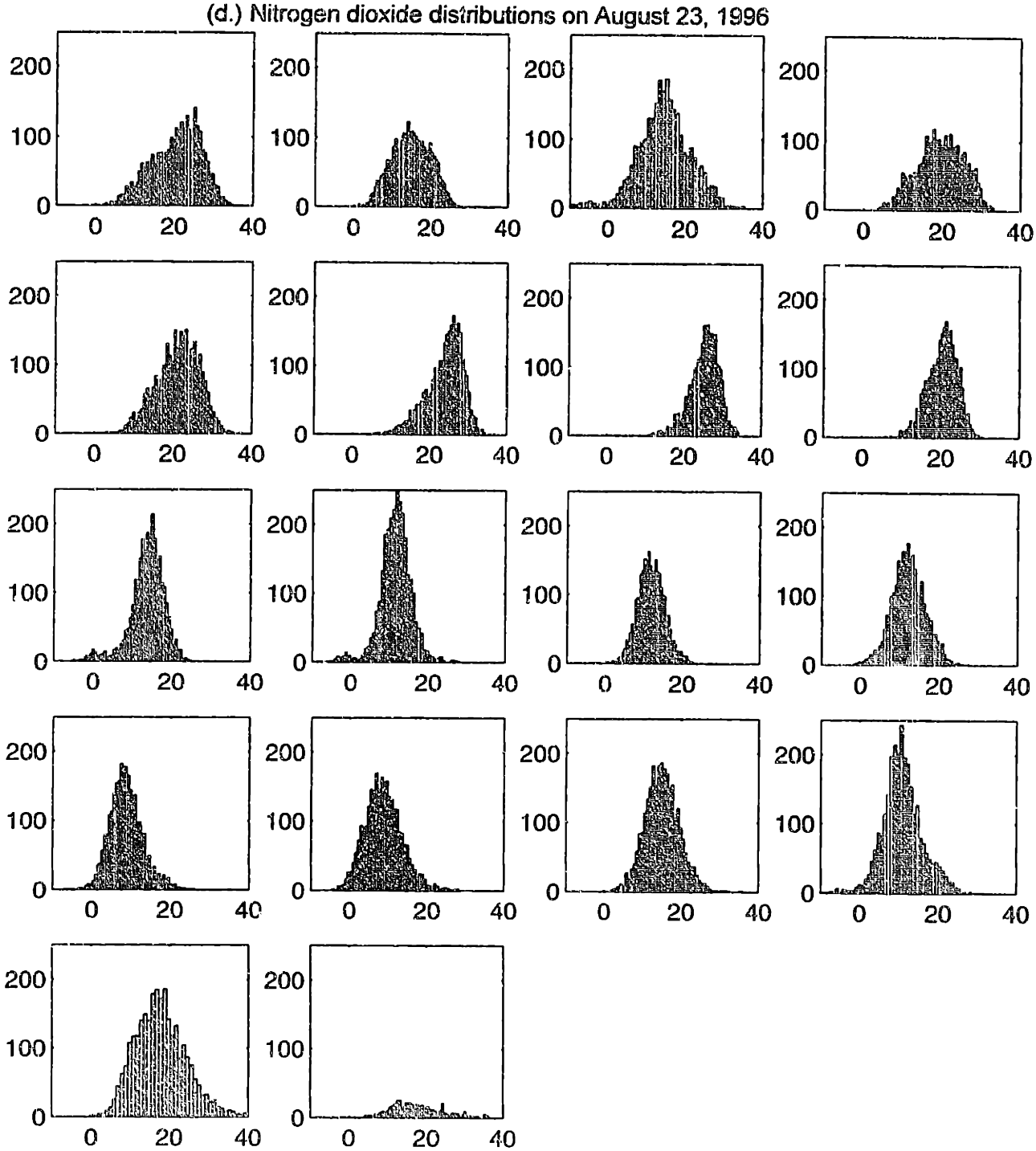


Figure 16-10. Hourly Nitrogen Dioxide Histograms, August 20 to 23, 1996

16.3.3 Autocorrelation and Power Spectra

The autocorrelation function measures the time-related properties in the data that are separated by fixed time delays in a random data stream. For a sine wave, the autocorrelation is also a sine wave. For random white noise, the autocorrelation function drops off very quickly from a maximum value at zero time lag, which is also the variance of the data set. Short time scale autocorrelation coefficients were used as a predictor to fill in the gaps in the concentration data (as discussed in Chapter 14). This section focuses specifically on lag times longer than a few seconds. Matlab was used to analyze the autocorrelation functions of the data streams of ozone, NO, and NO_2. In most cases, the hourly autocorrelation functions decayed rapidly with increasing time difference. However, many autocorrelation functions, especially for ozone, did not approach zero correlation with increasing time. The lack of decay in the autocorrelation function with time was likely to be an artifact due to the unremoved trends in the data. Therefore, the assumption that the raw data may be treated as piecewise stationary was a poor one at the one hour time scale. The standard procedures called only for the removal of mean, which was probably insufficient for some portions of the data that displayed strong periodicity.

Steps were taken to remove the long term "wobbles" in the data, including trends and periodic signals with long periods. Several hours of ozone data in the late evening of August 20, 1996, were completely dominated by a sinusoidal signal of period 1300 seconds and magnitude (peak to peak) 40 ppb. Since there was no evidence from the other concentration measurements or the meteorological measurements to support such wide and structured fluctuations in ozone, wobbles at this timescale were suspected to be caused by drifts in the fringe structure of the diode laser or in the electronic components of the instrument. To investigate the underlying short time scale fluctuations in ozone, a procedure was devised to remove these long time scale structures by first filtering and decimating the signal (by a factor of 256) and then interpolating between the decimated points using a cubic spline to obtain a smooth curve. This method was particularly good for elucidating the underlying autocorrelation function at time differences on the order of tens of seconds.

The power spectra of a signal is the Fourier Transform of the autocorrelation function, presenting the same information in a format that is relatively easy to interpret. The power spectrum is so called because the area under the frequency spectrum (S_k) between two frequencies (k_- and k_+) is the fraction of total variance in the signal between the two frequencies. According to Taylor's hypothesis for turbulent flows, when the time scale of changes in the mean of the measurement far exceeds the time scale of the turbulent fluctuations, the spatial characteristics of field turbulences are not changed significantly over the time lags involved. Therefore, the time frequency information can be transformed into a spatial length scale information. There are well-established theories regarding the spectral behavior of wind components and scalars such as temperature and inert trace species (Lumley and Panofsky, 1964). The spectra can be divided into three sections, the energy input range in the low frequencies, the viscous dissipation range in the high frequencies, and the inertial subrange in the middle frequencies where neither feeding nor dissipation are important. The spectra usually display a peak power in the energy input range. The "inertial subrange law" dictates that the power in the inertial subrange is related to the frequency by a power law:

$$S(k) = C\varepsilon^{2/3}k^{-5/3}. \qquad (16.23)$$

where ε is the rate of dissipation of turbulent kinetic energy per unit mass of air. The inertial subrange law describes the fluctuations in an atmospheric system due to the breakup of large eddies into smaller ones, without substantial dissipation of energy. The spectral behavior of a

active chemical species is complicated by source contribution as well as chemical production and removal processes.

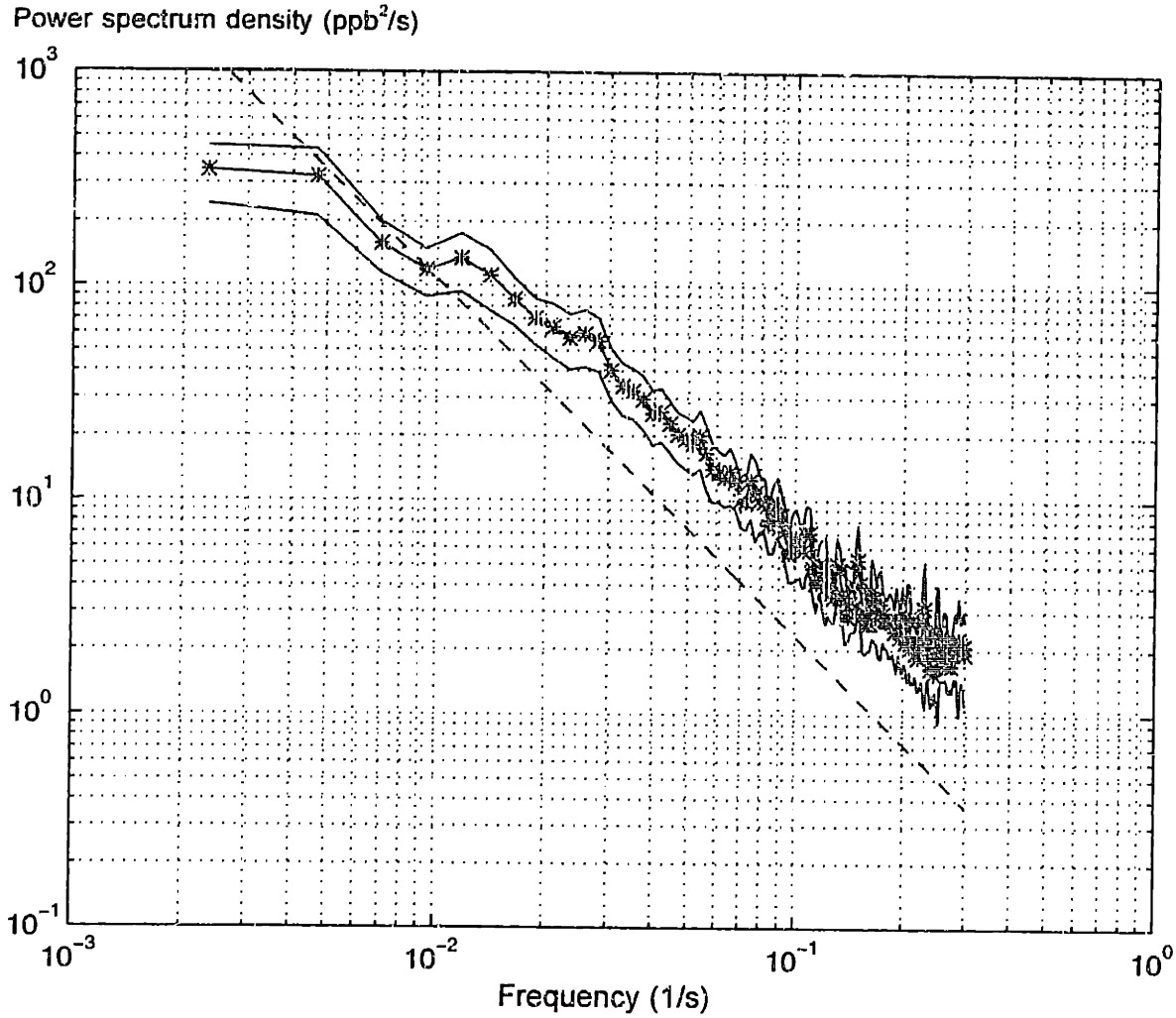
Power spectra were calculated by Matlab using fast fourier transform in sections of 256 points (10 minute sections), focusing on the high frequency contents of the data stream. A Hanning window was applied to the data to provide more stable estimates. The frequency resolution was limited by the data rate (d), with the Nyquist frequency defined by 1/(2d). A process known as aliasing occurs (Bendat and Piersol, 1986, Chapter 10) by which the power of frequencies higher than the Nyquist frequency is "folded back" onto the spectrum itself. Due to the short sections analyzed, stationarity was assumed and only the mean was removed. The daily averaged power spectra for ozone, NO, and NO_2 on August 20, 1996 are presented in Figures 16-11a, 16-12a, and 16-13a. The 95% confidence interval of the power spectra estimates are also given. Figures 16-11b, 16-12b, and 16-13b display the hourly spectral information.

There were significant differences in the frequency content of the concentration time series in ozone, NO, and nitrogen dioxide. Of the three species, ozone's spectra adhered most closely to the inertial subrange law, or the -5/3 power law in Equation 16.23, mimicking the behavior of a stable species. This seemed to indicate that the fluctuations in ozone concentrations were strongly affected by transport effects, and less so by chemical processes. Evidence of aliasing was observed at frequencies of 0.2 to 0.3 Hz, since the data rate of 1.67 seconds corresponds to a Nyquist frequency of 0.3 Hz.

The hourly power spectra in Figure 16-11b show that, except for a couple of hours during the morning, ozone spectra behaved consistently with the inertial subrange law throughout the day. The high frequency components (corresponding to time scales of 2 to 10 seconds) seemed to have a strong contribution to the total variance from 8 a.m. to 10 a.m. and again at 5 p.m. Although the time scale of the NO + O_3 reaction should be 30 seconds to 2 minutes for the levels of ozone observed, this reaction was postulated to be responsible for these shorter time-scale fluctuations when local NO emissions were significant.

Nitric oxide, a primary pollutant with a traffic source in close proximity, displayed a very different power spectrum from ozone. The spectral power as a function of frequency deviated significantly from the -5/3 power rule, with a very strong component in the higher frequencies. Two trend lines were included in Figures 16-12(a, b) with slopes of -5/3 (steeper line) and -2/3. The short time scale fluctuations in NO, with periods of less than 10 seconds, were not at all dominated by transport, since the observed slope was close to zero. Most likely, source emissions dominate the fluctuations of NO, which were probably also affected by rapid chemistry near the source. Much of the power in the high frequency range could also have been a result of aliasing, if the process time scale was less than the data rate of 1.2 seconds. The hour-by-hour display of the data (Figure 16-12b) shows that the characteristics of the power decay with frequency varied with time. Positive slopes were observed in the high frequency range in the evening, most likely due to aliasing of signals of frequencies higher than the Nyquist frequency. The amount of power associated with this frequency range also seemed to vary within the course of the day, with relatively higher powers observed in the morning.

(Figure 16-11a) Daily Average Ozone Power Spectral Density, August 20, 1996.



(b) Hourly Ozone Power Spectral Density (starting at 7 a.m.)

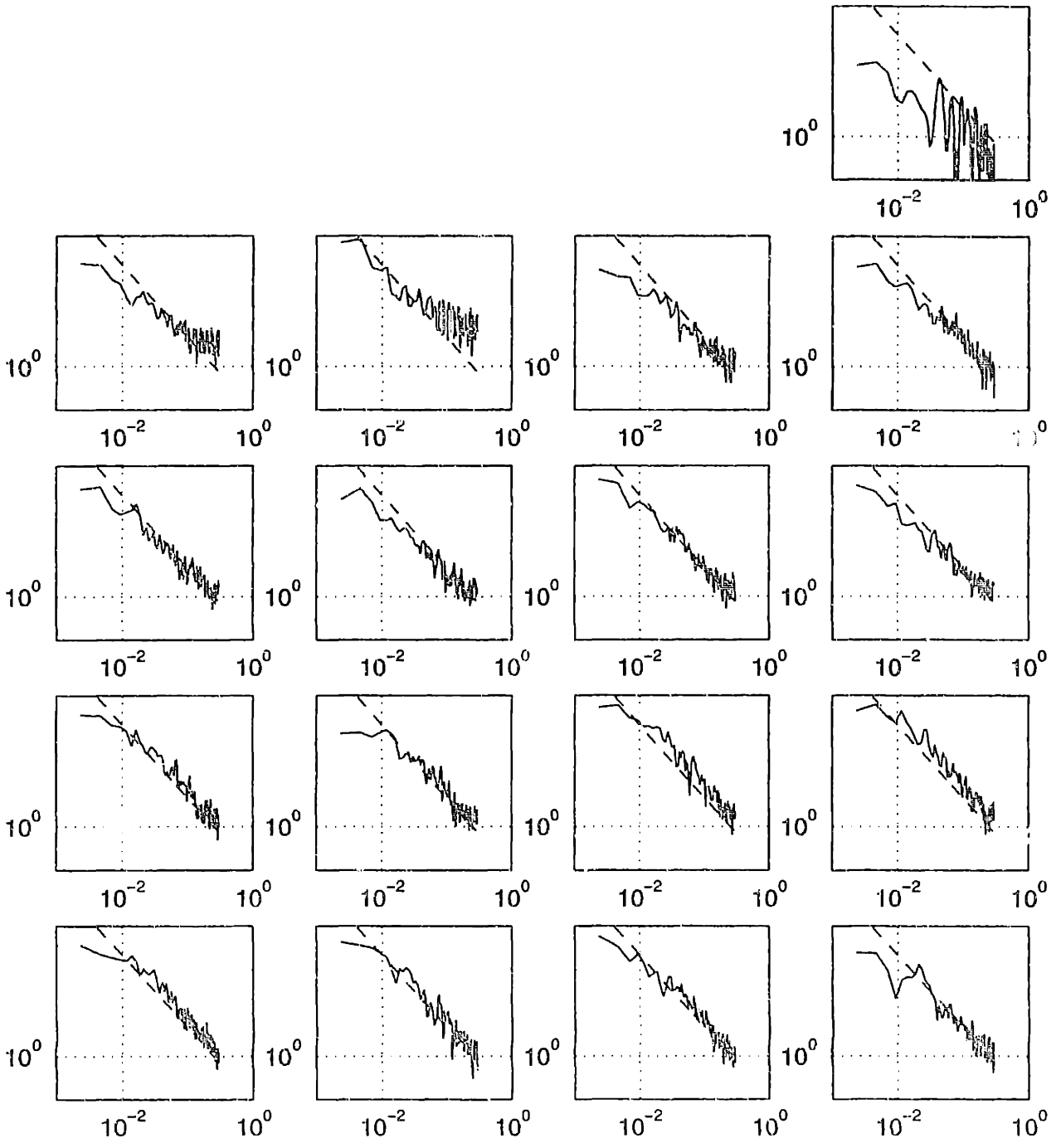
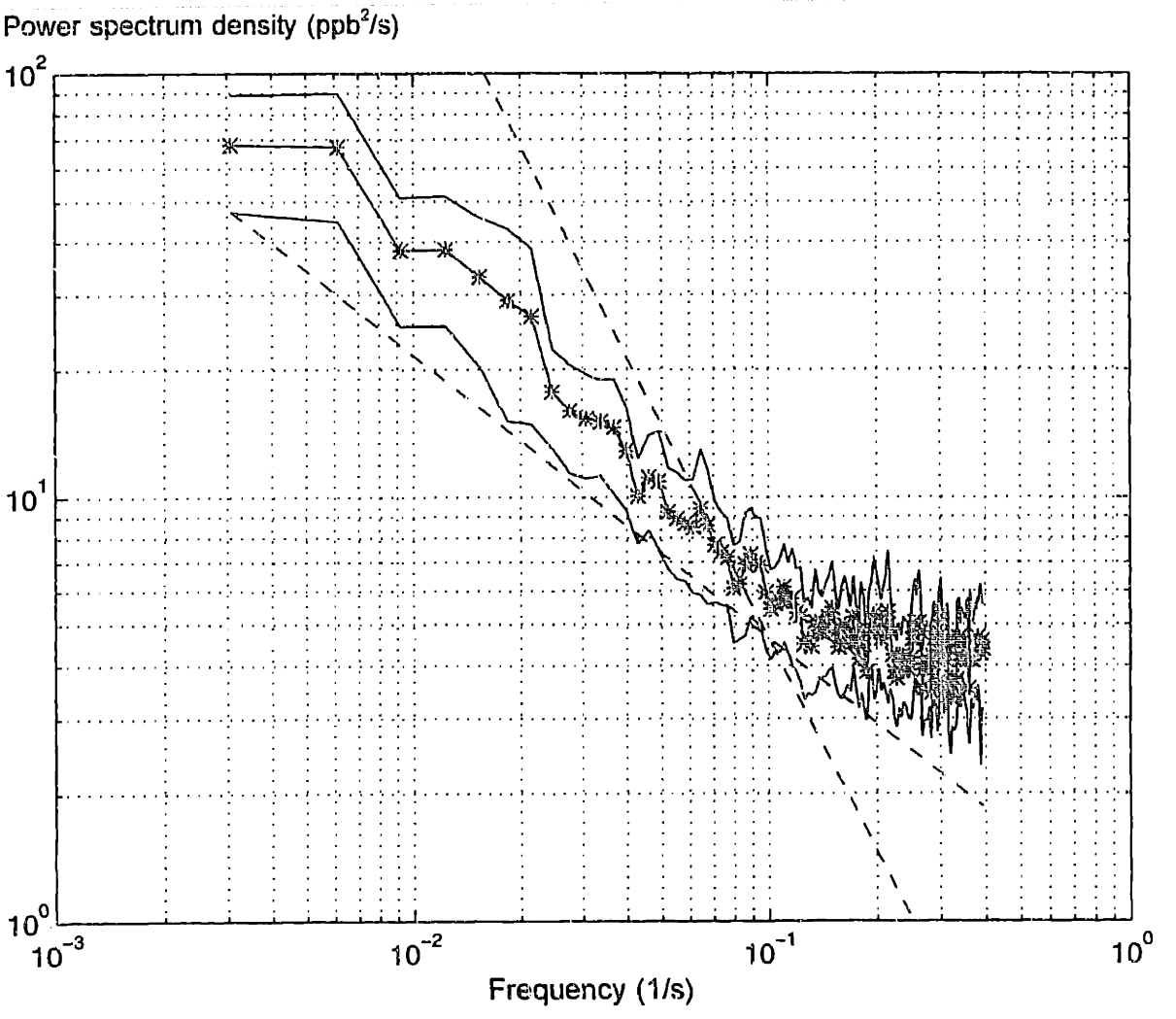


Figure 16-11. Ozone Power Spectrum on August 20, 1996
(a) Daily Average, (b) Hour-by-Hour. Trend line shown has slope of -5/3

(Figure 16-12a) Daily Average Nitric Oxide Power Spectral Density, August 20, 1996.



(b) Hourly Nitric Oxide Power Spectral Density (Starting at 8 a.m.)

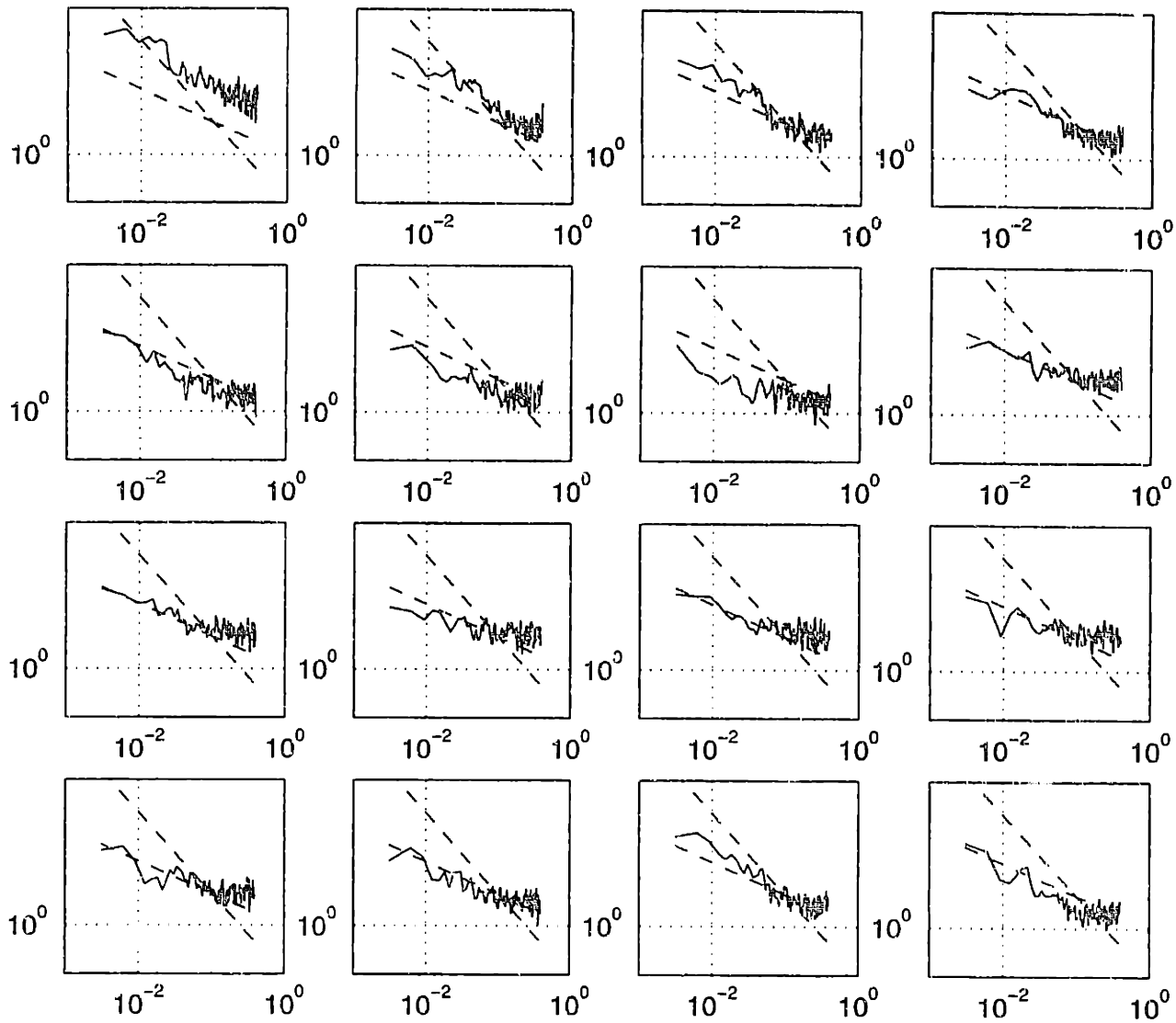
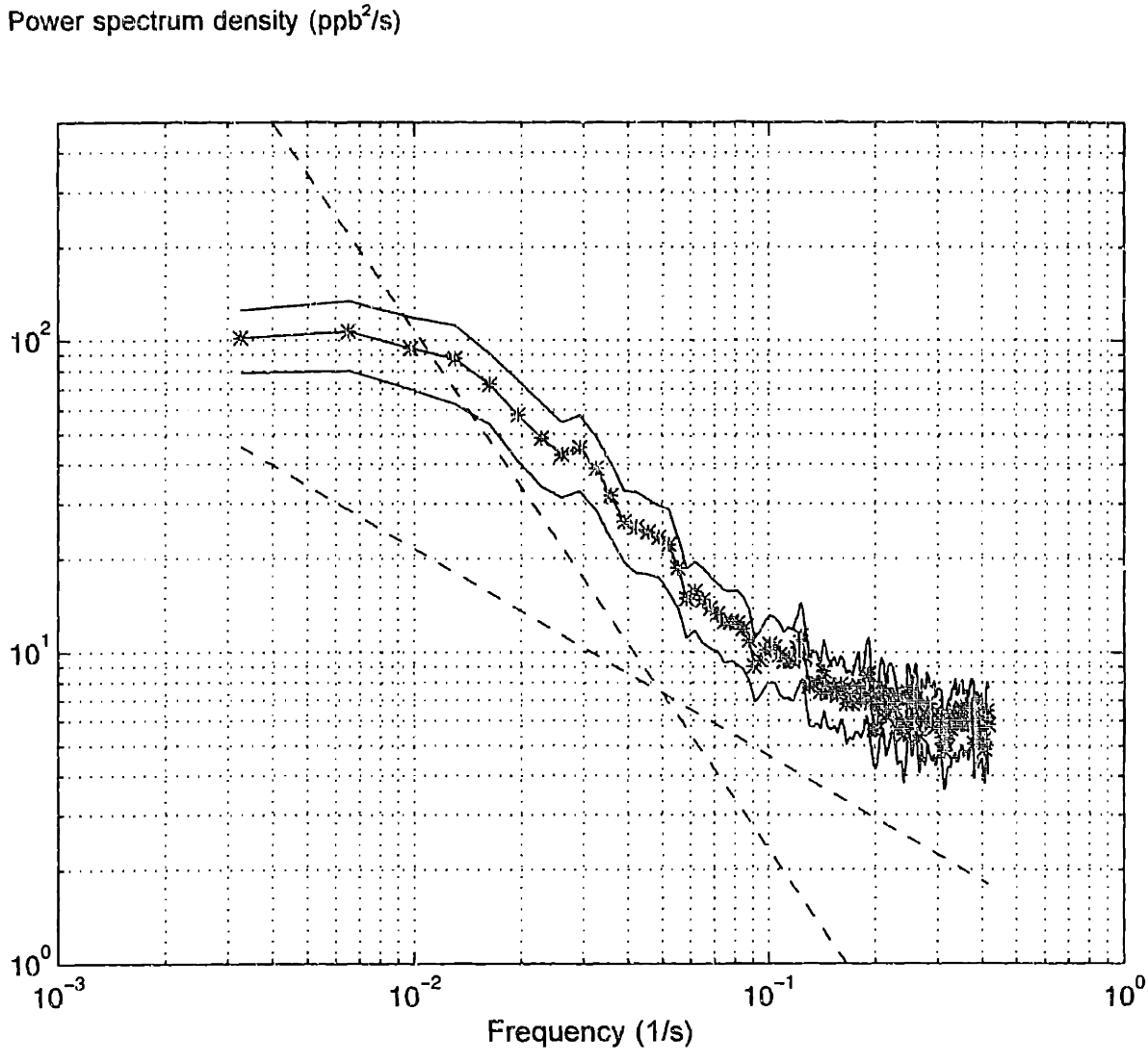


Figure 16-12. Nitric Oxide Power Spectral Density, August 20, 1996
(a) Daily Average, (b) Hour-by-Hour. Trend lines shown has slope of -5/3 (steeper line) and -2/3

(Figure 16-13a) Nitrogen Dioxide Power Spectral Density, August 20, 1996.



(b) Nitrogen Dioxide Power Spectral Density, August 20, 1996 (starting 8 a.m.)

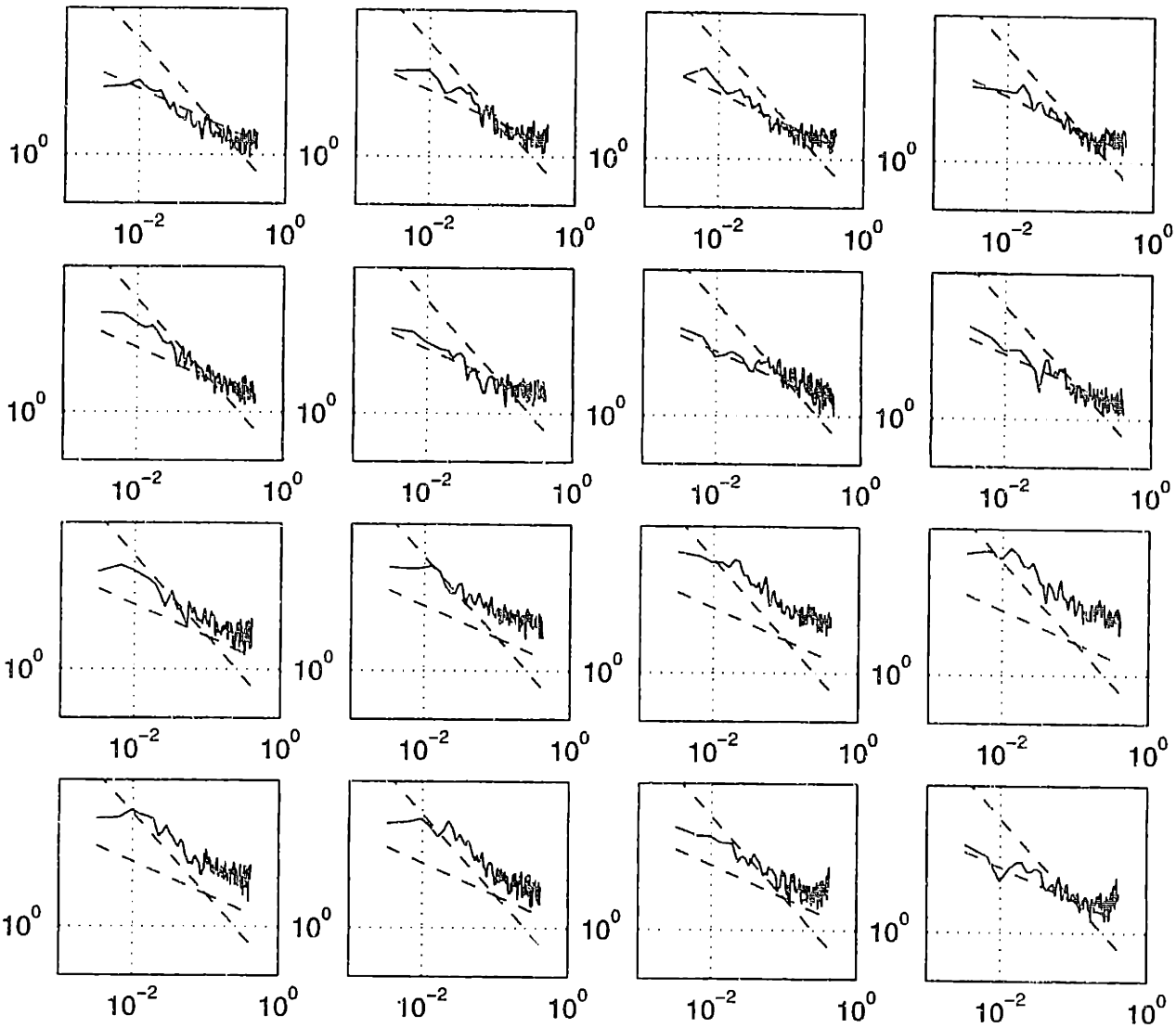


Figure 16-13. Nitrogen Dioxide Power Spectral Density, August 20, 1996
(a) Daily Average, (b) Hour-by-Hour. Trend lines shown has slope of -5/3 (steeper line) and -2/3

The behavior of the spectra of nitrogen dioxide (Figure 16-13a) resembled that of nitric oxide, with stronger high frequency fluctuations than expected of an inert scalar. A slightly steeper slope was observed for NO_2 than for NO at the low frequencies. Although vehicular emissions as a direct source of NO_2 was expected to be small compared to NO (10% of NO_x from diesel trucks, on the order of one percent from gasoline vehicles, Heywood, 1988), the fast fluctuations of this species was dominated by direct emission of NO_2 and rapid chemical

reactions at a close-by source of NO_x rather than transport as an inert species. As shown in Figure 16-13b, the slopes of the spectra varied throughout the day. Evidence of high frequency aliasing was observed late in the evening.

The power spectra of NO_x, the sum of NO and NO_2, should help elucidate the relative importance of the interconversion between NO and NO_2 vs. the source effects of NO and NO_2. If the source effects were small, NO_x would behave like an inert compound, and would have power spectra that adhered to the transport-dominated -5/3 rule. As shown in Figure 16-14, the NO_x spectrum displays a steeper slope than either the NO or the NO_2 spectrum. However, the slope still deviated from -5/3, and the power did not decay at the 0.4 to 0.1 Hz region, indicating aliasing from higher frequency signals. These observations indicated that although the interconversion between NO and NO_2 was significant at the suburban site close to a local commuter highway, the fluctuations, especially at the more rapid time-scale, were dominated by source effects of NO_x.

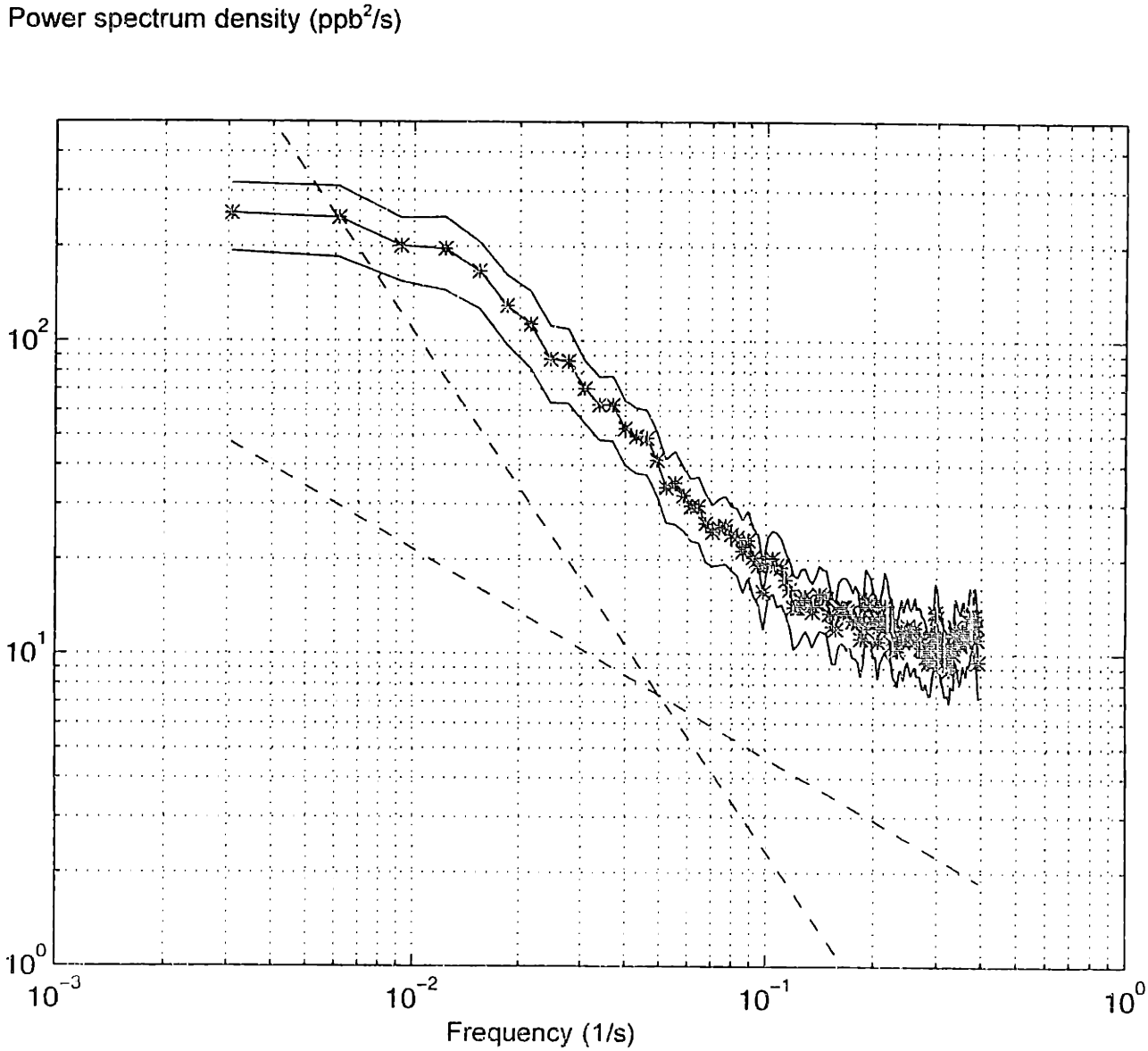


Figure 16-14. Power Spectrum of NO_x on August 20, 1996.

16.4 Effects of Averaging Time

In 1997, the EPA published an updated National Ambient Air Quality Standard (NAAQS) for ozone from a one-hour average of 120 ppb to a eight-hour standard of 80 ppb (EPA, 1997). This change raised some very interesting questions about the issue of averaging time. Due to the cyclical behavior of ozone, increasing the averaging time reduces the maximum value. Plotted in Figure 16-15 is a comparison for the daily maximum with different averaging time. Of the four days in the study period in August, 1996, the daily maximum decreased by to 12 to 25% when the averaging time was increased from one to eight hours. The largest difference was observed on a day when the ozone peak was relatively sharp. The difference between a eight-hour average and a one-hour average is not very dramatic. During the summer, the processes that form ozone are switched on early in the morning and ozone accumulate throughout the day. There are more than eight hours of day light in this season, and high ozone concentrations persist during a majority of that period of time. These findings seem to support the conclusion that suburban locations that are out of compliance with the one hour average, 120 ppb standard would also be out of compliance with the eight hour, 80 ppb standard.

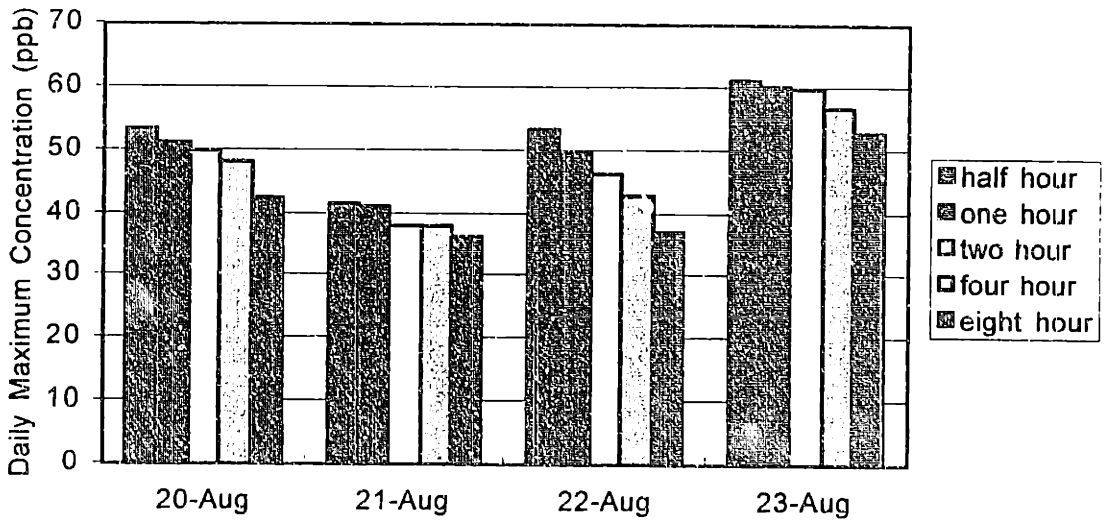


Figure 16-15. Daily Maximum Ozone Concentration as a Function of Averaging Time.

Another interesting issue regarding averaging time is related to sampling rate. The ozone data obtained from the TDL instrument (sampling rate of 1.67 seconds) was averaged to samples of one minute, a typical sampling rate of most commercially-available ozone monitors. As shown in Figure 16-16, the mean value can be calculated consistently from both sets of data. However, the minute-by-minute samples always underpredicted the standard deviation relative to the higher frequency data. The difference was from a few percent to 35%. For a truly random signal, the root mean square noise should decrease according to the square root of the averaging time, which would be a factor of 5 in this case. The observed decrease in standard deviation, therefore, was due to a filtering out of the high frequency structure of the data stream and not due to averaging of random error.

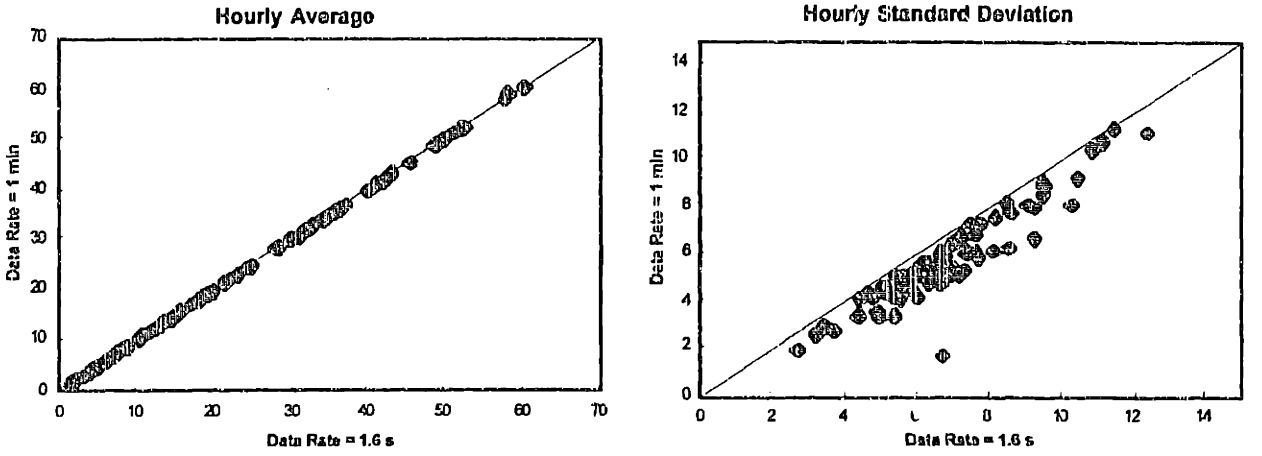


Figure 16-16. A Comparison of Statistical Quantities Predicted by Data of 1.67 Seconds vs. 1 Minute.

16.5 Conclusions

The key conclusions drawn from an analysis of short-time scale ambient concentration measurements are as follows:

The spread of the measurements varied significantly, from one hour to the next, and from one day to the next. Meteorological conditions that were conducive to mixing reduced the hourly variability of concentration measurements and vice versa. This variance due to atmospheric turbulence could provide a statistical basis for comparison with model predictions (See Figure 3.2) and for qualifying the agreement and disagreement of the two. (Is a 20% agreement good of bad?)

Contrary to popular belief, measurements collected at a stationary monitoring site were not always Normally distributed due to turbulent fluctuations. A variety of shapes were observed for the hourly ozone, NO, and NO_2 distributions. None of the standard forms (Gaussian, Lognormal, Weibull, or Gamma) was satisfactory in representing the observed distributions.

Highly skewed distributions and bimodal distributions were observed. In light of the variable shapes of the sample distribution function, the hourly mean alone may not the best representation of the data. For example, if the hourly ozone concentration distribution has two modes (Figure 16-8), the population may be exposed to higher levels of ozone more often than if the distribution was normal. Hence, a 90th percentile concentration maybe a more robust format for an air quality standard written for purpose of protecting public health.

Of the three pollutants measured, ozone's power spectrum adhered most closely to the inertial subrange law (Equation 16.23) for an inert tracer. The concentration time series for NO_x (and for both NO and NO_2) contained significant high frequency fluctuations from the local traffic source. This observation has significant implications towards the choice of monitoring sites and the utility of stationary data. Data recorded at a site with strong local source effects may not be representative of the greater geographical area at all. It may be hard to find a location that is free from strong NO_x source effects in an urban area. Regarding representativeness, long open-path measurements have definite advantages over point measurements.

The National Ambient Air Quality Standard (NAAQS) for ozone has always been written in the format of an averaged concentration over a period of time. In the past, a one-hour average has been the regulatory standard and a wealth of information was recorded in that format from monitoring stations around the country. While this presentation is appropriate for regulatory purposes, this format may not be optimal for model validation purposes. Measurements are frequently reported at a single monitoring station, while models predict a spatial average. Therefore, comparing a one-hour average model prediction and a one-hour average point measurement is like comparing apples and oranges and it is hard to define what constitutes a "good agreement." On the other hand, according to the Taylor hypothesis, the fluctuations in the time series reflect the spatial variability of instantaneous measurements in the area around the monitoring site. Most monitoring instruments are capable of higher sampling rate than one hour. The fluctuation / variability information should be recorded on a routine basis to facilitate a more informed comparison between point ambient measurements and grid cell average predictions from the models and to provide limits to how good the model predictions have to be for validation.

17. Correlations between Concentration and Meteorology Data

The last chapter presented an analysis of statistical properties of trace gas concentrations measured at a suburban site. This chapter investigates the relationship between different streams of concentration data, as well as the effect of meteorology on the concentration measurements.

17.1 General Observations for Concentration Correlations

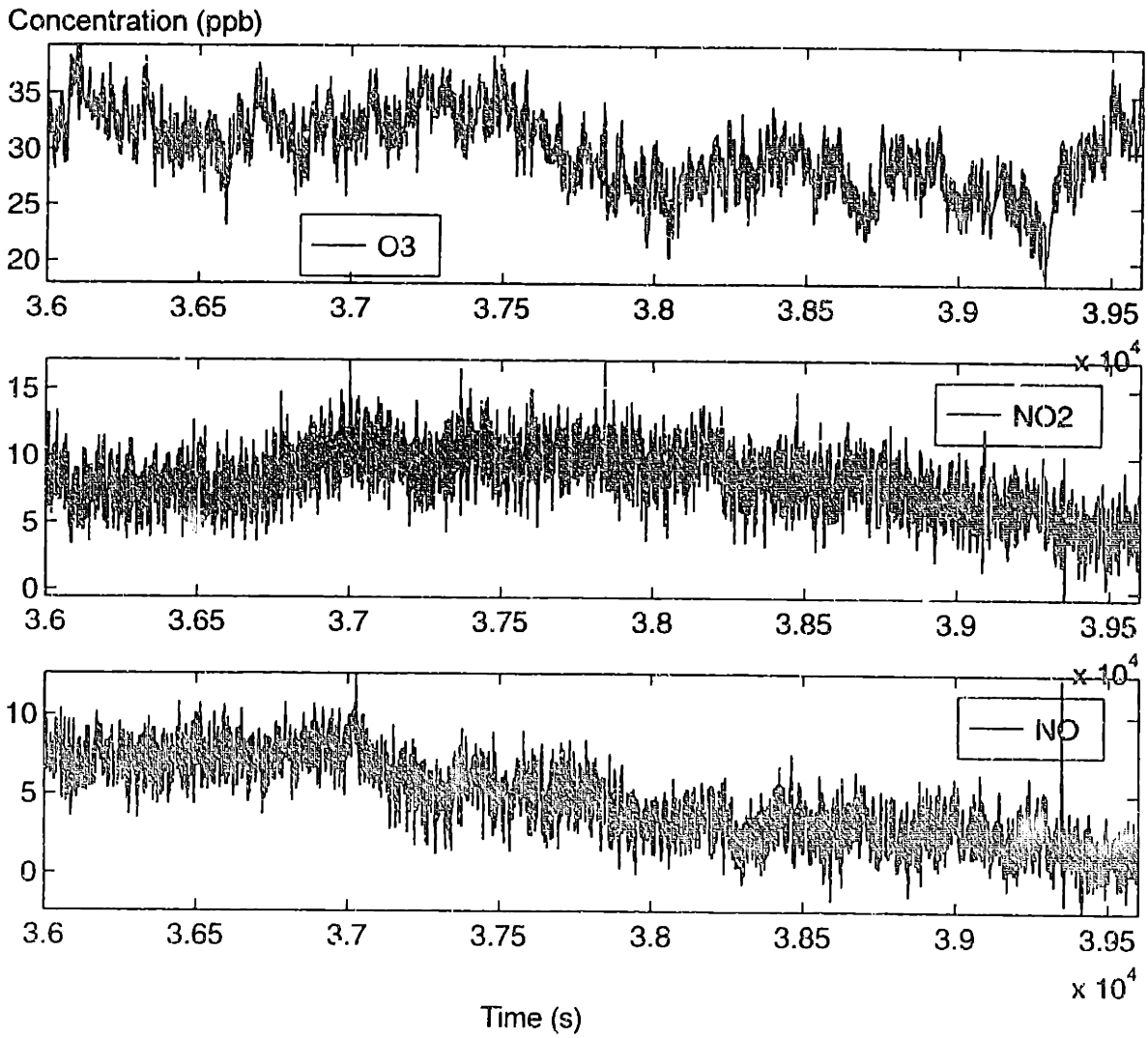


Figure 17-1. Ozone, NO, and NO_2, 10 to 11 a.m., August 22, 1996.

Several samples of data were chosen from the study period of August 20 to August 23, 1996 to illustrate the fluctuations and correlations of ozone, nitric oxide, and nitrogen dioxide. Figure 17-1 shows an excerpt from August 22, 1996 in the morning at 10 a.m. In general, the middle of the morning would be a period when ozone concentration built up. During this particular hour, however, ozone accumulation seemed to have "stalled". The ozone data record showed much more structure when compared to NO and NO_2. Both nitrogen oxides decreased steadily in the morning, with rapid fluctuations around the mean trend. No obvious correlation was observed between the three pollutants in the short time scale.

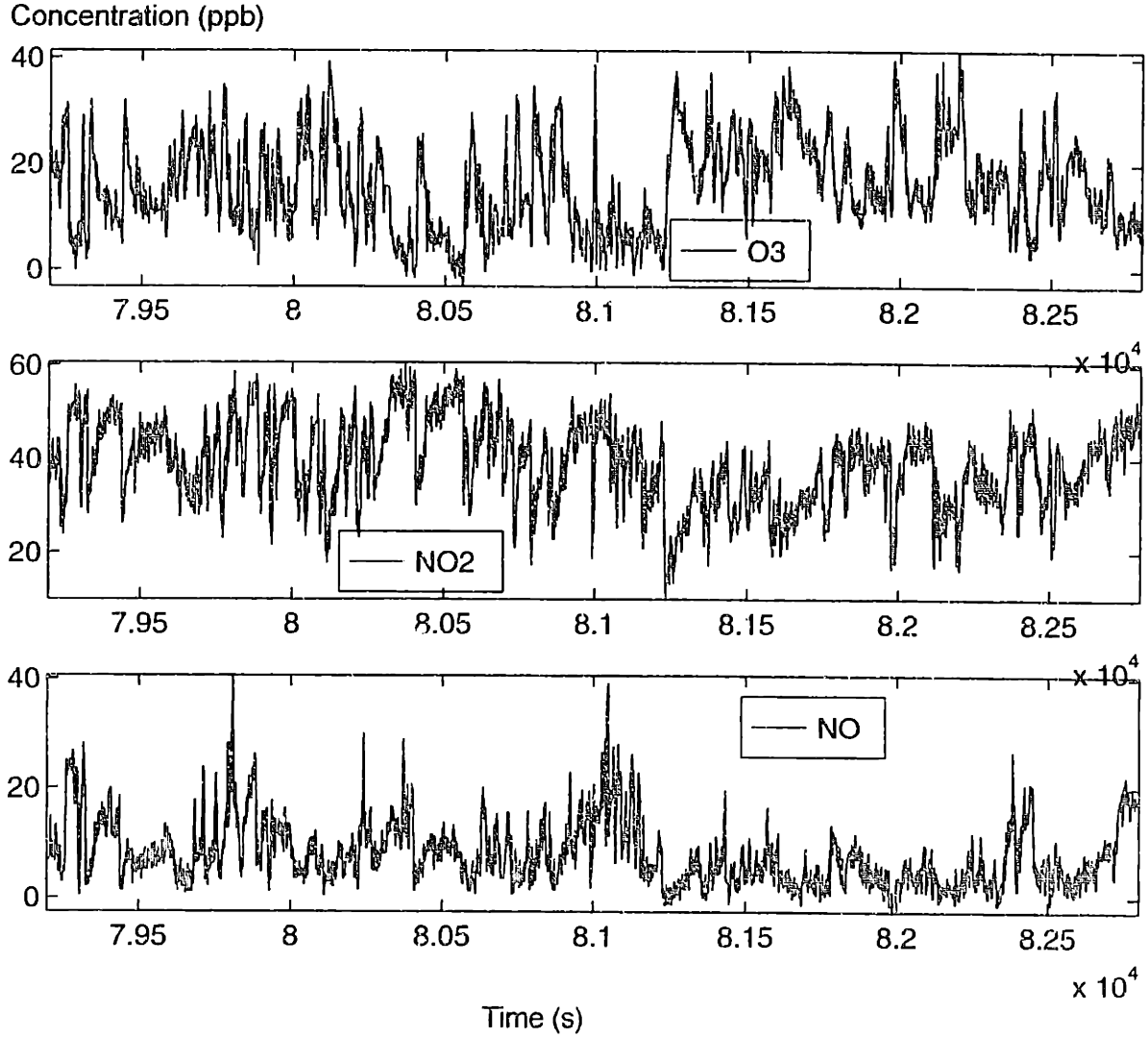


Figure 17-2. Ozone, NO, NO_2 10 to 11 p.m., August 22, 1996.

Figure 17-2 is sample record taken from the evening of August 22. All three pollutants were fluctuating wildly at this time. In contrast to Figure 17-1, Figure 17-2 shows clear anticorrelation between ozone and nitrogen dioxide measurements. The correlation between the two nitrogen oxides was also significant. The fluctuations were characterized by a sharp increase of ozone and a corresponding drop in NO_x concentration, followed by a slower decay of ozone back to a lower value accompanied by a movement of the NO_x reading in the

opposite direction. Nitric oxide and nitrogen dioxide measurements tracked well throughout this period. The distinguishing feature between the two nitrogen oxides was that NO_2 seemed to have a ceiling concentration, while NO did not. These observations were common to much of the data obtained during the evening and early morning.

Throughout the afternoon of summer days, the concentration of NO remain at the level of a few ppb, with occasionally "spikes." On August 23, these spikes were of the order of 10 to 15 ppb, as shown in Figure 17-3, which shows the concentration of ozone, nitric oxide, and nitrogen dioxide at 2 p.m. On other days, these afternoon spikes of NO could be significantly higher. For example, a few NO spikes that were observed in the afternoon of August 22, 96 reached 100 ppb, the NO level observed during the morning and evening rush hours. Though the level of nitrogen dioxide was also quite low during the afternoon, the anti-correlation between NO_2 and ozone was easily observable, as was the positive correlation between the two nitrogen oxides. In addition to the concentration difference, the afternoon and evening data also exhibited different fluctuation behavior. No "ceiling" value was observed for nitrogen dioxide during the afternoon.

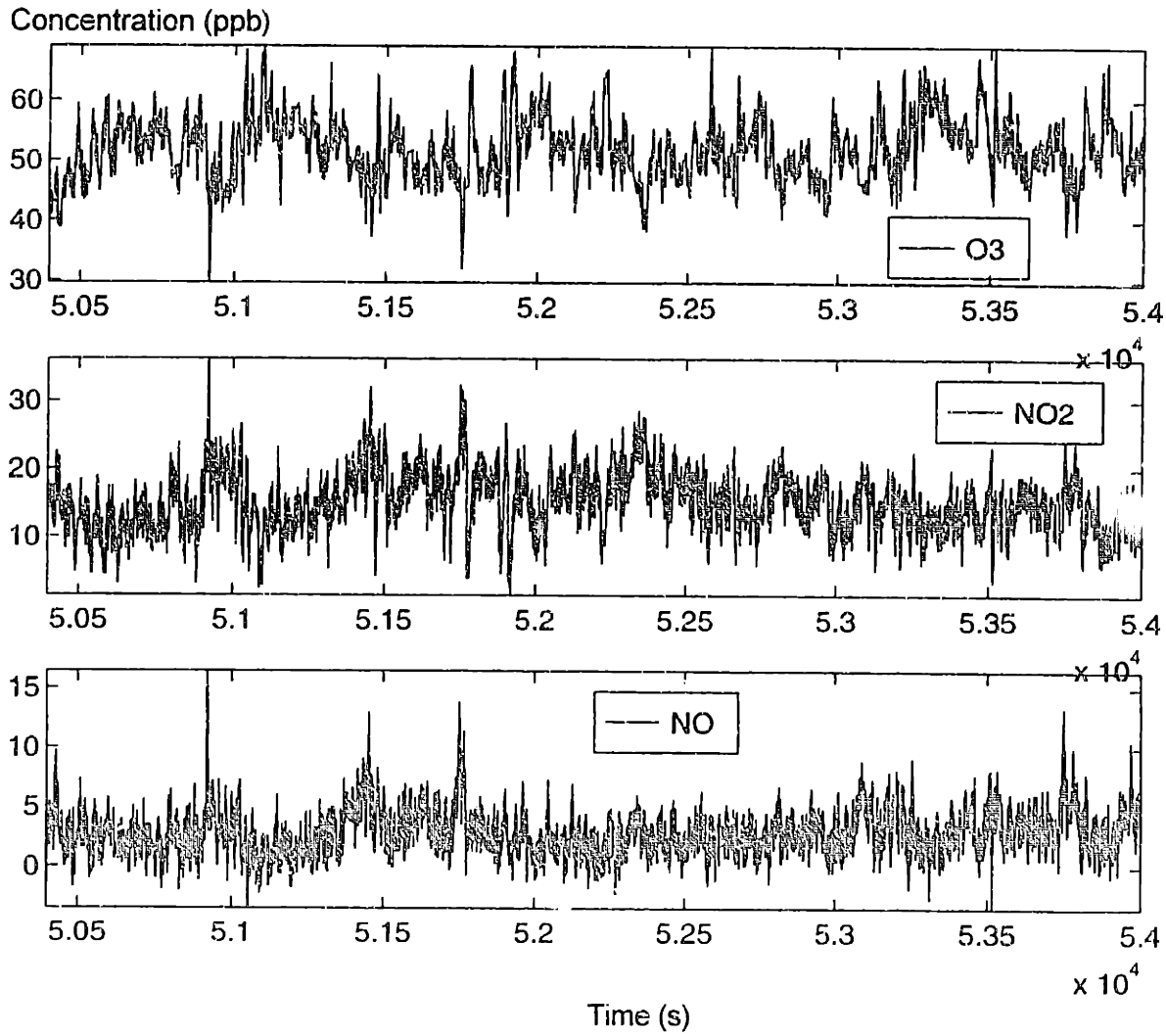


Figure 17-3. Ozone, NO, NO_2, 2 p.m., August 23, 1996

More insights could be gained about the interplay of the concentrations by replotting the concentration streams against one another. The procedures used for matching the time stamps of measurements recorded on different computer systems were described previously. Since NO_x measurements were obtained at a slightly higher sampling rate, they were interpolated onto the ozone sample points so that the concentrations could be plotted against one another. The following three figures display the corresponding data in Figures 17-1 through 17-3 by plotting the concentration of NO_2 against NO and against ozone.

The lack of short time-scale correlations between the pollutants of interest on the morning of August 22, 1996 is clearly shown in Figure 17-4. Both nitric oxide and nitrogen dioxide varied within a small range of about 0 to 12 ppb for NO and 2 to 15 for NO_2. There was no discernible correlation between the two species, except at very low NO concentrations (below 3 ppb), where NO_2 seemed to increase with increasing NO. Ozone varied between 20 and 40 ppb, independently of the concentration of NO_2. At this time of the day then, the formation of ozone should be the dominating process. The decrease of ozone which accompanied the decrease of NO and NO_2 was most unusual and was probably a result of meteorology and transport.

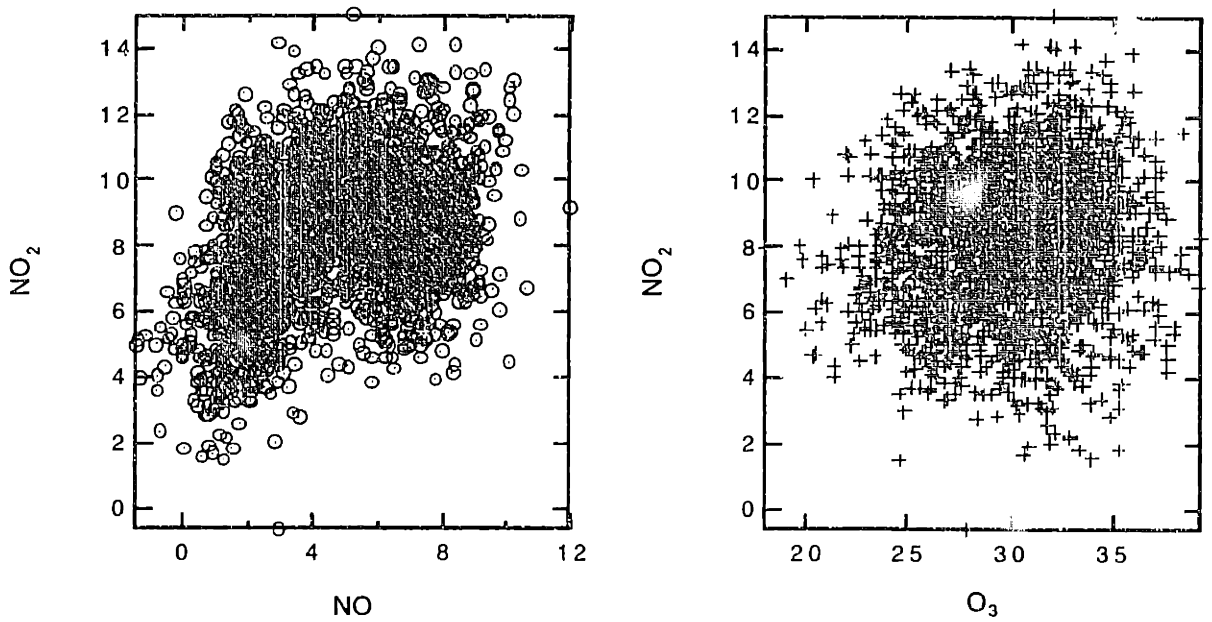


Figure 17-4. Scatter Plots of Concentrations, 10 - 11 a.m., August 22, 1996.
(a) NO_2 vs. NO. (b) NO_2 vs. O_3

The night-time correlations between ozone, NO, and NO_2 are shown in Figure 17-5. In the evening, the mixing time between different air parcels was long, and the pollutant concentrations were determined by the chemical reactions. After sunset, ozone ceased to be produced by the photolysis of NO_2, and the reaction $NO + O_3 \rightarrow NO_2$ dominated the behavior of the ozone / nitric oxide / nitrogen dioxide system. NO was continually emitted throughout the evening and converted all the available ozone into NO_2 through this titration reaction. Therefore, NO_2 increased with NO. The maximum amount of NO_2 formed was controlled by the amount of ozone available for the conversion reaction. In the early evening (Figure 17-5(a)), before the afternoon air trapped above the mixing layer had a chance to dissipate, a ceiling value for NO_2 was observed at about 50 ppb, which was also the maximum ozone level during the previous afternoon. The ozone-to-NO_2 conversion is

illustrated even more clearly by the anti-correlation between these two species shown in Figure 17-5(b). With a well-mixed common starting point of high ozone and low NO_x, for all air parcels, the conversion between NO_2 and ozone led to a relation with a slope of - 0.86 for the best fit line to the data, with an R^2 value of 0.72. The variation of O_3 could explain a large fraction of the variability of NO_2. Other processes (such as the formation of sink species) might have played a small role in removing NO_2 without consuming ozone.

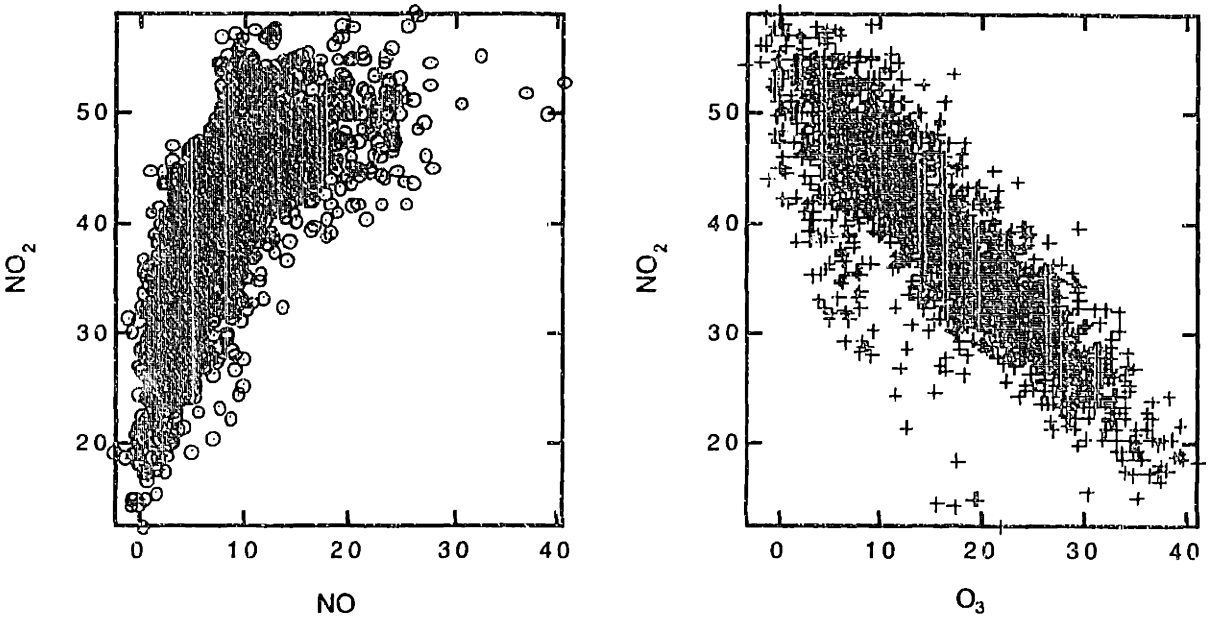


Figure 17-5. Scatter Plots of Concentrations. 10 - 11 p.m., August 22, 1996. (a) NO_2 vs. NO. (b) NO_2 vs. O_3

During the mid-afternoon, NO_2 and NO were positively correlated to one another, as shown in Figure 17-6(a). Even though there was considerable scatter in the data, the ratios of NO_2 to NO in most air parcels were between 3 and 7, with a mean of 6.2, which was generally consistent with, the assumption of photostationary state, which predicted a ratio of 6.24 (based on mean ozone = 52 ppb, k(300K) = 0.03 $ppb^{-1}min^{-1}$, j(scaled to UV) = 0.25 min^{-1}). The assumption of photostationary state will be evaluated for short time-scale data later in this chapter.

The tight anti-correlation of ozone and NO_2 observed in Figure 17-6(b) suggested that some systematic interconversion took place during the afternoon between the two species. The slope of a best fit line was -0.59, and the R^2 value was 0.64, both were somewhat lower than for the nighttime correlation. If a common background level existed for all air parcels, the correlation should have a slope of -1, regardless of the validity of photostationary state assumption, because Reactions 17-1 dictated that there is a negative one-to-one correspondence between ozone and NO_2; therefore, the sum of NO_2 and ozone should be conserved, as was in the evening of August 22, 96.

$$NO_2 + h\nu \longrightarrow NO + O_3$$
$$NO + O_3 \longrightarrow NO_2 + O_2, \qquad (17.1)$$

One possible explanation was that the chemical processes between ozone, NO, and NO_2 (in Reactions 1) were not the only factors controlling the concentrations of NO_2 and ozone.

For example, if other processes were important to the removal of NO_x without consuming ozone, the NO_2-O_3 correlation would have a "flatter" slope. Although the formation of sink species no doubt occur in the high NO_x concentration region near the highway source, the time scales of formation of PAN and nitric acid were from one to four hours and should not affect the interconversion between ozone, NO, and NO_2 occurring in the faster time scale.

The observed correlations may be explained by air parcels of different origins passing through the monitor that were not necessarily well-mixed. In that case, the correlations observed might have been a result of a collection of random samples, each of which contained concentrations of ozone, NO, and NO_2 that were controlled by photostationary state or another process. This explanation also seemed unlikely, partly due to the meteorology observations that were inconsistent with a poorly mixed system, and partly due to the improbability of a tight correlation from a random collection of samples from such a system.

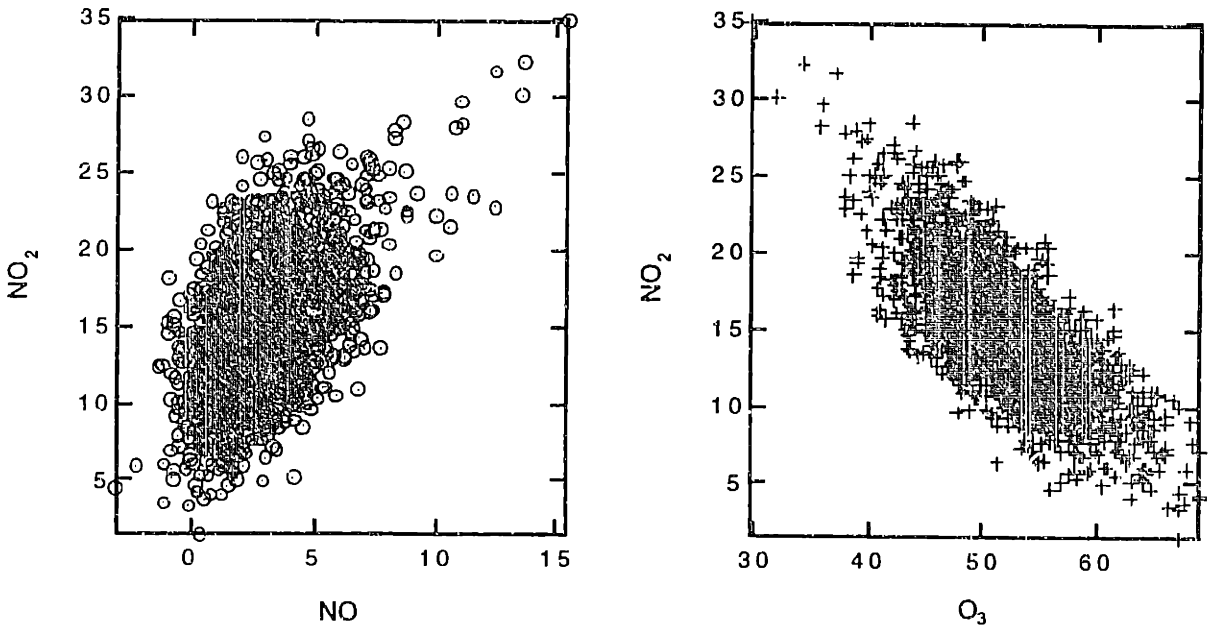


Figure 17-6. Scatter Plots of Concentrations. 2 - 3 p.m., August 23, 1996.
(a) NO_2 vs. NO. (b) NO_2 vs. O_3

Although a well-mixed background atmosphere was more consistent with meteorological observations, the relative amounts of ozone and NO_2 in each measurement depended on the trajectory and NO_x content of the air parcels. Ozone is continuously formed and destroyed by the photolysis of NO_2 and the reaction with NO. However, these reactions have time scales of a few minutes. Since a highway was close to the monitoring site, air parcels moving through the region might have received NO injection from this local source. The perturbation from photostationary state, coupled with the intrinsic fluctuations of a well-mixed background, was a probable explanation for the observations, especially if the transport conditions did not allow an equilibrium to be reached after the NO injection. This explanation was not confirmed because of the difficulty in determining the Lagrangian time scale of turbulence (as explained in Seinfeld, 1986, Chapter 13) from fixed point (Eulerian) turbulence data.

17.2 The effects of meteorology on ambient measurements

Much of the variability observed in the concentration measurements was due to micrometeorology fluctuations. The atmosphere sampled was not a homogenous mixture, but consisted of air parcels which alternatively stayed intact and exchanged pollutants and energy with other parcels. The relationship between the meteorology and concentration of air samples are investigated next.

17.2.1 General Observations.

The effects of meteorological parameters on short time-scale ambient measurements were extremely important in the evening when the mixing between different air parcels was minimal. Figure 17-7 shows the meteorological measurements taken by the sonic anemometer for the period of 10 to 11 p.m. on August 22, 1996, corresponding to the concentration measurements shown in Figure 17-2.

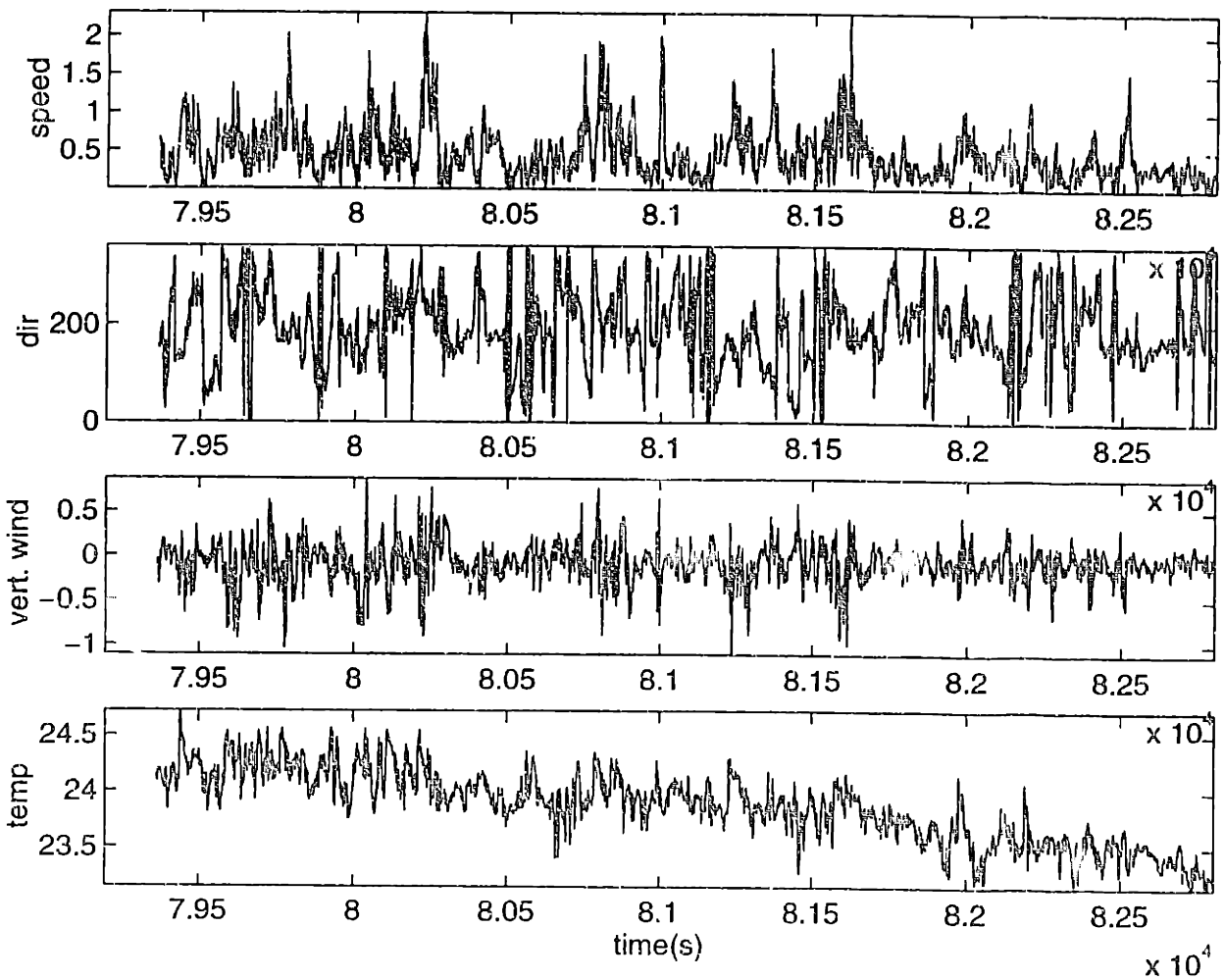


Figure 17-7. Meteorology Plots, 10 - 11 p.m., August 22, 1996.

Wind speeds were measured in ms^{-1}, temperature was measured in °C, and wind direction was expressed as an angle from 0 to 360 degrees, with 0° corresponding to the North direction, 90° to the East direction, etc. Ineffective mixing among air parcels was evident from the low wind speed during that period. Due to poor mixing at this time, each parcel passing by the sampling point was associated with micro-meteorological "identity" that was characteristic of the origin of the parcel. The average surface level temperature decreased continuously at this time of the evening as a result of radiative cooling. During the evening, spikes of elevated temperatures were indicative of air parcels which originated from above the mixed layer.

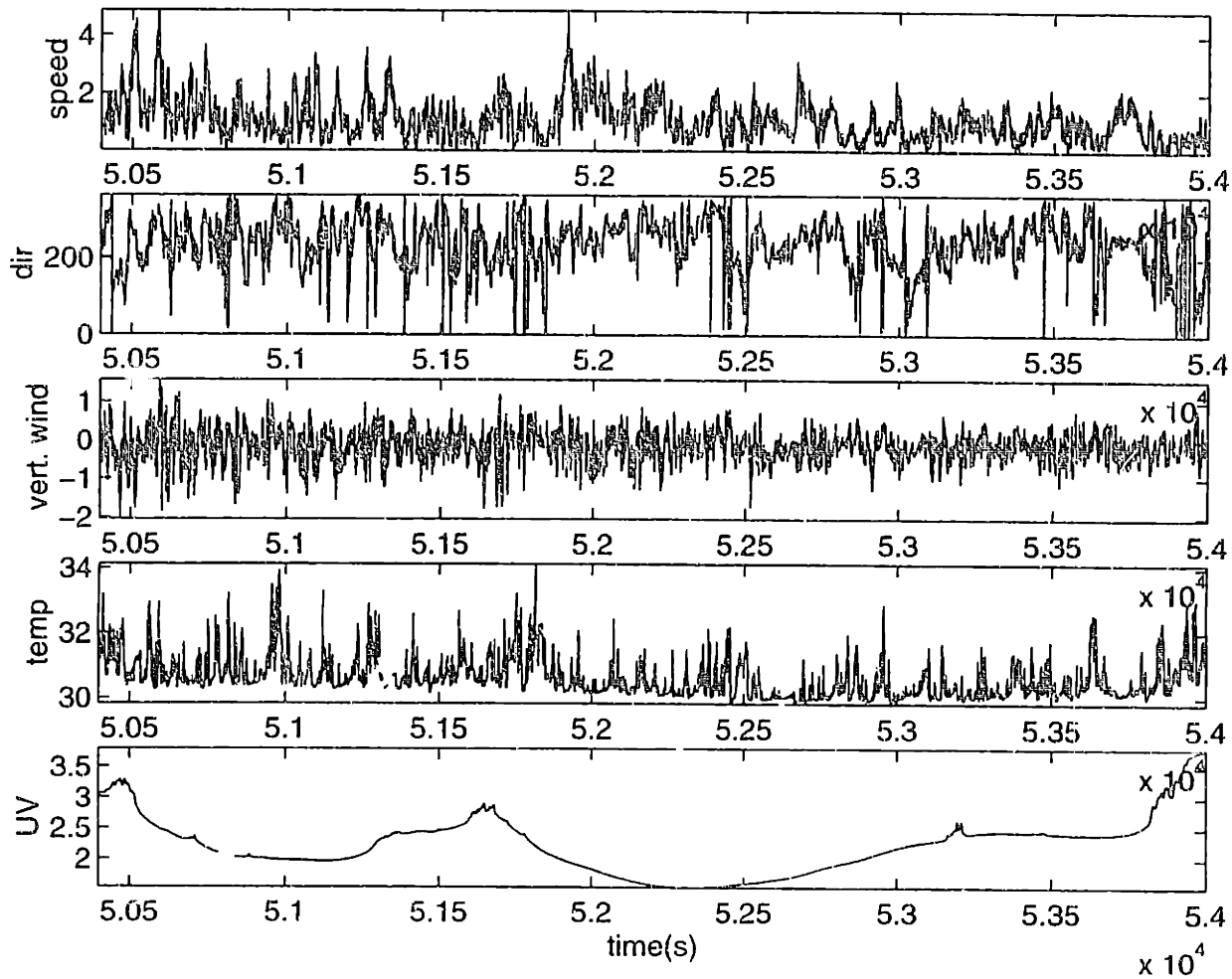


Figure 17-8. Meteorology Plots, 2 - 3 p.m., August 23, 1996.

During the day, the correlation between meteorology and concentrations was expected to be much less pronounced, because mixing was rapid in the boundary layer. As shown in Figure 17-8, higher averaged horizontal and vertical wind speeds were typically observed during the day, consistent with stronger mixing. At this time of the day, the wind was generally from the West. The temperature exhibited positive "spikes", or excursions, instead of random fluctuations around a mean trend. These warm temperature spikes may have originated from the warm roof surface, as they tended to be associated with upward wind velocities, indicating the possibility of surface effects on micrometeorological measurements in the afternoon. Ultra violet (UV) radiation intensity (mW/cm^2) was also measured by the radiometer described in Chapter 14. UV varied relatively slowly compared to the other meteorological

measurements. Clear sky conditions were not satisfied during this period due to the passage of clouds.

No strong correlations between meteorology (Figure 17-8) and concentrations (Figure 17-3) were immediately obvious from the short time-scale data. Therefore, while individual measurements could be analyzed during the evening to draw interesting conclusions for the evening data, a statistical approach was used for the data during the day to elucidate the possible relationship between some meteorological parameters and concentrations.

17.2.2 Evening Correlation between Concentration and Meteorology

As discussed in the previous section, strong correlations were observed between ozone, NO, and NO_2 in the evening of August 22, 1996. The NO_2-O_3 correlation can be explained in terms of the origins of poorly mixed air parcels, by correlating the concentration variables to the instantaneous temperature which was a good indicator of the air parcel's origin. Warmer air parcels were likely to come from aloft, where the afternoon air was trapped when the evening inversion set in. The correlations between the concentrations and temperature fluctuations (after the removal of the linear negative trend) are shown in Figure 17-9. Not having been exposed to high NO emissions, parcels from aloft carried with them pollutant concentrations characteristic of mid afternoon – high ozone, low NO_x. The mixing of air from aloft was primarily responsible for the fluctuations observed in the evening. Samples that contain relatively high ozone and low NO_2 are associated with positive temperature fluctuations. The time scale of this vertical mixing was expected to be quite small, as evident by the sharp increase in temperature and ozone, and the opposite movement of NO_2. As the air parcel slowly dissipates, the temperature and ozone concentration at the monitor decayed, and NO_2 rose back to the levels predominant in the boundary layer.

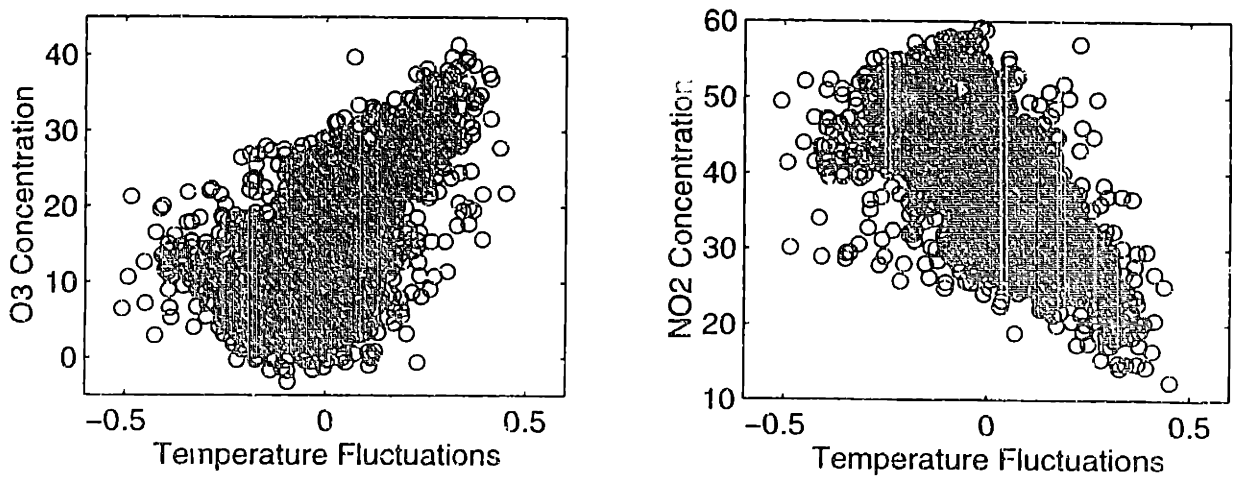


Figure 17-9. Scatter Plots of Concentration vs. Temperature, 10 p.m., August 22, 1996.

Although the mixing of air from aloft explained the correlations between the concentrations and temperature, there was no observable correlation between vertical wind speed and the observed concentrations. Due to the time scale of the mixing event, this relationship may be difficult to elucidate from the data. Although downward moving air parcels are expected to bring high ozone and low NO_2, the converse was not true. High ozone and low NO_2 were likely to linger long after the mixing event. To show this relation, the samples were divided

into two groups according to the vertical wind velocity measurements. A histogram of the ozone concentration was made for each group. Figure 17-10b shows that the 10% of the samples associated with the highest downward wind readings have a disproportionate number of high ozone measurements compared to the other 90% of the samples.

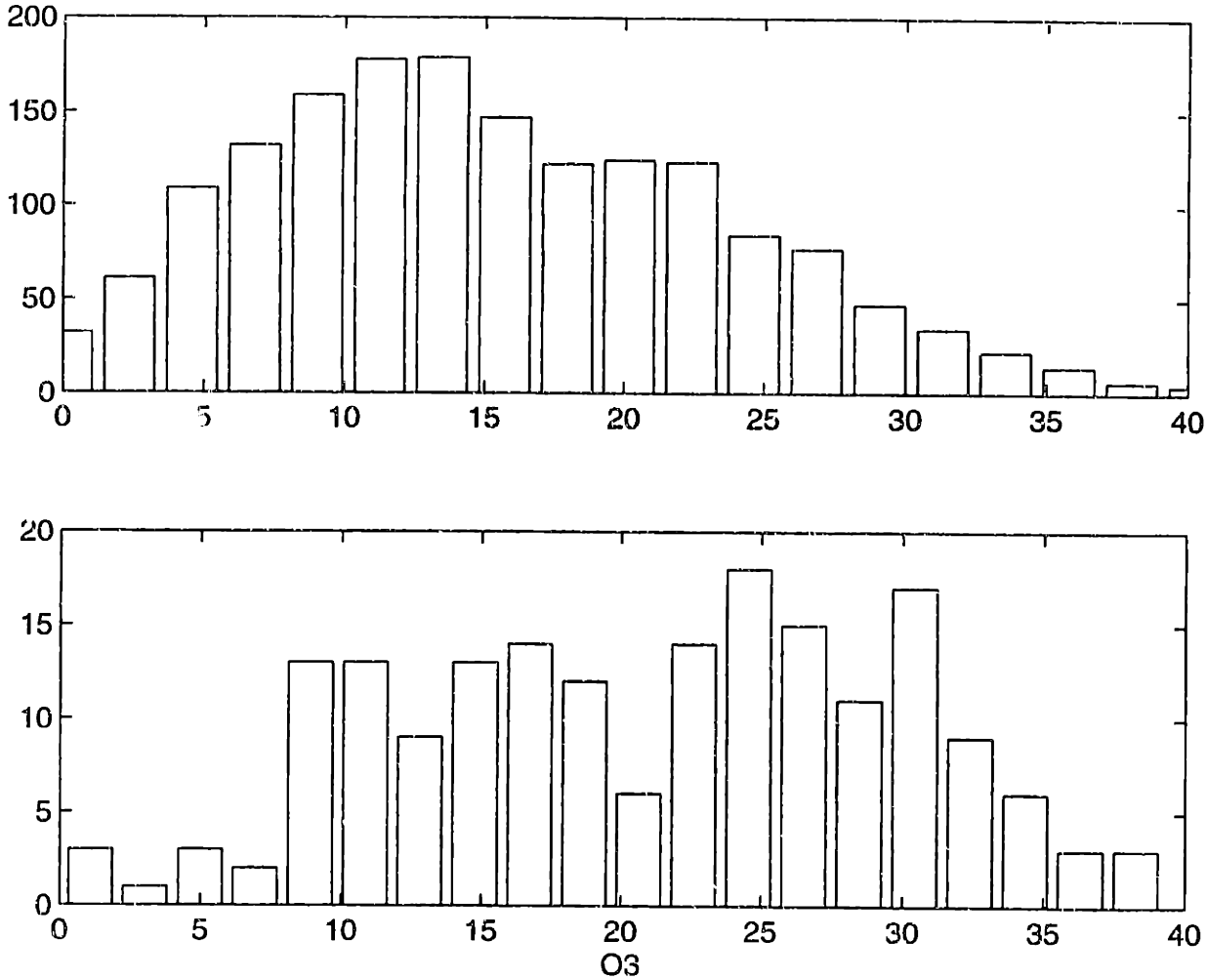


Figure 17-10. Two Histograms of O_3 Levels, 10 - 11 p.m., August 22, 1996
(a) 90% of Hourly Data: Samples Associated with Positive Vertical velocity and Small Negative Velocities, (b) 10% of Data: Samples with the Largest Downward Velocities

As shown previously in Figure 17-5, the concentrations of NO correlated with that of NO_2 at relatively low concentrations during the evening. Vertical mixing introduced into the boundary layer air with low NO and NO_2 concentrations. High temperatures were associated with low NO concentrations of air parcels from above the mixing layer. The negative trend in the correlation of NO concentration with temperature fluctuations, shown in Figure 17-11, is an indication of the effect of vertical mixing.

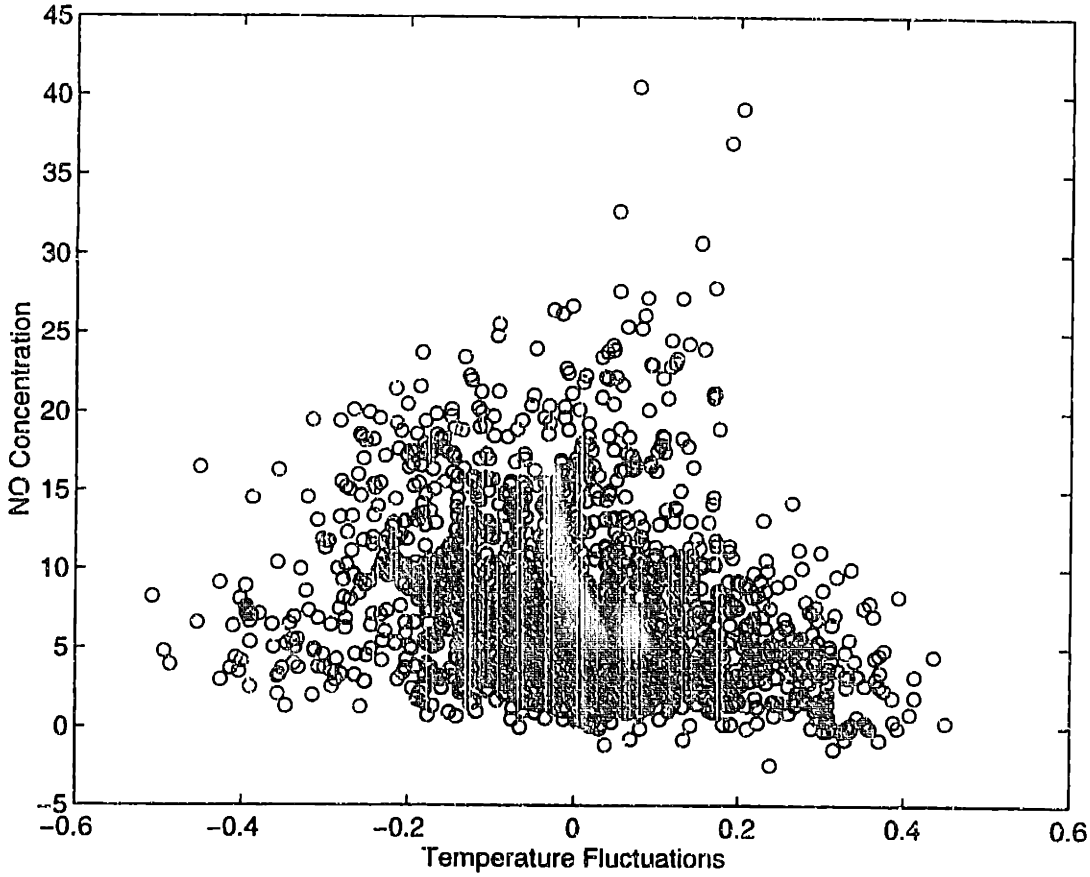


Figure 17-11. NO vs. Temperature Fluctuations, 10 - 11 p.m., August 22, 1996

However, another process was responsible for the relatively high NO concentrations that did not correlate well with the temperature of the air parcel. Emissions of nitrogen oxides from local traffic explained the "spikes" observed in the evening NO data that was not correlated to changes in NO_2 (relatively high) and O_3 (quite low). The (10%) subgroup of data points with the highest NO readings was shown to be associated with lower horizontal wind speeds (median wind speed = 0.27 m/s) than the rest of the samples (median wind speed = 0.4 m/s) (Figure 17-12). This seemed to indicate that slow advection of local emissions was a necessary condition for high NO concentration spikes, although low wind velocities were not always associated with high concentrations. No substantial differences were observed for the wind direction and vertical velocity between this group and the rest of the data points.

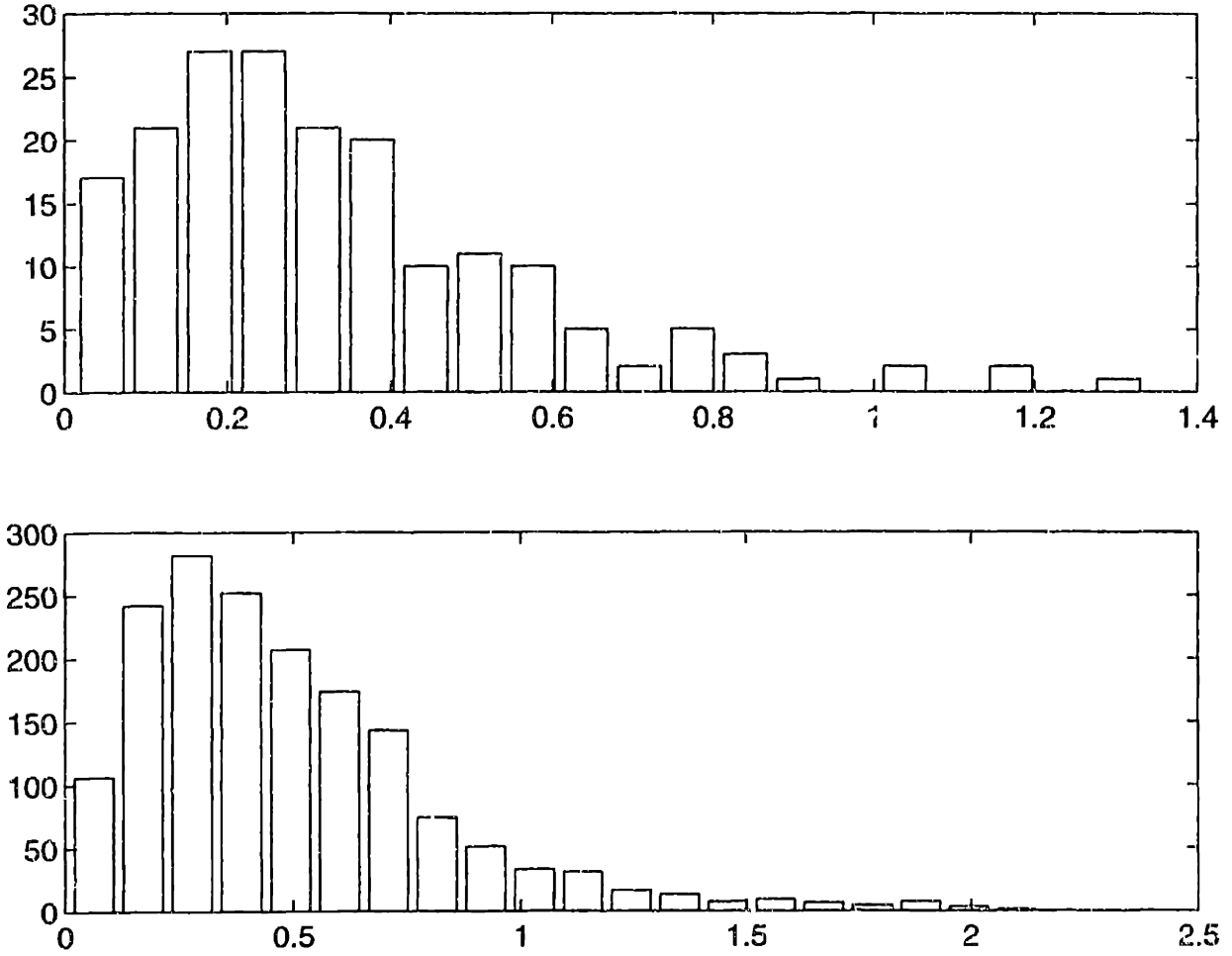


Figure 17-12. Two Histograms of wind speeds, 10 - 11 p.m., August 22, 1996 (a) wind speed associated with the highest 10% NO observations. (b) wind speed distribution associated with 90% lower NO observations

17.2.3 Afternoon Meteorology-Concentration Relationships

The processes that dominated the fluctuations of pollutants during the evening did not explain the concentrations of ozone, NO, and NO_2 during the day. While the effect of meteorology in the observed evening concentrations was clearly shown in the correlations of the species concentrations with temperature, these correlations were not observed during the day. Figure 17-13 shows the concentrations vs. temperature plots for sample of data at 2 p.m. on August 23, 1996. Since the atmosphere was better mixed in the afternoon, the temperature of an air parcel was not such a good indicator of its long-range origin as it was during the evening. Upward temperature excursions seemed to be associated with relatively low ozone and high NO_2 concentrations. It was postulated that higher temperatures increased the rate of the chemical reactions, such as $NO + O_3 \rightarrow NO_2 + O_2$, thus affecting the observed concentrations. Higher temperature was associated with upward wind movements. Near the concrete ground in an urban environment, higher temperature increased the rate of reaction, and produced more NO_2 at the expense of ozone.

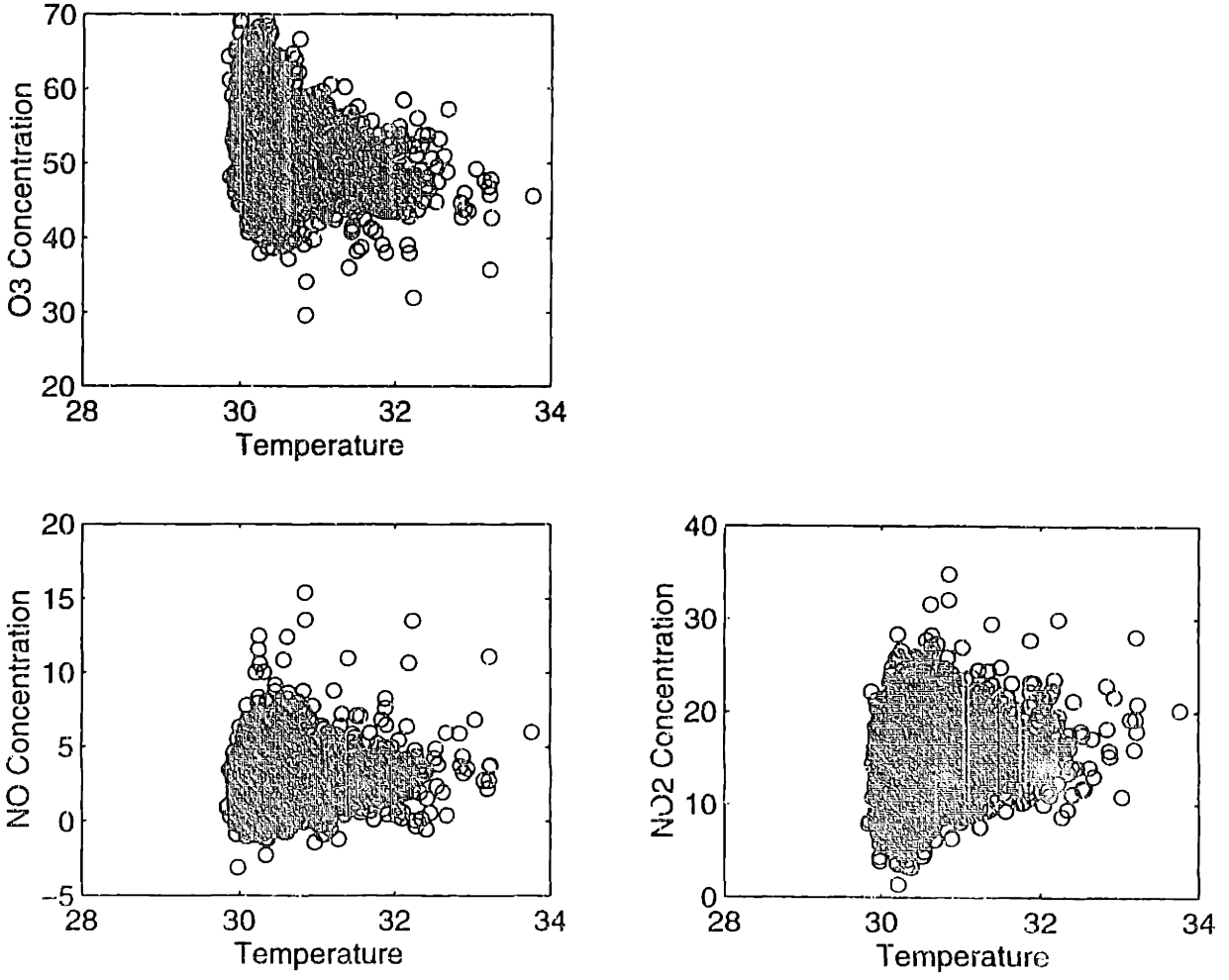


Figure 17-13. Scatter Plots of Temperature and Concentrations, 2 - 3 p.m., August 23, 1996

Wind speeds also played a role in the concentrations of species in the air parcels. Higher ozone seemed to be weakly correlated with higher wind speeds, which prevented the accumulation of locally-emitted NO, the titrating agent for ozone. High NO concentration, on the other hand occurred exclusively when the horizontal wind speed is below 2 ms^{-1}.

The correlations of NO, NO_2, and ozone, displayed in Figure 17-6, contained much scatter. Some of this scatter were results of air parcels of varying origins and thus varying amounts of ozone, NO, and NO_2 carried to the sampling point by meteorological fluctuations. This section addresses the validity of photostationary state applied to a suburban site by comparing the predicted relationship to observations of concentrations and meteorology. Under the assumption of photostationary state, the short time scale relations of NO, NO_2, and ozone are dictated by Reactions 17.1 and the concentration of ozone is related to the ratio of concentrations of NO_2 and NO:

$$[O_3] = \frac{j \cdot [NO_2]}{k \cdot [NO]}. \qquad (17.2)$$

The photolysis rate of NO_2, j, was determined by a simple scaling of the UV radiometer reading. This was only a crude approximation of the photolysis rate constant j, as the correspondence of total UV to j was only confirmed for the clear sky condition (Parish et al, 1986). The reaction rate, k, of NO + O_3, was calculated for each parcel of air from the temperature observed. It was determined previously that the average ratio of NO_2 to NO for the hour was smaller than that predicted by the photostationary state in Section 17.1. Figure 17-14 shows a plot of $j*[NO_2]$ vs. $k*[NO]*[O_3]$ from the observed data. The correlation was good, with a correlation coefficient of 0.59, better than the correlation of NO_2 to NO. The slope of the correlation was smaller than 1, indicating that the photostationary assumption was not very accurate at the short time-scale in general.

Figure 17-15 shows a distribution histogram of the ratio of $k \bullet [O_3] \circ [NO]$ to $j \circ [NO_2]$, which should be equal to one if the photostationary state was satisfied. The mean of the ratio was 1.67. Most of the deviation was towards low NO_2 to NO ratio for given ozone measurements, which indicated that the air parcels contained too much NO and / or ozone to explain the low readings of NO_2. It was argued in Section 17.1 that although the formation of nitric acid or PAN could explain the deviation from photostationary state and the low that NO_2 to NO ratio, these reactions occurred at a much longer time scale than the time scale of Reactions 17.1 to affect the equilibrium significantly. Since predominant wind was from the West where a commuter highway runs within 500 m, air parcels reaching the monitoring site received fresh NO emissions from highway vehicles in close proximity of the monitoring site. The time taken for the air parcels to reach the monitoring site from the highway, based on the averaged wind speed, was of the order of minutes, which was only slightly longer than the time scale for the Reactions in 17.1. The Lagrangian time scale of turbulence was not calculated, although it would have been a better indicator of atmospheric mixing. If the background air parcels were assumed to be fairly well-mixed and to be in photostationary state, a likely explanation for the observed deviation from the photostationary state was that air parcels passed through a local NO source and received a varying among of NO emission, but had not the time to reach a new photostationary state before reaching the monitor.

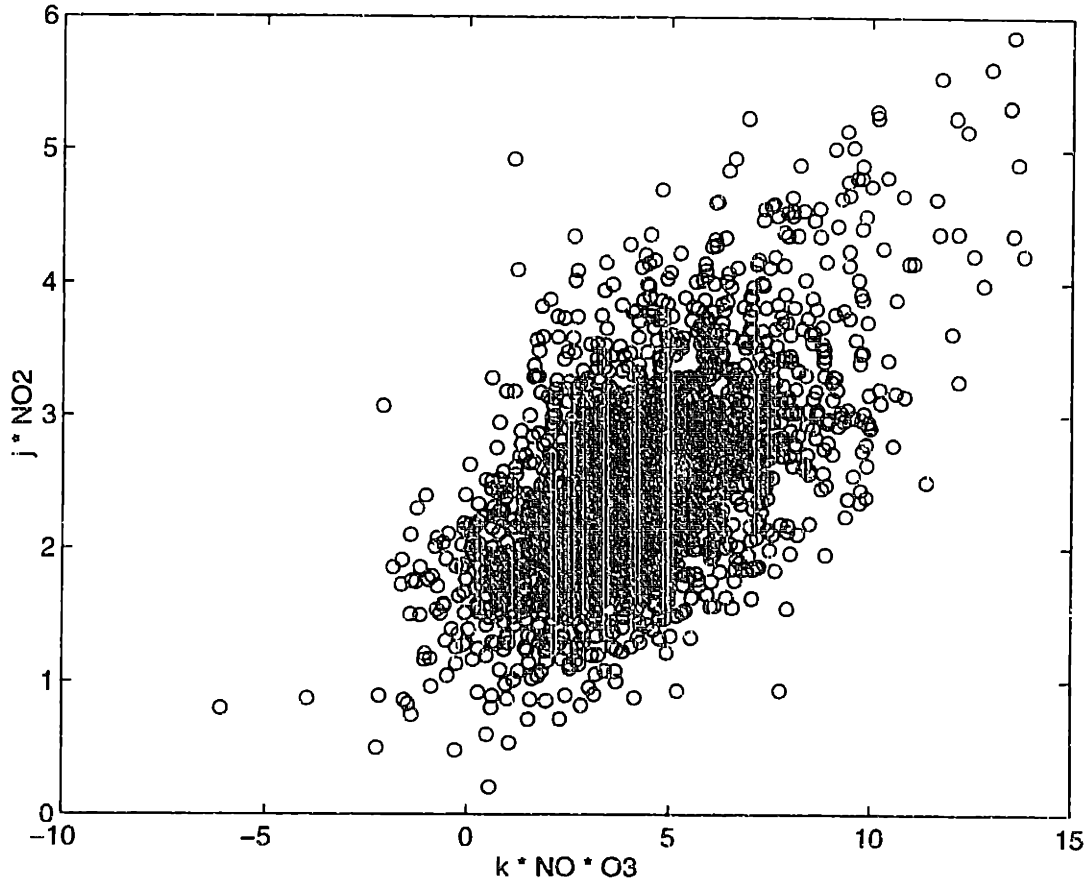


Figure 17-14. O_3 Measured vs. O_3 Calculated using Photostationary State Assumption Given NO, NO_2, and Meteorological Measurements, 2 - 3 p.m., August 23, 1996

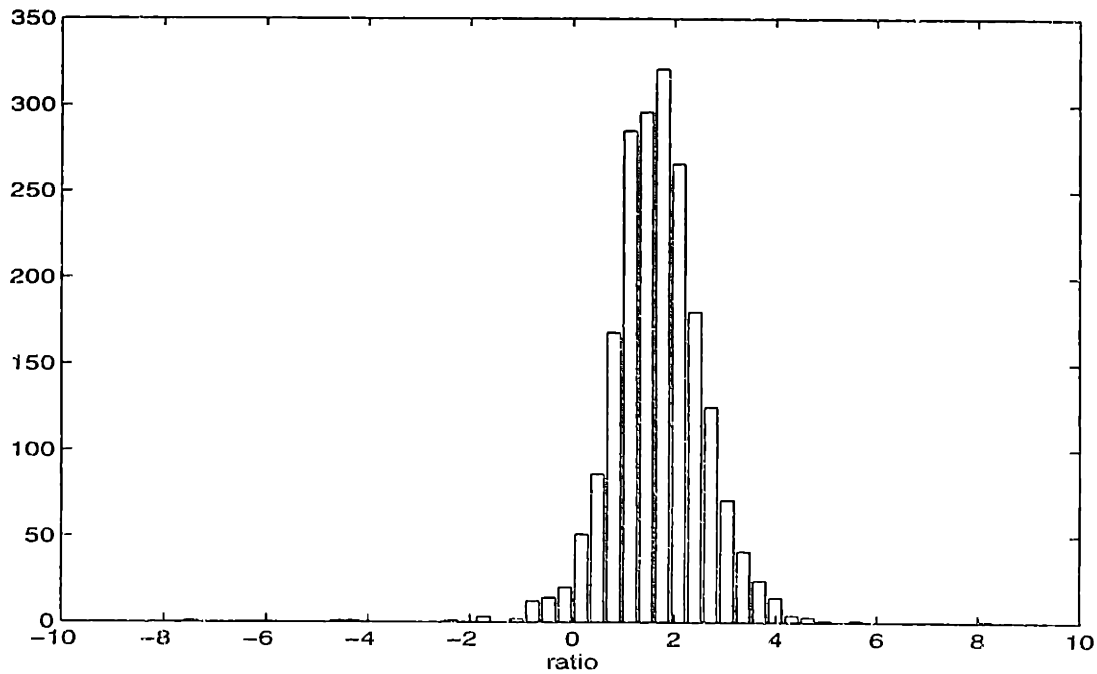


Figure 17-15. Histogram of Ratio $\frac{k[O_3][NO]}{j[NO_2]}$, 2 - 3p.m., August 23, 1996

17.2.4 Morning Observations

The data during mid-morning (10 a.m.) on August 22, 1996 (Figure 17-1, Figure 17-4) was unusual in a couple of ways: the small variance recorded with the data, and the different correlation behavior in the ozone-NO-NO_2 system. The small variance in the concentration time series (the standard deviations of ozone, NO, and NO_2 were 3.5, 2.6, and 2.5 ppb) was evidence for a well-mixed system, where parcels of air exchanged pollutants rapidly. The relatively high wind speeds observed during that period, averaging at 1.5 ms^{-1} with gusts of up to 4 ms^{-1}, also indicated strong mixing. The concentration trends of the species were also unexpected. While NO and NO_2 expectedly decreased during the morning after the peak hours of traffic, ozone, which had been increasing earlier in the morning, ceased to grow. There was very little observable correlations in the short-timescale fluctuations in the pollutant concentrations. As shown in Figure 17-16, photostationary state was not attained by the majority of the individual samples. If the mixing time-scale was indeed quite fast, pollutants such as NO_x would be transported in and out of the system before they had time to react to form ozone. The deviation of the observed concentrations of ozone, NO, and NO_2 could be explained by emissions of NO which were mixed in the local atmosphere before equilibrium was achieved.

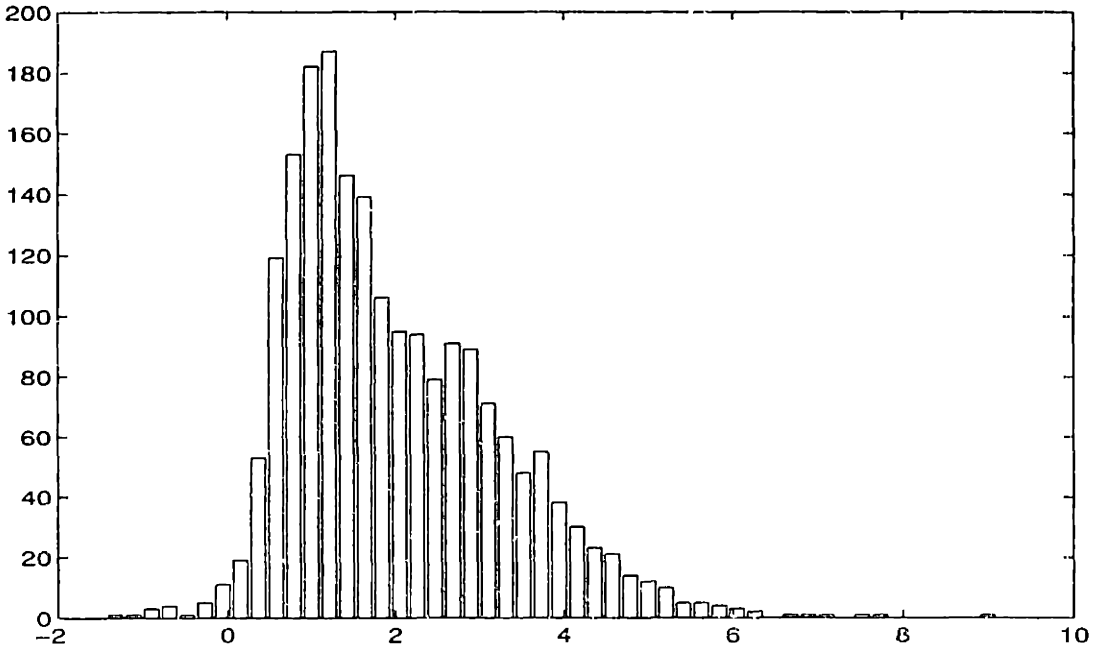


Figure 17-16. Histogram of Ratio $\frac{k[O_3][NO]}{j[NO_2]}$, 10 - 11 a.m., August 22, 1996

Micro-meteorological variations were expected to have little effect on the pollutant concentrations of a well-mixed system, where material transfer between different samples or air was fast compared to other processes that affected the concentration. This would explain a general lack of correlation between the concentrations of ozone, NO, and NO_2 and simultaneous measurements of micrometeorology, such as wind speed and direction. A

decrease in UV towards the end of the period did not result in a build up of NO_2. The only meteorological variable which seemed to have any influence ozone fluctuations was temperature, with higher temperature fluctuations weakly associated with lower ozone concentrations.

Although a slight shift in wind directions (from SE to East) was observed during the sampling period on August 22, 1996, none of the local meteorological variables underwent a definitive change that could explained the reduction in ozone during this mid morning hour. There could be a number of reasons, including the long-range effect of cloud cover in a well-mixed system. Another explanation for the decrease of ozone concentration was that the dilution effect was dominating over the chemistry. The reduction of pollutant concentrations would simply be a result of advection and dilution due to the increase of mixing height. The height of the mixed layer and the concentrations aloft were not measured in these experiments, and there was no evidence to support this theory.

17.2.5 Bimodal Distribution of Ozone

Several bimodal distributions were observed in the ozone hourly measurements. One such distribution was from the midday data on August 20, 1996. On the inspection of the time series data of ozone, NO, and NO_2, as shown in Figure 17-17, it could be seen that ozone increased in "steps" and resulted in the two modes at ~28 ppb and ~40 ppb. During this period, NO_2 stayed relatively constant and NO concentration decreased somewhat but exhibited trends that were only weakly correlated to that of ozone. The data stream could be divided into different sections, each of which exhibited relatively tight correlations between ozone, NO, and NO_2, but each had different slopes and intercepts for the correlations between the trace species.

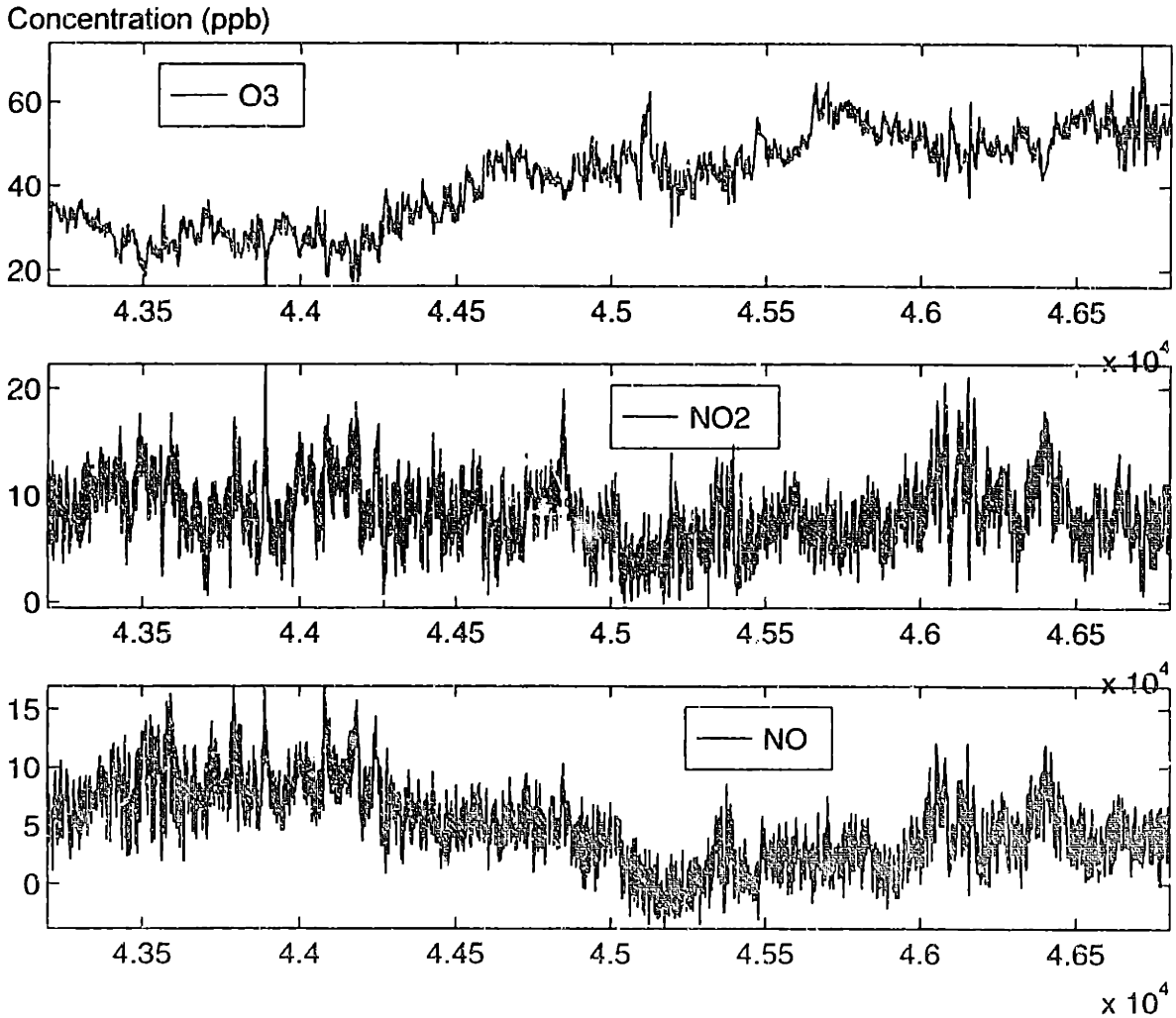


Figure 17-17. Ozone, NO, and NO_2, Noon - 1 p.m., August 20, 1996.

The first 20 minutes in the hour was associated with relatively low ozone level and higher NO concentrations. Relatively high ozone concentrations and lower NO concentrations were observed during the last 40 minutes of the hour. The meteorology data are plotted in Figure 17-18. As observed in Figure 17-18, the averaged wind speed decreased during the hour but

became somewhat gusty, and the calculated wind direction also became more varied as a result of the reduced wind speeds. There was a strong component of westerly wind throughout the hour. During the first twenty minutes, the majority of the samples were from a 135° sector between the West and SE direction. Then as the wind became lighter, the SE direction was no longer as dominant as before, and the wind direction distribution showed a mode in the South to West direction, with more scatter. A series of clouds passed above the monitor during the last 20 minutes of the data period. In a well-mixed system, air parcels under a cloud might not stay under the cloud for a long time. Therefore, clouds could have effects on the concentrations that were picked up by the ozone / NO / NO_2 monitor long before or after the radiometer registered the cloud's passage, depending on the wind and mixing characteristics of the area.

No obvious explanation for the bimodal behavior of ozone could be found in the accompanying meteorological measurements. Maybe subtle changes in meteorology affected the concentrations in a very profound way. On the other hand, the distribution of measurements of ozone and other pollutants was probably a very complex interplay between chemical and meteorological phenomena of very different time and length scales. The understanding that can be inferred from a stationary point measurement was quite limited in these situations.

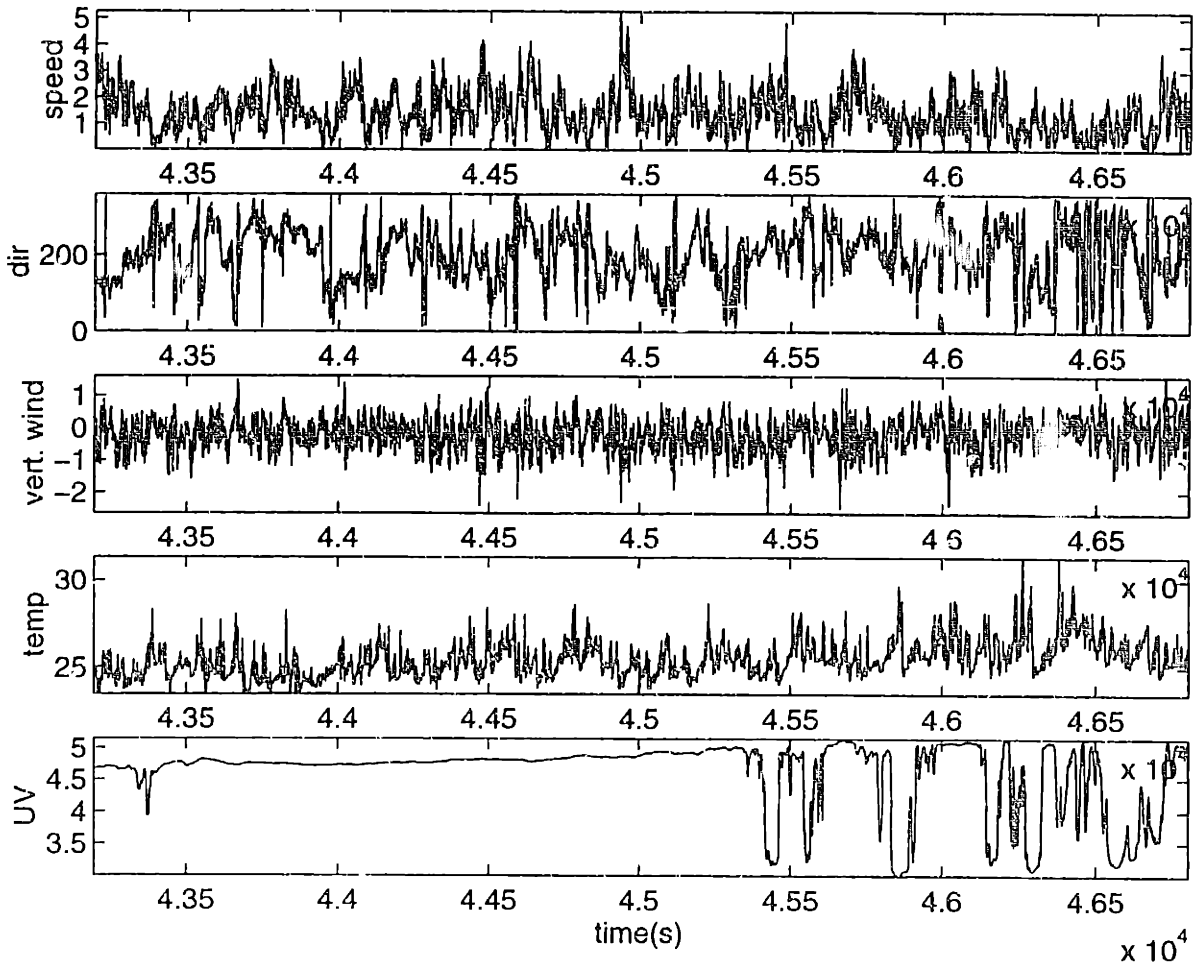


Figure 17-18. Meteorology Data on August 20, 1996 at noon.

17.3 Deposition fluxes

The eddy correlation method for measuring fluxes has been employed for methane and other trace gases using tunable diode laser spectroscopy (Anderson and Zahniser, 1991). The basic idea behind the method is that the turbulent fluxes of a species can be measured directly by the co-variance between the concentration and the vertical wind speed, as given by

$$\text{vertical flux} = w'c' = (w - \langle w \rangle) \cdot (c - \langle c \rangle), \qquad (17.3)$$

where w' and c' are the fluctuations of the vertical wind speed and concentration about their respective mean values. The symbol <.> denotes the temporal average of the quantity. The calculation of fluxes usually require measurements of faster than 1 Hz, preferably at 10 Hz, since turbulent fluctuations in the atmosphere extend to that frequency. Any conclusions regarding the fluxes in an urban environment drawn from this data, which was taken at less than or equal to 1 Hz, would be limited in accuracy. A standard technique to obtain flux information is to apply a correction factor based on the measurement of another scalar, such as temperature (Munger *et al.*, 1996), whose data rate was more suitable for flux measurements. The correction factor was determined by the ratio of a heat flux calculated using the high frequency data and that calculated from numerically-filtered data appropriate for the frequency of the concentration measurements.

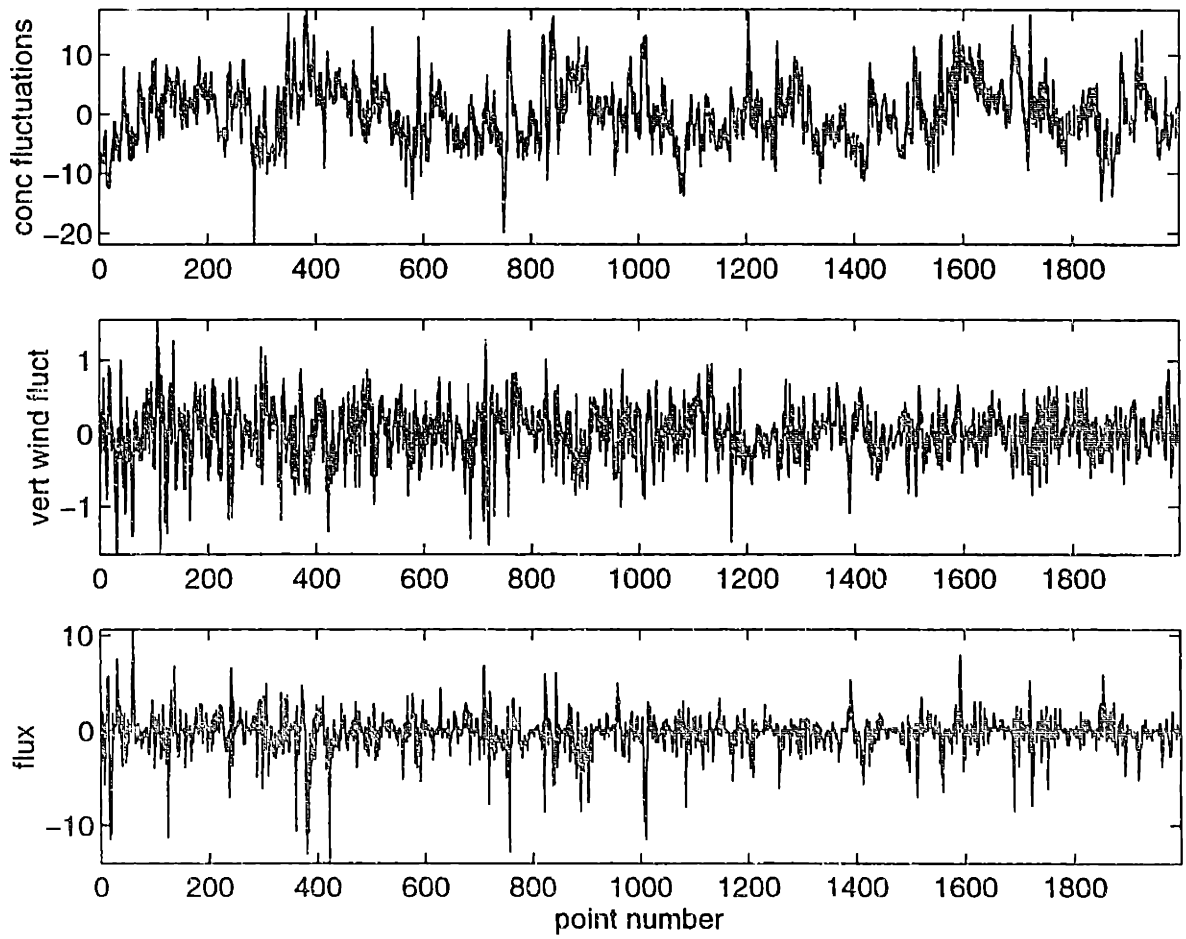


Figure 17-19. Calculation of O_3 Flux by the Eddy Correlation Method (2 - 3 p.m., August 23, 1996).

The instantaneous fluctuations of ozone and vertical wind velocities are shown in Figure 17-19. The averaged flux during the afternoon period shown here was -0.311 ppb * ms^{-1}, indicating that ozone flux is in the downward direction. From the heat flux calculation, it was determined that the flux was underestimated by about 11% because of the attenuated high frequency fluctuations. Therefore, the corrected ozone flux was -0.345 ppb * ms^{-1}.

The observed fluxes for NO and NO_2 were both positive at the sampling port, indicating a net upward movement of these trace species. The fluxes calculation is shown in Figure 17-20 for nitrogen dioxide. The average flux was found to be +0.203 ppb ms^{-1}. Nitric oxide had a flux of 0.005 ppb ms^{-1}, which is essentially zero because of experimental errors.

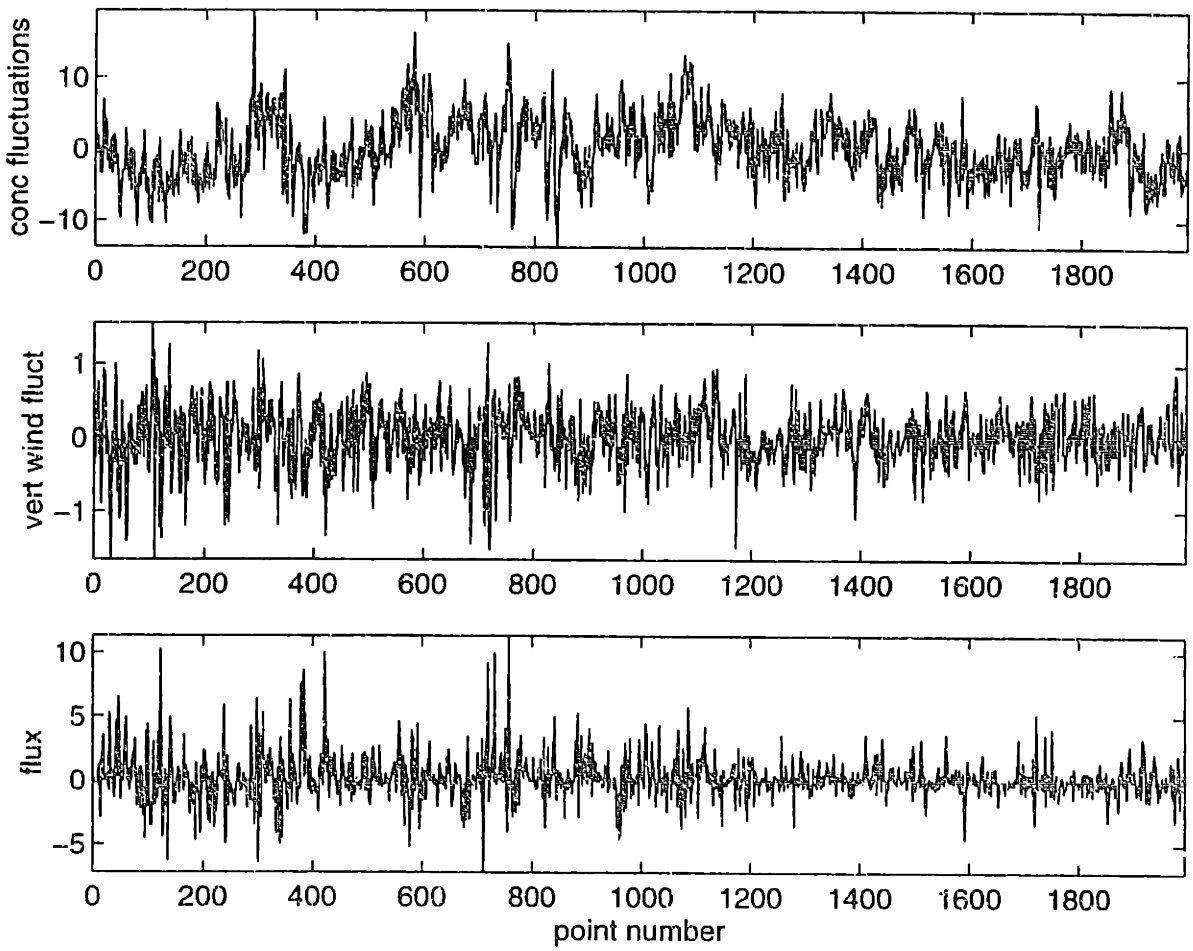


Figure 17-20. Calculation of NO_2 Flux by the Eddy Correlation Method (August 23, 1996, 2 - 3 p.m.).

For inert species, the deposition velocity, which is used in three-dimensional models to describe the dry removal processes of pollutants, can be calculated by dividing the average flux by the average concentration in the bulk atmosphere:

$$\text{deposition velocity} = \frac{\langle w' c' \rangle}{\langle c' \rangle}. \qquad (17.4)$$

However, the measured fluxes for ozone, NO, and NO_2, which were chemically active in the atmosphere, would contain contributions from both the deposition of pollutants to, or emission from the ground, and chemical conversion between the ground and the sampling port. In this case, the observed flux for ozone could provide an upper limit for the deposition velocity of ozone, assuming little chemical removal took place between the ground level and sampling port. The deposition velocity calculated from Equation 17.4 should be - 0.67 cm/s. To simplify the picture, NO_x (NO + NO_2) could be considered chemically inert, as NO and NO_2 interconvert. The reactions that form or destroy NO_x (formation of PAN and nitric acid) occur at a much longer timescale than deposition. NO_x had a flux of + 0.2085 ppb ms^{-1}, which translates to a deposition velocity of 1.16 cm/s (emission) at ground level. Note that the NO_x fluxes were not corrected by the heat flux, because the power behavior was different for the NO and NO_2 fluctuations than for ozone and temperature.

The night-time fluxes of ozone and NO_x on August 22, 1996 were smaller than the day-time values. The average ozone flux was -0.1041 ppb ms^{-1} * 1.0728 (correction factor) = -0.112 ppb ms^{-1}, which corresponded to a deposition velocity of -0.71 cm/s if ground level chemistry was ignored. The average NO_x fluxes was +0.1315 ppb ms^{-1} (emission).

These fluxes and deposition information should be interpreted with care. Although the range of ozone concentration fluctuations was +/- 20 ppb, due to the uncertainty of the tunable diode laser measurement of ozone (about 2 ppb), the fluxes were probably accurate to about 10% or so. As mentioned before, reaction with NO could be responsible for the removal of ozone between the monitor inlet and the ground. Therefore the absolute value of the deposition velocity of 0.7 was an upper limit for urban (concrete surfaces). The observed emission of NO_x in the ground level turbulent fluctuation correlation was an indication that the field of view of the monitor was more than the roof-top of the Aerodyne building. Since NO_x was not emitted from directly underneath the monitor, the upward fluxes observed were results of the abundant emission on roadways around the monitoring site

17.4 Conclusions

Several interesting conclusions were drawn regarding the interaction between pollutants and between pollutants and meteorology.

Correlations of ozone, NO, and NO_2 were observed during most of the monitoring period from August 20 to August 23, 1996. Ozone and NO_2's anticorrelation and the correlation of NO and NO_2 were very strong during the evening. Meteorological conditions in the evenings were generally quite still, contributing to long mixing times and resulting in the dominating effects of the $NO + O_3 \longrightarrow NO2 + O_2$ reaction on the observed concentrations of individual samples. During the day, weaker correlations were observed. The mixing time was faster so that the concentration fluctuations were determined by both chemical reactions and source/transport effects. There were also a few samples of examples in the morning when no short timescale correlations were observed in the fluctuations.

Micrometeorological measurements correlated with concentrations during the night. During the day, however, when mixing was more rapid, they contributed little to the understanding of the intrinsic fluctuations observed in and between concentration measurements. Concentration trends were frequently unexplained by observations of wind, temperature, and radiation. Although important information, like the mixing height, was missing in this experiment, the lack of definitive interpretation was as much a lack of data to support them as it was an indicator of the limitations of point measurements. When mixing was rapid, the local meteorological measurements did not provide a useful clue to the air parcel's history.

During the day, photostationary state between ozone, NO, and NO_2 was generally not reached at the suburban monitoring site at Billerica. The abundance of locally emitted nitric oxide that was not titrated away by ozone before it reached the monitor could be the reason for this.

Ozone is trapped above the mixing layer in the evening, as evident from the evening meteorology-concentration relationships. This phenomenon is likely to contribute to "multiday" carry-over during ozone episodes. There are also significant implications towards the effect of ozone transport in the northeast region. Ozone and other pollutants trapped aloft may have different transport fate and can have strong effects on air pollution at locations far removed from where the pollutants originally formed.

The flux measurements yielded a deposition velocity of about -0.7 cm^{-1} for ozone at a suburban location. This was an upper limit, because ground level chemistry could be responsible for some of the downward fluxes observed. It was certainly the right sign and was consistent with available information.

Even though the high time-resolution, multi-species data demonstrated quite a few interesting correlations and observations that were unavailable by traditional measurement techniques, the analyses conducted so far were rather rudimentary. Suggestions are given in the Future Directions chapter to extract more information from the time series measurements of concentrations and meteorological parameters.

18. Directions for Future Research

A systematic framework was proposed in this work to use the uncertainties in model predictions and experimental observations to provide for a more informed model validation process, as illustrated in Figure 3-3. The goals of this research were to quantify uncertainties in chemical mechanism and air quality model predictions due to uncertainties in reaction parameters, and to design an experiment to measure atmospheric variability. Several areas of future research are suggested in this chapter to expand on the current work.

The organization of this chapter follows that of the thesis. Areas of research to further the understanding of uncertainties in air quality models, and thus in air pollution control strategies are described first. The remainder of the chapter contains suggestions to improve the experimental design for the multi-colored tunable diode laser system described in Chapters 14 and 15. Further analyses using the available measurements are discussed in the last section.

18.1 Quantifying Uncertainties in Air Quality Models

A comprehensive uncertainty analysis of air quality models can be used not only to facilitate a more rigorous model validation process, but also to inform the design of air pollution control strategies. As shown in Table 1-1, reproduced here, many uncertain parameters may contribute to uncertain predictions of an air quality model. This work focused on the parameters used in the description of atmospheric chemical mechanisms. A thorough uncertainty analysis should be a coordinated effort between a number of research groups with different expertise. Some key areas of research are identified as follows.

Category	Specific details
Emissions	source strength (stationary and mobile sources) composition of source emissions
Meteorology	wind speed and direction, Temperature mixing height, Relative Humidity radiation estimates, Topology wet and dry deposition, Turbulent Diffusion Coefficients
Chemistry	reaction rates product distribution lumping representation
Observations Verification Data	initial and boundary conditions measurement error spatial / temporal variability of pollutant concentrations

Table 18-1. Uncertainties in Data and Observations. (Reproduction of Table 1-1)

18.1.1 Uncertainty Characterization of Parameters

Of the data used to model urban air quality, the reaction rate parameters are the only group for which uncertainties are routinely compiled by review panels (e.g., NASA, IUPAC) together with the data. When no estimate of the experimental error is given with the published measurement, or when conflicting information is reported, the review panels rely on expert judgment to assign an uncertainty factor to the reported measurement. Quantifying the uncertainties in the input is a vital aspect of a successful and informative uncertainty analysis. There is a critical need for standardized representation of uncertainties in the different data bases used for air quality modeling.

Perhaps the most uncertain aspect of air quality models lies with the emissions estimates. Recently, Harley *et al.* (1997) found that using an improved California inventory input, the CIT-Airshed model predicted significantly more ozone than the previous simulations. There is considerable disagreement about the sources (unburnt gasoline (Harley *et al., 1992*)? biogenics?) and magnitude of uncertainties in the inventory and in the composition profiles of organics. These issues need to be resolved to define the uncertain inputs for a systematic uncertainty analysis. Inverse modeling is being used in our research group (Adamkiewicz *et al.*, 1997) to infer the uncertainties in the emission fields from the discrepancies between model predictions and observed data.

18.1.2 Methodology Development (Parametric Uncertainty Propagation)

Although the Deterministic Equivalent Modeling Method (DEMM) has been applied successfully to propagate the uncertainties of a few hundred parameters related to the description of atmospheric chemistry, many more parameters need to be investigated within an air quality model. It is clear that even this most efficient of the uncertainty propagation methods will push on the limits of current computers. In the present study, an iterative procedure was used to successively increase the order of PCE representation and reduce the number of uncertain parameters, as illustrated in Figure 8-1. This procedure needs to be automated for investigation of large models.

An alternative strategy for representing a large number of parameters is to obtain the optimal polynomial chaos expansion (PCE) representation of the response surface using the fewest number of terms. The DEMM error calculation procedure (Section 5.5) provides a natural way to allocate the contribution to the error of representation. The error contribution may be used to decide what terms need to be included in the PCE. In this manner, higher order polynomial chaos will be used for parameters whose surface response is highly non-linear and lower order polynomial chaos will be used for parameters with low sensitivity. This procedure may allow more parameters to be tested simultaneously under the uncertainty framework with the same number of model runs. Parameters need not be eliminated completely, and synergistic effects may be elucidated more clearly. Both theoretical development of the optimal representation of a response surface, and numerical implementation into the DEMMUCOM software framework will be needed.

18.1.3 Structural Uncertainties

Structural uncertainties are always present in the modeling of complex systems where assumptions are made. Some assumptions are made to simplify the system so as to facilitate computation. In this case, structural uncertainties of the model can be quantified by comparison to the more complex model. However, other assumptions are made as the modeler's best guess of a phenomenon that is yet to be resolved, e.g., the parametric representation of aromatic compounds, or the parametric representation of cloud processes. Furthermore, different numerical procedures can be used in the advection scheme and the differential equation solver of the three-dimensional transport and transformation model that causes the predictions of the models to differ. These kinds of structural uncertainties need to be carefully documented in the manual of models, because they may dominate parametric uncertainties, resulting in erroneous model predictions even when parameters used in the model are perfectly defined. Since structural uncertainties of complex models cannot be easily quantified when some processes are yet unknown, the best alternative is to elucidate structural differences of models by comparing their responses to identical parametric uncertainties. An air quality model with a modular design (e.g., Byun, 1994) may provide a good host structure for such a comparison.

18.1.4 Uncertainty Analysis of Key Modules in Air Quality Models

Chemistry. The uncertainty analysis described in Chapter 10 for the CIT Airshed model considered only a subset of reaction parameters. The iterative approach (Figure 8-1) is recommended for completeness when the computational facility is available.

Recently, Harley *et al.* (1997) incorporated the SAPRC mechanism into the CIT model. A three-way uncertainty comparison between the CIT-SAPRC, the existing CIT (LCC mechanism), and the one-dimensional SAPRC mechanism may be enlightening. Although the SAPRC mechanism is considered a more detailed representation of atmospheric chemistry, the differences in terms of the uncertainty variance contributions need to be understood. Furthermore, if the same parameters contribute to uncertainties in both the one-dimensional and the three-dimensional model, a formal procedure can be designed to use the simpler model as a screening tool for parameters that should be investigated in detail.

Emissions. Two uncertainty factors were applied to the sources of NO_x and ROG uniformly in the uncertainty analysis of the three-dimensional CIT model (see Chapter 10). A spatially-resolved or "source-resolved" approach is preferable for identifying the priorities for measurements and research. The uncertainties in organic speciation are also an interesting topic in light of the current interest in ozone reduction strategies based on changing the reactivity of the organic emissions (such as gasoline reformulation).

Meteorology. Meteorological data are usually represented by spatially resolved fields, interpolated from limited observations. The uncertainty of the field representation is a combination of the errors of the meteorological measurements and the results of an interpolation scheme. Both aspects need to be addressed to estimate the effects of uncertain meteorological data on air quality model predictions. A mesoscale model can now be used instead of data interpolation to drive the meteorology routines in three-dimensional models (Pan, 1996). Uncertainties in the meteorology can be directly incorporated into the air quality model predictions.

18.1.5 Uncertainties in Control Strategies

The policy-relevance of this kind of uncertainty analysis is a frequently asked question. Since air quality models are used in the formulation of control strategies, uncertainties in the predictions of ozone necessarily lead to uncertainties in the efficacy of the control strategies. One solution may be to design control strategies using a probabilistic approach by taking into account the underlying uncertainties. For example, while a deterministic model is used to calculate the amount of controls needed to bring a base case scenario into compliance with a standard, an air quality model formulated to incorporate uncertainties may be used to compute the emissions reduction needed to bring 90% of the uncertain scenarios into compliance. A more comprehensive analysis should also address the uncertainties in the efficacy of control strategies in addition to model uncertainties. How uncertainty analysis fits into the scheme of policy design remains an active field of research in urban planning and political science departments.

18.2 Experiments Recommended Based on Uncertainty Analysis

Based on variance analyses that identified parameters whose uncertainties drive uncertainties in the model predictions, a series of experiments were recommended to improve the estimates of the key parameters and to improve the confidence of model predictions. (Please refer to Chapters 8 through 11 for these recommendations.) The next step is to communicate these findings to the experimentalists and establish an open channel for communication. Communication is a very important, but unfortunately, not very well emphasized aspect of research.

18.3 Ambient Monitoring using TDLAS

This section outlines the key design issues for improving the TDLAS instrument for a future measurement campaign involving simultaneous measurements of multiple species.

Laser Improvement. One of the major factors that limits the use of TDLAS as a continuous monitoring instrument is the radiation frequency stability of the diode lasers. New generations of lasers are being developed at JPL and Bell Labs with higher power, better stability and operation characteristics (Zahniser, 1996; Miller, 1997). With higher laser power, longer absorption path length can be used. In close cell measurements, both the detection limit and the sensitivity can be raised by increasing the absorption path. In open path measurements, increased power translates to increased range and sensitivity of the laser measurement.

Laser Scan Control. Current TDLAS operations require good temperature control. A plastic shell can be used to cover the instrument while it is in operation to insulate it from ambient temperature fluctuations. However, the optical pathways of the three-colored instrument need to be reconfigured. Another improvement can be made by tuning the peak locking control parameters according to the specific diode laser's emissions characteristics.

Other Species. In addition to ozone, NO, and NO_2, TDLAS can also be applied to measure other species of interest in urban air quality, such as carbon monoxide, formaldehyde, and benzene. Theoretically, more than three species can be monitored simultaneously within the same monitoring cell (Zahniser, 1996). There are several methods to achieve this: absorption features can be chosen so that two species may be measured by one laser, a second two-colored instrument may be used so that both absorption paths are time-shared, or additional optical pathways may be added within the same absorption cell. The simultaneous monitoring of one or more carbon species with ozone and NO_x may result in useful information for analyzing the history of the air sample.

Experimental Design Issues. Several issues need to be addressed in any future experiment to monitor the fluctuations of atmospheric pollutants. Surface effects were observed in the temperature reading from a 10-foot mast on the roof of the Aerodyne Research building. It is uncertain whether the surface temperature affected the concentrations and fluxes observed, but surface effects should be avoided by using a different location or a taller mast. The time lag due to sampling was rather long in current experimental set up. Although no evidence was found that implied axial mixing or chemical reaction within the sampling system, a better design should minimize the distance between the sampling port and the absorption cell. Alternatively, *in-situ* absorption may be used in an open path arrangement to measure a spatially averaged concentration.

Meteorological measurements have proven extremely useful in elucidating the processes that affect the concentration fluctuations of atmospheric pollutants. In the future, mixing layer height information should be collected concurrently with wind speeds, temperature, and radiation to characterize the stability of the atmosphere, which affects and the variability of pollutants.

18.4 Further Statistical Study of Ambient Variability Available Data

18.4.1 Sources of Concentration Fluctuations

A key follow-up to Part II of this research is to determine the exact roles of meteorology, source emissions, and chemistry on the variance and the concentration distribution of the ambient concentrations of pollutants. Joint probability densities is one way to elucidate the effects of meteorology on pollutant concentrations. It was found that concurrent meteorological measurements obtained at the site do not usually explain the observed fluctuations during the day when the mixing length scale is long. The longer-time effects can be analyzed by using cross correlation and spectral functions, which relate different data streams that are separated by time lags.

Strong source effect of the nearby highway was observed in the TDL measurements. One possible use of this data may be to infer the emissions of NO_x from the highway at different times of the day. The emissions and subsequent transport and chemical reactions of NO_x can be modeled using a turbulent reactive plume model (Georgopoulos and Seinfeld, 1986a,b) and the available meteorological data. Using the inverse modeling technique as described by Adamkiewicz *et al.* (1997), the source strength may be deduced from the ambient measurements. Such a source estimate will supplement current tunnel studies and emissions models by providing data on auto and truck emissions in real-world driving conditions.

18.4.2 Implication of Temporal Fluctuations Towards Spatial Variability

The relationships between temporal fluctuations and mixing conditions may provide a starting point for an estimate of spatial variability based on temporal fluctuations of a point measurement. By Taylor's hypothesis, the fluctuations in time can be translated into spatial variability assuming the parcels stay intact in the time scale it takes to reach the monitor. It is therefore important to determine the limits of applicability of the Taylor hypothesis for this kind of data.

19. Conclusions

Urban air pollution research is a multidisciplinary field, with strong experimental and modeling components. As the field matures and topics become more well-defined, there seemed to be a diverging trend between experimentalists conducting large scale field campaigns and running difficult laboratory experiments on the one side, and modelers writing ever more sophisticated models on the other side. Effective communication between the two groups is a difficult thing indeed.

During the past five years, the trend appears to be changing. Modelers are now involved in field campaigns right from the planning stage (e.g., in NARSTO-Northeast, Roberts *et al.*, 1995), and experimentalists treat modeling as an important component of data collection, quality assurance, and analysis (F.C. Fehsenfeld, presentation at EPA Center of Airborne Organics symposium, 1996 entitled "State of the Art Measurements - How to make them?" (EPA CAO, 1996)). The two groups are enjoying much more fruitful collaborations than ever. Ample opportunities still exist for modelers and experimentalists to work together in the field of air quality modeling to improve the understanding of the processes that produce photochemical smog.

Model validation is one activity which requires close collaboration between modelers and experimentalists. Traditionally, the flow of information is unidirectional: from field and laboratory measurements to models. Modelers simply wrote models based on the physical understanding, obtain whatever experimental data was available to run the models, and compare the results to field measurements. Statements such as "the model predicts the observed data to within 30%" were made without qualifying the acceptability of such an agreement. This research was conducted based on the premise that there are better ways to validate models that are used to formulate expensive compliance strategies than is currently practiced.

19.1 Value of Uncertainty Analyses

Part I of this thesis illustrated how a systematic uncertainty analysis can be used to put "error bounds" on model predictions based on parametric uncertainties of input variables and independently-measured parameters, such as reaction rate constants and product coefficients. The analysis showed the limits of predictability of current urban atmospheric mechanisms to be about 40 percent for ozone. Because of the significant uncertainties in the predictions due to the use of uncertain parameters, key mechanisms are not distinguishable from one another. The mechanism predictions are similar enough that any measured concentrations can probably be reproducible by all mechanisms using different values of the uncertain parameters within their range of uncertainties.

Uncertainties in the chemical descriptions are only one subset of the uncertain parameters in three dimensional air quality models. The effects of chemical uncertainties are smaller in such a model, with an error bound of roughly 25%. The uncertainty in the predictions is small compared to the observed discrepancies between data and model predictions at most monitoring stations This suggests that uncertainties in factors other than chemistry

dominate the uncertain predictions of models, assuming the measurements are reasonably accurate. As pointed out in the previous chapter, a comprehensive analysis of the air quality models requires the quantification of uncertainties in the emissions inventories and meteorological fields used to run the air quality models. A comprehensive uncertainty analysis should provide a basis for formulating sound control strategies for controlling urban air pollution.

The utility of uncertainty analyses reach beyond putting "error bars" on the model predictions. An important component of an uncertainty analysis is the analysis of variance, through which parameters whose uncertainties contribute to the uncertainties of the outcome of-interest are identified. This type of analysis can be used to guide experiments, thus providing a framework for modelers to feed information back to the experimentalists. By focusing of the variance-contributing parameters, modelers and experimentalists can collaborate to identify the areas that hinder the understanding of the atmospheric system and make measurements to improve the confidence of model predictions. The two-way flow of information is illustrated in Figure 19-1.

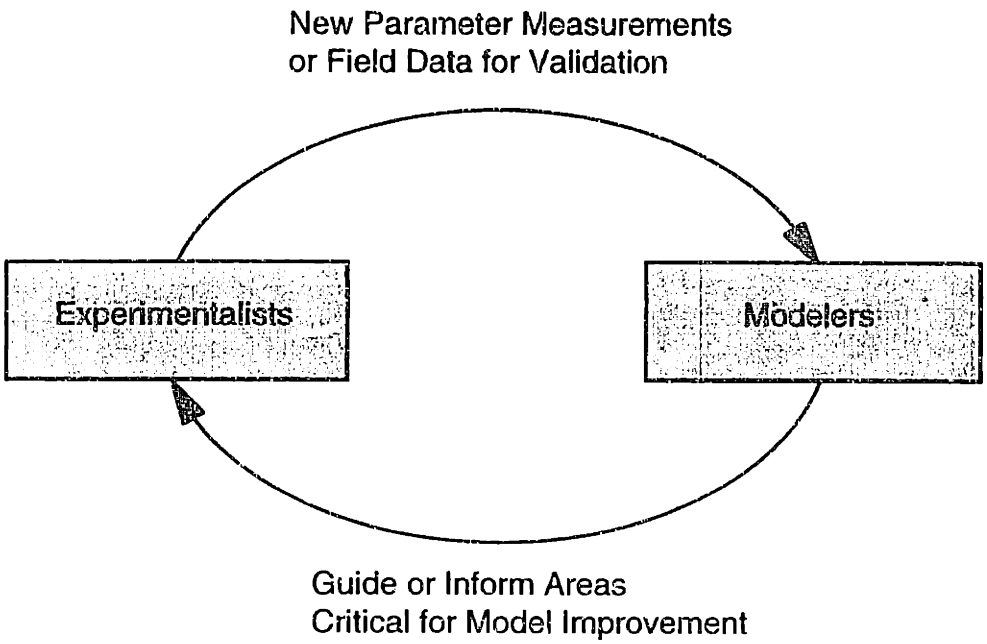


Figure 19-1. A Two-way Information Exchange between Experimentalists and Modelers.

The idea of merging models and measurements has been well-received by the research community. At a research meeting in 1996, some results of the parametric variance contributions to the uncertain predictions of the SAPRC mechanism (see Chapter 8) were presented. After the presentation, a member of the Harvard chemical kinetics group enthusiastically described their new results showing that the $HO+NO_2$ combination rate, one of the variance contributing inputs identified in the SAPRC mechanism, was over-estimated in most review panels' recommendations (Donahue *et al.*, 1997). Although it was subsequently found that the uncertainty of this reaction rate constant does not affect the uncertain predictions of three dimensional models as much as it affects the mechanism predictions, reducing the nitric acid formation rate by 20% results in a 5-10% increase in ozone predictions by AIRSHED within the Los Angeles domain. The over-estimation of this reaction rate may well be a part of the reason for the consistent underprediction of the

SCAQS data by AIRSHED. This kind of communication is exactly what an uncertainty analysis promotes, and what the research community needs.

Another important use of an uncertainty analysis is to inform research priorities, for both modeling and experimental studies. For example, this work showed that given the current uncertainties in the reaction rate constants and product parameters, the predictions of structurally dissimilar mechanisms cannot be distinguished. Therefore, the reduction of parametric uncertainties is a more urgent research goal at present than developing new chemical mechanisms that will be subjected to the same parametric uncertainties, and will likely produce results that are indistinguishable from current mechanisms. The modeling community has definitely become more specialized in recent years. Many models have been developed for many purposes, and modelers feel strongly about updating them when new information becomes available. However, different modelers may update different aspects of the same model, or unknowingly implement the same updates. The lack of a central channel where models and codes can be shared definitely contributed to time and resources being spent in a wasteful manner. The modular approach used in EPA's next generation air quality model, Models-3 (Byun, 1994), may provide a platform for which codes and subroutines can be shared easily. A more generous attitude towards sharing research results will also help conserve research resources within the modeling community .

In the field of atmospheric science, many discoveries and new understanding of the processes are inferred from measurement data. The addition of new details into models, or the incorporation of physical representation to replace parametric treatments, such as in the cases of isoprene and aromatics, should be encouraged. Parametric uncertainty analysis is a useful tool to identify model improvement opportunities and to aid resource allocation when the parameters can be measured and their uncertainties quantified. A good model can also be used in the inverse sense to provide estimates of the uncertainties and errors of the inputs.

19.2 Measuring Atmospheric Variability

The motivation for measuring the fluctuations of pollutants in the atmosphere was to quantify the variability associated with atmospheric point measurements, so as to provide a range of values (beyond the measurement error) against which a model prediction of a spatial average may be compared. This work achieved more than measuring fluctuations.

First and foremost, the construction and deployment of the tunable diode laser absorption spectroscopy (TDLAS) instrument was a proof of concept that simultaneous measurements can be obtained for multiple species in the same sample volume with high time resolution. The correlations and anticorrelations of ozone, nitric oxide, and nitrogen dioxide have never been observed before at the one-second time scale. Different correlation behaviors during the day and in the evening indicate that different processes control the concentrations within an air parcel within the time-scale of a few seconds.

The ability of TDLAS to measure NO_2 independently and unequivocally was also demonstrated. Traditional measurements rely on the conversion of NO_2 to NO, which suffer from artifacts because other nitrogen species are inevitably converted to NO. In light of the key role NO_2 plays in the formation and dissociation of nitrogen sink compounds (e.g., nitric acid, PAN), and as the precursor to ozone, accurate measurements of NO_2 are critical to the understanding of smog formation.

The measurement sensitivities were sufficient for measuring the ambient concentrations and fluctuations of these trace species. As suspected, the atmospheric variability of pollutants is a strong function of the species, the proximity to local sources, and the meteorological conditions. For example, the variability of ozone is highest when the meteorological conditions are not conducive to mixing. The nitric oxide time series, however, contain spikes of emissions from local traffic sources, and its variance changes with the concentration observed. Contrary to popular belief, atmospheric pollutant concentrations can not always be described by a Normal distribution. A variety of shapes are observed, including bimodal and multimodal distributions. The reasons for different distribution shapes were not investigated in in this work.

Although the TDL data were not used for model validation, the observed ambient fluctuations have implications towards the model validation process because they provide a measure of concentration variability around the site. This information may be useful for qualifying the agreement and disagreement between models and data. If atmospheric mixing is strong and the observed pollutant concentrations display low fluctuations of only a few percent, then a 30% agreement may be indicative that the error in the prediction is quite large. At urban or suburban monitoring stations, these kinds of variability are intrinsic to measurements. Since most monitoring instruments are capable of data rates faster than one hour, a simple report containing the standard deviation of the time series fluctuations with the observed mean value (and instrumental error) may provide valuable information needed to validate models in a more statistically robust manner.

TDLAS, and other emerging technologies, have promising applications in a large range of close-cell, long-path, and mobile measurements that will complement existing monitors in field campaigns. This study revealed the correlation between chemical concentrations in the individual air samples. Long-path measurements will provide grid-cell average concentrations to allow much more relevant model evaluation. Mobile monitors are useful in estimating spatial variability or in aircraft measurements. These applications should be exploited to aid developments in the field.

19.3 Application of Concepts

Uncertainties are endemic in modeling and data collection activities. It is important, therefore, to make use of the uncertainties and learn from them. This study proposed that, by quantifying the uncertainties associated with model predictions and measurements, the true state of the knowledge, as compiled in models, can be evaluated much more effectively. Decisions made from these models are better informed. Atmospheric chemistry was used as a case study. Part I of this work shows how to apply uncertainty analysis to understand the driving force behind uncertain predictions. Part II demonstrates the quantification of the variability in observations. There are many systems in which observation plays a significant role in the fundamental understanding of the processes, and in which models are written to compile the knowledge and sometimes to make decisions regarding the control of the system. Examples include risk assessment and management, global climate change, and many process design problems in engineering. The approach to incorporate uncertainty in informing decisions has wide applicability.

20. Appendix

20.1 Appendix A. A → B → C Files

A.1 Chemap Files

Chemap is a compiler designed to take as input a mechanism description and generate Fortran codes that describe the ordinary differential equations (ODE's) that represent the mechanism mathematically (Wyckoff, 1995). The input format for reactions and associated rate constants is simply

a A + b B --> c C; k = k1

where a, b, and c are the stoichiometric coefficients of species A, B, and C respectively. The rate constant k1 can either be a numerical value or a mathematical expression:

k1 = 100, or
k1 = 100 exp (-3.00/R/T).

In this case, constants such as temperature or the gas constant can also be used in the expression. Finally, initial conditions can be defined such as:

IC A = 100.

ODE's are generated for each species based on its reactions, the rate constants and stoichiometric coefficients. The ODE's are written in FORTRAN so that the subroutine is compatible with available numerical solvers. The following files contain the mechanism description (A1.1), and parser-generated FORTRAN codes that describe the ordinary equations for the chemical mechanism (A1.2). These codes can be used as subroutines for ODE solvers such as LSODE (Hindmarsh, 1983).

Reference:

A.C. Hindmarsh, "ODEPACK, A Systematized Collection of ODE Solvers", in Stepleman R.S. *et al.* Eds. Scientific Computing (IMACS Transaction in Scientific Computation v. 1) North Holland, Amsterdam, 1983 (LSODE pp. 55 -- 64)

P.S. Wyckoff, "Numerical Solution of Differential Equations Through Empirical Eigenfunction Expansions," Ph. D. Thesis, MIT, 1995

A.1.1 Chemap Input File

```
# (# sign indicate a comment line which is ignored by the compiler)
#
#################################################################################
#                                                                               #
# This is the Chemap input file for the mechanism:                              #
#                                                                               #
#    k1     k2                                                                  #
# A --> B --> C                                                                 #
#                                                                               #
# k1 and k2 are uncertain rate constants.                                       #
#################################################################################


# The format of the reaction description is A ----> B, where
# A is the reactant, B is the product. Chemap will obtain the list of active
# species from the mechanism description.  Any stoichiometric coefficients
# can also be incorporated in the reaction listing.  A semicolon is used to
# separate different statements, and k is used to describe the reaction rate
# constant.

A  ----> B     ;       k = k1; k1 = uk1
B  ----> C     ;       k = k2; k2 = uk2


# Uncertain parameters are declared here.  Uncertain parameters are
# to be read from either a file or the standard input.

Uncertain uk1 uk2


# Initial conditions are declared using the keyword IC followed by
# the species name, an equal sign, and the numerical number.

IC A = 100
IC B = 0
IC C = 0


# End of input file
```

A.1.2 Chemap-generated Subroutines

Chemap generates codes that can be used with an ODE solver such as LSODE. A subroutine contains the differential equations. Two header files are needed to declare the variables and the initial conditions

A.1.2.1 Derivs.f

```
c
c Automatically generated derivs (FORTRAN) routine.
c
c Specification of Rate Constants
c List of Ordinary differential equations all species
c
c
      subroutine derivs(neq, t, y, ydot, nuk, uk)

      implicit none
      integer neq, nuk
      real t, y, ydot, uk
      dimension y(neq), ydot(neq), uk(nuk)
```

```
      include 'derivs.h'

      k2 = uk(2)
      k1 = uk(1)


c     *** C ***
      ydot(1) = k2*y(2)

c     *** B ***
      ydot(2) = -(k2*y(2))+k1*y(3)

c     *** A ***
      ydot(3) = -(k1*y(3))


      return
      end

c End of derivs.f file
```

A.1.2.2 Derivs.h

```
c  Header file to declare rate constants

      real k2
      real k1

c  End derivs.h file
```

A.1.2.3 Derivs_names.h

```
c  Header file to declare names of species
c  and their initial conditions


static char *lsodey_names[ 3] = {
    "C",
    "B",
    "A",
};
double initlsodey[ 3] = {
     0.0000000000000000e+00,
     0.0000000000000000e+00,
     1.0000000000000000e+02,
};

c  End derivs_names.h file
```

A.2 Run-time Configuration File

```
#
# Configuration file for A --> B --> C mechanism.
# Controls the input/output of the ODE solver
#

#
# Starting time.
#
time 0.0

#
# Times at which to print concentrations.
#
print 0.1 0.25 0.5 0.75 1 1.25 1.5 2 3 4 5 6 7 8 9 10
```

```
#
# Species to report, in this order.  Acceptable names of species are listed
# in the file "derivs_names.h", which also lists the initial conditions to
# be used, as specified by menmap through the mechanism file.
#
Report A B C

# If the initial condition is uncertain, the next line should be de-commented
# uncertain ICs A

#
# List of uncertain coefficients.  First line specifies how many there
# will be on each line (one line's worth to be used per run).  Successive
# lines each are a set of coefficients.  Uncertain initial conditions are
# specified after uncertain rate coefficients.  (For example, if the initial
# condition of A is uncertain, each line will have three values.)  Model will
# be run to completion (maximum print time) once per line.
#

uk 2

0.5 2.0
# Nominal Value: k1 = 0.5, k2 = 2.0


# End of configuration file
```

A.3 Base Case Results

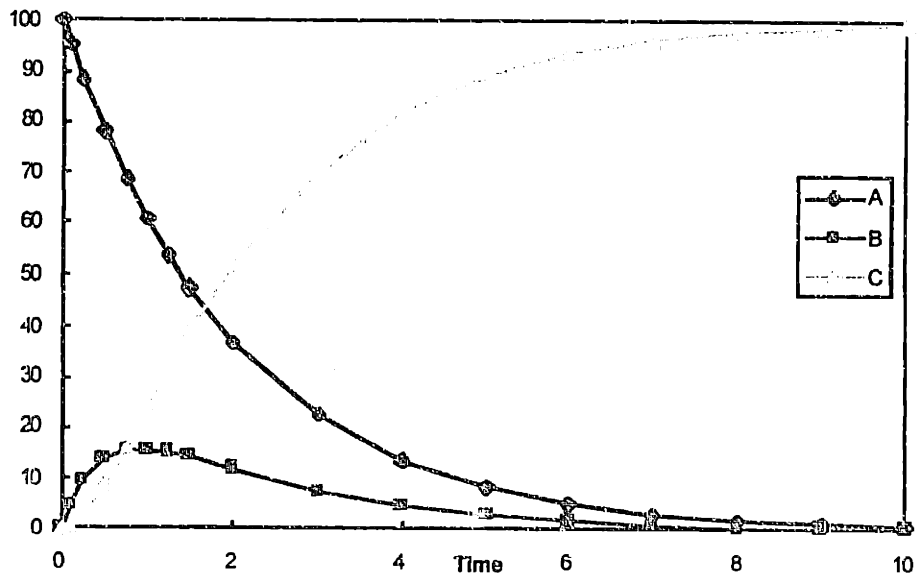


Figure 20-1. Base Case Results A → B → C Mechanism.
(k_1 = 0.5, k_2 = 2.0)

A.4 Demmucom Input

```
#
# Demmucom Input file for A --> B --> C
# with two uncertain rate constants.
#


# Declare uncertain parameter distributions

Uncertain inputs
    uk1 = normal(0.5, 0.1),
    uk2 = normal(2.0, 0.5);


# Declare Uncertain model output.

Uncertain outputs

    A01, B01, C01, A025, B025, C025, A05, B05, C05, A075, B075, C075, A1, B1, C1, A125,
B125, C125, A15, B15, C15, A2, B2, C2, A3, B3, C3, A4, B4, C4, A5, B5, C5, A6, B6, C6,
A7, B7, C7, A8, B8, C8, A9, B9, C9, A10, B10, C10;


# Specify DEMM (collocation) controls

Do Collocation With
   model: ABCmlsode,
   approximation: first,
   sampling file : sampling.pnt,
   number of points : 5000,
   options: looping, no header;


# end Demmucom input file
```

20.2 Appendix B: SAPRC Mechanism

The Statewide Air Pollution Research Center (SAPRC) atmospheric mechanism (Carter, 1990) was designed for ozone modeling and control strategy design. The precursors to the SAPRC mechanism were the "ALW" mechanism of Atkinson *et al.*(1982) and the "ADOM" mechanism (Lurmann *et al.*,1986). The current version of SAPRC has been verified against the results of 550 smog chamber experiments in four different chambers (Carter and Lurmann, 1991), and was found to predict maximum ozone yields and NO oxidation rates in most chamber experiments to within 30%. The main feature of SAPRC mechanism is that it includes the assignment of over 100 individual organic compounds using generalized species and reactions derived based on those compounds being represented. Most other photochemical mechanisms represent emitted species using a limited number of lumped species with fixed mechanistic parameters. Key characteristics of the SAPRC Mechanism are summarized in Table 20-1.

Number of Species	58
Number of Reactions	153
Inorganic Reactions	42
Inorganic Photolysis Reactions	7
Organic Reactions	106
Organic Photolysis Reactions	13
Primary Organic Species	User chosen. This study: 3 alkane and aromatic (AAR) species, ethene, 2 other olefins
Primary Organic Reactions	15
Lumping	Lumped reaction parameters based on individual species lumped into the group

Table 20-1. Key Characteristics of the SAPRC mechanism.

Web Address:
http://cert.ucr.edu/~carter/

References:

Atkinson, R., Lloyd, A.C., and Winges, L., "An Updated Chemical Mechanism for Hydrocarbon/NO_x/SO_2 Photo-oxidations Suitable for Inclusion in Atmospheric Simulation Models," *Atmos Environ.*, **16**:1314 (1982)

Carter, W.P.L., "A Detailed Mechanism for the Gas-Phase Atmospheric Reactions of Organic Compounds," *Atmos. Environ.*, **24A**, 481-518 (1990)

Carter, W.P.L. and F. W. Lurmann, "Evaluation of a Detailed Gas Phase Atmospheric Reaction Mechanism Using Environmental Chamber Data," *Atmos. Environ.*, **25A**, 2771-2806 (1991)

Lurmann F.W., Lloyd A.C., and Atkinson, R., "A Chemical Mechanism for Use in Long-range Transport/Acid Deposition Computer Modeling," *J. Geophy. Res.*, **91**, 10905-10936 (1986)

B.1 SAPRC Mechanism Species

Model Species	Description
Active Inorganic Species	
O3	Ozone
NO	Nitric Oxide
NO2	Nitrogen Dioxide
NO3	Nitrate Radical
N2O5	DiNitrogen Pentoxide
HNO3	Nitric Acid
HONO	Nitrous Acid
HNO4	Peroxynitric Acid
(SO2	Sulfur Dioxide)
Inorganic Radical Species	
HO2	Peroxyl Radical
HO	Hydroxyl Radical
O	Ground State Oxygen Radical $O(^3P)$
O1D2	Excited State Oxygen Radical $O(^1D)$
Inorganic Product Only Species	
CO	Carbon Monoxide
CO2	Carbon Dioxide
(H2SO4	Sulfuric Acid)
H2	Hydrogen
LOSTC	Lost Carbon. (accounts for carbon balance)
LOSTN	Lost Nitrogen (accounts for nitrogen balance, represents C_1– C_3 organic nitrates and dinitrophenols, whose reactions are ignored)
Active Organic Species	
HCHO	Formaldehyde
CCHO	Acetaldehyde
PAN	Peroxy Acetyl Nitrate
RCHO	Propionaldehyde and lumped higher aldehydes
PPN	Peroxy Propionyl Nitrate and higher PAN - analogues
ACET	Acetone
MEK	Methyl Ethyl Ketone and lumped higher ketones
RNO3	Lumped organic trates
GLY	Glyoxal
GPAN	PAN-analogue fromed from glyoxal
MGLY	Methyl Glyoxal
PHEN	Phenol
CRES	Cresols and Other Alkyl Phenols
BALD	Benzaldehyde and other Aromatic Aldehydes
PBZN	Peroxyl Benzoyl Nitrate
NPHE	Nitrophenols and other Aromatic Nitro compounds
AFG1	Unknown Aromatic Fragmentation Product #1 (formed from benzene, tetralin, and naphthalenes)
AFG2	Unknown Aromatic Fragmentation Product #2 (formed from aromatics containing alkyl groups)
ROOH	Chemical Operator used to represent reactions at hydroperoxy groups)
ETHE	Ethene

AARn	n-th lumped group used to represent emitted alkanes and/or aromatics
OLEn	n-th lumped group used to represent emitted higher alkenes
Organic Radical Species	
RO2	Total Alkyl Peroxy Radicals
RCO3	Total Peroxyl Acyl Radicals
HOCOO	Intermediate Formed in the HCHO + HO2 Reaction
CCOO2	Peroxy Acetyl Radicals
C2COO2	Higher Peroxyacyl Radicals
BZCOO2	Peroxy Benzoyl Radicals
HCOCOO2	Peroxyacyl Radical fromed from Glyoxyal
BZO	Phenoxy Radicals
BZNO2O	Phenoxyl-type Radicals containing Nitro Groups
O3OLSB	Chemical operator used to account for the oxidation of SO2 by ozone-alkene reaction intermediates
RO2R	Chemical operator used to represent NO-to-NO2 conversion with generation of a HO2 radical
RO2N	Chemical operator used to represent NO consumption and alkyl nitrate formation
RO2NP	Chemical operator used to represent NO consumption and nitrophenol formation
RO2XN	Chemical operator used to represent NO sink reactions
R2O2	Chemical operator used to represent extra NO-to-NO2 conversions
Constant species	
O2	Oxygen
M	Third Body (Air)
HV	Light Factor (1.0 = normal light intensity)
H2O	Water

Table 20-2. List of Species in the SAPRC mechanism.

B.2 SAPRC Chemap Input file (Reactions and Rates)

```
#
# saprc.p.map 6/10/95
# Carter's SAPRC Mechanism with 19 Uncertain Photolysis Rates
# All concentrations are described in molec/cm3 units
#

# No steady state Species
# steady state species in Carter's original implementation
# O + O1D2 + RO2R + HO + RO2N + R2O2 + RO2XN + HOCOO + CCOO2 +
# C2COO2 + HCOCOO2 + BZCOO2 + BZO + BZNO2O + RO2NP
# e.g., O1D2=k16*O3/(k17*H2O+k18*M)

#
# Notes:
#     Constants species can be specified in the Chemap environment using statements
such as:
#     HV = 1
#     No ODE will be generated for these species.  The species name will be replaced by
its
#     numerical value in derivs.f, the ODE file.
#


# NOx-air reactions

NO2 + HV --> NO + O                              ;k=k1; k1=uk1
#photolysis; k  (zenith angle 0) = 0.00898

O + O2 + M --> O3 + M                            ;k=k2
k2=6.000000e-34*(T/300)^(-2.300000e+00)*exp(-(0.000000e+00/R/T))

O + NO2 --> NO + O2                              ;k=k3
k3=6.500000e-12*(T/300)^(0.000000e+00)*exp(-(-2.380000e-01/R/T))

O + NO2 --> NO3 + M                              ;k=k4
k4=((9.000000e-32*(T/300)^(-2.000000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((9.000000e-
32*(T/300)^(-2.000000e+00)*exp(-(0.000000e+00/R/T)))*P/(2.200000e-11))))*6.000000e-
01^(1/(1+(log10((9.000000e-32*(T/300)^(-2.000000e+00)*exp(-
(0.000000e+00/R/T)))*P/(2.200000e-11))/1.000000e+00)^2))
#Troe formula
O3 + NO --> NO2 + O2                             ;k=k5
k5=2.000000e-12*(T/300)^(0.000000e+00)*exp(-(2.782000e+00/R/T))

O3 + NO2 --> O2 + NO3                            ;k=k6
k6=1.400000e-13*(T/300)^(0.000000e+00)*exp(-(4.968000e+00/R/T))

NO + NO3 --> 2 NO2                               ;k=k7
k7=1.700000e-11*(T/300)^(0.000000e+00)*exp(-(-2.980000e-01/R/T))

NO + NO + O2 --> 2 NO2                           ;k=k8
k8=3.300000e-39*(T/300)^(0.000000e+00)*exp(-(-1.050000e+00/R/T))

NO2 + NO3 --> N2O5                               ;k=k9
k9=((2.200000e-30*(T/300)^(-4.300000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((2.200000e-
30*(T/300)^(-4.300000e+00)*exp(-(0.000000e+00/R/T)))*P/(1.500000e-12*(T/300)^(-
5.000000e-01)*exp(-(0.000000e+00/R/T))))))*6.000000e-01^(1/(1+(log10((2.200000e-
30*(T/300)^(-4.300000e+00)*exp(-(0.000000e+00/R/T)))*P/(1.500000e-12*(T/300)^(-
5.000000e-01)*exp(-(0.000000e+00/R/T))))/1.000000e+00)^2));

N2O5  --> NO2 + NO3                              ;k=k10
k10=k9*9.090000e+26*(T/300)^(0.000000e+00)*exp(-(2.226000e+01/R/T));
# Equilibrium equation; RCON8 = coefficient multiplied to previous reaction

N2O5 + H2O --> 2 HNO3                            ;k=k11;
k11=1.000000e-21*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))
# 298K only

NO2 + NO3 --> NO + NO2 + O2                      ;k=k12
```

```
k12=2.500000e-14*(T/300)^(0.000000e+00)*exp(-(2.440000e+00/R/T))

NO3 + HV --> NO + O2                                    ;k=k13; k13=uk2
#photolysis; k   = 0.0187

NO3 + HV --> NO2 + O                                    ;k=k14; k14=uk3
#photolysis; k   = 0.169

O3 + HV --> O + O2                                      ;k=k15; k15=uk4
#photolysis; k   = 0.000499

O3 + HV --> O1D2 + O2                                   ;k=k16; k16=uk5
#photolysis; k   = 3.78e-05

O1D2 + H2O --> 2 HO                                     ;k=k17
k17=2.200000e-10*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

O1D2 + M --> O + M                                      ;k=k18
k18=1.919000e-11*(T/300)^(0.000000e+00)*exp(-(-2.510000e-01/R/T))

HO + NO --> HONO                                        ;k=k19
k19=((7.000000e-31*(T/300)^(-2.600000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((7.000000e-
31*(T/300)^(-2.600000e+00)*exp(-(0.000000e+00/R/T)))*P/(1.500000e-11*(T/300)^(-
5.000000e-01)*exp(-(0.000000e+00/R/T))))))*6.000000e-01^(1/(1+(log10((7.000000e-
31*(T/300)^(-2.600000e+00)*exp(-(0.000000e+00/R/T)))*P/(1.500000e-11*(T/300)^(-
5.000000e-01)*exp(-(0.000000e+00/R/T))))/1.000000e+00)^2))
#Troe formula

HONO + HV --> HO + NO                                   ;k=k20; k20=uk6
#photolysis; k   = 0.0163

HO + NO2 --> HNO3                                       ;k=k21
k21=((2.600000e-30*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((2.600000e-
30*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T)))*P/(2.400000e-11*(T/300)^(-
1.300000e+00)*exp(-(0.000000e+00/R/T))))))*6.000000e-01^(1/(1+(log10((2.600000e-
30*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T)))*P/(2.400000e-11*(T/300)^(-
1.300000e+00)*exp(-(0.000000e+00/R/T))))/1.000000e+00)^2))
#Troe formula

HO + HNO3 --> H2O + NO3                                 ;k=k22
k22=6.450000e-15*(T/300)^(0.000000e+00)*exp(-(-1.652000e+00/R/T))
# 1 ATM ONLY.

# HNO3 + HV --> HO + NO2                                ;(IGNORED)

HO + CO --> HO2 + CO2                                   ;k=k23
k23=2.400000e-13*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))
#1 ATM ONLY

HO + O3 --> HO2 + O2                                    ;k=k24
k24=1.600000e-12*(T/300)^(0.000000e+00)*exp(-(1.870000e+00/R/T))

HO2 + NO --> HO + NO2                                   ;k=k25
k25=3.700000e-12*(T/300)^(0.000000e+00)*exp(-(-4.800000e-01/R/T))

HO2 + NO2 --> HNO4                                      ;k=k26
k26=((1.800000e-31*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((1.800000e-
31*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T)))*P/(4.700000e-12*(T/300)^(-
1.400000e+00)*exp(-(0.000000e+00/R/T))))))*6.000000e-01^(1/(1+(log10((1.800000e-
31*(T/300)^(-3.200000e+00)*exp(-(0.000000e+00/R/T)))*P/(4.700000e-12*(T/300)^(-
1.400000e+00)*exp(-(0.000000e+00/R/T))))/1.000000e+00)^2))
#Troe formula

HNO4  --> HO2 + NO2                                     ;k=k27
k27=k26*4.760000e+26*(T/300)^(0.000000e+00)*exp(-(2.166000e+01/R/T))
#Equilibrium equation

#HNO4 + HV -->                                          ;(IGNORED)

HNO4 + HO --> H2O + NO2 + O2                            ;k=k28
k28=1.300000e-12*(T/300)^(0.000000e+00)*exp(-(-7.550000e-01/R/T))
```

```
HO2 + O3 --> HO + 2 O2                                  ;k=k29
k29=1.100000e-14*(T/300)^(0.000000e+00)*exp(-(9.940000e-01/R/T))

HO2 + HO2 --> HO2H + O2                                 ;k=k30
k30=2.200000e-13*(T/300)^(0.000000e+00)*exp(-(-1.230000e+00/R/T))

HO2 + HO2 + M --> HO2H + O2                             ;k=k31
k31=1.900000e-33*(T/300)^(0.000000e+00)*exp(-(-1.950000e+00/R/T))

HO2 + HO2 + H2O --> HO2H + O2 + H2O                     ;k=k32
k32=3.100000e-34*(T/300)^(0.000000e+00)*exp(-(-5.600000e+00/R/T))

HO2 + HO2 + M + H2O --> HO2H + O2 +H2O+ M               ;k=k33
k33=2.683000e-54*(T/300)^(0.000000e+00)*exp(-(-6.320000e+00/R/T))

#Original Form of Rxn 33, concentration of third body incoporated into rate expression
#HO2 + HO2 + H2O --> HO2H + O2 + H2O
#k=6.600000e-35*(T/300)^(0.000000e+00)*exp(-(-6.320000e+00/R/T))
#(1 ATM ONLY)

NO3 + HO2 --> HNO3 + O2                                 ;k=k34
k34=2.200000e-13*(T/300)^(0.000000e+00)*exp(-(-1.230000e+00/R/T))

NO3 + HO2 + M --> HNO3 + O2                             ;k=k35
k35=1.900000e-33*(T/300)^(0.000000e+00)*exp(-(-1.950000e+00/R/T))

NO3 + HO2 + H2O --> HNO3 + O2 + H2O                     ;k=k36
k36=3.100000e-34*(T/300)^(0.000000e+00)*exp(-(-5.600000e+00/R/T))

NO3 + HO2 + H2O + M--> HNO3 + O2 + H2O + M  k=k37
k37=2.683000e-54*(T/300)^(0.000000e+00)*exp(-(-6.320000e+00/R/T))

HO2H + HV --> 2 HO                                      ;k=k38;  k38=uk7
#photolysis; k   = 7.53e-06

HO2H + HO --> HO2 + H2O                                 ;k=k39
k39=3.300000e-12*(T/300)^(0.000000e+00)*exp(-(3.970000e-01/R/T))

HO + HO2 --> H2O + O2                                   ;k=k40
k40=4.600000e-11*(T/300)^(0.000000e+00)*exp(-(-4.570000e-01/R/T))


# RO2 reactions

RO2 + NO --> NO                                         ;k=k41
k41=4.200000e-12*(T/300)^(0.000000e+00)*exp(-(-3.600000e-01/R/T))

RCO3 + NO --> NO                                        ;k=k42
k42=5.100000e-12*(T/300)^(0.000000e+00)*exp(-(-3.970000e-01/R/T))

#RO2 + NO2 --> NO2                                      ;(IGNORED)

RCO3 + NO2 --> NO2                                      ;k=k43
k43=((1.950000e-28*(T/300)^(-4.000000e+00)*exp(-(0.000000e+00/R/T))*P)/(1+((1.950000e-
28*(T/300)^(-4.000000e+00)*exp(-(0.000000e+00/R/T)))*P/(8.400000e-12))))*2.700000e-
01^(1/(1+(log10((1.950000e-28*(T/300)^(-4.000000e+00)*exp(-
(0.000000e+00/R/T)))*P/(8.400000e-12))/1.000000e+00)^2))
#Troe formula

RO2 + HO2 --> HO2                                       ;k=k44
k44=3.400000e-13*(T/300)^(0.000000e+00)*exp(-(-1.590000e+00/R/T))

RCO3 + HO2 --> HO2                                      ;k=k45
k45=3.400000e-13*(T/300)^(0.000000e+00)*exp(-(-1.590000e+00/R/T))

RO2 + RO2 --> M                                         ;k=k46
k46=1.000000e-15*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

RO2 + RCO3 --> M                                        ;k=k47
k47=1.860000e-12*(T/300)^(0.000000e+00)*exp(-(-1.053000e+00/R/T))

RCO3 + RCO3 --> M                                       ;k=k48
```

```
k48=2.800000e-12*(T/300)^(0.000000e+00)*exp(-(-1.053000e+00/R/T))

RO2R + NO --> NO2 + HO2                          ;k=k49; k49 = k41

RO2R + HO2 --> ROOH                              ;k=k50; k50 = k44

RO2R + RO2 --> RO2 + 0.5 HO2                     ;k=k51; k51 = k46

RO2R + RCO3 --> RCO3 + 0.5 HO2                   ;k=k52; k52 = k47

RO2N + NO --> RNO3                               ;k=k53; k53 = k41

RO2N + HO2 --> ROOH + MEK + 1.5 LOSTC                    ;k=k55; k55 = k44

RO2N + RO2 --> RO2 + 0.5 HO2 + MEK + 1.5 LOSTC
k=k55; k55 = k46

RO2N + RCO3 --> RCO3 + 0.5 HO2 + MEK + 1.5 LOSTC
k=k56; k56 = k47

R2O2 + NO --> NO2                                ;k=k57; k57 = k41

R2O2 + HO2 --> M                                 ;k=k58; k58 = k44

R2O2 + RO2 --> RO2                               ;k=k59; k59 = k46

R2O2 + RCO3 --> RCO3                             ;k=k60; k60 = k47

RO2XN + NO --> LOSTN                             ;k=k61; k61 = k41

RO2XN + HO2 --> ROOH                             ;k=k62; k62 = k44

RO2XN + RO2 --> RO2 + 0.5 HO2                            ;k=k63; k63 = k46

RO2XN + RCO3 --> RCO3 + HO2                      ;k=k64; k64 = k47


#. ROOH Reactions

ROOH + HV --> HO2 + HO                                   ;k=k65; k65=uk8
#photolysis; k   = 7.388e-06

HO + ROOH --> HO                                 ;k=k66
k66=1.180000e-12*(T/300)^(0.000000e+00)*exp(-(-2.540000e-01/R/T))

HO + ROOH --> RO2R + RO2                         ;k=k67
k67=1.790000e-12*(T/300)^(0.000000e+00)*exp(-(-4.350000e-01/R/T))


# Formaldehyde reactions

HCHO + HV --> 2 HO2 + CO                         ;k=k68; k68=uk9
#photolysis; k   = 3.007e-05

HCHO + HV --> H2 + CO                            ;k=k69; k69=uk10
#photolysis; k   = 4.305e-05

HCHO + HO --> HO2 + CO + H2O                     ;k=k70
k70=1.125000e-12*(T/300)^(2.000000e+00)*exp(-(-1.288000e+00/R/T))

HCHO + HO2 --> HOCOO                             ;k=k71
k71=9.700000e-15*(T/300)^(0.000000e+00)*exp(-(-1.242000e+00/R/T))

HOCOO --> HO2 + HCHO                             ;k=k72
k72=2.400000e+12*(T/300)^(0.000000e+00)*exp(-(1.391000e+01/R/T))

HOCOO + NO --> LOSTC + NO2 + HO2                 ;k=k73; k73=k41

HCHO + NO3 --> HNO3 + HO2 + CO                   ;k=k74
k74=2.800000e-12*(T/300)^(0.000000e+00)*exp(-(5.000000e+00/R/T))
```

```
# Alkane products reactions

CCHO + HO --> CCO2 + H2O + RCO3                    ;k=k75
k75=5.550000e-12*(T/300)^(0.000000e+00)*exp(-(-6.180000e-01/R/T))

CCHO + HV --> CO + HO2 + HCHO + RO2R + RO2 ;k=k76; k76=uk11
#photolysis; k   = 5.94e-06

#CCHO + HV --> CH4 + CO                            ;(IGNORED)

CCHO + NO3 --> HNO3 + CCO2 + RCO3                  ;k=k77
k77=1.400000e-12*(T/300)^(0.000000e+00)*exp(-(3.696000e+00/R/T))

RCHO + HO --> C2CO2 + RCO3                         ;k=k78
k78=8.500000e-12*(T/300)^(0.000000e+00)*exp(-(-5.000000e-01/R/T))

RCHO + HV --> CCHO + RO2R + RO2 + CO + HO2 ;k=k79; k79=uk12
#photolysis; k   = 2.14e-05

NO3 + RCHO --> HNO3 + C2CO2 + RCO3                 ;k=k80
k80=1.400000e-12*(T/300)^(0.000000e+00)*exp(-(3.696000e+00/R/T))

ACET + HO --> 0.8 MGLY + 0.8 RO2R + 0.2 R2O2 + 0.2 HCHO + 0.2 CCO2 + 0.2 RCO3 + RO2
k=k81; k81=1.920000e-13*(T/300)^(2.000000e+00)*exp(-(-1.100000e-01/R/T))

ACET + HV --> CCO2 + HCHO + RO2R + RCO3 + RO2
k=k82; k82=uk13 #photolysis; k   = 8.17e-07

MEK + HO --> H2O + 0.5 CCHO + 0.5 HCHO + 0.5 CCO2 + 0.5 C2CO2 + RCO3 + 1.5 R2O2 +
1.5 RO2
k=k83; k83=2.920000e-13*(T/300)^(2.000000e+00)*exp(-(-8.230000e-01/R/T))

MEK + HV --> CCO2 + CCHO + RO2R + RCO3 + RO2
k=k84; k84=uk14 #photolysis; k   = 1.58e-06

RNO3 + HO --> NO2 + 0.155 MEK + 1.05 RCHO + 0.48 CCHO + 0.16 HCHO + 0.11 LOSTC + 1.39
R2O2 + 1.39 RO2
k=k85; k85=2.191000e-11*(T/300)^(0.000000e+00)*exp(-(1.408000e+00/R/T))

MGLY + HV --> HO2 + CO + CCO2 + RCO3               ;k=k86; k86=uk15
#photolysis; k   = 0.000148

MGLY + HO --> CO + CCO2 + RCO3                     ;k=k87
k87=1.720000e-11*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

MGLY + NO3 --> HNO3 + CO + CCO2 + RCO3     ;k=k88; k88 = k77

CCO2 + NO --> CO2 + NO2 + HCHO + RO2R + RO2        ;k=k89; k89 = k42

CCO2 + NO2 --> PAN                                 ;k=k90; k90 = k43

CCO2 + HO2 --> ROOH + CO2 + HCHO                   ;k=k91; k91 = k45

CCO2 + RO2 --> RO2 + 0.5 HO2 + CO2 + HCHO ;k=k92; k92 = k47

CCO2 + RCO3 --> RCO3 + HO2 + CO2 + HCHO    ;k=k93; k93 = k48

PAN --> CCO2 + NO2 + RCO3                          ;k=k94
k94=((6.300000e-02*exp(-(2.540600e+01/R/T))*P)/(1+((6.300000e-02*exp(-
(2.540600e+01/R/T)))*P/(2.200000e+16*exp(-(2.669800e+01/R/T)))))*2.700000e-
01^(1/(1+(log10((6.300000e-02*exp(-(2.540600e+01/R/T)))*P/(2.200000e+16*exp(-
(2.669800e+01/R/T))))/1.000000e+00)^2))
#Troe formula

C2CO2 + NO --> CCHO + RO2R + CO2 + NO2 + RO2
k=k95; k95 = k42

C2CO2 + NO2 --> PPN                                ;k=k96
k96=8.400000e-12*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

C2CO2 + HO2 --> ROOH + CCHO + CO2                  ;k=k97; k97 = k45
```

```
C2COO2 + RO2 --> RO2 + 0.5 HO2 + CCHO + CO2;k=k98; k98 = k47

C2COO2 + RCO3 --> RCO3 + HO2 + CCHO + CO2  ;k=k99; k99 = k48

PPN --> C2COO2 + NO2 + RCO3                ;k=k100
k100=1.600000e+17*(T/300)^(0.000000e+00)*exp(-(2.796600e+01/R/T))


# Aromatic products reactions

GLY + HV  --> 0.8 HO2 + 0.45 HCHO + 1.55 CO;k=k101; k101=uk16
#photolysis; k  = 4.62e-05

GLY + HV --> 0.13 HCHO + 1.87 CO           ;k=k102; k102=uk17
#photolysis; k  = 6.44e-05

GLY + HO --> 0.6 HO2 + 1.2 CO + .4 HCOCCO2 + 0.4 RCO3
k=k103; k103=1.140000e-11*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

GLY + NO3 --> HNO3 + .6 HO2 + 1.2 CO + 0.4 HCOCOO2 + 0.4 RCO3
k=k104; k104 = k77

HCOCOO2 + NO --> NO2 + CO2 + CO + HO2              ;k=k105; k105 = k42

HCOCOO2 + NO2 --> GPAN                             ;k=k106; k106 = k43

GPAN --> HCOCOO2 + NO2 + RCO3                      ;k=k107; k107 = k94

HCOCOO2 + HO2 --> ROOH + CO2 + CO          ;k=k108; k108 = k45

HCOCOO2 + RO2 --> RO2 + .5 HO2 + CO2 + CO  ;k=k109; k109 = k47

HCOCOO2 + RCO3 --> RCO3 + HO2 + CO2 + CO   ;k=k110; k110 = k48

HO + PHEN --> 0.15 RO2NP + 0.85 RO2R + 0.2 GLY + 4.7 LOSTC + RO2
k=k111; k111=2.630000e-11*(T/300)^(0.000000e+00)*exp(-(0.000000e+00/R/T))

NO3 + PHEN --> HNO3 + BZO                  ;k=k112; k112=3.6e-12

HO + CRES --> 0.15 RO2NP + 0.85 RO2R + 0.2 MGLY + 5.5 LOSTC + RO2
k=k113; k113=4.2e-11

NO3 + CRES --> HNO3 + BZO + LOSTC          ;k=k114; k114=2.1e-11

BALD + HO --> BZCOO2 + RCO3                ;k=k115; k115=1.29e-11

BALD + HV --> 7 LOSTC                      ;k=k116; k116=uk18
#photolysis; k  = 4.30e-05

BALD + NO3 --> HNO3 + BZCOO2               ;k=k117
k117=1.400000e-12*(T/300)^(0.000000e+00)*exp(-(3.747000e+00/R/T))

BZCOO2 + NO --> BZO + CO2 + NO2 + R2O2 + RO2       ;k=k118; k118 = k42

BZCOO2 + NO2 --> PBZN                      ;k=k119; k119 = 8.4e-12

BZCOO2 + HO2 --> ROOH + CO2 + PHEN         ;k=k120; k120 = k45

BZCOO2 + RO2 --> RO2 + .5 HO2 + CO2 + PHEN ;k=k121; k121 = k47

BZCOO2 + RCO3 --> RCO3 + HO2 + CO2 + PHEN  ;k=k122; k122 = k48

PBZN --> BZCOO2 + NO2 + RCO3               ;k=k123
k123 = 1.6e+15*exp(-25.90/R/T)

BZO + NO2 --> NPHE                         ;k=k124
k124 = 1.300000e-11*(T/300)^(0.000000e+00)*exp(-(-5.960000e-01/R/T))

BZO + HO2 --> PHEN                         ;k=k125; k125 = k44

BZO --> PHEN                               ;k=k126; k126=1.0e-3
```

```
NPHE + NO3 --> HNO3 + BZNO2O                ;k=k127; k127=3.6e-12

BZNO2O + NO2 --> 2 LOSTN + 6 LOSTC          ;k=k128; k128 = k124


# Dinitrophenol

BZNO2O + HO2 --> NPHE                       ;k=k129; k129 = k44

BZNO2O --> NPHE                             ;k=k130; k130 = k126

HO + AFG1 --> HCOCOO2 + RCO3                ;k=k131; k131=1.14e-11

AFG1 + HV --> HO2 + HCOCOO2 + RCO3          ;k=k132; k132 = uk19
#photolysis; k  = 5.29e-05

HO + AFG2 --> C2COO2 + RCO3                 ;k=k133; k133=1.72e-11

AFG2 + HV --> HO2 + CO + CCOO2 + RCO3              ;k=k134; k134=10*k132
# defined in Carter version and k132 = 0.1*k134

RO2NP + NO --> NPHE                         ;k=k135; k135 = k41

RO2NP + HO2 --> ROOH + 6 LOSTC              ;k=k136; k136 = k44

RO2NP + RO2 --> RO2 + 0.5 HO2 + 6 LOSTC     ;k=k137; k137 = k46

RO2NP + RCO3 --> RCO3 + HO2 + 6 LOSTC              ;k=k138; k138 = k47

# Ethene

ETHE + HO -->  0.22 CCHO +  1.56 HCHO + RO2R + RO2
k=k139; k139=1.960000e-12*(T/300)^(0.000000e+00)*exp(-(-8.700000e-01/R/T))

ETHE + O3 --> HCHO + 0.37 O3OLSB + 0.44 CO + 0.56 LOSTC + 0.12 HO2
k=k140; k140=1.200000e-14*(T/300)^(0.000000e+00)*exp(-(5.226000e+00/R/T))

ETHE + O --> HCHO + CO + HO2 + RO2R + RO2  ;k=k141
k141=1.040000e-11*(T/300)^(0.000000e+00)*exp(-(1.574000e+00/R/T))

ETHE + NO3 --> NO2 + 2.0 HCHO + R2O2 + RO2 ;k=k142
k142=5.430000e-12*(T/300)^(0.000000e+00)*exp(-(6.043000e+00/R/T))


# Lumped alkane-aromatic reactions

HO + AAR1 --> 9.141922e-01 RO2R + 3.837124e-02 RO2N + 7.463920e-03 RO2XN +
0.000000e+00 RO2NP + 3.997259e-02 HO2 + 3.016226e-01 R2O2 + 1.261650e+00 RO2 +
1.267282e-01 HCHO + 3.087707e-01 CCHO + 1.594395e-01 RCHO + 2.530871e-01 ACET +
2.323413e-01 MEK + 2.892522e-02 CO + 0.000000e+00 CO2 + 1.104737e-02 PHEN
+0.000000e+00 CRES + 0.000000e+00 BALD + 7.718203e-02 GLY + 0.000000e+00 MGLY +
2.293734e-02 AFG1 + 0.000000e+00 AFG2 + 2.033840e-01 LOSTC
k=k143;k143=2.755085e-12*exp(-(2.250031e-01/R/T))

HO + AAR2 --> 8.249258e-01 RO2R + 1.187701e-01 RO2N + 2.029170e-03 RO2XN +
0.000000e+00 RO2NP + 5.427408e-02 HO2 + 5.834542e-01 R2O2 + 1.529179e+00 RO2 +
1.609208e-03 HCHO + 1.590588e-01 CCHO + 1.996797e-01 RCHO + 1.545300e-01 ACET +
6.580250e-01 MEK + 8.875062e-03 CO + 1.801978e-04 CO2 + 0.000000e+00 PHEN +5.427408e-
02 CRES + 1.774345e-02 BALD + 2.463208e-02 GLY + 2.734579e-02 MGLY + 0.000000e+00 AFG1
+ 8.558605e-02 AFG2 + 9.382650e-01 LOSTC
k=k144; k144=2.652212e-12*exp(-(-4.855759e-01/R/T))

HO + AAR3 --> 7.963319e-01 RO2R + 6.211264e-02 RO2N + 2.565418e-04 RO2XN +
0.000000e+00 RO2NP + 1.412987e-01 HO2 + 1.849867e-01 R2O2 + 1.043688e+00 RO2 +
3.831410e-03 HCHO + 8.734376e-03 CCHO + 5.315202e-02 RCHO + 0.000000e+00 ACET +
2.65 898e-01 MEK + 1.047743e-04 CO + 1.947650e-03 CO2 + 0.000000e+00 PHEN +1.412987e-
01 CRES + 2.844189e-02 BALD + 5.283478e-02 GLY + 3.643928e-01 MGLY + 0.000000e+00 AFG1
+ 5.032837e-01 AFG2 + 3.765364e+00 LOSTC
k=k145;k145=9.705207e-12*exp(-(-5.763993e-01/R/T))

# Lumped olefin reactions
```

```
OLE1 + HO --> 0.913685 RO2R + 8.6315e-02 RO2N + 1.0 RO2 + 0.913685 HCHO + 0.312453
CCHO + 0.601232 RCHO + 0.933738 LOSTC
k=k146;  k146=2.198977e-12*(T/300)^(0.000000e+00)*exp(-(-1.570000e+00/R/T))

OLE1 + O3 --> 0.06 HO + 0.165 HO2 + 0.135 RO2R + 0.135 RO2 + 0.551296 HCHO +  0.26969
CCHO + 0.329015 RCHO + 0.186062 MEK +  0.295 CO + 1.59063 LOSTC + 0.285 O3OLSB
k=k147;  k147=2.024443e-15*(T/300)^(0.000000e+00)*exp(-(3.071000e+00/R/T))
# O3OLSB is ozonolysis product with so2 or so3
# So2.rxn not included in this study (not in ftp files)

OLE1 + O --> 0.4 HO2 +  0.5 RCHO + 0.5 MEK +  1.2076 LOSTC
k=k148;  k148=4.532975e-12*(T/300)^(0.000000e+00)*exp(-(5.500000e-02/R/T))

OLE1 + NO3 --> 1.0 NO2 + 1.0 R2O2 + 1.0 RO2 +  1.0 HCHO + 3.4197e-01 CCHO + 6.5803e-01
RCHO + 1.04957 LOSTC    ;k=k149;  k149=2.206743e-12*(T/300)^(0.000000e+00)*exp(-
(3.114000e+00/R/T))

OLE2 + HO --> 0.942622 RO2R + 5.7378e-02 RO2N + 1.0 RO2 + 0.2630387 HCHO + 0.8244172
CCHO + 0.6111018 RCHO + 0.0135266 ACET + 0.09988 MEK + 0.754432 LOSTC
k=k150;  k150=4.826658e-12*(T/300)^(0.000000e+00)*exp(-(-1.557000e+00/R/T))

OLE2 + O3 --> 0.111201 HO + 0.176591 HO2 + 0.206954 RO2R + 0.0276364 R2O2 + 0.23459
RO2 + 0.00921215 CCOO2 + 0.00921215 C2COO2 + 0.0184243 RCO3 + 0.279846 HCHO + 0.543487
CCHO + 0.32415 RCHO + 0.007175 ACET + 0.390929 MEK + 0.001435 MGLY + 0.175568 CO +
1.07616 LOSTC + 0.20386 O3OLSB
k=k151;  k151=1.826542e-15*(T/300)^(0.000000e+00)*exp(-(1.439000e+00/R/T))

OLE2 + O --> 0.4 HO2 +  0.5 RCHO + 0.5 MEK +  1.7266 LOSTC
k=k152;  k152=1.012937e-11*(T/300)^(0.000000e+00)*exp(-(-6.130000e-01/R/T))

OLE2 + NO3 --> 1.0 NO2 +  1.0 R2O2 + 1.0 RO2 + 2.7905e-01 HCHO + 8.7406e-01 CCHO +
6.483e-01 RCHO + 1.4350e-02 ACET + 1.0596e-01 MEK + 0.78764 LOSTC
k=k153;  k153=2.488163e-13*(T/300)^(0.000000e+00)*exp(-(-8.070000e-01/R/T))


Uncertain uk1 uk2 uk3 uk4 uk5 uk6 uk7 uk8 uk9 uk10 uk11 uk12 uk13 uk14 uk15 uk16 uk17
uk18 uk19


# declare constants reaction species
O2=5.1414e18
M=2.46e19
H2O=4.92e17
HV=1.0

# declare constant parameters (used in rate constant expressions)
T=300.0
R=0.001987
P=2.46e19


# declare initial conditions
IC NO2=8.9052e11
IC NO=2.64814e12
IC CO=4.9938e13
IC ETHE=2.4846e11
IC HCHO=2.5756e11
IC CCHO=1.9317e11
IC AAR1=6.636e11
IC AAR2=7.2155e11
IC AAR3=2.435e11
IC OLE1=1.8694e11
IC OLE2=1.0051e11


# End of saprc.p.map
```

B.3 All City Average NMOC Composition.

The base case scenario used in this study was chosen to mimic the maximum incremental reactivity (MIR) conditions in a box model. An "all-city average" (Jeffries *et al.*, 1989) ROG composition profile was used to represent typical ground level ROG emissions in urban areas. The composition is listed in Table 20-3, where aldehydes were assumed to account for 5% of the mixture. This organic profile was used throughout this study and lumped differently for different mechanisms.

Reference:

Jeffries, H.E., Sexton, K.G., Arnold, J.R., Kale T.L., "Validation Testing of New Mechanisms with Outdoor Chamber Data, Volume 2: Analysis of VOC Data for the CB4 and CAL Photochemical Mechanisms", Final Report, EPA 600/3-89-010b, 1989

Alkanes		Aromatics		Alkenes	
ETHANE	0.020898	BENZENE	0.003460	ETHENE	0.019293
PROPANE	0.015588	TOLUENE	0.008268	PROPENE	0.004964
2-ME-C3	0.0069685	C2-BENZ	0.001241	C4-OLE1	0.001741
N-C4	0.0165335	M-XYLENE	0.001947	1-BUTENE	0.002509
2-ME-C4	0.0142225	P-XYLENE	0.001947	13-BUTDE	0.000671
N-C5	0.0068745	O-XYLENE	0.001547	T-2-BUTE	0.001028
BR-C6	0.000896	I-C3-BEN	0.000144	C-2-BUTE	0.000801
22-DM-C4	0.000390	N-C3-BEN	0.000288	C5-OLE1	0.001095
CYCC5	0.000640	C9-BEN2	0.002281	1-PENTEN	0.000545
23-DM-C4	0.001017	C9-BEN1	0.000325	2M-1-BUT	0.000827
2-ME-C5	0.003854	C9-BEN3	0.000097	ISOPRENE	0.000680
3-ME-C5	0.002762	123-TMB	0.002911	C5-OLE2	0.002008
N-C6	0.002972	C10-BEN1	0.001284	2M-2-BUT	0.000112
BR-C7	0.005220	C10-BEN2	0.001636	C6-OLE1	0.000737
CYC-C6	0.001681	C10-BEN3	0.000914	C6-OLE2	0.000943
CYCC6	0.000850	C10-BEN4	0.000183	1-HEXENE	0.000213
BR-C8	0.004937	C11-BEN1	0.000075	C7-OLE1	0.001156
N-C7	0.001131	C11-BEN2	0.000742	C8-OLE1	0.000400
CYC-C7	0.000845	C11-BEN3	0.000742	C7-OLE2	0.000101
BR-C9	0.001945	C12-BEN2	0.000187	C8-OLE2	0.000023
N-C8	0.000599	C12-BEN3	0.000187	C9-OLE1	0.000716
CYC-C8	0.000097	C12-BEN1	0.000002	A-PINENE	0.000611
N-C9	0.000461			C10-OLE1	0.000032
BR-C10	0.001395			C11-OLE1	0.000408
N-C10	0.000636				
BR-C11	0.000558	Aldehydes		Others	
BR-C12	0.000129	FORMALD	0.020	ACETYLEN	0.011496
BR-C13	0.000003	ACETALD	0.015	INERT	0.008021

Table 20-3. All City Averaged Organic Composition, Normalized to 1 ppmC.

B.4 Base Case Simulation

The base case simulation was chosen to test aspects of uncertainties within the chemical mechanism itself: rate constants, product coefficients, and initial conditions. To isolate the effects of the mechanistic parameters, features of the box model used were kept to a minimum. The model is run as if to simulate a constant volume of air with a set of initial conditions left to "cook" under bright sunlight for 10 hours, with no dilution, no entrainment, no emissions, no deposition, no change in mixing height, and no change in actinic flux (all photolysis rates were calculated with zenith angle = 0).

The initial conditions used in the base case study are shown in Table 20-4. The total amount of organics present was 0.126 ppbv, and the initial concentration of NO_x was 0.144 ppbv. These conditions were chosen to resemble the Maximum Incremental Reactivity (MIR) condition of most organics (Carter, 1991). The composition of the organics was assumed to the "All City Average" (Jeffris, 1989), which was reproduced in Table 6.4 in the Appendix. The model used in this study employs 6 species of primary organic compounds, as discussed in Section 1. Ethene was represented explicitly. Alkanes and aromatics are lumped together into three AAR classes The most reactive class, AAR3, contained mostly aromatic species. The least reactive class, AAR1, represented most of the straight chain alkanes. Due to the lumping procedure applied to the organic compounds, described in the previous section, the total organics represented in the initial condition used in the model was 0.106 ppbv.

Species	Concentration (molec/cm^3)	(ppb)
NO	2.65e12	107.7
NO2	8.91e11	36.2
HCHO	2.58e11	10.4
CCHO	1.93e11	7.8
CO	4.99e11	20.3
ETHE	2.48e11	10.1
AAR1	6.64e11	27.0
AAR2	7.22e11	29.3
AAR3	2.44e11	10.0
OLE1	1.87e11	7.6
OLE2	1.01e11	4.1

Table 20-4. Initial condition used in the base case simulation of SAPRC.

Using LSODE to solve the model, the instantaneous concentrations of species-of-interest can be output at regular intervals. The results of the base case simulation are shown in Figures 20-2 to 20-5. The numerical values for the concentrations of key species are listed in Table 20-5.

Time	O3 (ppm)	NO (ppm)	NO_2 (ppm)	HCHO (ppb)	CCHO (ppb)	HO (ppt)	PAN (ppb)	HNO_3 (ppb)
2	0.018	0.068	0.065	12.52	10.32	0.091	0.553	8.82
4	0.041	0.038	0.080	12.66	11.07	0.077	1.18	22.09
6	0.067	0.023	0.080	11.49	10.99	0.078	1.74	35.72
8	0.096	0.015	0.073	10.26	10.48	0.089	2.39	49.86
10	0.127	0.009	0.061	9.257	9.620	0.111	3.23	64.77
Time	AAR1 (ppb)	AAR2 (ppb)	AAR3 (ppb)	ETH (ppb)	OLE1 (ppb)	OLE2 (ppb)		
2	26.18	26.68	6.61	8.80	4.56	1.03		
4	25.42	24.29	4.43	7.63	2.64	0.150		
6	24.73	22.27	3.06	6.62	1.50	0.011		
8	24.01	20.27	2.05	5.64	0.778	3e-4		
10	23.17	18.10	1.27	4.63	0.344	3e-6		

Table 20-5. Base Case Time Series of Key Species (SAPRC)

Figure 20-2 shows the evolution of ozone, nitric oxide, and nitrogen dioxide. As is typical of box models, nitrogen dioxide concentration exhibited a peak about 5 hours into the simulation. Since there was no source terms in the simulation, NOx (NO plus NO2) ultimately decreased towards the end of the simulation. Ozone concentration increased throughout the simulation, reaching a concentration of 127 ppb in 10 hours. Since neither organics nor nitrogen oxides was depleted in the time of simulation, ozone was produced throughout the simulation and no ozone peak was observed.

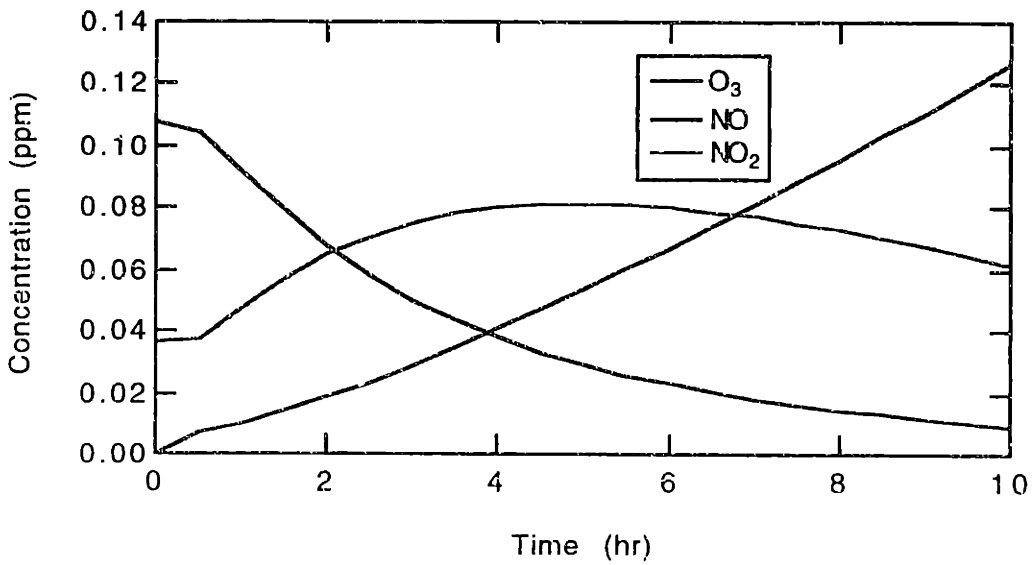


Figure 20-2. SAPRC Base Case Dynamics of Ozone, NO, and NO_2

Figure 20-3 shows the time series of the classes of primary organic compounds. The most reactive category, OLE2, which represented most of the internal alkenes, was depleted four hours into the simulation. Two other classes, OLE1 (terminal alkenes) and AAR3 (mostly aromatics), were also mostly reacted by the end of the simulation. Ethene and the less reactive alkanes decayed in a mostly linear fashion. In terms of absolute amount reacted, AAR2, the most abundant of the organic classes, contributed more than any other classes.

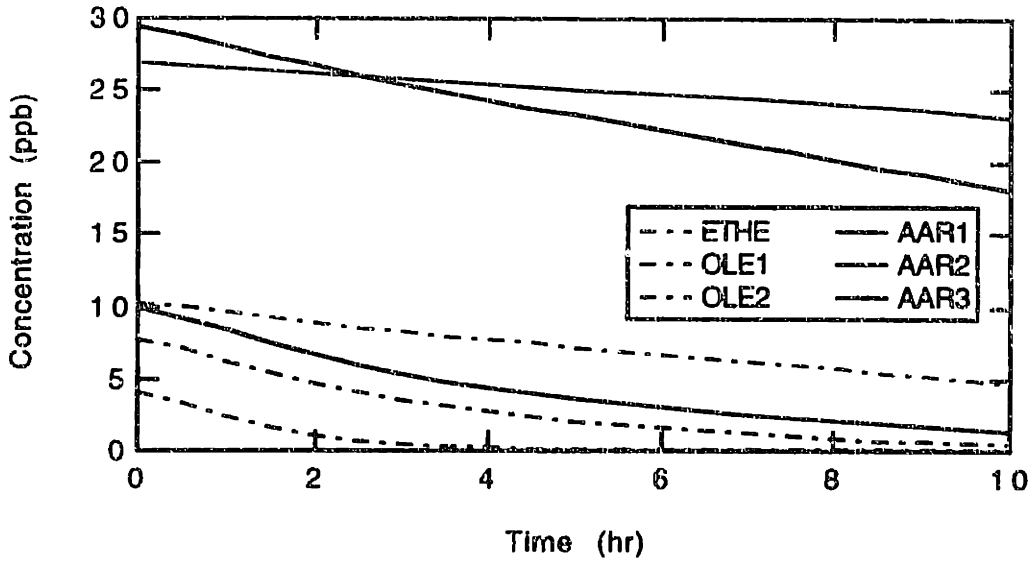


Figure 20-3. SAPRC Base Case Dynamics of Organic Source Species.

Figure 20-4 shows the accumulation of nitric acid and PAN throughout the simulation. Even though both compounds can dissociate back into their precursors, both serve as sink compounds for NOx in the base case scenario. Nitric acid formation is a particular important process because it was an effective sink for both NO2 and hydroxyl radical. 65 ppb of nitric acid and 3 ppb of PAN were formed in ten hours.

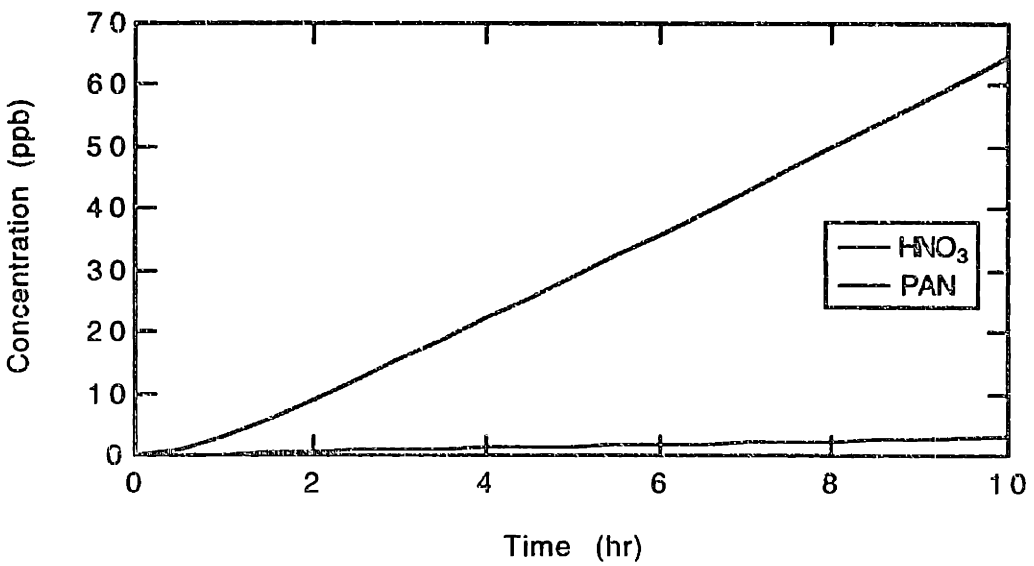


Figure 20-4. SAPRC Base Case Dynamics of Sink Species.

The concentrations of the radical species HO and HO2 are shown in Figure 20-5. Both radicals exhibited concentrations typical of polluted atmospheres. The maximum level of hydroxyl radicals was $5x10^6$ molecules/cm^3, and peroxyl radical was about ten times as

abundant. The accumulation of radicals towards the end of the simulation is probably an artifact due to inadequate removal processes described in the box model.

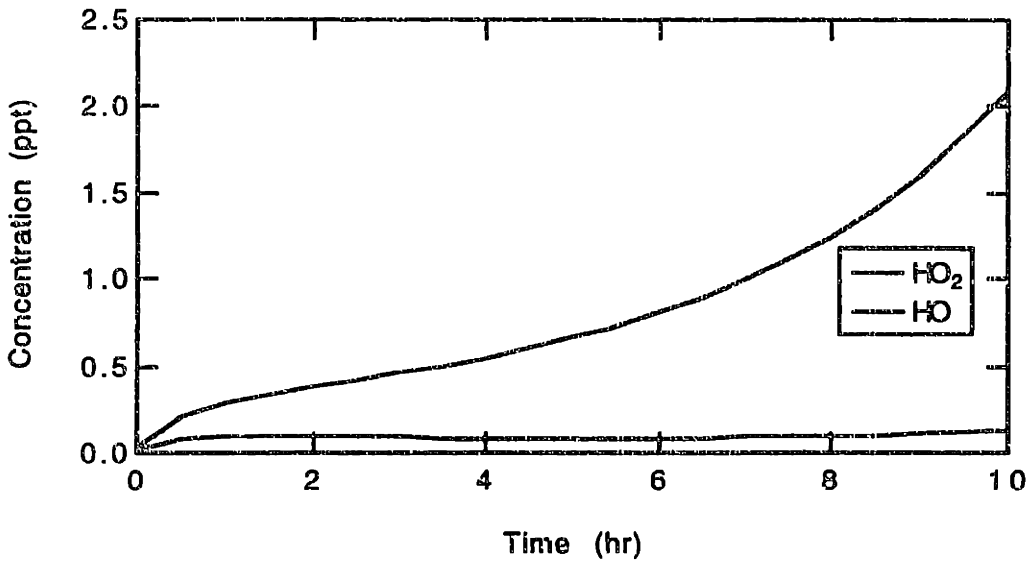


Figure 20-5. SAPRC Base Case Dynamics of Key Radicals.

B.5 Demmucom Input Scripts

The input files for uncertainty analysis of the SAPRC mechanism are included in this subsection of Appendix B. The method of carrying out a comprehensive uncertainty analysis for a complex mechanism was discussed in the main text. Using a "divide and conquer" approach, the parameters used in the SAPRC mechanism were divided into groups such as inorganic parameters, secondary and primary organic parameters. (See text.) The uncertainty input files are included here for (1) inorganic reactions, (2) secondary organic reactions, (3) alkane reaction rate and product coefficients, (4) aromatic reactions, (5)olefin reactions, (6) composition, and (7) the combined study of all varianice-contributing input parameters as identified previously.

Each input file contains a description of the uncertain parameters (distribution type and parameters), a list of uncertain outcomes expected from the model, and other commands and information for Demmucom.

B.5.1 Inorg.dmc

```
# file name: inorg.dmc
# input date: July 6, 1996
# description: Demmucom input script for inorganic reactions in the SAPRC mechanism
# uncertainty compilation references: Stockwell et al., 1994

Uncertain inputs
j1 = lognormal(0.00898167, 1.3),
k2 = lognormal(6.0e-34, 1.1),                #fall off
# fall off reaction parameters determined by fitting to Monte Carlo sampling of rate
expression
k3 = lognormal(9.68971e-12, 1.15),
k4 = lognormal(1.55705e-12, 1.1),    #fall off
k5 = lognormal(1.88008e-14, 1.2),
k6 = lognormal(3.36235e-17, 1.15),
k7 = lognormal(2.80259e-11, 1.32),
k8 = lognormal(1.92089e-38, 1.26),
k9 = lognormal(1.25732e-12, 1.53),           #fall off
keq10 = lognormal(5.50584e+10, 1.5),
k11 = lognormal(1.000000e-24, 10.0),         #carter k = 1e-21; stockwell discrete PDF
k12 = lognormal(4.17113e-16, 1.5),
j13 = lognormal(0.0187, 2.0),
j14 = lognormal(0.169333, 2.0),
j15 = lognormal(0.000499333, 1.1),
j16 = lognormal(3.78e-05, 1.4),
k17 = lognormal(2.2e-10, 1.26),
k18 = lognormal(2.92377e-11, 1.26),
k19 = lognormal(4.81886e-12, 1.29),          #fall off
j20 = lognormal(0.0016275, 1.6),
k21 = lognormal(1.132455e-11, 1.28),         #fall off
k22 = lognormal(1.03072e-13, 1.3),
k23 = lognormal(2.400000e-13, 1.3),
k24 = lognormal(6.94565e-14, 1.41),
k25 = lognormal(8.27771e-12, 1.5),
k26 = lognormal(1.36845e-12, 1.12),          #fall off
keq27 = lognormal(7.88864e+10, 5.0),
k28 = lognormal(4.61324e-12, 1.5),
k29 = lognormal(2.07589e-15, 1.58),
k30 = lognormal(1.73201e-12, 1.41),
k31 = lognormal(5.00547e-32, 1.41),
k32 = lognormal(3.72646e-30, 1.41),
k33 = lognormal(1.07924e-49, 1.41),
j38 = lognormal(7.53e-06, 1.4),
k39 = lognormal(1.69541e-12, 1.26),
k40 = lognormal(9.9017e-11, 1.3);
```

```
Uncertain outputs
O3, NO, NO2, HO;

Do Collocation with
model: mlsode,
dynamic: starting = 0.0, step = 1.67, interval = 6,
    # 7 time points; output format (dynamic option):
    # O3(t = 0)      NO(t = 0)      NO2(t = 0)      HO(t = 0)
    # O3(t = 1.67)   NO(t = 1.67)   NO2(t = 1.67)   HO(t = 1.67)
    # . . .
    # O3(t = 10.0)   NO(t = 10.0)   NO2(t = 10.0)   HO(t = 10.0)
approximation: linear,
sampling file : samp.pnt,
number of points: 5000,
options: looping, no header;


# end inorg.dmc
```

B.5.2 Sec.dmc

```
# file name: sec.dmc
# date: June 17, 1996
# description: Demmucom input script for secondary organic reactions in the SAPRC
mechanism
# uncertain reaction rate compilation references: Stockwell et al., 1994
# uncertain product coefficients, see text


Uncertain inputs

# reaction rates
k41 = lognormal (7.683e-12, 2.0),    # RO2 + NO
k42 = lognormal (9.92678e-12, 2.0), # RCO3 + NO
k43 = lognormal (7.20102e-12, 2.0), # RCO3 + NO2
k44 = lognormal (4.89654e-12, 2.0), # RO2 + HO2
k46 = lognormal (1.0e-15, 2.0),      # RO2 + RO2
k47 = lognormal (1.08814e-11, 2.0), # RO2 + RCO3
k48 = lognormal (1.63807e-11, 2.0), # RCO3 + RCO3
k65 = lognormal (7.38833e-06, 1.8), # ROOH + hv
k66 = lognormal (1.80691e-12, 1.5), # ROOH + HO
k67 = lognormal (3.71344e-12, 1.5), # ROOH + HO (RO2)
k68 = lognormal (3.00667e-05, 1.4),         # HCHO + hv (r)
k69 = lognormal (4.305e-05, 1.4),    # HCHO + hv (s)
k70 = lognormal (9.76197e-12, 1.26),        # HCHO + HO
k71 = lognormal (7.79189e-14, 5.0), # HCHO + HO2
k72 = lognormal (176.175, 2.0),      # HOCOO --> HO2 + HCHO
k74 = lognormal (6.37322e-16, 2.0), # HCHO + NO3
k75 = lognormal (1.5651e-11, 1.41),         # CCHO + HO
k76 = lognormal (5.94e-06, 1.4),     # CCHO + hv
k77 = lognormal (2.84035e-15, 2.0), # CCHO + NO3
k78 = lognormal (1.96652e-11, 1.41),        # RCHO + HO
k79 = lognormal (2.14e-05, 1.4),     # RCHO + hv
k80 = lognormal (2.84035e-15, 2.0), # RCHO + NO3
k81 = lognormal (2.3091e-13, 1.58), # ACET + HO
k82 = lognormal (8.17e-07, 1.5),     # ACET + hv
k83 = lognormal (1.16142e-12, 1.3), # MEK + HO
k84 = lognormal (1.58117e-06, 1.5),         # MEK + hv
k85 = lognormal (2.06457e-12, 1.2), # RNO3 + HO
k86 = lognormal (0.00014845, 1.6),  # MGLY + hv
k87 = lognormal (1.72e-11, 2.0),     # MGLY + HO
k94 = lognormal (0.000669667, 2.0), # PAN --> CCOO2 + NO2
k96 = lognormal (8.4000e-12, 2.0),  # C2COO2 + NO2
   # k96 is assumed to be independent and different from k43
k100 = lognormal (0.000674866, 1.9),        # PPN --> C2COO2 + NO2
k101 = lognormal (4.62e-05, 1.5),    # GLY + hv (r)
k102 = lognormal (6.44167e-05, 1.5),        # GLY + hv (s)
k103 = lognormal (1.1400e-11, 2.0), # GLY + HO
```

```
k111 = lognormal (2.6300e-11, 1.25),        # PHEN + HO
k112 = lognormal (3.6e-12, 2.0),     # PHEN + NO3
k113 = lognormal (4.2e-11, 1.25),    # CRES + HO
k114 = lognormal (2.1e-11, 2.0),     # CRES + NO3
k115 = lognormal (1.29e-11, 1.41),   # BALD + HO
k116 = lognormal (4.30167e-05, 1.5),        # BALD + hv
k117 = lognormal (2.60744e-15, 2.0),        # BALD + NO3
k119 = lognormal (8.4e-12, 2.0),     # BZCOO2 + NO2
   # k119 is assumed independent and different from k43
k123 = lognormal (0.000215984, 1.9),        # PBZN --> BZCOO2 + NO2
k124 = lognormal (3.53317e-11, 2.0),        # BZO + NO2
k126 = lognormal (1.0e-3, 2.0),      # BZO --> PHEN
k127 = lognormal (3.6e-12, 1.58),    # NPHE + NO3
k132 = lognormal (5.28667e-05, 3.0),        # AFG1 + hv

# product coefficients
c51 = uniform (0.35, 0.62),          # RO2 + RO2 --> RO2 + c HO2 (nominal c = 0.5)
   # c51 also used in Reactions 55, 63, 137
c52 = uniform (0.4, 0.6),            # RO2 + RCO3 --> RCO3 + c HO2(nominal c = 0.5)
   # c52 also used in Reactions 56, 64, 91, 98, 109, 121, 138
c81 = uniform (0.44 , 0.95),         # product split in ACET + HO (nominal 0.8)
c831 = uniform (0.16, 0.83),         # product split in MEK + HO (nominal 0.5)
c832 = lognormal (1.5, 1.5),         # RO2 yield in MEK + HO (nominal 1.5)
c851 = uniform (0.11, 0.2),          # MEK yield in RNO3 + HO (nominal 0.155)
c852 = uniform (0.32, 1.78),         # RCHO yield in RNO3 + HO (nominal 1.05)
c853 = uniform (0.001, 0.96),              # CCHO yield in RNO3 + HO (nominal 0.48)
c854 = uniform (0.001, 0.32),              # HCHO yield in RNO3 + HO (nominal 0.16)
c855 = lognormal (1.39, 1.5),              # RO2 yield in RNO3 + HO (nominal 1.39)
c1011 = uniform (0.4, 1.2),          # HO2 yield in GLY + hv (nominal 0.8)
c1012 = uniform(0.3, 0.6),           # product split in GLY + hv (nominal 0.45)
c102 = uniform (0.08, 0.2),          # product split in GLY + hv (nominal 0.13)
c103 = uniform (0.12, 0.75),         # product split in GLY + HO (nominal 0.4)
  # also for GLY + NO3 (Reaction 104)
c1111 = uniform (0.03, 0.47),              # product split in PHEN + HO (nominal 0.15)
c1112 = lognormal (0.2, 1.5);              # GLY yield in PHEN + HO (nominal 0.2)
  # coefficients in PHEN + HO Reaction 111 also used for CRES + HO (Reactions 113)

Uncertain Outputs
O3, NO, NO2, HO;

Do Collocation with
Model: mlsode,
Dynamic: starting = 0.0, step = 1.67, interval = 6,  # 7 outputs
approximation: linear,
sampling file: sampling.pnt,
number of points: 5000,
options: looping, no header;

# end sec.dmc
```

B.5.3 Alkane.dmc

```
# file name: alkane.dmc
# date: March 22, 1996
# description: Demmucom input script for alkane parameters (pre-lump)
# (excluding concentrations)
# Uncertainty in aromatic reaction rate constants (Stockwell et al., 1994)
# Uniform distribution for assignment based on Carter's uncertainty classification,
see text
# Coefficients for CO and CO2 are not treated as uncertain


Uncertain inputs
# Ethane
ethane_k = lognormal (2.74e-13, 1.2),
ethane_rr = constant (1.0),
ethane_a2 = constant (1.0),          # the only possible product
# Acetylene (product range based on experimental measurements)
acetylene_k = lognormal (7.82e-13, 1.2),
```

```
acetylene_rh = uniform (0, 0.6),            # nominal 0.3
acetylene_gl = uniform (0.4, 1),            # nominal 0.7 range from Atkinson
acetylene_rr = constant (0.7),              # rr = gl
# acetylene_co = constant (0.3),
# Propane
propane_k = lognormal (1.223e-12, 1.2),
propane_xn = constant (0.03880),
# propane_rr = constant (0.96120),
propane_k3 = uniform (0.569, 0.736),        # nominal 0.65843
propane_a3 = constant (0.30277),            # compete with k3
# N-C4
nc4_k = lognormal (2.555e-12, 1.2),
nc4_nr = uniform (0.0618, 0.0934),          # nominal 0.07613
# nc4_rr = constant (0.92387),              # mechanism = 1 - nr
nc4_r2 = uniform (0.3605, 0.4362),          # nominal 0.39654
nc4_a1 = uniform(0.000645,0.000781),        # nominal 0.00071
nc4_a2 = uniform (0.5191, 0.6282),          # nominal 0.57105
nc4_a3 = uniform (0.1272, 0.1538),          # nominal 0.13981
nc4_k4 = uniform (0.4848, 0.5864),          # nominal 0.53331
# 2-Me-C3
twmec3_k = lognormal(2.386e-12, 1.2),
twmec3_nr = uniform (0.011, 0.065), # nominal 0.02704
# twmec3_rr = constant(0.97296),
twmec3_r2 = uniform (0.5494, 0.866),        # nominal 0.74360
twmec3_a1 = constant (0.74360),             # a1 = k3 = r2
twmec3_k3 = constant (0.74360),
twmec3_a3 = constant (0.22935),             # a3 = rr - r2
# N-C5
nc5_k = lognormal(3.957e-12, 1.2),
nc5_nr = uniform (0.0639, 0.2144),          # nominal 0.12010
nc5_xn = constant(0.00026),
# nc5_rr = constant(0.87964),
nc5_r2 = uniform (0.3108, 0.9519),          # nominal 0.54397
nc5_a1 = uniform (0.00411, 0.0126), # nominal 0.00720
nc5_a2 = uniform (0.0459, 0.1406),          # nominal 0.08033
nc5_a3 = uniform (0.0983, 0.3011),          # nominal 0.17203
nc5_k4 = uniform (0.5307, 1.6252),          # nominal 0.92868, impose upper limit?
# 2-Me-C4
twmec4_k = lognormal(3.998e-12, 1.2),
twmec4_nr = uniform (0.02676, 0.1466),      # nominal 0.06430
twmec4_xn = constant (0.00238),
# twmec4_rr = constant(0.93332),
twmec4_r2 = uniform(0.3668, 1.4673),        # nominal 0.73365
twmec4_a1 = uniform(0.0002, 0.00078),       # nominal 0.00039
twmec4_a2 = uniform(0.3072, 1.2289),        # nominal 0.61446
twmec4_k3 = uniform(0.3057, 1.2226),        # nominal 0.61130
twmec4_a3 = uniform(0.06626, 0.265),        # nominal 0.13251
twmec4_k4 = uniform(0.1513, 0.6053),        # nominal 0.30265
# Cyclo-pentane
cycc5_k = lognormal(5.616e-12, 1.2),
cycc5_nr = uniform (0.0552, 0.2674),        # nominal 0.12739
# cycc5_rr = constant(0.87261),
cycc5_r2 = constant (1.74521),              # r2 = 2 a3
cycc5_a3 = constant(0.87261),               # a3 = rr (ring opening path assumed)
cycc5_k4 = uniform (0.109, 0.4363), # nominal 0.21815
# cycc5_co = constant(0.87261),
# N-C6
nc6_k = lognormal(5.359e-12, 1.2),
nc6_nr = uniform (0.10198, 0.31235),        # nominal 0.18509
nc6_xn = constant (0.00022),
# nc6_rr = constant(0.81469),
nc6_r2 = uniform (0.422, 1.291),            # nominal 0.73775
nc6_a2 = uniform(0.0115, 0.0351),           # nominal 0.02007
nc6_a3 = uniform (0.0598, 0.183),           # nominal 0.10457
nc6_k4 = uniform (0.648, 1.984),            # nominal 1.13374, impose upper limit?
# 2-Me-C5
twmec5_k = lognormal(5.400e-12, 1.2),
twmec5_nr = uniform (0.0529, 0.2585),       # nominal 0.12248
twmec5_xn = constant (0.00494),
# twmec5_rr = constant (0.87257),
twmec5_r2 = uniform (0.375, 1.499), # nominal 0.74948
twmec5_a1 = uniform (0.00294,0.0117),       # nominal 0.00587
```

```
twmec5_a2 = uniform (0.0116, 0.0462),      # nominal 0.02312
twmec5_k3 = uniform (0.1113, 0.4452),      # nominal 0.22260
twmec5_a3 = uniform (0.2725, 1.09), # nominal 0.54498, impose upper limit?
twmec5_k4 = uniform (0.362, 1.448), # nominal 0.72390, impose upper limit?
# 3-Me-C5
thmec5_k = lognormal(5.761e-12, 1.2),
thmec5_nr = uniform(0.0482, 0.2404),       # nominal 0.11238
# thmec5_rr = constant(0.88762),
thmec5_r2 = uniform(0.43, 1.72),           # nominal 0.85990
thmec5_a1 = uniform (0.0024, 0.0096),      # nominal 0.00479
thmec5_a2 = uniform(0.262, 1.047),         # nominal 0.52326
thmec5_a3 = uniform(0.0444,0.177),         # nominal 0.08871, impose upper limit?
thmec5_k4 = uniform(0.501, 2.01),          # nominal 1.00276, impose upper limit?
# 2,2,Dimethyl-C4
twtwdmc4_k = lognormal(1.843e-12, 1.2),
twtwdmc4_nr = uniform(0.0675, 0.3116),     # nominal 0.15327
# twtwdmc4_rr = constant(0.84673),
twtwdmc4_r2 = uniform(0.48, 1.92),         # nominal 0.96013
twtwdmc4_a1 = uniform (0.1476,0.5905),     # nominal 0.29523
twtwdmc4_a2 = uniform (0.1514, 0.6054),    # nominal 0.30272
twtwdmc4_k3 = uniform (0.1476,0.5905),     # nominal 0.29523
twtwdmc4_a3 = uniform(0.1859, 0.7438),     # nominal 0.37189
twtwdmc4_k4 = uniform(0.2709, 1.0836),     # nominal 0.54180
# 2,3,Dimethyl-C4
twthdmc4_k = lognormal(5.441e-12, 1.2),
twthdmc4_nr = uniform(0.0313, 0.114),      # nominal 0.06066
twthdmc4_xn = constant (0.03864),
# twthdmc4_rr = constant(0.90070),
twthdmc4_r2 = uniform(0.5394, 1.652),      # nominal 0.94396
twthdmc4_k3 = uniform(0.9052, 2.772),      # nominal 1.58407
twthdmc4_a3 = uniform(0.0731, 0.224),      # nominal 0.12799
twthdmc4_k4 = uniform(0.0549,0.168),       # nominal 0.09599
# Cyclohexane
cycc6_k = lognormal(8.412e-12, 1.2),
cycc6_nr = uniform(0.0874, 0.375),         # nominal 0.19325
# cycc6_rr = constant(0.80675),
cycc6_r2 = uniform(0.176, 0.704),          # nominal 0.35200
cycc6_a1 = uniform (0.00167, 0.00668),     # nominal 0.00334
cycc6_a3 = uniform(0.166, 0.665),          # nominal 0.33270
cycc6_k4 = uniform(0.408, 1.632),          # nominal 0.81603
# cycc6_c2 = constant(0.00334),
# N-C7
nc7_k = lognormal(6.760e-12, 1.4),
nc7_nr = uniform (0.1542, 0.4217),         # nominal 0.26720
nc7_xn = constant (0.00009),
# nc7_rr = constant(0.73271),
nc7_r2 = uniform (0.4154, 1.272),          # nominal 0.72689
nc7_a3 = uniform (0.0317, 0.0972),         # nominal 0.05554
nc7_k4 = uniform (0.709, 2.171),           # nominal 1.24067
# 3-Me-C6
thmec6_k = lognormal(7.163e-12, 1.4),
thmec6_nr = uniform(0.082, 0.358),         # nominal 0.18245
thmec6_xn = constant (0.00207),
# thmec6_rr = constant(0.81548),
thmec6_r2 = uniform(0.421, 1.684),         # nominal 0.84190
thmec6_a1 = uniform (0.00004, 0.00016),    # nominal 0.00008
thmec6_a2 = uniform (0.0637,0.255), # nominal 0.12738
thmec6_k3 = uniform(0.165, 0.658),         # nominal 0.32902
thmec6_k4 = uniform(0.5596, 2.238), # nominal 1.11916
# Methyl-cyclohexane
mecycc6_k = lognormal(1.022e-11, 1.4),
mecycc6_nr = uniform(0.1212, 0.3556),      # nominal 0.21623
# mecycc6_rr = constant(0.78377),
mecycc6_r2 = uniform(0.53, 1.623),  # nominal 0.92752
mecycc6_a1 = uniform (0.0526, 0.161),      # nominal 0.09197
mecycc6_a2 = uniform (0.00039, 0.0012),    # nominal 0.00068
mecycc6_a3 = uniform(0.266,  0.815),       # nominal 0.46556
mecycc6_k4 = uniform(0.564, 1.727), # nominal 0.98703
# mecycc6_co = constant(0.00252),
# mecycc6_c2 = constant(0.04574),
# N-C8
nc8_k = lognormal(8.162e-12, 1.4),
```

```
nc8_nr = uniform (0.20, 0.499),            # nominal 0.33272
# nc8_rr = constant(0.66728),
nc8_r2 = uniform (0.403, 1.235),           # nominal 0.70560
nc8_a3 = uniform(0.0009, 0.00277),  # nominal 0.00158
nc8_k4 = uniform (0.762, 2.333),           # nominal 1.33337
# 4-Me-C7
fomec7_k = lognormal(8.565e-12, 1.4),
fomec7_nr = uniform(0.1145, 0.4466),       # nominal 0.24424
fomec7_xn = constant (0.00247),
# fomec7_rr = constant(0.75329),
fomec7_r2 = uniform(0.4015, 1.606), # nominal 0.80303
fomec7_a3 = uniform(0.1762, 0.705), # nominal 0.35248
fomec7_k4 = uniform(0.602, 2.408),  # nominal 1.20384
# Ethyl-cyclohexane
etcycc6_k = lognormal(1.217e-11, 1.4),
etcycc6_nr = uniform(0.126, 0.474), # nominal 0.26495
# etcycc6_rr = constant(0.73505),
etcycc6_r2 = uniform(0.641, 2.564), # nominal 1.28213
etcycc6_a1 = uniform (0.0932, 0.373),      # nominal 0.18643
etcycc6_a2 = uniform (0.146, 0.568),       # nominal 0.29283
etcycc6_a3 = uniform(0.173, 0.694), # nominal 0.34693
etcycc6_k4 = uniform(0.405, 1.622), # nominal 0.81075
# etcycc6_c0 = constant(0.00951),
# etcycc6_c2 = constant(0.18537),
# N-C9
nc9_k = lognormal(9.564e-12, 1.4),
nc9_nr = uniform (0.229, 0.544),           # nominal 0.37319
# nc9_rr = constant(0.62681),
nc9_r2 = uniform (0.385, 1.179),           # nominal 0.67347
nc9_a3 = uniform (0.00076, 0.00233),       # nominal 0.00133
nc9_k4 = uniform (0.742, 2.273),           # nominal 1.29895
# 4-Ethyl-C7
foetc7_k = lognormal(1.052e-11, 1.4),
foetc7_nr = uniform(0.1295, 0.4815),       # nominal 0.27108
foetc7_xn = constant (0.00212),
# foetc7_rr = constant(0.72680),
foetc7_r2 = uniform(0.4018, 1.6073),       # nominal 0.80367
foetc7_a1 = uniform (0.00076, 0.00302),    # nominal (0.00151),
foetc7_a2 = uniform (0.0297,0.1186),       # nominal (0.05931),
foetc7_a3 = uniform(0.1515, 0.6059),       # nominal 0.30296
foetc7_k4 = uniform(0.5833, 2.333), # nominal 1.16669
# N-C10
nc10_k = lognormal(1.097e-11, 1.4),
nc10_nr = uniform (0.208, 0.622),          # nominal 0.39688
# nc10_rr = constant(0.60312),
nc10_r2 = uniform(0.329, 1.318),           # nominal 0.65879
nc10_a3 = uniform(0.00057, 0.00228),       # nominal 0.00114
nc10_k4 = uniform(0.6304, 2.522),          # nominal 1.26076
# 4-Propyl-C7
foprc7_k = lognormal(1.192e-11, 1.4),
foprc7_nr = uniform (0.126, 0.563), # nominal 0.30125
foprc7_xn = constant (0.00229),
# foprc7_rr = constant(0.69646),
foprc7_r2 = uniform (0.3099, 1.9369),      # nominal 0.77476
foprc7_a2 = uniform(0.0016, 0.01),  # nominal 0.00400
foprc7_a3 = uniform(0.1311, 0.8195),       # nominal 0.32779
foprc7_k4 = uniform(0.4558, 2.8486),       # nominal 1.13944 impose upper limit?
# 3,5,Diethyl-C8
thfidec8_k = lognormal(1.428e-11, 1.4),
thfidec8_nr = uniform(0.09822, 0.4950),    # nominal 0.24629
# thfidec8_rr = constant(0.75371),
thfidec8_r2 = uniform(0.509, 3.182),       # nominal 1.27289
thfidec8_a1 = uniform (0.00834,0.0521),    # nominal 0.02085
thfidec8_a2 = uniform (0.0214, 0.1338),    # nominal 0.05353
thfidec8_a3 = uniform(0.0361, 0.225),      # nominal 0.09013
thfidec8_k4 = uniform(0.7448, 4.6552),     # nominal 1.86208 impose upper limit?
# 3,6,Diethyl-C8
thsidec8_k = lognormal(1.569e-11, 1.4),
thsidec8_nr = uniform(0.1084, 0.5226),     # nominal 0.26735
# thsidec8_rr = constant(0.73265),
thsidec8_r2 = uniform(0.5402, 3.3763),     # nominal 1.35050
thsidec8_a1 = uniform(0.0008, 0.0052),     # nominal 0.00206
```

```
thsidec8_a2 = uniform(0.1686, 1.054),      # nominal 0.42158
thsidec8_a3 = uniform(0.0049, 0.0307),     # nominal 0.01227
thsidec8_k4 = uniform(0.6589, 4.118),      # nominal 1.64723 impose upper limit?
# 3,7,Diethyl-C9
thsedec9_k = lognormal(1.709e-11, 1.4),
thsedec9_nr = uniform(0.1174, 0.5449),     # nominal 0.28524
# thsedec9_rr = constant(0.71476),
thsedec9_r2 = uniform(0.4904, 3.065),      # nominal 1.22602
thsedec9_a1 = uniform(0.0006,0.0038),      # nominal 0.00153
thsedec9_a2 = uniform(0.0033,0.0209),      # nominal 0.00834
thsedec9_a3 = uniform(0.0446, 0.2787),     # nominal 0.11149
thsedec9_k4 = uniform(0.7278, 4.5486);     # nominal 1.81942 impose upper limit?

Uncertain outputs
       O3, NO, NO2, HO;

Do Collocation With
  model: alkalump,
  Dynamic: starting = 0.0, step = 1.67, interval = 6,
# 7 outputs
  approximation: linear,
  sampling file : sampling.pnt,
  number of points : 5000,
  options: nonlinkable, no header;

# end alkane.dmc
```

B.5.4 Aro.dmc

```
# file name: Aro.dmc
# date: April 2, 1996
# description: Demmucom input script for aromatic parameters (pre-lump)
# (excluding concentrations)
# Stockwell assigned general UF 1.3 to aromatic reaction rate constants (Stockwell et
al., 1994)
# uncertain product coefficients described using normal or uniform distribution.
# Normal distribution for assignment based on actual experimental error
# Uniform distribution for assignment based on Carter's uncertainty classification,
see text


Uncertain inputs

# BENZENE default k and mechanism
    benzene_k = lognormal(1.2844e-12, 1.3),
    benzene_gl = normal (0.207, 0.010),    # gl = glyoxal
    benzene_ph = normal (0.236, 0.022),    # ph = phenol
    # benzene_rh = constant(0.236),        # rh = HO2 ; correlated to ph in mechanism:
rh = ph
    benzene_u1 = uniform(0.327, 0.735),    # u1 = AFG1 ; nominal = 0.49
    # benzene_rr = constant(0.764),        # rr = R2O2; correlated to rh: rr = 1 - rh

# TOLUENE default k and mechanism
    toluene_k = lognormal(5.90627e-12, 1.3),
    toluene_bz = normal(0.085, 0.0128),    # bz = benzene
    toluene_cr = normal(0.26, 0.0172),     # cr = cresol
    # toluene_rh = constant(0.26),         # mechanism: rh = cr
    toluene_gl = normal(0.118, 0.0158),
    toluene_mg = normal(0.131, 0.0187),    # mg = methyl-glyoxal
    toluene_u2 = uniform(0.273, 0.615),    # u2 = AFG2; nominal = 0.41
    # toluene_rr = constant(0.74),         # rr = 1 - rh

# C2-BENZ default k
    c2benz_k = lognormal(7.1e-12, 1.3),

# I-C3-BEN default k
    ic3ben_k = lognormal(6.5e-12, 1.3),

# N-C3-BEN default k
```

```
    nc3ben_k = lognormal(6.0e-12, 1.3),

# M-XYLENE default k and mechanism
    mxylene_k = lognormal(2.36e-11, 1.3),
    mxylene_bz = normal(0.04, 0.004),
    mxylene_cr = normal (0.18, 0.0342),     # impose lower limit.
    # mxylene_rh = constant(0.18),          # mechanism: rh = cr
    mxylene_gl = normal(0.108, 0.0124),
    mxylene_mg = normal(0.370, 0.024),
    mxylene_u2 = uniform(0.444, 0.999),     # nominal 0.666
    # mxylene_rr = constant(0.82),          # rr = 1 - rh

# O-XYLENE default k
    oxylene_k = lognormal(1.37e-11, 1.3),

# P-XYLENE default k
    pxylene_k = lognormal(1.43e-11, 1.3),

# 135-TMB default k and mechanism
    one35tmb_k = lognormal(5.75e-11, 1.3),
    one35tmb_bz = uniform(0.02, 0.045),     # nominal 0.03
    # one35tmb_rh = constant(0.18),         # mechanism: rh = cr = mxylene cr/rh
    one35tmb_mg = normal(0.62, 0.017),
    one35tmb_u2 = uniform(0.4, 0.9),              # nominal 0.60
    # one35tmb_rr = constant(0.82),         # rr = 1 - rh

# 123-TMB default k
    one23tmb_k = lognormal(3.27e-11, 1.3);


Uncertain outputs
        O3, NO, NO2, HO;

Do Collocation With
  model: arolump,
  Dynamic: starting = 0.0, step = 1.67, interval = 6,
  approximation: linear,
  sampling file : sampling.pnt,
  number of points : 5000,
  options: nonlinkable, no header;


#end aro.dmc
```

B.5.5 Ole.dmc

```
# file name: ole.dmc
# date: May 12, 1996
# description: Demmucom input script for olefin parameters (pre-lump)
# (excluding concentrations)
# Include parameters for structural-rxn relationship used to generate product coef.

Uncertain inputs
       ETHENE_kOH = lognormal(8.435e-12, 1.15),
       ETHENE_kO3 = lognormal(1.869e-18, 1.25),
       ETHENE_kNO3 = lognormal(2.231e-16, 2.0),
       ETHENE_kOA = lognormal(7.418e-13, 1.2),
       ETHENE_OHA2 = uniform(0.184, 0.261),
       ETHENE_O3RH = uniform(0.109, 0.132),
       ETHENE_NO3A1 = uniform(1.0 ,2.0),
       PROPENE_PN = constant(0.0),
       PROPENE_kOH = lognormal(2.60034e-11, 1.2),
       PROPENE_kO3 = lognormal(1.18505e-17, 1.5),
       PROPENE_kNO3 = lognormal(9.79037e-15, 2.),
       PROPENE_kOA = lognormal(4.0058e-12, 1.2),

       ONE_BUTENE_PN = constant(0.0),
       ONE_BUTENE_kOH = lognormal(3.10703e-11, 1.2),
       ONE_BUTENE_kO3 = lognormal(1.14759e-17, 1.5),
```

```
ONE_BUTENE_kNO3 = lognormal(1.30239e-14, 2.),
ONE_BUTENE_kOA = lognormal(4.21506e-12, 1.2),

T2BUTENE_PN = constant(0.0),
T2BUTENE_kOH = lognormal(6.29767e-11, 1.2),
T2BUTENE_kO3 = lognormal(2.05588e-16, 1.5),
T2BUTENE_kNO3 = lognormal(3.92265e-13, 2.),
T2BUTENE_kOA = lognormal(2.33711e-11, 1.2),

C2BUTENE_PN = constant(0.0),
C2BUTENE_kOH = lognormal(5.58006e-11, 1.2),
C2BUTENE_kO3 = lognormal(1.32943e-16, 1.5),
C2BUTENE_kNO3 = lognormal(3.46893e-13, 2.),
C2BUTENE_kOA = lognormal(1.79472e-11, 1.2),

TWOM_1_BUT_PN = constant(0.0),
TWOM_1_BUT_kOH = lognormal(5.99484e-11, 1.2),
TWOM_1_BUT_kO3 = lognormal(1.25705e-17, 1.5),
TWOM_1_BUT_kNO3 = lognormal(3.39537e-13, 2.),
TWOM_1_BUT_kOA = lognormal(1.52611e-11, 1.2),

TWOM_2_BUT_PN = constant(0.0),
TWOM_2_BUT_kOH = lognormal(8.60268e-11, 1.2),
TWOM_2_BUT_kO3 = lognormal(4.3131e-16, 1.5),
TWOM_2_BUT_kNO3 = lognormal(9.4116e-12, 2.),
TWOM_2_BUT_kOA = lognormal(4.72925e-11, 1.2),

ONE3_BUTDE_PN = constant(0.0),
ONE3_BUTDE_kOH = lognormal(6.58688e-11, 1.2),
ONE3_BUTDE_kO3 = lognormal(7.92554e-18, 1.5),
ONE3_BUTDE_kNO3 = lognormal(1.03384e-13, 2.),
ONE3_BUTDE_kOA = lognormal(2.1e-11, 1.2),

ISOPRENE_PN = constant(0.0),
ISOPRENE_kOH = lognormal(9.96804e-11, 1.25),
ISOPRENE_kO3 = lognormal(1.49853e-17, 1.5),
ISOPRENE_kNO3 = lognormal(6.85391e-13, 2.),
ISOPRENE_kOA = lognormal(6.0e-11, 1.2),

A_PINENE_PN = constant(0.00),
A_PINENE_kOH = lognormal(5.31343e-11, 1.25),
A_PINENE_kO3 = lognormal(1.00097e-16, 1.5),
A_PINENE_kNO3 = lognormal(6.09767e-12, 2.),
A_PINENE_kOA =lognormal(3.0e-11, 1.2),

ONE_PENTEN_PN = uniform(0.0426, 0.219),
ONE_PENTEN_kOH = lognormal(3.10447e-11, 1.2),
ONE_PENTEN_kO3 = lognormal(1.11159e-17, 1.5),

ONE_HEXENE_PN = uniform(0.142, 0.337),
ONE_HEXENE_kOH = lognormal(3.66114e-11, 1.2),
ONE_HEXENE_kO3 = lognormal(1.21495e-17, 1.5),

TWO_C5_OLE_PN = uniform (0.0416, 0.219),
TWO_C5_OLE_kOH = lognormal(6.55686e-11, 1.2),
TWO_C5_OLE_kO3 = lognormal(2.67891e-16, 1.5),

ONE_C7_OLE_PN = uniform(0.1098, 0.526),

ONE_C8_OLE_PN = uniform(0.141, 0.596),

ONE_C9_OLE_PN = uniform(0.158, 0.628),

ONEC10_OLE_PN = uniform(0.170, 0.648),

NEC11_OLE_PN = uniform(0.179, 0.662),

TWO_C6_OLE_PN = uniform(0.0882, 0.466),

TWO_C7_OLE_PN = uniform(0.1098, 0.526),

TWO_C8_OLE_PN = uniform(0.141, 0.596),
```

```
#inputs for O3 and O rxn generation

        O3U1RH = uniform(0.08, 0.18),
        O3U2K4 = uniform(0.187, 0.42),
        O3U2A1 = uniform(0.2, 0.45),
        O3U2RR = uniform(0.18, 0.41),
        O3U2RH = uniform(0.14, 0.32),
        O3U2OH = uniform(0.08, 0.18),
        O3U3K4 = uniform(0.28, 0.63),
        O3U4OH = uniform(0.077, 0.43),
        O3U5A1f = uniform(0.25 , 0.75),
        O3U5R2p = uniform(0.04, 0.25),
        OAUA3 = uniform (0.4, 0.6),
        OAUK4 = uniform (0.0 , 1.0),
        OAURH = uniform (0.2, 0.4);

Uncertain outputs
         O3, NO, NO2, HO;

Do Collocation With
  model: oletest,
  Dynamic: starting = 0.0, step = 1.67, interval = 6,   # 7 outputs
  approximation: third,
  sampling file : sampling.pnt,
  number of points : 5000,
  options: nonlinkable, saveall(savetest.all), no header;


# end ole.dmc
```

B.5.6 Conc.dmc (Organics only)

```
# file name: conc.dmc
# date: Apr 29, 1996
# description: Demmucom input script for initial concentrations (pre-lump)
# A 20% uncertainty is assumed for each species

Uncertain inputs

# alkane species
ETHANE      = lognormal(0.020898, 1.2),
PROPANE     = lognormal(0.015588, 1.2),
TW_ME_C3    = lognormal(0.0069685, 1.2),
N_C4        = lognormal(0.0165335, 1.2),
TW_ME_C4    = lognormal(0.0142225, 1.2),
N_C5        = lognormal(0.0068745, 1.2),
TW2_DM_C4   = lognormal(0.000390, 1.2),
CYCC5       = lognormal(0.000640, 1.2),
TW3_DM_C4   = lognormal(0.001017, 1.2),
TW_ME_C5    = lognormal(0.00475, 1.2),
TH_ME_C5    = lognormal(0.002762, 1.2),
N_C6        = lognormal(0.002972, 1.2),
TH_ME_C6    = lognormal(0.005220, 1.2),
CYCC6       = lognormal(0.002531, 1.2),
FO_ME_C7    = lognormal(0.004937, 1.2),
N_C7        = lognormal(0.001131, 1.2),
ME_CYCC6    = lognormal(0.000845, 1.2),
FO_ET_C7    = lognormal(0.001945, 1.2),
N_C8        = lognormal(0.000599, 1.2),
ET_CYCC6    = lognormal(0.000097, 1.2),
N_C9        = lognormal(0.000461, 1.2),
FO_PR_C7    = lognormal(0.001395, 1.2),
N_C10       = lognormal(0.000636, 1.2),
TH5_DE_C7   = lognormal(0.000558, 1.2),
TH6_DE_C8   = lognormal(0.000129, 1.2),
TH7_DE_C9   = lognormal(0.000003, 1.2),
ACETYLEN    = lognormal(0.011496, 1.2),
```

```
# aromatic species
BENZENE     = lognormal(0.003460, 1.2),
C2_BENZ     = lognormal(0.001241, 1.2),
P_XYLENE    = lognormal(0.001947, 1.2),
O_XYLENE    = lognormal(0.001547, 1.2),
I_CTH_BEN   = lognormal(0.000144, 1.2),
N_C3_BEN    = lognormal(0.000288, 1.2),
ONE23_TMB   = lognormal(0.002911, 1.2),
TOLUENE     = lognormal(0.009954, 1.2),
M_XYLENE    = lognormal(0.006793, 1.2),
ONE35_TMB   = lognormal(0.002123, 1.2),

# olefin species
PROPENE     = lognormal(0.004964, 1.2),
ONE_BUTENE  = lognormal(0.00425, 1.2),
ONE3_BUTDE  = lognormal(0.000671, 1.2),
T_2_BUTE    = lognormal(0.001028, 1.2),
C_2_BUTE    = lognormal(0.000801, 1.2),
ONE_PENTEN  = lognormal(0.00164, 1.2),
TW_M_1_BUT  = lognormal(0.000827, 1.2),
ISOPRENE    = lognormal(0.000680, 1.2),
TW_C5_OLE   = lognormal(0.002008, 1.2),
TW_M_2_BUT  = lognormal(0.000112, 1.2),
TW_C6_OLE   = lognormal(0.000943, 1.2),
ONE_HEXENE  = lognormal(0.00095, 1.2),
ONE_C7_OLE  = lognormal(0.001156, 1.2),
ONE_C8_OLE  = lognormal(0.000400, 1.2),
TW_C7_OLE   = lognormal(0.000101, 1.2),
TH_C8_OLE   = lognormal(0.000023, 1.2),
ONE_C9_OLE  = lognormal(0.000716, 1.2),
A_PINENE    = lognormal(0.000611, 1.2),
ONEC10_OLE  = lognormal(0.000032, 1.2),
ONEC11_OLE  = lognormal(0.000408, 1.2),

# explicit species
ETHENE      = lognormal(0.019293, 1.2),
HCHO        = lognormal(0.020, 1.2),
CCHO        = lognormal(0.015, 1.2),
INERT       = constant(0.008021);

Uncertain outputs
        O3, NO, NO2, HO;

Do Collocation With
  model: conc,
  Dynamic: starting = 0.0, step = 1.67, interval = 6,   # 7 outputs
  approximation: linear,
  sampling file : sampling.pnt,
  number of points : 5000,
  options: nonlinkable, no header;

# end conc.dmc
```

B.5.7 Allp.dmc

```
# file name: allp.dmc
# date: July 14, 1996
# description: Demmucom input script for all variance-contributing parameters
# as identified from uncertainty analysis of subgroups

Uncertain inputs
# inorganic parameters
        jNO2 = lognormal(0.00898167, 1.3),
        kO3pNO = lognormal(1.68008e-14, 1.2),
        jO3 = lognormal(3.78e-5, 1.4),
        kOSDpH2O = lognormal(2.2e-10, 1.26),
        kOSD = lognormal(2.92377e-11, 1.26),
        kHOpNO = lognormal(4.81886e-12,1.29),
        jHONO = lognormal(0.0016275, 1.6),
```

```
        kHOpNO2 = lognormal (1.132455e-11, 1.28),
        kHO2pCO = lognormal (2.400000e-13, 1.3),
# organic parameters
        kRCO3pNO = lognormal(9.92678e-12, 2),
        kRCO3pNO2 = lognormal(7.20102e-12, 2),
        jHCHOR = lognormal(3.00667e-05, 1.4),
        jHCHOM = lognormal(4.305e-05, 1.4),
        kCCHOpHO = lognormal(1.5651e-11, 1.41),
        jCCHO = lognormal(5.94e-6, 1.4),
        jRCHO = lognormal(2.14e-05 , 1.4),
        jMGLY = lognormal (0.00014845, 1.6),
        jAFG = lognormal (5.28667e-05, 3),
# organic composition
        tmb123 = lognormal(0.002911, 1.2),
        mxylene = lognormal(0.006793, 1.2),
        tmb135 = lognormal(0.002123, 1.2),
# aromatic parameters
        toluk  = lognormal(5.90627e-12, 1.3),
        mxyk = lognormal(2.36e-11, 1.3),
        mxyu2 = uniform (0.444, 0.999),              #nominal 0.666
        tmb13k = lognormal(5.75e-11, 1.3),
        tmb13u2 = uniform (0.4, 0.9),                #nominal 0.6
        tmb12k = lognormal(3.27e-11, 1.3),
# initial conditons
icethe=lognormal(2.4846e11, 1.2),
ichcho=lognormal(2.5756e11, 1.2),
icccho=lognormal(1.9317e+11, 1.2),
icaar1=lognormal(6.636e+11, 1.2),
icaar2=lognormal(7.2155e+11, 1.2),
icaar3=lognormal(2.435e11 , 1.2),
icole1= lognormal(1.8694e+11, 1.2),
icole2= lognormal(1.0051e+11, 1.2),
icno = lognormal (2.64814e12, 1.2),
icno2 = lognormal(8.9052e+11, 1.2);

Uncertain outputs
        O3, NO, NO2, HO;

Do Collocation With
   model: /usr/users/pun/cork/airchem/saprc/col_all/modeldir/mlsode,
   dynamic: starting = 0.0, step = 1.67, interval = 6,
   approximation: linear,
   sampling file : samp.pnt,
   number of points : 5000,
   options: nonlinkable, no header;


# end allp.dmc
```

20.3 Appendix C: RADM Mechanism and Supplement

In 1986, a chemical mechanism was created for the EPA/NCAR Regional Acid Deposition Model (RADM) by simplifying several state-of-the-art chemical mechanisms (Stockwell, 1986). Due to run-time memory constraints, the first generation RADM mechanism used only 25 species in explicit ordinary differential equations. Additional species that were judged to have lifetimes shorter than atmospheric transport time scales were included as local equilibrium species.

The second generation RADM mechanism was published in 1990 (Stockwell *et al.*, 1990) which included three classes of higher alkanes, two classes of higher alkenes, and isoprene. Furthermore, RADM2 also included more complete treatments for aromatic chemistry, ketones and dicarbonyl species, and peroxyl radical-peroxyl radical reactions. The RADM2 mechanism was independently reviewed against 550 smog chamber experiments (Carter and Lurmann, 1989). Key characteristics of the RADM2 mechanism, used in this study and referred to as "RADM" in this document, are listed in Table 20-6.

Number of Species	60
Number of Reactions	155
Inorganic Reactions	33
Inorganic Photolysis Reactions	9
Organic Reactions	106
Organic Photolysis Reactions	12
Primary Organic Species	CH4, ETH, 3 higher alkanes, OL2, 2 higher alkenes, ISO, TOL, XYL
Primary Organic Reactions	19
Lumping	Reactivity Weighted

Table 20-6. Key Characteristics of the RADM2 Mechanism.

Complete listings of the species and reactions in RADM are included in Appendix C.1 and C.2. Appendix C.3 contains details about the primary organic lumping procedure used in this study. Appendix C.4 presents the base case results of the RADM mechanism.

To improve RADM's ability to predict oxidant concentrations in the presence of biogenic organic compounds, an extension was developed to supplement RADM with reactions of isoprene, the most important of the biogenic hydrocarbon (Zimmerman and Poppe, 1996). The complete description of isoprene chemistry involved 39 new species and 153 additional reaction to the RADM mechanism. A condensed version was used in this study. Table 20-7 summarizes the key parameters in the condensed supplement. Sections C.6 and C.7 list respectively the isoprene-related species and reactions. The base case used to probe the isoprene mechanism is included in Appendix C.8.

Number of Additional Species	6
Reactions Modified in RADM2	ISO + HO, ISO + O3, ISO + NO3 (to form isoprene-specific products)
Number of Additional Reactions	16

Table 20-7. Key Characteristics of the Isoprene Supplement.

Web Address:
http://airsite.unc.edu/atmchemifu/stockwell.html

Reference:

Carter, W.P.L. and Lurmann, F.W., "Evaluation of the RADM gas-phase Chemical Mechanism," US EPA cooperative agreement C$-813448-01-0, Statewide Air Pollution Research Center, Univ. Of California, Riverside, 1989

Stockwell, W.R., "A Homogeneous Gas Phase Mechanism for Use in a Regional Acid Deposition Model," *Atmos. Environ.* **20**: 1615-1632 (1986)

Stockwell, W.R., Middleton, P., Chang, J.S., Tang, X, "The Second Generation Regional Acid Deposition Model Chemical Mechanism for Regional Air Quality Modeling," *J. Geophy. Res.* **95**: 16343-16367 (1990)

Zimmerman, J. and Poppe, D., "A Supplement for the RADM2 Chemical Mechanism: the Photooxidation of Isoprene," *Atmos. Environ.* **30**: 1225-1269 (1996)

C.1 RADM Species

Model Species	Description
Inorganic Species	
NO2	Nitrogen Dioxide
NO	Nitric Oxide
HONO	Nitrous Acid
NO3	Nitrogen Trioxide
N2O5	Nitrogen Pentoxide
HNO4	Pernitric Acid
HNO3	Nitric Acid
O3	Ozone
H2O2	Hydrogen Peroxide
SO2	Sulfur Dixoide
CO	Carbon Monoxide
Inorganic Radicals	
O3P	Ground State Oxygen Atom
O1D	Excited State Oxygen Atom
HO	Hydroxyl Radical
HO2	Hydroperoxyl Radical
Primary Organic Compounds	

CH4	Methane
ETH	Ethane
HC3	Alkanes with HO rate constant between 2.7e-13 and 3.4e-12
HC5	Alkanes with HO rate constant between 3.4e-12 and 6.8e-12
HC8	Alkanes with HO rate constant greater than 6.8e-12
OL2	Ethene
OLT	Terminal Alkenes
OLI	Internal Alkenes
ISO	Isoprene
TOL	Toluene and Less Reactive Aromatics
XYL	Xylene and More Reactive Aromatics
Organic Products	
CSL	Cresol and Other Hydroxy-substituted Aromatics
HCHO	Formaldehyde
ALD	Acetaldehyde and Higher Aldehydes
KET	Ketones
GLY	Glyoxal
MGLY	Methyl-glyoxal
DCB	Unsaturated Dicarbonyls
PAN	Peroxyacetyl Nitrate and Higher PAN's
TPAN	H(CO)CH=CHCO3NO2
ONIT	Organic Nitrate
OP1	Methyl Hydrogen Peroxide
OP2	Higher Organic Peroxides
PAA	Peroxyacetic Acid
ORA1	Formic Acid
ORA2	Acetic Acid and Higher Acids
Organic Radicals	
MO2	Methyl Peroxyl Radical
ETHP	Peroxyl Radical from ETH
HC3P	Peroxyl Radical from HC3
HC5P	Peroxyl Radical from HC5
HC8P	Peroxyl Radical from HC8
OL2P	Peroxyl Radical from OL2
OLTP	Peroxyl Radical from OLT
OLIP	Peroxyl Radical from OLI
TOLP	Peroxyl Radical from TOL
XYLP	Peroxyl Radical from XYL
ACO3	Acetyl Peroxyl Radical
KETP	Peroxyl Radical from KET
TCO3	H(CO)CH=CHCO3
OLN	NO3-Alkene adduct
XNO2	Operator for Organic Nitrate Formation by Lumped Organic Species
XO2	Operator for NO-to-NO2 Conversion by Lumped Organic Species

Table 20-8. List of Species in the RADM mechanism.

C.2 RADM Chemap Input (Reactions and Rates)

The following is the Chemap implementation of the RADM mechanism.

```
#
# RADM.map with 10 uncertain photolysis rates
# Input date: 10/26/96
# Units:  molecule/cm3 - s units
#

# Acknowledgment:
# adapted from RADM2.mech file from Jerry Gibson, US EPA
# model 3 development in National Exposure Research Laboratory.

# Reference: Stockwell, Middleton, Chang, Tang, JGR 95:16343 (1990)

# Notes:
# 1.  replaced all 7.70E-14  @ -1300.0; by k86
#                   4.20E-12  @ -180.0;  by k58
# 2.  changed default j values to zenith angle = 0
# 3.  Following are comments from RADM2.mech file (EPA version)
#     RADM2 Mechanism originally based on OZIPR version and modified
#     to be consistent with the RADM2 mechanism embedded in
#     version h26.f of HR-RADM.  Reaction labels below correspond
#     to the reaction numbers in the HR-RADM model. (Note: Reaction 40
#     in the original RADM mechanism is represented by reactions 40 and
#     40a below.) Jeff & Jerry 23 Aug 96



O3P + O2     --> O3
k=k1  ;k1  = 6.00E-34*(T/300)^-2.3 * M
# from reference.  model 3 file has M in rxn

O3P + NO2    --> NO
k=k2  ;k2  = 6.50E-12 * exp (-( -120.0/T))

O1D + N2     --> O3P
k=k3  ;k3  = 1.80E-11 * exp (-( -110.0/T))

O1D + O2     --> O3P
k=k4  ;k4  = 3.20E-11 * exp (-( -70.0/T))

O1D + H2O    --> 2.0 HO
k=k5  ;k5  = 2.20E-10

O3  +   NO   --> NO2
k=k6  ;k6  = 2.00E-12 * exp (-(1400.0/T))

O3  +   HO   --> HO2
k=k7  ;k7  = 1.60E-12 * exp (-( 940.0/T));

O3  +   HO2 --> HO
k=k8  ;k8  = 1.10E-14 * exp (-( 500.0/T))

HO2 +   NO   --> NO2 + HO
k=k9  ;k9  = 3.70E-12 * exp (-( -240.0/T))

HO2 +   NO2 --> HNO4
k=k10 ;k10 = (k10not* M/(1 + k10not* M / k10inf)) * 0.6^(1 / (1+(log10(k10not* M /
k10inf))^2))
k10not = 1.8E-31*(T/300)^-3.2
k10inf = 4.7E-12*(T/300)^-1.4

HNO4         --> HO2 + NO2
k=k11 ;k11 = 4.76E26 * exp (-(10900.0/T)) * k10
# taken from paper, consistent with Carter's SAPRC
# model 3 file 2.1E-27 @ -10900.0 'E< 10>;
# numbers are inverse of paper, E<10> is some equilibrium definition
```

```
HO2 + HO2    --> H2O2
k=k12 ;k12 = 2.20E-13 * exp (-(-620.0/T)) + 1.90E-33 * M * exp (-(-980.0/T))
# taken from reference paper
# model 3 file %3 # 2.20E-13@-620.0  & 1.90E-33@-980.0;

HO2 + HO2 + H2O --> H2O2
k=k13 ;k13 = 3.08E-34 * exp (-(-2820.0/T)) + 2.66E-54 * M * exp (-(-3180.0/T))
# taken from reference paper
# %3 # 3.08E-34@-2820.0 & 2.66E-54@-3180.0;

H2O2 + HO       --> HO2
k=k14 ;k14 = 3.30E-12 * exp (-( 200.0/T))

NO  +  HO    --> HONO
k=k15 ;k15 = (k15not* M/(1 + k15not* M / k15inf)) * 0.6^(1 / (1+(log10(k15not* M /
k15inf))^2))
k15not = 7.0E-31 * (T/300)^-2.6
k15inf = 1.5E-11 * (T/300)^-0.5
# 7.0E-31^-2.6 &1.5E-11^-0.5;

NO  +  NO  + O2 --> 2.0 NO2
k=k16 ;k16 = 3.30E-39 * exp (-( -530.0/T))

O3  +  NO2  --> NO3
k=k17 ;k17 = 1.40E-13 * exp (-( 2500.0/T))

NO3 +  NO   --> 2.0 NO2
k=k18 ;k18 = 1.70E-11 * exp (-( -150.0/T))

NO3 +  NO2  --> NO + NO2
k=k19 ;k19 = 2.50E-14 * exp (-( 1230.0/T))

NO3 +  HO2  --> HNO3
k=k20 ;k20 = 2.50E-12

NO3 +  NO2  --> N2O5
k=k21 ;k21 = (k21not* M/(1 + k21not* M / k21inf)) * 0.6^(1 / (1+(log10(k21not* M /
k21inf))^2))
k21not = 2.20E-30*(T/300)^-4.3
k21inf = 1.5E-12*(T/300)^-0.5

N2O5        --> NO2 + NO3
k=k22 ;k22 = 9.09E26 * exp (-(11200.0/T)) * k21
# from reference paper
# 1.10E-27 @ -11200.0 *E<21>; some equilibrium definition numbers are inverse

N2O5+  H2O  --> 2.0 HNO3
k=k23 ;k23 = 2.00E-21

HO  +  NO2  --> HNO3
k=k24 ;k24 = (k24not* M/(1 + k24not* M / k24inf)) * 0.6^(1 / (1+(log10(k24not* M /
k24inf))^2))
k24not = 2.60E-30*(T/300)^-3.2
k24inf = 2.4E-11*(T/300)^-1.3

HO  +  HNO3 --> NO3
k=k25 ;k25 = k25zero + k25three / (1 + k25three/k25two)
k25zero = 7.20E-15* exp (-(-785/T))
k25two = 4.1E-16* exp (-(-1440/T))
k25three = 1.9E-33* exp (-(-725/T)) * M
# from reference
# model 3 file %2 #7.2E-15@-785&4.1E-16@-1440&1.9E-33@-725;

HO  +  HNO4 --> NO2
k=k26 ;k26 = 1.30E-12 * exp (-( -380.0/T))

HO  +  HO2  -->
k=k27 ;k27 = 4.60E-11 * exp (-( -230.0/T))

HO  +  SO2  --> SULF + HO2
k=k28 ;k28 = (k28not* M/(1 + k28not* M / k28inf)) * 0.6^(1 / (1+(log10(k28not* M /
k28inf))^2))
```

```
k28not = 3.00E-31*(T/300)^-3.3
k28inf = 1.5E-12*(T/300)^0.0

CO  +  HO   --> HO2
k=k29 ;k29 = 1.50E-13 * (1. + 2.439E-20 * M)
# from ref.
# %1 #1.5000E-13;

CH4 + HO      --> MO2
k=k30 ;k30 = 6.95E-18 * exp (-( 1280.0/T)) * T^2
# ref: HO + CH4 --> MO2 + H2O; k = T^2 * 6.95E-18 exp(-1280/T)
# (X 300 SQUARED)        # 28.3^2   @ 1280.0;

ETH +  HO   --> ETHP
k=k31 ;k31 = 1.37E-17 * T^2 * exp (-( 444.0/T))
# ref: ETH +  HO   --> ETHP k = 1.37E-17 * exp (-444/T) * T^2
# (X 300 SQUARED)        # 1.233E-12^2 @ 444.0;

HC3 +  HO   --> 0.83 HC3P + 0.17 HO2 + 0.009 HCHO + 0.075 ALD + 0.025 KET
k=k32 ;k32 = 1.59E-11 * exp (-( 540.0/T))

HC5 +  HO   --> HC5P + 0.250 XO2
k=k33 ;k33 = 1.73E-11 * exp (-( 380.0/T))

HC8 +  HO   --> HC8P + 0.750 XO2
k=k34 ;k34 = 3.64E-11 * exp (-( 380.0/T))

OL2 +  HO   --> OL2P
k=k35 ;k35 = 2.15E-12 * exp (-( -411.0/T))

OLT +  HO   --> OLTP
k=k36 ;k36 = 5.32E-12 * exp (-( -504.0/T))

OLI +  HO   --> OLIP
k=k37 ;k37 = 1.07E-11 * exp (-( -549.0/T))

TOL +  HO   --> 0.75 TOLP + 0.250 CSL + 0.250 HO2
k=k38 ;k38 = 2.10E-12 * exp (-( -322.0/T))

XYL +  HO   --> 0.83 XYLP + 0.170 CSL + 0.170 HO2
k=k39 ;k39 = 1.89E-11 * exp (-( -116.0/T))

CSL +  HO   --> 0.10 HO2  + 0.900 XO2 + 0.900 TCO3
k=k40 ;k40 = 4.00E-11

CSL +  HO   --> CSL
k=k41 ;k41 = 0.9*k40

HCHO + HO   --> HO2 + CO
k=k42 ;k42 = 9.00E-12

ALD +  HO   --> ACO3
k=k43 ;k43 = 6.87E-12 * exp (-( -256.0/T))

KET +  HO   --> KETP
k=k44 ;k44 = 1.20E-11 * exp (-( 745.0/T))

GLY +  HO   --> HO2 +  2.000 CO
k=k45 ;k45 = 1.15E-11

MGLY + HO   --> ACO3 + CO
k=k46 ;k46 = 1.70E-11

DCB +  HO   --> TCO3
k=k47 ;k47 = 2.8E-11

OP1 +  HO   --> 0.5 MO2 + 0.500 HCHO +0.500 HO
k=k48 ;k48 = 1.00E-11

OP2 +  HO   --> 0.5 HC3P + 0.500 ALD + 0.500 HO
k=k49 ;k49 = 1.00E-11
```

```
PAA +  HO   --> ACO3
k=k50 ;k50 = 1.00E-11

PAN +  HC   --> HCHO + NO3 + XO2
k=k51 ; k51 = 0.5 * k31
# reference k = 6.85E-18 * T^2 * exp (-( 444.0/T)) = half of ethane
# model 3 file: (x 300 sq)     # 6.1650E-13^2  @ 444.0;

ONIT + HO   --> HC3P + NO2
k=k52 ;k52 = 1.55E-11 * exp (-( 540.0/T))

ISO +  HO   --> OLTP
k=k53 ;k53 = 2.55E-11 * exp (-( -409.0/T))

ACO3 + NO2  --> PAN
k=k54 ;k54 = 2.80E-12 * exp (-( -181.0/T))

PAN         --> ACO3 + NO2
k=k55 ;k55 = 1.95E+16 * exp (-( 13543.0/T))

TCO3 + NO2  --> TPAN
k=k56 ;k56 = 4.70E-12

TPAN        --> TCO3 + NO2
k=k57 ;k57 = 1.95E+16 * exp (-( 13543.0/T))

MO2 +  NO   --> HCHO + HO2 + NO2
k=k58 ;k58 = 4.20E-12 * exp (-( -180.0/T))

HC3P + NO   --> 0.75 ALD + 0.25 KET + 0.09 HCHO + 0.036 ONIT + 0.964 NO2 + 0.964 HO2
k=k59 ;k59 = k58

HC5P + NO   --> 0.38 ALD + 0.69 KET + 0.08 ONIT + 0.92 NO2 + 0.92 HO2
k=k60 ;k60 = k58

HC8P + NO   --> 0.35 ALD + 1.06 KET + 0.04 HCHO + 0.24 ONIT + 0.76 NO2 + 0.76 HO2
k=k61 ;k61 = k58

OL2P + NO   --> 1.60 HCHO + HO2 + NO2 + 0.20 ALD
k=k62 ;k62 = k58

OLTP + NO   --> ALD + HCHO + HO2 + NO2
k=k63 ;k63 = k58

OLIP + NO   --> HO2 + 1.45 ALD + 0.28 HCHO + 0.1 KET + NO2
k=k64 ;k64 = k58

ACO3 + NO   --> MO2 + NO2
k=k65 ;k65 = k58

TCO3 + NO   --> NO2 + 0.920 HO2 + 0.890 GLY + 0.110 MGLY + 0.050 ACO3 + 0.950 CO +
2.000 XO2
k=k66 ;k66 = k58

TOLP + NO   --> NO2 + HO2 + 0.17 MGLY + 0.16 GLY + 0.70 DCB
k=k67 ;k67 = k58

XYLP + NO   --> NO2 + HO2 + 0.45 MGLY + 0.806 DCB
k=k68 ;k68 = k58

ETHP + NO   --> ALD  + HO2 + NO2
k=k69 ;k69 = k58

KETP + NO   --> MGLY + NO2 + HO2
k=k70 ;k70 = k58

GLN  + NO   --> HCHO + ALD + 2.0 NO2
k=k71 ;k71 = k58

HCHO + NO3  --> HO2 + HNO3 + CO
k=k72 ;k72 = 6.00E-13 * exp (-( 2058.0/T))
```

```
ALD +  NO3  --> ACO3 + HNO3
k=k73 ;k73 = 1.40E-12 * exp (-( 1900.0/T))

GLY +  NO3  --> HNO3 + HO2 + 2.000 CO
k=k74 ;k74 = 6.00E-13 * exp (-( 2058.0/T))

MGLY + NO3  --> HNO3 + ACO3 + CO
k=k75 ;k75 = 1.40E-12 * exp (-( 1900.0/T))

DCB +  NO3  --> HNO3 + TCO3
k=k76 ;k76 = 1.40E-12 * exp (-( 1900.0/T))

CSL +  NO3  --> HNO3 + XNO2 + 0.500 CSL
k=k77 ;k77 = 2.20E-11

OL2 +  NO3  --> OLN
k=k78 ;k78 = 2.00E-12 * exp (-( 2923.0/T))

OLT +  NO3  --> OLN
k=k79 ;k79 = 1.00E-11 * exp (-( 1895.0/T))

OLI +  NO3  --> OLN
k=k80 ;k80 = 3.23E-11 * exp (-( 975.0/T))

ISO +  NO3  --> OLN
k=k81 ;k81 = 5.81E-13

OL2 +  O3   --> HCHO + 0.400 ORA1 + 0.420 CO + 0.120 HO2
k=k82 ;k82 = 1.20E-14 * exp (-( 2633.0/T))

OLT +  O3   --> 0.53 HCHO + 0.50 ALD + 0.33 CO + 0.20 ORA1 + 0.20 ORA2 + 0.23 HO2 +
0.22 MO2 + 0.10 HO
k=k83 ;k83 = 1.32E-14 * exp (-( 2105.0/T))

OLI +  O3   --> 0.18 HCHO + 0.72 ALD + 0.10 KET + 0.23 CO + 0.06 ORA1 + 0.29 ORA2 +
0.26 HO2 + 0.14 HO + 0.31 MO2
k=k84 ;k84 = 7.29E-15 * exp (-( 1136.0/T))

ISO +  O3   --> 0.53 HCHO + 0.500 ALD + 0.33 CO + 0.20 ORA1 + 0.20 ORA2 + 0.23 HO2 +
0.22 MO2 + 0.10 HO
k=k85 ;k85 = 1.23E-14 * exp (-( 2013.0/T))

HO2 +  MO2  --> OP1
k=k86 ;k86 = 7.70E-14 * exp (-( -1300.0/T))

HO2 +  ETHP --> OP2
k=k87 ;k87 = k86  # 7.70E-14 * exp (-( -1300.0/T))

HO2 +  HC3P --> OP2
k=k88 ;k88 = k86

HO2 +  HC5P --> OP2
k=k89 ;k89 = k86

HO2 +  HC8P --> OP2
k=k90 ;k90 = k86

HO2 +  OL2P --> OP2
k=k91 ;k91 = k86

HO2 +  OLTP --> OP2
k=k92 ;k92 = k86

HO2 +  OLIP --> OP2
k=k93 ;k93 = k86

HO2 +  KETP --> OP2
k=k94 ;k94 = k86

HO2 +  ACO3 --> PAA
k=k95 ;k95 = k86
```

```
HO2 +  TOLP --> OP2
k=k96 ;k96 = k86

HO2 +  XYLP --> OP2
k=k97 ;k97 = k86

HO2 +  TCO3 --> OP2
k=k98 ;k98 = 7.70E-14 * exp (-( -1300.0/T))

HO2 +  OLN  --> ONIT
k=k99 ;k99 = 7.70E-14 * exp (-( -1300.0/T))

MO2 +  MO2  --> 1.5 HCHO + HO2
k=k100;k100= 1.90E-13 * exp (-( -220.0/T))

MO2 +  ETHP --> 0.75 HCHO + HO2 + 0.75 ALD
k=k101;k101= 1.40E-13 * exp (-( -220.0/T))

MO2 +  HC3P --> 0.84 HCHO + 0.770 ALD + 0.260 KET + 1.00 HO2
k=k102;k102= 4.20E-14 * exp (-( -220.0/T))

MO2 +  HC5P --> 0.77 HCHO + 0.41 ALD + 0.75 KET + 1.00 HO2
k=k103;k103= 3.40E-14 * exp (-( -220.0/T))

MO2 +  HC8P --> 0.80 HCHO + 0.46 ALD + 1.39 KET + 1.00 HO2
k=k104;k104= 2.90E-14 * exp (-( -220.0/T))

MO2 +  OL2P --> 1.55 HCHO + 0.350 ALD + HO2
k=k105;k105= 1.40E-13 * exp (-( -220.0/T))

MO2 +  OLTP --> 1.25 HCHO + 0.750 ALD + HO2
k=k106;k106= 1.40E-13 * exp (-( -220.0/T))

MO2 +  OLIP --> 0.89 HCHO + 0.725 ALD + HO2 + 0.55 KET
k=k107;k107= 1.70E-14 * exp (-( -220.0/T))

MO2 +  KETP --> 0.75 HCHO + 0.750 MGLY + HO2
k=k108;k108= 1.70E-14 * exp (-( -220.0/T))

MO2 +  ACO3 --> HCHO + 0.5 HO2 + 0.5 MO2 + 0.50 ORA2
k=k109;k109= 9.60E-13 * exp (-( -220.0/T))

MO2 +  TOLP --> HCHO + 0.17 MGLY + 0.16 GLY + 0.70 DCB + 2.0 HO2
k=k110;k110= 1.70E-14 * exp (-( -220.0/T))

MO2 +  XYLP --> HCHO + 0.45 MGLY + 0.806 DCB + 2.000 HO2
k=k111;k111= 1.70E-14 * exp (-( -220.0/T))

MO2 +  TCO3 --> 0.50 HCHO + 0.445 GLY + 0.055 MGLY + 0.50 ORA2 + 0.025 ACO3 + 0.460
HO2 + 0.475 CO + XO2
k=k112;k112= 9.60E-13 * exp (-( -220.0/T))

MO2 +  OLN  --> 1.75 HCHO + 0.50 HO2 + ALD + NO2
k=k113;k113= 1.70E-14 * exp (-( -220.0/T))

ETHP + ACO3 --> ALD + 0.5 HO2 + 0.5 MO2 + 0.5 ORA2
k=k114;k114= 3.40E-13 * exp (-( -220.0/T))

HC3P + ACO3 --> 0.77 ALD + 0.26 KET + 0.5 HO2 + 0.50 MO2 + 0.5 ORA2
k=k115;k115= 1.00E-13 * exp (-( -220.0/T))

HC5P + ACO3 --> 0.41 ALD + 0.75 KET + 0.5 HO2 + 0.50 MO2 + 0.5 ORA2
k=k116;k116= 8.40E-14 * exp (-( -220.0/T))

HC8P + ACO3 --> 0.46 ALD + 1.39 KET + 0.5 HO2 + 0.50 MO2 + 0.5 ORA2
k=k117;k117= 7.20E-14 * exp (-( -220.0/T))

OL2P + ACO3 --> 0.80 HCHO + 0.6 ALD + 0.5 HO2 + 0.50 MO2 + 0.5 ORA2
k=k118;k118= 3.40E-13 * exp (-( -220.0/T))

OLTP + ACO3 --> ALD + 0.5 HCHO + 0.5 HO2 + 0.5 MO2 + 0.5 ORA2
k=k119;k119= 3.40E-13 * exp (-( -220.0/T))
```

```
OLIP + ACO3 --> 0.725 ALD + 0.55 KET + 0.14 HCHO + 0.5 HO2 + 0.50 MO2 + 0.5 ORA2
k=k120;k120= 4.20E-14 * exp (-( -220.0/T))

KETP + ACO3 --> MGLY + 0.5 HO2 +  0.5 MO2 + 0.5 ORA2
k=k121;k121= 4.20E-14 * exp (-( -220.0/T))

ACO3 + ACO3 --> 2.0 MO2
k=k122;k122= 1.19E-12 * exp (-( -220.0/T))

ACO3 + TOLP --> MO2 + 0.170 MGLY + 0.16 GLY + 0.70 DCB + HO2
k=k123;k123= 4.20E-14 * exp (-( -220.0/T))

ACO3 + XYLP --> MO2 + 0.45 MGLY + 0.806 DCB + HO2
k=k124;k124= 4.20E-14 * exp (-( -220.0/T))

ACO3 + TCO3 --> MO2 + 0.92 HO2 +  0.89 GLY +  0.11 MGLY + 0.05 ACO3 + 0.95 CO + 2.0
XO2
k=k125;k125= 1.19E-12 * exp (-( -220.0/T))

ACO3 + OLN  --> HCHO + ALD + 0.5 ORA2 + NO2 + 0.5 MO2
k=k126;k126= 4.20E-14 * exp (-( -220.0/T))

OLN +  OLN  --> 2.0 HCHO + 2.0 ALD +  2.0 NO2
k=k127;k127= 3.60E-16 * exp (-( -220.0/T))

XO2 +  HO2  --> OP2
k=k128;k128= k86

XO2 +  MO2  --> HCHO + HO2
k=k129;k129= 1.70E-14 * exp (-( -220.0/T))

XO2 +  ACO3 --> MO2
k=k130;k130= 4.20E-14 * exp (-( -220.0/T))

XO2 +  XO2  -->
k=k131;k131= 3.60E-16 * exp (-( -220.0/T))

XO2 +  NO   --> NO2
k=k132;k132= k58

XNO2 + NO2  --> ONIT
k=k133;k133= k58

XNO2 + HO2  --> OP2
k=k134;k134= k86

XNO2 + MO2  --> HCHO + HO2
k=k135;k135= 1.70E-14 * exp (-( -220.0/T))

XNO2 + ACO3 --> MO2
k=k136;k136= 4.20E-14 * exp (-( -220.0/T))

XNO2 + XNO2 -->
k=k137;k137= 3.60E-16 * exp (-( -220.0/T))

NO2   + HV    --> O3P + NO
k=k138; k138 = uk1
#k138 = 7.72E-3 (zen. angle = 0; jNO2 = 0.00898167)

O3    + HV    --> O1D
k=k139; k139 = uk2
#k139= 1.70E-5 (zen. angle = 0; jO3O1D = 3.78e-05)

O3    + HV    --> O3P
k=k140; k140 = uk3
#k140= 4.22E-4 (zen. angle = 0; jO3O3P = 0.000499333)

HONO  + HV    --> HO + NO
k=k141;k141 = uk4
#k141= 1.39E-3 (zen. angle = 0; jHONO = 0.0016275)
```

```
HNO3   + HV    --> HO + NO2
k=k142; k142=4.67e-7 #OZIPR value
#k142= 2.75E-7

HNO4   + HV     --> HO2 + NO2
k=k143;k143=6.85e-6   # OZIPR value
#k143= 4.58E-6

NO3    + HV     --> NO
k=k144;k144 = uk5
#k144= 2.02E-2 (zen. angle = 0; jNO3NO = 0.0187)

NO3    + HV     --> NO2 + O3P
k=k145;k145 = uk6
#k145= 1.64E-1 (zen angle = 0; jNO3NO2 = 0.169333)

H2O2   + HV     --> 2.0 HO
k=k146;k146=uk7
#k146= 5.67E-6 (zen angle = 0; jH2O2 = 7.53e-06)

HCHO   + HV     --> CO
k=k147;k147= uk8
#k147= 3.32E-5 (zen angle = 0; jHCHOM = 4.305e-05)

HCHO   + HV     --> HO2 + HO2 + CO
k=k148;k148 = uk9
#k148= 2.10E-5 (zen angle = 0; jHCHOR = 3.00667e-05)

ALD    + HV     --> MO2 + HO2 + CO
k=k149;k149 = uk10
#k149= 3.55E-6 (zen angle = 0; jALD(CCHO) = 5.94e-06)

OP1    + HV    --> HCHO + HO2 + HO
k=k150;k150=7.39e-6 # same as SAPRC for ROOH
#k150= 5.57E-6

OP2    + HV    --> ALD + HO2 + HO
k=k151;k151= k150

PAA    + HV     --> MO2 + HO
k=k152; k152=2.18e-6 # use OZIPR value
#k152= 1.59E-6

KET    + HV     --> ACO3 + ETHP
k=k153;k153=9.62e-7   # use OZIPR value
#k153= 6.07E-7

GLY    + HV     --> 0.13 HCHO + 1.870 CO
k=k154; k154 = 6.44167e-5    #same as SAPRC
#k154= 4.98E-5

GLY    + HV     --> 0.45 HCHO + 1.550 CO + 0.800 HO2
k=k155;k155 = 4.62e-5 #same as SAPRC
#k155= 5.17E-5

MGLY   + HV     --> ACO3 + HO2 + CO
k=k156;k156=1.4845e-4        #same as SAPRC
#k156= 1.44E-4 (zen ang - 40)

DCB    + HV     --> 0.98 HO2 + 0.020 ACO3 + TCO3
k=k157;k157=5.28667e-4       #same as SAPRC
#k157= 4.30E-4 (zen ang - 40)

ONIT + HV     --> 0.20 ALD + 0.800 KET + HO2 + NO2
k=k158; k158=1.27e-6 #using OZIPR value
#k158 = 7.95E-7 (zen ang - 40)

T = 300
M = 2.46e19
CH4 = 2.2 * 2.46E13
# Airshed used 2.2 ppm
# clean troposphere (Seinfeld) 1.65 ppm
```

```
HV = 1
O2 = 0.2 * 2.46e19
N2 = 0.8 * 2.46e19
H2O = 4.92e17

uncertain uk1 uk2 uk3 uk4 uk5 uk6 uk7 uk8 uk9 uk10

IC NO2 = 8.9052e11
IC NO = 2.64814e12

IC CO = 4.9938e13

IC HCHO = 2.5756e11
IC ALD = 1.9317e11

IC ETH = 2.69e11
IC HC3 = 4.51e11
IC HC5 = 4.09e11
IC HC8 = 2.51e11
IC OL2 = 2.48416e11
IC OLT = 1.91e11
IC OLI = 9.66e10
IC TOL = 1.63e11
IC XYL = 1.97e11


# end of RAMD.map
```

C.3 RADM Lumping (IC)

A two-step lumping procedure was employed in this study. The first was to group all organic compounds into a 32-category system, listed in Table 20-9, on a mole-to-mole basis. These 32 categories were then allocated to the RADM model species using the reactivity lumping method described in the text (Section 9.2). An OH rate constant was assumed for each category and the reactivity factor are shown in Table 20-10.

Category	Description	Average Carbon No.
1	Methane	1
2	Ethane	2
3	Propane	3
4	Alkanes (2.5e3 - 5.0e3 ppm^{-1} min^{-1})	4.02
5	Alk. nes (5.0e3 - 1.0e4 ppm^{-1} min^{-1})	5.52
6	Alkanes (1.0e4 - 2.0e4 ppm^{-1} min^{-1})	7.74
7	Alkanes (>2e4 ppm^{-1} min^{-1})	12.38
8	Alkane / Aromatic Mix	10.00
9	Ethene	2
10	Propene	3
11	Alkene (Primary)	4.74
12	Alkene (Internal)	4.85
13	Alkene (Prim / Intern)	5.89
14	Benzene	6.00
15	Aromatic (<2 Reactivity)	7.26
16	Aromatic (>2 Reactivity)	8.88
17	Phenol and Cresol	6.60
18	Styrenes	6.96
19	Formaldehyde	1
20	Higher aldehyde	2.44
21	Acetone	3.00
22	Higher Ketones	4.8
23	Organic Acids	2.29
24	Acetylene	2
25	Haloalkanes	2.00
26	Unreactive	1.45
27	Others (< 0.25 Reactivity)	1.88
28	Others (0.25 - 0.5 Reactivity)	2.10
29	Others (0.5 - 1.0 Reactivity)	3.43
30	Others (> 1.0 Reactivity)	2.88
31	Unidentified	
32	Uassigned	

Table 20-9. Emission Categories

RADM Class	Emission Category	HO Rate Constant (1000/ppm-min)		Reactivity Factor
		Category	RADM	
CH4	Methane	0.001	0.001	1.0
ETH	Ethane	0.041	0.041	1.0
HC3	Propane	0.18	0.386	0.519
	Alkane (0.25 - 0.5)	0.369		0.964
	Acetylene	0.115		0.343
	Haloalkanes	0.025		0.078
	Others (< 0.25)	0.137		0.404
	Others (0.25 - 0.5)	0.496		1.215
HC5	Alkanes (0.5 - 1.0)	0.668	0.715	0.956
	Others (0.5 - 1.0)	0.801		1.075
HC8	Alkanes (1.0 - 2.0)	1.315	1.507	0.945
	Alkanes (> 2.0)	2.339		1.141
	0.91 * Alkane / Aromatic Mix	2.018		1.101
	Others (>1.0 React)	1.552		1.011
OL2	Ethene	1.242	1.242	1.0
OLT	Propene	3.187	4.190	1.0
	Alkenes (primary)	4.714		1.0
	0.5 * Alkene Prim. / Int. Mix	5.11		1.0
	Styrene	4.99		1.0
OLI	Alkenes (internal)	9.881	9.82	1.0
	0.5 * Alkene Prim. / Int. Mix	9.11		1.0
TOL	Benzene	0.185	0.908	0.293
	Aromatics (< 2 react)	0.919		1.0
	Styrene	0.902		1.0
XYL	Aromatics (> 2 react)	4.086	4.090	1.0
	0.09 * Alkane / Aromatic Mix	3.73		1.0
CSL	Phenols and Cresols	5.62	5.62	1.0
HCHO	Formaldehyde	1.33	1.33	1.0
ALD	Higher Aldehydes	2.39	2.39	1.0
KET	Acetone	0.035	0.147	0.253
	Higher Ketones	0.268		1.0
ORA2	Organic Acids	0.088	0.088	1.0

Table 20-10. Allocation of Emission Categories to RADM Model Classes.

Reference:

Middleton, P, Stockwell, W.R., and Carter, W.P.L., "Aggregation and Analysis of Volatile Organic compound Emissions for Regional Modeling," *Atmos. Environ.* 24A:1107-1133 (1990)

C.4 Base Case Simulation

The results of the base case simulation using the RADM mechanism summarized in Table 20-11 and are shown in Figures 20-6 through 20-9. These are very similar to the results of SAPRC (Appendix B).

Time	O3 (ppm)	NO (ppm)	NO2 (ppm)	HCHO (ppb)	ALD (ppb)	HO (ppt)	HO2 (ppt)	PAN (ppb)	HNO3 (ppb)
2	0.01855	0.061	0.063	12.89	13.17	0.096	0.397	0.571	9.418
4	0.0399	0.034	0.076	13.038	13.59	0.0798	0.554	1.297	22.55
6	0.0640	0.021	0.075	11.97	12.64	0.082	0.813	1.964	35.99
8	0.0897	0.014	0.068	10.79	11.31	0.095	1.249	2.694	50.08
10	0.1176	0.0087	0.058	9.741	9.755	0.120	2.074	3.586	64.99
Time	ETH (ppb)	HC3 (ppb)	HC5 (ppb)	HC8 (ppb)	TOL (ppb)	XYL (ppb)	OL2 (ppb)	OLT (ppb)	OLI (ppb)
2	10.88	17.56	15.34	8.620	5.989	5.068	8.754	4.743	0.942
4	10.84	16.876	14.25	7.377	5.456	3.323	7.622	2.881	0.136
6	10.80	16.26	13.30	6.380	5.002	2.241	6.639	1.711	0.010
8	10.75	15.60	12.33	5.438	4.545	1.452	5.664	0.921	2.9E-04
10	10.69	14.85	11.24	4.4798	4.047	0.858	4.654	0.425	2.7E-06

Table 20-11. RADM Base Case Time Series of Key Species.

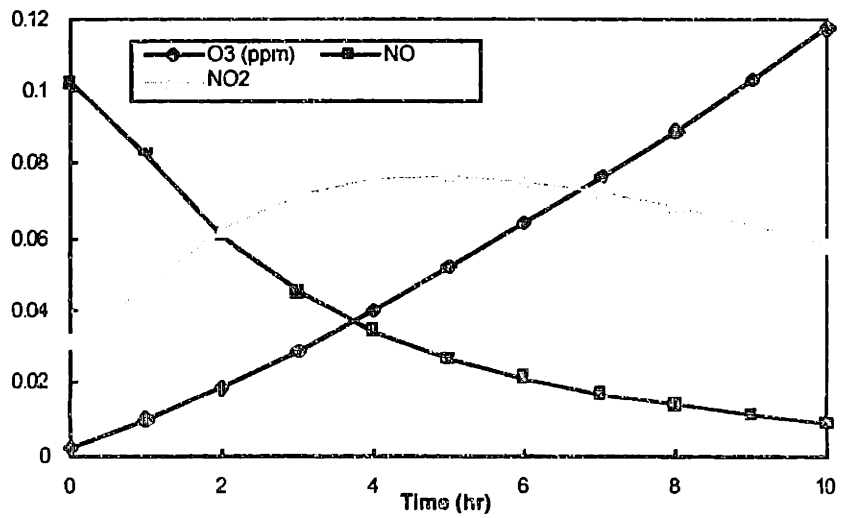
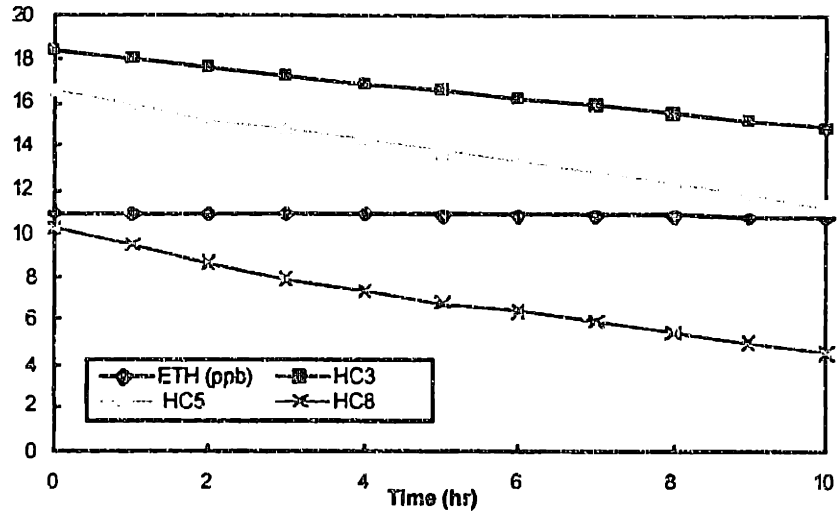


Figure 20-6. RADM Base Case Simulation for Ozone, NO, and NO_2.



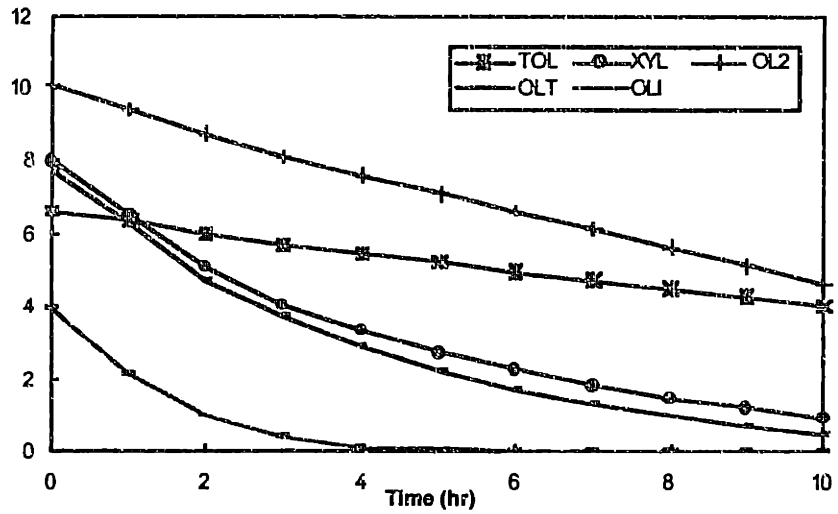


Figure 20-7. RADM Base Case Simulation (a) Alkanes, (b) Aromatic and Olefins

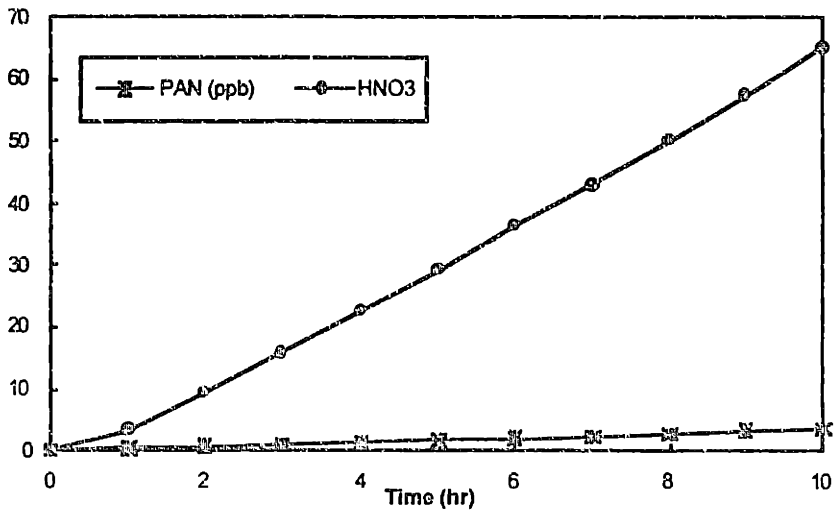


Figure 20-8. RADM Base Case Simulation for Sink Species.

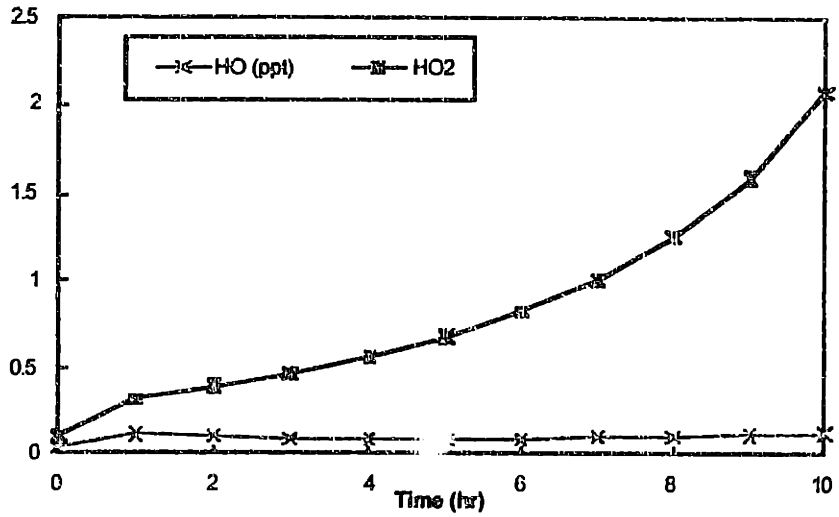


Figure 20-9. RADM Base Case Simulation for Key Radicals.

C.5 Demmucom Input Scripts for Uncertainty Analysis of RADM Mechanism

Case studies were performed to analyze parametric uncertainties of the RADM mechanism. The Demmucom input scripts are included here as a record of the uncertain inputs for each modeling experiment and the associated uncertainties. Appendix C.5.1 is the input for the photolysis rate constant case study, Appendix C.5.2 contains the input file for the primary organic case study, and an input script of all important parameters is included in Appendix C.5.3.

C.5.1 phk.dmc

```
# file name: phk.dmc
# input date:  Jan 17, 1997
# description: Demmucom input script for photolysis case study of the RADM mechanism
# uncertainty compilation: same as SAPRC + a few described in text (Chapter)

Uncertain inputs

  jNO2       = lognormal (0.00898167 , 1.3),
  jO3O1D     = lognormal (3.78e-05   , 1.4),
  jO3O3P     = lognormal (4.99333e-4 , 1.1),
  jHONO      = lognormal (1.6275e-3  , 1.6),
  jHNO3      = lognormal (4.67e-7    , 1.2),   #my UF
  jHNO4      = lognormal (6.85e-6    , 1.6),   #my UF
  jNO3NO     = lognormal (0.0187     , 2.0),
  jNO3NO2    = lognormal (0.169333   , 2.0),
  jH2O2      = lognormal (7.53e-06   , 1.4),
  jHCHOM     = lognormal (4.305e-05  , 1.4),
  jHCHOR     = lognormal (3.00667e-05, 1.4),
  jALD       = lognormal (5.68e-06   , 1.4),   #UF 1.4 for CCHO & RCHO in SAORC
  jOP        = lognormal (7.39e-6    , 1.8),   #equivalent to ROOH in SAPRC
  jPAA       = lognormal (2.18e-6    , 2.1),   #my UF
  jKET       = lognormal (9.62e-7    , 1.5),   #UF 1.5 for MEK & ACET in SAPRC
  jGLYM      = lognormal (6.44167e-05, 1.5),
  jGLYMR     = lognormal (4.62e-5    , 1.5),
```

```
   jMGLY      = lognormal (0.00014845 , 1.6),
   jDCB       = lognormal (5.28667e-04, 3.0),  #equivalent to AFG in SAPRC
   jONIT     = lognormal (1.27e-6     , 1.3);  #my UF

Uncertain outputs
  O3, NO, NO2, HO;

Do Collocation With
   model: /usr/users/pun/cork/airchem/RADM/col_phk/mlsode,
   dynamic: starting = 0.0, step = 1.0, interval = 10,
   approximation: linear,
   sampling file : samp.pnt,
   number of points : 5000,
   options: no header, looping;

# end phk.dmc
```

C.5.2 primorg.dmc

```
# file name primorg.dmc
# input date: March 18, 1997
# description: primary organic reactions uncertainty analysis.
#       (1) rate of initial attack of organic by HO, O3, NO3
#       (2) product coefficients of second generation radicals and carbonyl products
# uncertainty compilation:
#    rate similar to SAPRC study,  product representative of individual species in
class

Uncertain inputs

kETHHO    = lognormal (2.8068e-13, 1.2),
kHC3HO    = lognormal (2.6283e-12, 1.2),
kHC5HO    = lognormal (4.8746e-12, 1.3),
kHC8HO    = lognormal (1.0256e-11, 1.4),
kOL2HO    = lognormal (8.461e-12 , 1.15),
kOLTHO    = lognormal (2.8545e-11, 1.2),
kOLIHO    = lognormal (6.6703e-11, 1.2),
kTOLHO    = lognormal (6.1427e-12, 1.3),
kXYLHO    = lognormal (2.7822e-11, 1.3),
kOL2NO3   = lognormal (1.1737e-16, 2.0),
kOLTNO3   = lognormal (1.8060e-14, 2.0),
kOLINO3   = lognormal (1.2524e-12, 2.0),
kOL2O3    = lognormal (1.8515e-18, 1.25),
kOLTO3    = lognormal (1.1838e-17, 1.5),
kOLIO3    = lognormal (1.6527e-16, 1.5),

# HC3 + HO rxn
  hc3p32 =  uniform (0.765, 0.880), # nominal value 0.83
  hcho32 =  uniform (0.0072, 0.0113),       # nominal 0.009
  ald32  =  uniform (0.06, 0.09375),        # nominal 0.075
  ket32  =  uniform (0.02, 0.03125),        # nominal 0.025
# HC5 + HO rxn
  xo233  =  uniform (0.143, 0.438), # nominal 0.25
# HC8 + HO rxn
  xo234  =  uniform (0.3, 1.88),    # nominal 0.75
# TOL + HO rxn
  tolp38 =  uniform (0.632, 0.84),  # nominal 0.75
# XYL + HO rxn
  xylp39 =  uniform (0.736, 0.895),         # nominal 0.83
# CSL + HO rxn
  ho240  =  uniform (0.0597, 0.1628),       # nominal 0.1
# HC3P + NO rxn
  onit59 =  uniform (0.0243, 0.053),        # nominal 0.036
  ald59  =  uniform (0.6, 0.94),    # nominal 0.75
  ket59  =  uniform (0.2, 0.31),    # nominal 0.25
  hcho59 =  uniform (0.072, 0.1125),        # nominal 0.09
# HC5P + NO rxn
  onit60 =  uniform (0.0336, 0.1786),       # nominal 0.08
  ald60  =  uniform (0.19, 0.76),   # nominal 0.38
  ket60  =  uniform (0.345, 1.38),  # nominal 0.69
# HC8P + NO rxn
```

```
  onit61 =  uniform (0.095, 0.4865),       # nominal 0.24
  ald61  =  uniform (0.14, 0.875),  # nominal 0.35
  ket61  =  uniform (0.424, 2.65),  # nominal 1.06
  hcho61 =  uniform (0.016, 0.1),   # nominal 0.04
# OL2P + NO rxn
  hcho62 =  uniform (1.524, 1.667), # nominal 1.6
# OLTP + NO rxn
  ald63  =  uniform (0.4, 2.5),     # nominal 1.0
  hcho63 =  uniform (0.4, 2.5),     # nominal 1.0
# OLIP + NO rxn
  ald64  =  uniform (0.58, 3.625),  # nominal 1.45
  hcho64 =  uniform (0.112, 0.7),   # nominal 0.28
  ket64  =  uniform (0.04, 0.25),   # nominal 0.1
# TOLP + NO rxn
  ho267  =  uniform (0.667, 1.5),   # nominal 1.0
  mgly67 =  uniform (0.113, 0.255), # nominal 0.17
  gly67  =  uniform (0.1067, 0.24), # nominal 0.16
  dcb67  =  uniform (0.467, 1.05),  # nominal 0.7
# XYLP + NO rxn
  ho268  =  uniform (0.67, 1.5),    # nominal 1.0
  mgly68 =  uniform (0.3, 0.675),   # nominal 0.45
  dcb68  =  uniform (0.537, 1.209), # nominal 0.806
# OL2 + O3 rxn
  ora182 =  uniform (0.3636, 0.44), # nominal 0.4
  ho282  =  uniform (0.109, 0.132),        # nominal 0.12
# OLT + O3 rxn
  hcho83 =  uniform (0.353, 0.795),        # nominal 0.53
  ald83  =  uniform (0.333, 0.75),  # nominal 0.5
  or183  =  uniform (0.133, 0.3),   # nominal 0.2
  or283  =  uniform (0.133, 0.3),   # nominal 0.2
  ho283  =  uniform (0.153, 0.345),        # nominal 0.23
  mo283  =  uniform (0.147, 0.33),  # nominal 0.22
  ho83   =  uniform (0.067, 0.15),  # nominal 0.1
# OLI + O3 rxn
  hcho84 =  uniform(0.072, 0.45),   # nominal 0.18
  ald84  =  uniform(0.288, 1.8 ),   # nominal 0.72
  ket84  =  uniform(0.04, 0.25),    # nominal 0.1
  or184  =  uniform(0.024, 0.15),   # nominal 0.06
  or284  =  uniform(0.116, 0.725),  # nominal 0.29
  ho284  =  uniform(0.104, 0.65),   # nominal 0.26
  ho84   =  uniform(0.056, 0.35),   # nominal 0.14
  mo284  =  uniform(0.124, 0.775),  # nominal 0.31

ICETH = lognormal (2.69e11, 1.2),     # fit to lumped classes
ICHC3 = lognormal (4.59e11, 1.1),   # 20% uncertainty of individual cpd
ICHC5 = lognormal (4.16e11, 1.1),   # IC uncertainty
ICHC8 = lognormal (2.55e11, 1.08),
ICOL2 = lognormal (2.48416e11, 1.2),
ICOLT = lognormal (1.94e11, 1.085),
ICOLI = lognormal (9.83e10 , 1.07),
ICTOL = lognormal (1.655e11 , 1.15),
ICXYL = lognormal (2.01e11 , 1.1),
ICHCHO = lognormal (2.5756e11, 1.2),
ICALD = lognormal (1.9317e11, 1.2);

Uncertain outputs
  O3, NO, NO2, HO;

Do Collocation With
  model: modeldir/mlsode,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: linear,
  sampling file : samp.pnt,
  number of points : 5000,
  options: no header, looping;

# end primorg.dmc
```

C.5.3 allcity.dmc

```
# file name: allcity.dmc
# input date: April 14, 1997
# description:  uncertainty analysis of parameters identified in photolysis case study
#                and primary organic case study, plus other variance-contributing
inputs
#               identified in the SAPRC uncertainty study.
# uncertainty compilation: from SAPRC


Uncertain inputs
   kO3NO =  lognormal (1.8807e-14, 1.2),
   kHONO2 = lognormal (1.1325e-11, 1.28),
   kXYLHO = lognormal (2.7822e-11, 1.3),
   kACO3NO2 = lognormal (5.119e-12, 2.0),
   jNO2 = lognormal (0.00898167, 1.3),
   jO3O1D = lognormal (3.78e-05, 1.4),
   jHCHOM = lognormal (4.305e-05, 1.4),
   jHCHOR = lognormal (3.00667e-05, 1.4),
# OLTP + NO
   ald63 =  uniform (0.4, 2.5),              #1.0
   hcho63 = uniform (0.4, 2.5),              #1.0
# XYLP + NO
   ho268 =  uniform (0.67, 1.5),             #1.0
   dcb68 =  uniform (0.537, 1.209),           #0.806
   ICETH = lognormal (2.69e11, 1.2), # 20% uncertainty
   ICHC3 = lognormal (4.51e11, 1.2),
   ICHC5 = lognormal (4.09e11, 1.2),
   ICHC8 = lognormal (2.51e11, 1.2),
   ICOL2 = lognormal (2.48416e11, 1.2),
   ICOLT = lognormal (1.91e11, 1.2),
   ICOLI = lognormal (9.66e10, 1.2),
   ICTOL = lognormal (1.63e11, 1.2),
   ICXYL = lognormal (1.97e11, 1.2),
   ICHCHO = lognormal (2.5756e11, 1.2),
   ICALD = lognormal(1.9317e11, 1.2),
   ICNO = lognormal (2.64814e12, 1.2),
   ICNO2 = lognormal (8.9052e11, 1.2);

Uncertain outputs
     O3, NO, NO2, HO;
     #new       HCHO, PAN, HNO3, HO2;

Do Collocation With
   model: mlsode,
   dynamic: starting = 0.0, step = 1.0, interval = 10,
   approximation: linear,
   sampling file : samp.pnt,
   number of points : 5000,
   options: no header, looping;


# end allcity.dmc
```

C.6 Isoprene Supplement Species

Species	Description
Stable Species	
MAC	Methacrolein
MVK	Methyl vinyl ketone
Radical Species	
ISOP	Surrogate isoprene-derived peroxyl radical
ISON	Isoprene-NO3 adduct
MACP	OHCC(CH3)(O2)CH2OH + OHCC(CH3)(OH)CH2O2
MVKP	CH3C(O)CH(O2)CH2OH + CH3C(O)CH(OH)CH2O2

Table 20-12. Additional Species Used in RADM's Isoprene Reaction Supplement

C.7 Isoprene Supplement Reactions and Rates

```
# file name: RADMC.map
# input date: Jan 24, 97
# RADM + isoprene supplement
# ISO + HO, ISO + NO3, ISO + O3 are the only reactions changed from RADM.map
# There are sixteen extra reactions for ISO-derived products
# Ref : Zimmermann & Poppe, Atm. Env. 30:1255-1269 (1996)

# Reactions 1 through 52, 54 through 80, 82 through 84, 86 through 158
# are the same as RADM.map in Appendix C.2.  These reactions are deleted from
RADMC.map
# Betty Pun     June 7, 97


# ...
# Reactions 1 - 52 same as RADM.map
# ...

ISO +  HO   --> ISOP
k=k53 ;k53 = uk1
# Reaction 53 was ISO +  HO   --> OLTP in RADM.map
# with k53 = 2.55E-11 * exp (-( -409.0/T)) = 9.97e-11

# ...
# Reactions 54 - 80 same as RAMD.map
# ...

ISO +  NO3  --> ISON
k=k81 ; k81 = uk2
# Reaction 81 was ISO +  NO3  --> OLN in RADM.map
# with k81 = 5.81E-13
# Zimmermann's paper:k81 = 3.03e-12 * exp (-(446/T)) = 6.852e-13

# ...
# Reactions 82 - 84 same as RADM.map
# ...


ISO +  O3   --> 0.55 MAC + 0.21 MVK + 0.3 O3P + 0.5 HO + 0.89 HCHO + 0.11 HO2 + 0.18
CO + 0.07 OLT + 0.07 ORA1 + 0.15 ORA2 + 0.09 MO2 + 0.01 GLY
k=k85 ;k85 = uk3
# Reaction 85 was ISO +  O3   -->
#      0.53 HCHO + 0.500 ALD + 0.33 CO + 0.20 ORA1 + 0.20 ORA2 + 0.23 HO2 + 0.22 MO2 +
0.10 HO
# in RADM.map, with k85 = 1.23E-14 * exp (-( 2013.0/T)) = 1.499e-17

# ...
```

```
# Reactions 86 - 158 same as RADM.map
# ...

MAC + HO --> 0.5 MACP + 0.8 ACO3 + 0.5 HCHO
k=k159;k159 = 1.86e-11 exp(175./T)

MVK + HO --> MVKP
k=k160;k160 = 4.13e-12 exp(452./T)

MAC + O3 --> 0.76 MGLY + 0.23 O3P + 0.46 HO + 0.7 HCHO + 0.11 ORA1 + 0.23 ORA2
k=k161;k161 = 4.4e-15 exp(-2500./T)
#k161 = 1.05e-18 at 300K consistent with Grosjean et al. (93)  EST 27:830

MVK + O3 --> 0.76 MGLY + 0.23 O3P + 0.46 HO + 0.7 HCHO + 0.11 ORA1 + 0.24 MO2 + 0.24
CO
k=k162;k162 = 8.00e-15 exp(-2000./T)
#k162 = 5.09e-18 at 300K

ISOP + NO --> 0.87 NO2 + 0.87 HO2 + 0.38 MVK + 0.28 MAC + 0.62 HCHO + 0.13 ONIT +
0.013 GLY + 0.05 MGLY + 0.02 ORA2 + 0.08 ALD + 0.04 KET + 0.25 XO2
k=k163;k163 = 4.2e-12 exp(180./T)
# k163 = k58?? This reaction could be assigned to be the same as MO2 + NO and HC3P +
NO etc

MACP + NO --> 0.8 KET + 0.2 HCHO + 0.2 MGLY + HO2 + NO2
k=k164;k164 = k163

MVKP + NO --> 0.7 ACO3 + 0.8 ALD + 0.3 HCHO + 0.3 HO2 + 0.3 MGLY + NO2
k=k165;k165 = k163
# The condensation procedure for this reaction is unclear in the Zimmermann's paper
# Explicit reaction in Table 2 different.  I don't understand the products
# k = 4.2e-12 exp(180/T) =k163 in Table 2
# but 4.02e-12 exp (180/T) in Table 3.  Suspect Table 3 typo.

ISON + NO --> 2 NO2 + HCHO + 0.5 MVK + 0.5 MAC
k=k166; k166 = k163

ISOP + HO2 --> OP2
k=k167; k167 = 7.4e-14 exp(1300./T)
#k167 = k86

MACP + HO2 --> OP2
k=k168; k168 = k167

MVKP + HO2 --> OP2
k=k169; k169 = k167

ISOP + ISOP --> 0.34 MAC + 0.42 MVK + 0.3 OLT + 0.06 DCB + 0.1 OLI + 1.3 HO2 + 0.55
HCHO + 0.1 MGLY
k=k170; k170 = 1.0e-13

MVK + NO3 --> NO2 + MO2 + MGLY
k=k171; k171 = 6.0e-14

MAC + NO3 --> 0.5 HNO3 + 0.5 NO2 + 0.8 ACO3 + 0.5 MGLY
k=k172; k172 = 1.0e-14

MVKP + ISOP --> 0.3 MAC + 0.3 MVK + 0.6 MGLY + HCHO + 1.2 HO2 + 0.1 DCB + 0.4 OLT +
0.3 OLI
k=k173; k173 = 5.0e-14

MACP + ISOP --> 0.3 MAC + 0.3 MVK + 0.6 MGLY + HCHO + 1.2 HO2 + 0.1 DCB + 0.4 OLT +
0.3 OLI
k=k174; k174 = 5.0e-14


T = 300
M = 2.46e19
CH4 = 2.2 * 2.46E13  # 2.2 ppm
HV = 1
O2 = 0.2 * 2.46e19
N2 = 0.8 * 2.46e19
```

```
H2O = 4.92e17

uncertain uk1 uk2 uk3

# All city base case
IC NO2 = 8.9052e11
IC NO = 2.64814e12

IC CO = 4.9938e13

IC HCHO = 2.5756e11
IC ALD = 1.9317e11

IC ETH = 2.69e11
IC HC3 = 4.51e11
IC HC5 = 4.09e11
IC HC8 = 2.51e11

IC OL2 = 2.48416e11
IC OLT = 1.91e11
IC OLI = 9.66e10

IC TOL = 1.63e11
IC XYL = 1.97e11

# end RADMC.map
```

C.8 Base Case Used for Isoprene Studies

A different base case was used for the uncertainty analyses of Isoprene reaction and parameters. The isoprene composition in the all city average base case is so low that the contribution of uncertainties in isoprene reactions to uncertainties in model predictions were expected to be negligible. Maintaining the same ROG/NO_x ratio as in the base case used previously (See Appendix C.4), the isoprene composition is raised to 10% of the hydrocarbons.

Organic Compounds	IC (molec/cm3)	IC (ppb)
ETH	2.42e11	9.84
HC3	4.06e11	16.50
HC5	3.68e11	14.96
HC8	2.26e11	9.19
TOL	1.47e11	5.98
XYL	1.78e11	7.24
OL2	2.24e11	9.11
OLT	1.72e11	6.99
OLI	7.91e10	3.22
ISO	2.36e11	9.59
HCHO	2.58e11	10.49
ALD	1.93e11	7.85

Table 20-13. Base Case Organic Initial Conditions for Isoprene Uncertainty Analysis

The dynamics of ozone production predicted using a simple box model was different from those depicted in Appendix C.4 using the more reactive mixture. In fact, ozone production was enhanced by almost 40% (50ppb) over the all city average base case scenario in 10 hours for the RADM mechanism where isoprene was modeled explicitly after a terminal alkene. The differences in NO and NO_2 predictions are only a few ppb's. The ozone production of the Zimmerman case was even higher than the RADM2 mechanism. Figure 20-10 and Table 20-14 show the base case results.

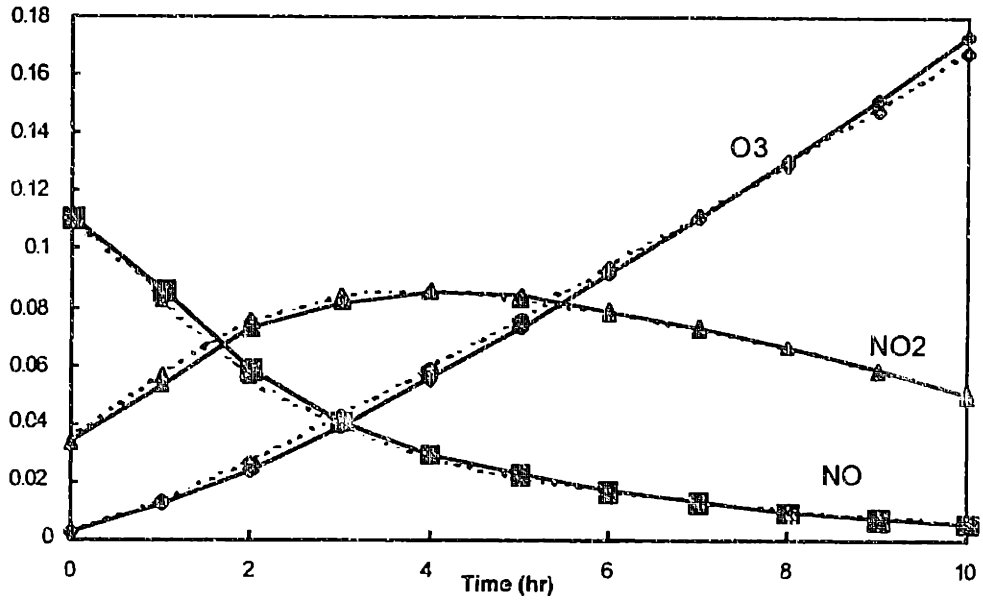


Figure 20-10. Dynamics of 10% Isoprene Base Case.
(—RADM2 predictions with isoprene represented explicitly;
— Predictions using Zimmerman's isoprene supplement)

Time	RADM2			RADM-Zimmerman		
	O_3 (ppm)	NO (ppm)	NO_2 (ppm)	O_3 (ppm)	NO (ppm)	NO_2 (ppm)
0	0.0025	0.1101	0.0338	0.0025	0.1101	0.0338
1	0.0132	0.0824	0.0564	0.0119	0.0859	0.0532
2	0.0264	0.0551	0.0756	0.0239	0.0588	0.0729
3	0.0418	0.0384	0.0836	0.0387	0.0409	0.0824
4	0.0585	0.0279	0.0851	0.0555	0.0294	0.0850
5	0.0756	0.0210	0.0830	0.0732	0.0218	0.0834
6	0.0931	0.0161	0.0787	0.0917	0.0165	0.0792
7	0.1108	0.0125	0.0729	0.1108	0.0126	0.0733
8	0.1290	0.0097	0.0661	0.1306	0.0096	0.0663
9	0.1480	0.0075	0.0587	0.1513	0.0073	0.0583
10	0.1681	0.0057	0.0506	0.1734	0.0054	0.0497

Table 20-14. Model Predictions of Key Species

C.9 Demmucom Input Scripts for Uncertainty Analysis of Isoprene Reactions

C.9.1 Isoprene Rate and Parameters

```
# file name: Zim_par.dmc
# input date: Feb 23, 1997
# description:  Uncertainty analysis of rates and product parameters associated with
isoprene reactions
# uncertainty assignment: see text

Uncertain inputs

   kISOHO = lognormal (9.97e-11, 1.2),
   kISONO3 = lognormal (5.81e-13, 2.5),
   kISOO3 = lognormal (1.499e-17, 2.0),
   kMACHO = lognormal (3.333e-11, 1.4),
   kMVKHO = lognormal (1.863e-11, 1.4),
   kMACO3 = lognormal (1.058e-18, 2.0),
   kMVKO3 = lognormal (5.091e-18, 2.0),
   kRO2NO = lognormal (7.653e-12, 2.0),
   kRO2HO2 = lognormal (5.639e-12, 2.0),
   kRO2RO2 = lognormal (1.0e-13, 2.0),
   kMVKNO3 = lognormal (6.0e-14, 2.0),
   kMACNO3 = lognormal (1.0e-14, 2.0),
   kMVKPISOP = lognormal (5.0e-14, 2.0),
   kMACPISOP = lognormal (5.0e-14, 2.0),

# ISO + O3
   mac85  = uniform (0.39, 0.67),    #nominal value 0.55
   mvk85  = uniform (0.15, 0.26),    #0.21
   o3p85  = uniform (0.25, 0.65),    #0.3
   ho85   = uniform (0.18, 0.83),    #0.5
   hcho85 = uniform (0.80, 0.96),    #0.89
   ho285  = uniform (0.06, 0.16),    #0.11
   co85   = uniform (0.17, 0.19),    #0.18
   olt85  = uniform (0.00, 0.07),    #0.07
   or185  = uniform (0.00, 0.12),    #0.07
   or285  = uniform (0.00, 0.15),    #0.15
   mo285  = uniform (0.07, 0.11),    #0.09
   gly85  = uniform (0.00, 0.03),    #0.01
# MAC + HO
   macp59 = uniform (0.34, 0.66),    #0.5
# MAC + O3
   mgy61  = uniform (0.54, 0.92),    #0.76
   o3p61  = uniform (0.19, 0.50),    #0.23
   hcho61 = uniform (0.62, 0.76),    #0.7
   or161  = uniform (0.00, 0.19),    #0.11
   or261  = uniform (0.19, 0.50),    #0.23
# MVK + O3
   mgly62 = uniform (0.54, 0.92),    #0.76
   o3p62  = uniform (0.19, 0.50),    #0.23
   hcho62 = uniform (0.62, 0.76),    #0.7
   or162  = uniform (0.00, 0.19),    #0.11
   mo262  = uniform (0.19, 0.50),    #0.24
# ISOP + NO
   no263  = uniform (0.74, 1.00),    #0.87
   mvk63  = uniform (0.00, 0.88),    #0.38
   mac63  = uniform (0.00, 0.88),    #0.28
   hcho63 = uniform (0.00, 0.88),    #0.62
   gly63  = uniform (0.00, 0.026),   #0.013
   mgy63  = uniform (0.00, 0.10),    #0.05
   or263  = uniform (0.00, 0.04),    #0.02
   ald63  = uniform (0.00, 0.16),    #0.08
   ket63  = uniform (0.00, 0.08),    #0.04
   xo263  = uniform (0.00, 0.50),    #0.25
# MACP + NO
```

```
   ket64  = uniform (0.74, 0.88),    #0.8
# MVKP + NO
   ald65  = uniform (0.56, 0.84),    #0.7
# ISON + NO
   mvk66  = uniform (0.25, 0.75),    #0.5
# ISOP + ISOP
   mac70  = uniform (0.00, 1.20),    #0.34
   mvk70  = uniform (0.00, 1.60),    #0.42
   olt70  = uniform (0.00, 0.70),    #0.3
   dcb70  = uniform (0.00, 0.20),    #0.06
   oli70  = uniform (0.00, 0.40),    #0.1
   ho270  = uniform (1.20, 1.80),    #1.3
   hcho70 = uniform (0.00, 1.60),    #0.55
   mgy70  = uniform (0.00, 0.80),    #0.1
# MAC + NO3
   hno372 = uniform (0.25, 0.75),    #0.5
# MVKP + ISOP
   mac73  = uniform (0.00, 0.60),    #0.3
   mvk73  = uniform (0.00, 0.60),    #0.3
   mgy73  = uniform (0.30, 0.90),    #0.6
   hcho73 = uniform (0.60, 1.20),    #1.0
   ho273  = uniform (0.60, 1.80),    #1.2
   dcb73  = uniform (0.05, 0.15),    #0.1
   olt73  = uniform (0.20, 0.60),    #0.4
   oli73  = uniform (0.15, 0.45),    #0.3
# MACP + ISOP
   mac74  = uniform (0.00, 0.60),    #0.3
   mvk74  = uniform (0.00, 0.60),    #0.3
   mgy74  = uniform (0.30, 0.90),    #0.6
   hcho74 = uniform (0.60, 1.20),    #1.0
   ho274  = uniform (0.60, 1.80),    #1.2
   dcb74  = uniform (0.05, 0.15),    #0.1
   olt74  = uniform (0.20, 0.60),    #0.4
   oli74  = uniform (0.15, 0.45),    #0.3

   ICETH = constant (2.42e11),
   ICHC3 = constant (4.06e11),
   ICHC5 = constant (3.68e11),
   ICHC8 = constant (2.26e11),
   ICOL2 = constant (2.24e11),
   ICOLT = constant (1.72e11),
   ICOLI = constant (7.91e10),
   ICTOL = constant (1.47e11),
   ICXYL = constant (1.78e11),
   ICISO = normal (2.36e11, 4.72e10);  # 20% uncertainty

Uncertain outputs
  O3, NO, NO2, HO;

Do Collocation With
  model: RADMCmlsode,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: third,
  sampling file : samp.pnt,
  number of points : 5000,
  options: no header, looping;

# end Zim_par.dmc
```

C.9.2 Important Isoprene Rate and Parameters and Others

```
# file name: Zim_all.dmc
# input date: Mar 26, 1997
# description: demmucom input.  iso + a few other parameters

Uncertain inputs
  kO3NO =  lognormal (1.8807e-14, 1.2),
  kHONO2 = lognormal (1.1325e-11, 1.28),
```

```
   kXYLHO = lognormal (2.7822e-11, 1.3),
   kISOHO = lognormal (9.97e-11, 1.2),
   kACO3NO2 = lognormal (5.119e-12, 2.0),
   kISOO3 = lognormal (1.499e-17, 2.0),
   jNO2 = lognormal (0.00898167, 1.3),
   jO3O1D = lognormal (3.78e-05, 1.4),
   jHCHOM = lognormal (4.305e-05, 1.4),
   jHCHOR = lognormal (3.00667e-05, 1.4),
   kMACHO = lognormal (3.333e-11, 1.4),
   kMVKHO = lognormal (1.863e-11, 1.4),
   kMACO3 = lognormal (1.058e-18, 2.0),
   kMVKO3 = lognormal (5.091e-18, 2.0),
   kRO2NO = lognormal (7.653e-12, 2.0),

# OLTP + NO
   ald63 =  uniform (0.4, 2.5),      #1.0
   hcho63 = uniform (0.4, 2.5),      #1.0
# XLYP + NO
   ho268 =  uniform (0.67, 1.5),     #1.0
   dcb68 =  uniform (0.537, 1.209),  #0.806
# ISO + O3
   o3p85  = uniform (0.25, 0.65),    #0.3
   ho85   = uniform (0.18, 0.83),    #0.5
# MAC + HO
   macp59 = uniform (0.34, 0.66),    #0.5
# MVK + O3
   o3p62  = uniform (0.19, 0.50),    #0.23
# ISOP + NO
   no2163  = uniform (0.74, 1.00),   #0.87
   mvk163  = uniform (0.00, 0.88),   #0.38
   mac163  = uniform (0.00, 0.88),   #0.28
   hcho163 = uniform (0.00, 0.88),   #0.62
   mgy163  = uniform (0.00, 0.10),   #0.05
   ald163  = uniform (0.00, 0.16),   #0.08
   xo2163  = uniform (0.00, 0.50),   #0.25
# MACP + NO
   ket64  = uniform (0.74, 0.88),    #0.8
# MVKP + NO
   ald65  = uniform (0.56, 0.84),    #0.7

   ICETH = lognormal (2.42e11, 1.2), # 20% uncertainty
   ICHC3 = lognormal (4.06e11, 1.2),
   ICHC5 = lognormal (3.68e11, 1.2),
   ICHC8 = lognormal (2.26e11, 1.2),
   ICOL2 = lognormal (2.24e11, 1.2),
   ICOLT = lognormal (1.72e11, 1.2),
   ICOLI = lognormal (7.91e10, 1.2),
   ICTOL = lognormal (1.47e11, 1.2),
   ICXYL = lognormal (1.78e11, 1.2),
   ICISO = lognormal (2.36e11, 1.2),
   ICHCHO = lognormal (2.5756e11, 1.2),
   ICALD = lognormal(1.9317e11, 1.2),
   ICNO = lognormal (2.64814e12, 1.2),
   ICNO2 = lognormal (8.9052e11, 1.2);

Uncertain outputs
  O3, NO, NO2, HO;

Do Collocation With
  model: mlsode,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: fifth,
  sampling file : samp.pnt,
  number of points : 5000,
  options: no header, looping;

# end Zim_all.dmc
```

20.4 Appendix D: CB4 Mechanism

The Carbon Bond Mechanisms (CBMs) utilize an organic grouping approach that is much different from the lumped molecules used in the Regional Acid Deposition Model mechanism and the Statewide Air Pollution Research Center mechanism. CBMs lump organic compounds according to the bond type: carbon single bond, carbon double bounds, or carbonyl bonds. This structual-lumping technique categorizes the reactions of similar carbon bonds, and is much different from the molecular-lumping approach that groups the reactions of entire molecules. The main advantage of the structural-lumping approach is that fewer surrogate categories are needed to represent bond groups, leading to easier implementation in large air quality models.

The original CBM mechanism, developed in 1977, contained 35 reactions of 20 species. In 1982, the third generation CBM mechanism was published. CBM-III was expanded to a detailed mechanism (CBM-EX) that contained 71 species and 164 reactions. The version used in this study, CBM-IV (CB4), was condensed from the expanded mechanism, and was tested against about 170 smog chamber experiments. Some characteristics of CB4 are listed in Table 20-15.

Number of Species	32
Number of Reactions	81
Inorganic Reactions	36
Inorganic Photolysis Reactions	6
Organic Reactions	45
Organic Photolysis Reactions	5
Primary Organic Species	CH4, PAR, ETH, OLE, TOL, XYL, ISO
Lumping	Lumped Structure Approach

Table 20-15. Key Characteristics of the CB4 Mechanism.

Reference:

Gery, M.W., Whitten, G,Z., Killus, J.P., "Development and Testing of the CBM-IV for Urban and Regional Modeling", EPA Contract No. 68-02-4136, EPA Report No. EPA/600/3-88/012, PB88-180039 (1988)

Gery, M.W., Whitten, G,Z., Killus, J.P., Dodge, M.C., "A Photochemical Kinetics Mechanism for Urban and Regional Scale Computer Modeling", *J. Geophy. Res.*, **94**: 12925-12956 (1989)

D.1 CB4 species

Model Species	Description
Inorganic Species	
NO	Nitric Oxide
NO2	Nitrogen Dioxide
NO3	Nitrogen Trioxide
N2O5	Dinitrogen Pentoxide
HONO	Nitrous Acid
HNO3	Nitric Acid
PNA	Peroxynitric Acid (HO2NO2)
O1D	Oxygen Atom (Singlet)
O	Oxygen Atom (Triplet)
OH	Hydroxyl Radical
O3	Ozone
HO2	Hydroperoxyl Radical
H2O2	Hydrogen Peroxide
CO	Carbon Monoxide
Organic Species	
FORM	Formaldehyde (CH2=O)
ALD2	Higher Molecular Weight Aldehydes (RCHO, R > H)
C2O3	Peroxylacyl Radicals (CH3C(O)O2•)
PAN	Peroxylacyl Nitrate (CH3C(O)OONO2)
PAR	Paraffin Carbon Bond (C-C)
ROR	Secondary Organic Oxy Radical
OLE	Olefinic Carbon Bond (C=C)
ETH	Ethene
TOL	Toluene
CRES	Cresol and Higher Molecular Weight Phenols
TO2	Toluene-Hydroxyl Radical Adduct
CRO	Methyl-Phenoxyl Radical
OPEN	Higher Molecular Weight Aromatic Oxidation Ring Fragment
XYL	Xylene
MGLY	Methyl-Glyoxal
ISOP	Isoprene
XO2	NO-to-NO2 Operation
XO2N	NO-to-Nitrate Operation

Table 20-15. List of Species in the CB4 Mechanism

D.2 CB4 Chemap Input (rates and reactions)

```
# file name: CB4.map
# input date: Oct 29, 1996
# Description: CB4 mechanism with 5 uncertain photolysis rates
# Source: adapted from EPA National Exposure Research Laboratory's CB4.mech file
# ref: Gery, Whitten, Killus, Dodge JGR 94: 12925 (1989)

# Notes:
# 1. Chemap does not accept -ve product coefficient in equation.  Changed to:
#       A + B --> c C + D
#       c = -1.0
# 2. When a compound is both a product and a reactant, Chemap treats
#       incorrectly when it is the first appearance of species in file.
#       Hence Reaction 52 is split up.
# 3. Hand-checked derivs.f file against CB4.map reactions
# 4. Sulfur Reactions (84, 85) not used in this study.
# 5. Reaction 52 was modified from
#       PAR + OH --> 0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR - 0.11 PAR
#    to:
#       PAR + OH --> 0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR
#       PAR -->   (k = 0.11 * previous k)
# 6. Reaction 83 XO2  + HO2  --> was not in paper


NO2  + HV -->  NO +  O
k=k1 ; k1 = j1

O + O2   -->  O3
k=k2 ; k2 = (k2not * M/(1+ k2not * M / k2inf)) * 0.6^(1/(1+(log10(k2not * M /
k2inf))^2))
k2not = 6.0E-34 * (T/300)^-2.3
k2inf = 2.8E-12

O3  +   NO   -->  NO2
k=k3 ; k3 = 1.8E-12 * exp(-(1370/T))
# (UP-NA87) comment from EPA file.

O   +  NO2   -->  NO
k=k4 ; k4 = 9.3E-12
# (UP-NA87)

O   +  NO2   -->  NO3
k=k5 ; k5 = (k5not * M/(1+ k5not * M / k5inf)) * 0.6^(1/(1+(log10(k5not * M /
k5inf))^2))
k5not = 9.0E-32 * (T/300)^-2.0
k5inf = 2.2E-11

O   +   NO   -->  NO2
k=k6 ; k6 = (k6not * M/(1+ k6not * M / k6inf)) * 0.6^(1/(1+(log10(k6not * M /
k6inf))^2))
k6not = 9.0E-32 * (T/300)^-1.5
k6inf = 3.0E-11

O3  +  NO2   -->  NO3
k=k7 ; k7 = 1.2E-13 * exp(-(2450/T))
# (UP-NA87)

O3 + HV  -->  O
k=k8 ; k8 = 0.053 * j1

O3 + HV  -->  O1D
k=k9 ; k9 = j2

O1D   -->  O
k=k10 ; k10= 1.91E+08 * exp(-( -390/T))
# EPA file has (?) comment.  This is a combined reaction of M = N2 or O2

H2O +  O1D   -->  2.0 OH
k=k11 ; k11= 2.2E-10
```

```
O3  +   OH    -->  HO2
k=k12 ; k12= 1.6E-12 * exp(-(940/T))

O3  +  HO2    -->  OH
k=k13 ; k13= 1.4E-14 * exp(-(580/T))
# (UP-NA87)

NO3 + HV --> 0.89 NO2 + 0.89 O +  0.11 NO
k=k14 ; k14= 33.9 *ji

NO3 +  NO-->  2.0 NO2
k=k15 ; k15= 1.3E-11 * exp(-(-250/T))
# (UP-NA87)

NO3  + NO2 -->  NO + NO2
k=k16 ; k16= 2.5E-14 * exp(-(1230/T))

NO3  + NO2 --> N2O5
k=k17 ; k17= 5.346E-13 * exp(-(-256/T))
# (UP-NA87)

N2O5 + H2O -->   2.0 HNO3
k=k18 ; k18= 1.3E-21

N2O5   -->   NO3   NO2
k=k19 ; k19= 3.52E+14 * exp(-(10897/T))
# (UP-NA87)

NO  +  NO -->   2.0 NO2
k=k20 ; k20= 1.76E-20 * exp(-(-530/T))
# (Add O2) bi-molecular rxn in ref.

NO   + NO2 + H2O -->   2.000 HONO
k=k21 ; k21= 4.4E-40

OH   + NO -->   HONO
k=k22 ; k22= (k22not * M/(1+ k22not * M / k22inf)) * 0.6^(1/(1+(log10(k22not * M /
k22inf))^2))
k22not = 6.7E-31 * (T/300)^-3.3
k22inf = 3.0E-11 * (T/300)^-1.0

HONO + HV -->   OH + NO
k=k23 ; k23= 0.1975 * j1

OH  +  HONO  -->   NO2
k=k24 ; k24= 6.6E-12

HONO + HONO  -->   NO + NO2
k=k25 ; k25= 1.0E-20

OH   + NO2   -->   HNO3
k=k26 ; k26= (k26not * M/(1+ k26not * M / k26inf)) * 0.6^(1/(1+(log10(k26not * M /
k26inf))^2))
k26not = 2.6E-30 * (T/300)^-3.2
k26inf = 2.4E-11 * (T/300)^-1.3

OH   + HNO3  -->   NO3
k=k27 ; k27= k27zero + k27thr / (1 + k27thr/k27two)
k27zero = 7.2E-15 * exp(-(-785/T))
k27two = 4.1E-16 * exp(-(-1440/T))
k27thr = 1.9E-33 * exp(-(-725/T)) * M

HO2  + NO -->   OH + NO2
k=k28 ; k28= 3.7E-12 * exp(-(-240/T))

HO2  + NO2   -->   PNA
k=k29 ; k29= (k29not * M/(1+ k29not * M / k29inf)) * 0.6^(1/(1+(log10(k29not * M /
k29inf))^2))
k29not = 2.3E-31 * (T/300)^-4.6
k29inf = 4.2E-12 * (T/300)^0.2
```

```
PNA -->    HO2 +    NO2
k=k30 ; k30= 4.79E13 * exp(-(10121/T))

OH    + PNA    -->    NO2
k=k31 ; k31= 1.3E-12 * exp(-(-380/T))

HO2    +HO2    -->    H2O2
k=k32 ; k32= 5.9E-14 * exp(-(-1150/T))

HO2 +  HO2 + H2O  -->    H2O2
k=k33 ; k33= 2.2E-38 * exp(-(-5800/T))

H2O2 + HV -->  2 OH
k=k34 ; k34= 0.189 * j4              # value according to Carter and OZIPR
#k=k34 ; k34= 0.255 * j4

OH  +  H2O2  -->  HO2
k=k35 ; k35= 3.1E-12 * exp(-(187/T))
# (UP-NA87)

OH +    CO -->    HO2
k=k36 ; k36= 2.2E-13
# taken from ref
# EPA file read %1 1.50E-13
# from RADM %1 means 1.50E-13 * (1. + 2.439E-20 * M)

FORM + OH -->    HO2 + CO
k=k37 ; k37= 1.0E-11

FORM + HV    -->  2 HO2  +  CO
k=k38 ; k38= j3

FORM + HV    -->    CO
k=k39 ; k39= j4

FORM + O-->    OH + HO2 + CO
k=k40 ; k40= 3.0E-11 * exp(-(1550/T))
# (UP-NA87)

FORM + NO3    -->    HNO3 + HO2 + CO
k=k41 ; k41= 6.3E-16

ALD2 + O-->  C2O3 + OH
k=k42 ; k42= 1.177E-11 * exp(-(986/T))

ALD2 + OH -->    C2O3
k=k43 ; k43= 6.9E-12 * exp(-(-250/T))

ALD2 + NO3    -->    C2O3 + HNO3
k=k44 ; k44= 2.50E-15

ALD2 + HV -->  XO2 + 2 HO2 + CO + FORM
k=k45 ; k45= j5

C2O3 + NO -->  NO2 + XO2 + FORM + HO2
k=k46 ; k46= 5.36E-12 * exp(-( -250/T))

C2O3 + NO2    -->    PAN
k=k47 ; k47= 7.98E-20 * exp(-( -5500/T))

PAN -->    C2O3 +  NO2
k=k48 ; k48= 9.360E+16 * exp(-(14000/T))

C2O3 + C2O3  -->  2 XO2 +  2 FORM + 2 HO2
k=k49 ; k49= 2.50E-12

C2O3 + HO2    -->  0.79 FORM + 0.79 XO2 + 0.79 HO2 + 0.79 OH
k=k50 ; k50= 6.50E-12

# There is an error in Equation 51 and Rate
# OH + CH4 -->    XO2 + FORM  + HO2
# k=k51 ; k51= 1.09E+02 * exp(-(1710/T))
```

```
# Original equation for this k is OH --> XO2 + FORM + HO2, assumed CH4 = 1.85ppm
OH + CH4 -->   XO2 + FORM  + HO2
k=k51 ; k51= 2.4e-12 * exp(-(1710/T))

#original Reaction 52
#PAR + OH --> 0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR - 0.11 PAR
#k=k52 ; k52= 8.14E-13
#split into Raction 52 and Reaction 53 as follows

PAR + OH --> 0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR
k=k52 ; k52= 8.14E-13

PAR -->
k=k53; k53 = 0.11 * k52

ROR   -->  1.1 ALD2 + 0.96 XO2 + 0.94 HO2 + c54 PAR + 0.04 XO2N + 0.020 ROR
k=k54 ; k54= 1.04E+15 * exp(-( 8000/T))
c54 = (-2.1)

ROR   -->   HO2
k=k55 ; k55= 1.591E+03

ROR  + NO2  -->
k=k56 ; k56= 1.488E-11

O + OLE  -->  0.63 ALD2 + 0.38 HO2  + 0.28 XO2 + 0.3 CO + 0.2 FORM + 0.02 XO2N + 0.22
PAR + 0.2 OH
k=k57 ; k57= 1.188E-11 * exp(-( 324/T))

OH   + OLE  -->  FORM + ALD2 + XO2 + HO2 + c58 PAR
k=k58 ; k58= 5.24E-12 * exp(-( -504/T))
c58 = (-1.0)

O3 + OLE --> 0.5 ALD2 + 0.74 FORM + 0.33 CO + 0.44 HO2 + 0.22 XO2 + 0.1 OH + c59 PAR
k=k59 ; k59= 1.42E-14 * exp(-( 2105/T))
c59 = (-1.0)

NO3 + OLE --> 0.91 XO2 + 0.09 XO2N + FORM + ALD2 + c60 PAR + NO2
k=k60 ; k60= 7.68E-15
c60 = (-1.0)

O   + ETH --> FORM + 0.7 XO2 + CO + 1.7 HO2  + 0.3 OH
k=k61 ; k61= 1.04E-11 * exp(-( 792/T))

OH   + ETH  -->  XO2 + 1.56 FORM + HO2 + 0.22 ALD2
k=k62 ; k62= 2.03E-12 * exp(-( -411/T))

O3   + ETH  -->  FORM + 0.42 CO +  0.12 HO2
k=k63 ; k63= 1.256E-14 * exp(-(2633/T))

OH + TOL --> 0.08 XO2 + 0.36 CRES + 0.44 HO2 + 0.56 TO2
k=k64 ; k64= 2.1E-12 * exp(-( -322/T))

TO2  + NO   --> 0.9 NO2 + 0.9 HO2 + 0.9 OPEN
k=k65 ; k65= 8.12E-12

TO2   --> CRES  +  HO2
k=k66 ; k66= 4.17

OH   + CRES --> 0.4 CRO + 0.6 XO2 + 0.6 HO2 + 0.3 OPEN
k=k67 ; k67= 4.127E-11

NO3  + CRES --> CRO   + HNO3
k=k68 ; k68= 2.2E-11

CRO  + NO2  -->
k=k69 ; k69= 1.35E-11

OPEN + HV   --> C2O3   +HO2 +CO
k=k70 ; k70= 8.4 * j3        # Carter and OZIPR
#k=k70 ; k70= 9.04 * j3
```

```
OPEN + OH    --> XO2 + 2 CO + 2 HO2 + C2O3 + FORM
k=k71 ; k71= 2.98E-11

OPEN + O3 --> 0.03 ALD2 + 0.62 C2O3 + 0.7 FORM + 0.03 XO2 + 0.69 CO + 0.08 OH + 0.76
HO2 + 0.2 MGLY
k=k72 ; k72= 5.433E-17 * exp(-( 500/T))

OH    +XYL    --> 0.7 HO2 + 0.5 XO2 + 0.2 CRES + 0.8 MGLY + 1.1 PAR + 0.3 TO2
k=k73 ; k73= 1.66E-11 * exp(-(-116/T))

OH   + MGLY   --> XO2   +   C2O3
k=k74 ; k74= 1.76E-11

MGLY + HV    --> C2O3   + HO2   + CO
k=k75 ; k75= 8.96 * j3              # Carter and OZIPR
#k=k75 ; k75= 9.64 * j3

O   + ISOP    --> 0.6 HO2 + 0.8 ALD2 + 0.55 OLE + 0.5 XO2 + 0.5 CO + 0.45 ETH   + 0.90
PAR
k=k76 ; k76= 1.83E-11

OH   +   ISOP   --> XO2 + FORM + 0.67 HO2 + 0.13 XO2N + ETH + 0.4 MGLY + 0.2 C2O3 + 0.2
ALD2
k=k77 ; k77= 9.6E-11

O3   +   ISOP   --> FORM + 0.4 ALD2 + 0.55 ETH + 0.2 MGLY + 0.1 PAR + 0.06 CO + 0.44 HO2
+ 0.1 OH
k=k78 ; k78= 1.22E-17

NO3 +   ISOP   -->   XO2N
k=k79 ; k79= 3.18E-13

XO2   + NO -->    NO2
k=k80 ; k80= 8.12E-12

XO2N + NO -->
k=k81 ; k81= 6.77E-13

XO2   + XO2    -->
k=k82 ; k82= 1.725E-14 * exp(-(-1300/T))

XO2   + HO2   -->
k=k83 ; k83= 7.71E-14 * exp(-(-1300.0/T))

SO2 + OH--> SULF + HO2
k=k84 ; k84= 7.515E-13 * exp(-(-150/T))

SO2--> SULF
k=k85 ; k85= 5.416E-20


j1 = uk1
j2 = uk2
j3 = uk3
j4 = uk4
j5 = uk5

uncertain uk1 uk2 uk3 uk4 uk5

HV = 1
T = 300
M = 2.46E19
H2O = 4.92e17
O2 = 0.2 * 2.46e19
CH4 = 1.65 * 2.46E13

IC NO2 = 8.9052e11
IC NO = 2.64814e12
IC CO = 4.9938e13

IC OLE = 2.0986e11
IC PAR = 6.9439e12
```

```
IC TOL =  1.4785e11
IC XYL = 1.9913e11
IC FORM = 2.6817e11
IC ALD2 = 3.2993e11
IC ETH = 2.4842e11
IC ISOP = 8.7557e9

# end CB4.map
```

D.3 CB4 lumping

In the structural lumping procedure employed in CB4, each compound is represented by the bonds of its structure. For example, 1-butene is one OLE bond and 2 PAR bonds. Table 20-15 lists the compounds present in the organic mixture used in the uncertainty studies of the CB4 mechanism and the assignment of each compound into the modeled classes. The initial conditions used for the base case are determined by summing the product of the amount and the assignment of the individual compounds.

Individual Cpds	Equivalent Cpd	SARAOD	Amount	Assignment (Each Compound)								
				OLE	PAR	TOL	XYL	FORM	ALD2	ETH	ISOP	UNR
FORMALD		43502	2.58E+11	0	0	0	0	1	0	0	0	0
ACETALD		43503	1.93E+11	0	0	0	0	0	1	0	0	0
ETHANE		43202	2.69E+11	0	0.4	0	0	0	0	0	0	1.6
PROPANE		43204	2.01E+11	0	1.5	0	0	0	0	0	0	1.5
2-ME-C3		43214	8.97E+10	0	4	0	0	0	0	0	0	0
N-C4		43212	2.13E+11	0	4	0	0	0	0	0	0	0
2-ME-C4		98132	1.83E+11	0	5	0	0	0	0	0	0	0
N-C5		43220	8.85E+10	0	5	0	0	0	0	0	0	0
BR-C6	2-ME-C5	43229	1.15E+10	0	6	0	0	0	0	0	0	0
22-DM-C4		43291	5.02E+09	0	6	0	0	0	0	0	0	0
CYCC5		43242	8.24E+09	0	5	0	0	0	0	0	0	0
23-DM-C4		98001	1.31E+10	0	6	0	0	0	0	0	0	0
2-ME-C5		43229	4.96E+10	0	6	0	0	0	0	0	0	0
3-ME-C5		43230	3.56E+10	0	6	0	0	0	0	0	0	0
N-C6		43231	3.63E+10	0	6	0	0	0	0	0	0	0
BR-C7	3-ME-C6	43295	6.72E+10	0	7	0	0	0	0	0	0	0
CYC-C6		43248	2.16E+10	0	6	0	0	0	0	0	0	0
CYCC6		43248	1.09E+10	0	6	0	0	0	0	0	0	0
BR-C8	4-ME-C7	43297	6.36E+10	0	8	0	0	0	0	0	0	0
N-C7		43232	1.46E+10	0	7	0	0	0	0	0	0	0
CYC-C7	ME-CYCC6	43261	1.09E+10	0	7	0	0	0	0	0	0	0
BR-C9	4-ET-C7	90084	2.5E+10	0	9	0	0	0	0	0	0	0
N-C8		43233	7.71E+09	0	8	0	0	0	0	0	0	0
CYC-C8	ET-CYCC6	43268	1.25E+09	0	8	0	0	0	0	0	0	0
N-C9		43235	5.94E+09	0	9	0	0	0	0	0	0	0
BR-C10	4-PR-C7	43109	1.8E+10	0	10	0	0	0	0	0	0	0
N-C10		43238	8.19E+09	0	10	0	0	0	0	0	0	0
BR-C11	35-DE-C7	43110	7.18E+09	0	11	0	0	0	0	0	0	0
BR-C12	36-DE-C8	43111	1.66E+09	0	12	0	0	0	0	0	0	0
BR-C13	37-DE-C9	43112	36627920	0	13	0	0	0	0	0	0	0
ETHENE		43203	2.48E+11	0	0	0	0	0	0	1	0	0
PROPENE		43205	6.39E+10	1	1	0	0	0	0	0	0	0
C4-OLE1	1-BUTENE	43213	2.24E+10	1	2	0	0	0	0	0	0	0

1-BUTENE		43213	3.23E+10	1	2	0	0	0	0	0	0	0
13-BUTDE		43218	8.64E+09	2	0	0	0	0	0	0	0	0
T-2-BUTE		43216	1.32E+10	0	0	0	0	0	2	0	0	0
C-2-BUTE		43217	1.03E+10	0	0	0	0	0	2	0	0	0
C5-OLE1	1-PENTENE	43224	1.41E+10	1	3	0	0	0	0	0	0	0
1-PENTEN		43224	7.02E+09	1	3	0	0	0	0	0	0	0
2M-1-BUT		43225	1.06E+10	0	4	0	0	1	0	0	0	0
ISOPRENE		43243	8.76E+09	0	0	0	0	0	0	0	1	0
C5-OLE2	2-C5-OLE	43227	2.59E+10	0	1	0	0	0	2	0	0	0
2M-2-BUT		43228	1.44E+09	1	3	0	0	0	0	0	0	0
C6-OLE1	1-HEXENE	43245	9.49E+09	1	4	0	0	0	0	0	0	0
C6-OLE2	2-C6-OLE	43246	1.21E+10	0	2	0	0	0	2	0	0	0
1-HEXENE		43245	2.74E+09	1	4	0	0	0	0	0	0	0
C7-OLE1	1-C7-OLE	98005	1.49E+10	1	5	0	0	0	0	0	0	0
C8-OLE1	1-C8-OLE	43265	5.15E+09	1	6	0	0	0	0	0	0	0
C7-OLE2	2-C7-OLE	98041	1.3E+09	0	5	0	0	0	1	0	0	0
C8-OLE2	3-C8-OLE	90013	2.96E+08	0	4	0	0	0	2	0	0	0
C9-OLE1	1-C9-OLE	43267	9.22E+09	1	7	0	0	0	0	0	0	0
A-PINENE		98025	7.87E+09	0.5	6	0	0	0	1.5	0	0	0
C10-OLE1	1C10-OLE	43268	4.12E+08	1	8	0	0	0	0	0	0	0
C11-OLE1	1C11-OLE	43269	5.25E+09	1	9	0	0	0	0	0	0	0
BENZENE		45201	4.46E+10	0	1	0	0	0	0	0	0	5
TOLUENE		45202	1.06E+11	0	0	1	0	0	0	0	0	0
C2-BENZ		45203	1.6E+10	0	1	1	0	0	0	0	0	0
M-XYLENE		45205	2.51E+10	0	0	0	1	0	0	0	0	0
P-XYLENE		45206	2.51E+10	0	0	0	1	0	0	0	0	0
O-XYLENE		45204	1.99E+10	0	0	0	1	0	0	0	0	0
I-C3-BEN		98043	1.85E+09	0	2	0	1	0	0	0	0	0
N-C3-BEN		45108	3.71E+09	0	2	1	0	0	0	0	0	0
C9-BEN2	M-XYLENE	45205	2.94E+10	0	0	0	1	0	0	0	0	0
C9-BEN1	TOLUENE	45202	4.18E+09	0	0	1	0	0	0	0	0	0
C9-BEN3	135-TMB	45207	1.25E+09	0	1	0	1	0	0	0	0	0
123-TMB		45225	3.75E+10	0	1	0	1	0	0	0	0	0
C10-BEN1	TOLUENE	45202	1.65E+10	0	0	1	0	0	0	0	0	0
C10-BEN2	M-XYLENE	45205	2.11E+10	0	0	0	1	0	0	0	0	0
C10-BEN3	135-TMB	45207	1.18E+10	0	1	0	1	0	0	0	0	0
C10-BEN4	135-TMB	45207	2.36E+09	0	1	0	1	0	0	0	0	0
C11-BEN1	TOLUENE	45202	9.66E+08	0	0	1	0	0	0	0	0	0
C11-BEN2	M-XYLENE	45205	9.55E+09	0	0	0	1	0	0	0	0	0
C11-BEN3	135-TMB	45207	9.55E+09	0	1	0	1	0	0	0	0	0
C12-BEN2	M-XYLENE	45205	2.41E+09	0	0	0	1	0	0	0	0	0
C12-BEN3	135-TMB	45207	2.41E+09	0	1	0	1	0	0	0	0	0
C12-BEN1	TOLUENE	45202	25751947	0	0	1	0	0	0	0	0	0
ACETYLEN		43206	1.48E+11	0	1	0	0	0	0	0	0	1
INERT		0	1.03E+11	0	0	0	0	0	0	0	0	1

Table 20-15. Structural Lumping of Organic Species in the CB4 Mechanism

D.4 CB4 Base Case Simulation

The same organic and NO_x concentrations were used in the CB4 base case as those used for the SAPRC and RADM mechanisms. The results are shown in Figures 20-11 through 20-14, and summarized in Table 20-16. The dynamics of key species predicted by the CB4 mechanism were very similar to those depicted in Appendices B and C for the other mechanisms. The most significant difference is the amount of the PAR compound, which was allocated from alkanes, alkenes and aromatics. In fact, about 35 ppb of PAR was consumed during the course of 10 hours, far exceeding the other source compounds.

Time	O3	NO	NO2	FORM	ALD2	OH	HO2
2	0.0187	0.0708	0.0624	13.79	14.99	0.0958	0.3460
4	0.0411	0.0402	0.0780	14.55	15.05	0.0832	0.5116
6	0.0674	0.0243	0.0778	13.44	14.33	0.0863	0.7678
8	0.0964	0.0152	0.0697	12.01	13.22	0.1014	1.224
10	0.1292	0.0093	0.0573	10.70	11.85	0.1319	2.194
Time	PAN	HNO3	PAR	ETH	OLE	TOL	XYL
2	0.658	8.593	274.7	9.107	5.295	5.450	5.486
4	1.431	22.14	267.2	7.944	3.092	4.925	3.665
6	2.097	36.52	260.4	6.880	1.747	4.473	2.500
8	2.856	51.67	253.2	5.816	0.8837	4.020	1.635
10	3.876	67.73	244.6	4.706	0.3715	3.522	0.9663

Table 20-16. CB4 Base Case Time Series Results of Key Species

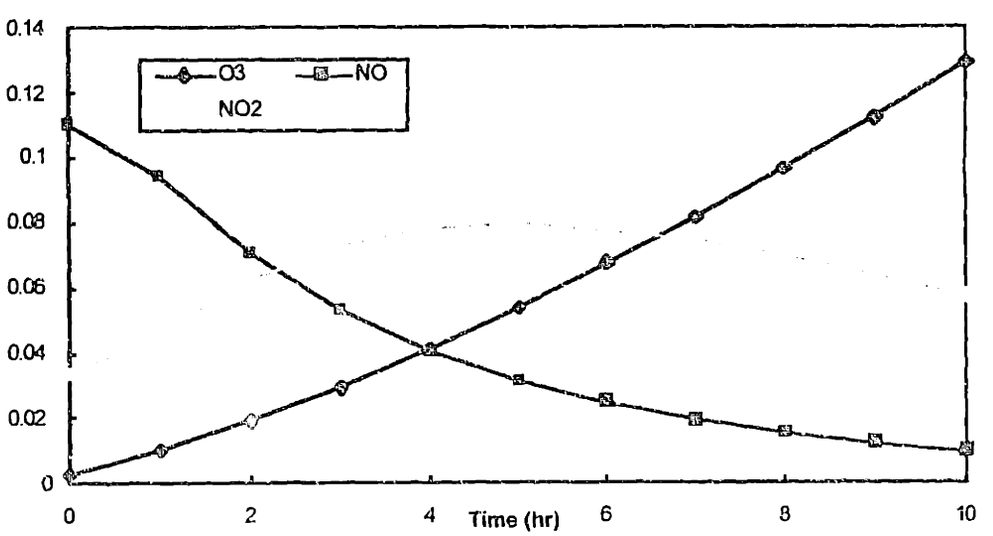


Figure 20-11. CB4 Base Case Simulation for Ozone, NO, NO_2.

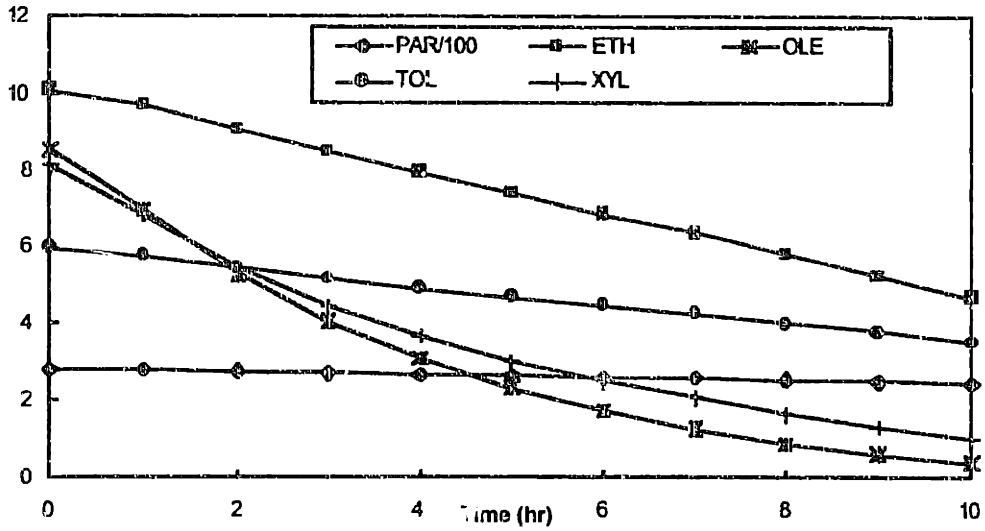


Figure 20-12. CB4 Base Case Simulation for Organic Source Species.

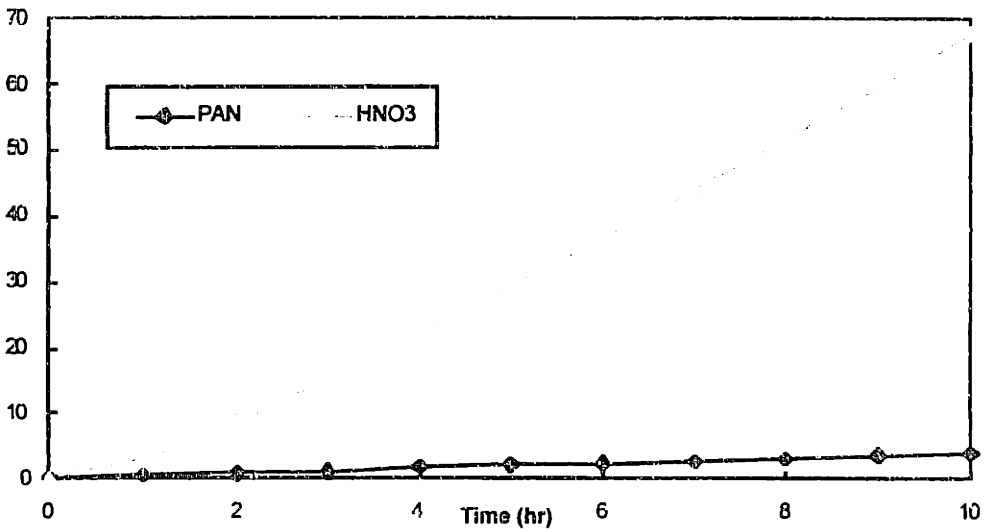


Figure 20-13. CB4 Base Case Simulation for Sink Species.

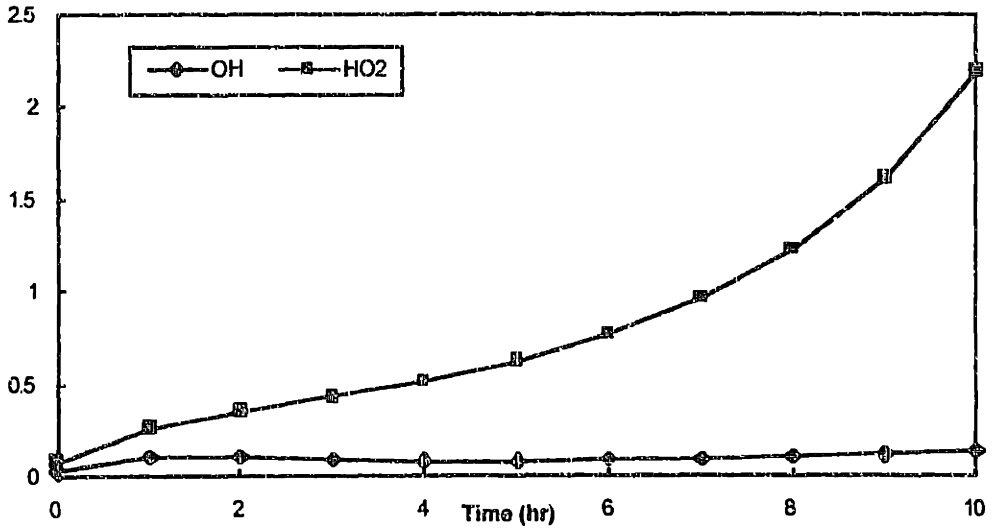


Figure 20-14. CB4 Base Case Simulation for Inorganic Radical Species.

D.5 CB4 Uncertainty Analysis Inputs

Case studies were performed to analyze the uncertainties in the parameters used in the CB4 mechanism. The Demmucom input files included in this section are records of the uncertain parameters considered as well as the assigned uncertainty characteristics of these parameters. Appendix D.5.1 is the input for the photolysis rate constant case study. Appendix D.5.2 lists the uncertain inputs in the primary organic parameters case study. The input script for a combined uncertainty analysis of all variance contributing parameters identified in SAPRC is included in Appendix D.5.3.

D.5.1 phk.dmc

```
# file name: phk.dmc
# input date: 3/31/97
# Description: Demmucom input file for CB4 photolysis rates uncertainty analysis
# Uncertainty allocation: same as SAPRC, relative factors allocated same uncertainty
#                          factor as photolysis rates of associated compounds

Uncertain Inputs

# independent photolysis rate constants
jNO2 = lognormal (0.009822, 1.3),
jO3OSD = lognormal (4.55e-05, 1.4),
jHCHOR = lognormal (3.00667e-05, 1.4),
jHCHOM = lognormal (4.305e-05, 1.4),
jALD = lognormal (5.68e-06, 1.4),

# factors used to describe relative rates
rO3O3P = lognormal (0.053, 1.1),
rNO3 = lognormal (33.9, 3.0),
rHONO = lognormal (0.1975, 1.6),
rH2O2 = lognormal (0.255, 1.4),
rOPEN = lognormal (8.4, 3.0),
rMGLY = lognormal (8.96, 1.6);

Uncertain Outputs

O3, NO, NO2, HCHO;

Do Collocation with
  model: mlsode,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: linear,
  sampling file : samp.pnt,
  number of points : 5000,
  options: no header, looping;

# end phk.dmc
```

D.5.2 Porg.dmc

```
# file name: porg.dmc
# input date: April 11, 1997
# description: CB4 uncertainty analysis of primary organics: rates + product
coefficients
# Uncertainty Assignment:
#  rate uncertainty factor: typical of species in lumped group
#  product coefficient: typical of species in lumped group,
#                        uniform with range = nominal multiply/divide factor
# 95 Uncertain Inputs, including 14 rate constants and 8 IC


Uncertain Inputs
```

```
        kPAR_HO = lognormal(8.14E-13 , 1.3),
        kOLE_O = lognormal (4.03439e-12, 1.2),
        kOH_OLE = lognormal (2.81155e-11, 1.2),
        kO3_OLE = lognormal (1.27347e-17, 1.5),
        kNO3_OLE = lognormal (7.68E-15, 2.0),
        kO_ETH = lognormal (7.42157e-13,1.2),
        kOH_ETH = lognormal(7.98876e-12, 1.15),
        kO3_ETH = lognormal (1.9379e-18, 1.25),
        kOH_TOL = lognormal (6.14274e-12, 1.2),
        kHO_XYL = lognormal (2.44363e-11, 1.3),
        kO_ISO = lognormal (1.83E-11, 1.5),
        kOH_ISO = lognormal (9.6E-11, 1.2),
        kO3_ISO = lognormal (1.22E-17, 2.0),
        kNO3_ISO = lognormal (3.18E-13, 2.5),
# PAR + HO
        xn_52 = uniform (0.0695, 0.23),          #0.13
        ho2_52 = uniform (0.0629, 0.193),        #0.11
        ror_52 = uniform (0.434, 1.33),          #0.76
#ROR
        ald_54 = uniform (0.629, 1.925),         #1.1
        ho2_54 = uniform (0.537, 1.645),         #0.94
        par_54 = uniform (-3.68, -1.20),         #-2.1
        xn_54 = uniform (0.02, 0.077),           #0.04
        ror_54 = uniform (0.011, 0.035),         #0.02
#O + OLE
        ald_57 = uniform (0.42, 0.945),          #0.63
        ho2_57 = uniform (0.25, 0.57),           #0.38
        xo2_57 = uniform (0.19, 0.42),           #0.28
        co_57 =uniform (0.2, 0.45),              #0.3
        form_57 = uniform (0.13, 0.3),           #0.2
        xn_57 = uniform (0.013, 0.03),           #0.02
        par_57 = uniform (0.15, 0.33),           #0.22
        ho_57 = uniform (0.13, 0.3),             #0.2
#OH + OLE
        for_58 = uniform (0.667, 1.5),           #1.0
        ald_58 = uniform (0.667, 1.5),           #1.0
        x_58 = uniform (0.667, 1.5),             #1.0
        ho2_58 = uniform (0.667, 1.5),           #1.0
        par_58 = uniform (-1.5, -0.667),         #-1.0
#O3 + OLE
        ald_59 = uniform (0.333,0.75),           #0.5
        for_59 = uniform (0.493, 1.11),          #0.74
        co_59 = uniform (0.22, 0.495),           #0.33
        ho2_59 = uniform (0.293, 0.66),          #0.44
        x_59 = uniform (0.15, 0.33),             #0.22
        oh_59 = uniform (0.067, 0.15),           #0.1
        par_59 = uniform (-1.5, -0.667),         #-1.0
#NO3 + OLE
        xn_60 = uniform (0.0535, 0.1475),        #0.09
        for_60 = uniform (0.667, 1.5),           #1.0
        ald_60 = uniform (0.667, 1.5),           #1.0
        par_60 = uniform (-1.5, -0.667),         #-1.0
#O + ETH
        form_61 = uniform (0.9, 1.1),                    #1.0
        xo2_61 = uniform (0.3, 1.0),                     #0.7
        ho2_61 = uniform (1.0, 2.4),                     #1.7
        oh_61  = uniform (0.0, 0.6),             #0.3
#OH + ETH
        ald_62 = uniform (0.184, 0.261),         #0.22
#O3 + ETH
        co_63 = uniform (0.38, 0.46),                    #0.42
        ho2_63 = uniform (0.109, 0.132),         #0.12
#OH + TOL
        xo2_64  = uniform (0.053, 0.12),         #0.08
        cres_64 = uniform (0.24, 0.54),          #0.36
        ho2_64 = uniform (0.293, 0.66),          #0.44
        to2_64 = uniform (0.373, 0.84),          #0.56
#TO2 + NO
        no2_54 = uniform (0.84, 0.94),           #0.9
#OH + XYL
        ho2_73 = uniform (0.467, 1.05),          #0.7
```

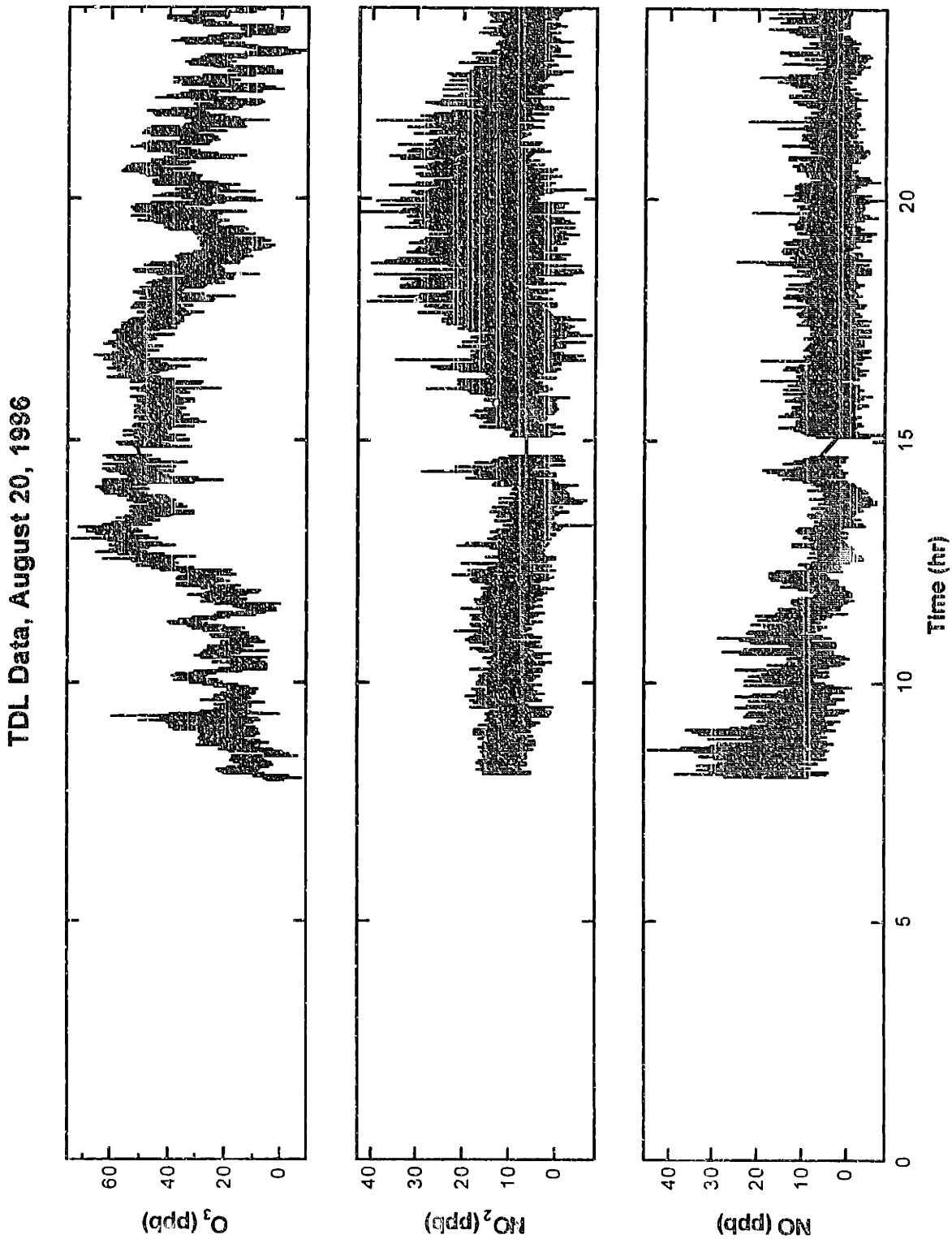
```
        xo2_73 = uniform (0.33, 0.75),              #0.5
        cre_73 = uniform (0.13, 0.3),               #0.2
        mg_73 = uniform (0.53, 1.20),                       #0.8
        par_73 = uniform (0.73, 1.65),              #1.1
        to2_73 = uniform (0.2, 0.45),                       #0.3
#O   + ISOP
        ho2_76 = uniform (0.343, 1.05),             #0.6
        ald_76 = uniform (0.457, 1.4),              #0.8
        ole_76 = uniform (0.314, 0.9625),           #0.55
        xo2_76 = uniform (0.286, 0.875),            #0.5
        co_76 = uniform (0.286, 0.875),             #0.5
        eth_76 = uniform (0.257, 0.789),            #0.45
        par_76 = uniform (0.514, 1.575),            #0.90
#OH + ISOP
        xo2_77 = uniform (0.57, 1.75),              #1.0
        form_77 = uniform (0.57, 1.75),             #1.0
        ho2_77 = uniform (0.38, 1.17),              #0.67
        xn_77 = uniform (0.074, 0.23),              #0.13
        eth_77 = uniform (0.57, 1.75),              #1.0
        mgly_77 = uniform (0.23, 0.7),              #0.4
        c2o3_77 = uniform (0.11, 0.35),             #0.2
        ald_77 = uniform (0.11, 0.35),              #0.2
#O3 + ISOP
        form_78 = uniform (0.57, 1.75),             #1.0
        ald2_78 = uniform (0.23, 0.7),              #0.4
        eth_78 = uniform (0.314, 0.96),             #0.55
        mgly_78 = uniform (0.11, 0.35),             #0.2
        par_78 = uniform (0.057, 0.175),            #0.1
        co_78 = uniform (0.0343, 0.105),            #0.06
        ho2_78 = uniform (0.25, 0.77),              #0.44
        oh_78 = uniform (0.057, 0.175),             #0.1

#concentration uncertainty due to lumping only.  FORM and ALD2 assumed const,
#non-carbonyl species lumped onto them and led to uncertainty

        IC_OLE = lognormal (2.1026e+11, 1.073),     #2.0986e11
        IC_PAR = lognormal (7.060883e+12, 1.045),   #6.9439e12
        IC_TOL = lognormal (1.4955e+11, 1.15),      #1.4785e11
        IC_XYL = lognormal (2.024641e+11, 1.07),    #1.9913e11
        IC_FORM = lognormal (2.6810e+11, 1.008),    #2.6817e11
        IC_ALD2 = lognormal (3.3136e+11, 1.04),     #3.2993e11
        IC_ETH = lognormal (2.4842e+11, 1.2),               #2.4842e11
        IC_ISOP = lognormal (8.7557e+09, 1.2);      #8.7557e9


Uncertain Outputs
        O3, NO, NO2, FORM;

Do Collocation with

  model: CB4,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: linear,
  sampling file : samp.pnt,
  number of points : 5000,
  options: looping, no header;

#end porg.dmc
```

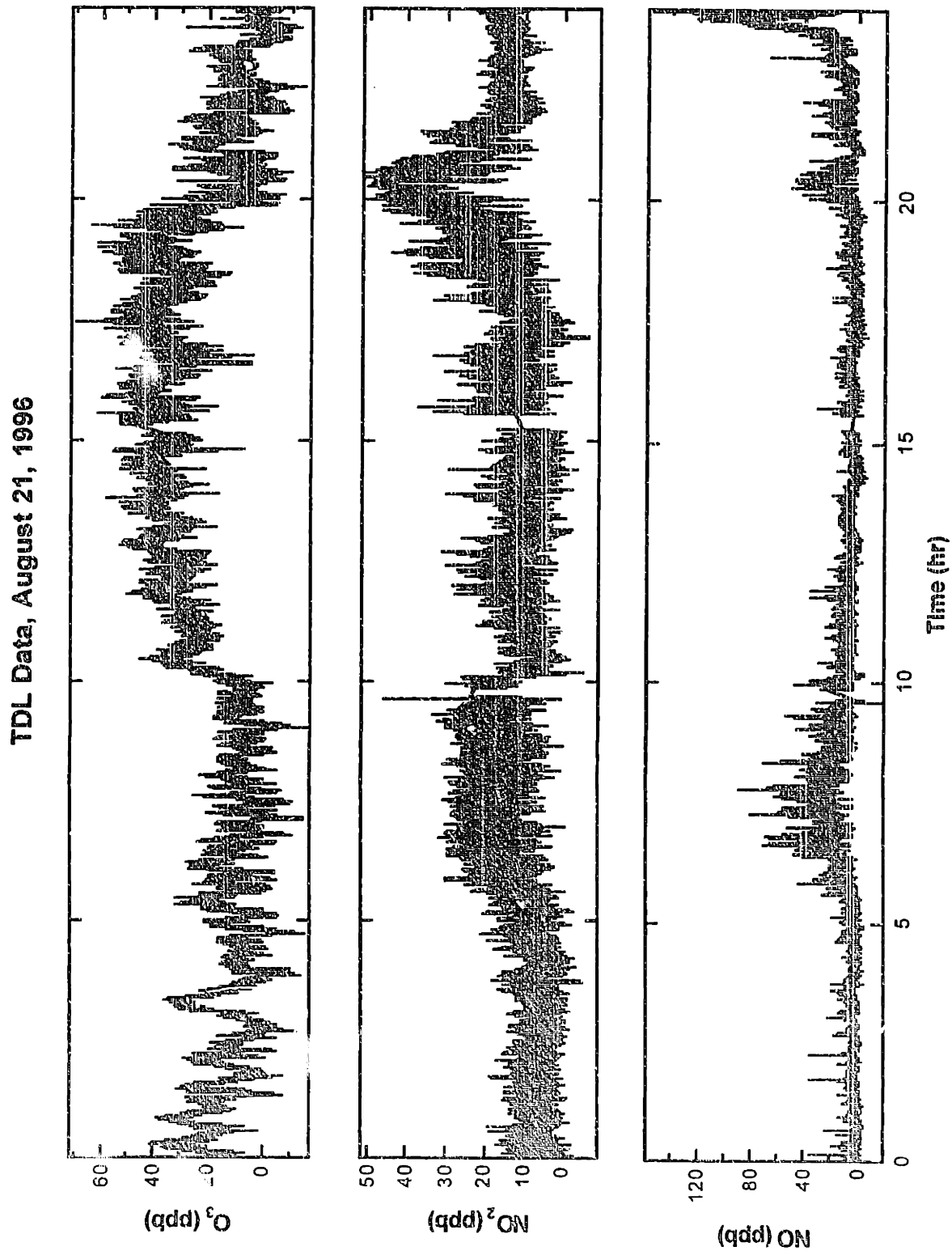
D.5.3 Cb4all.dmc

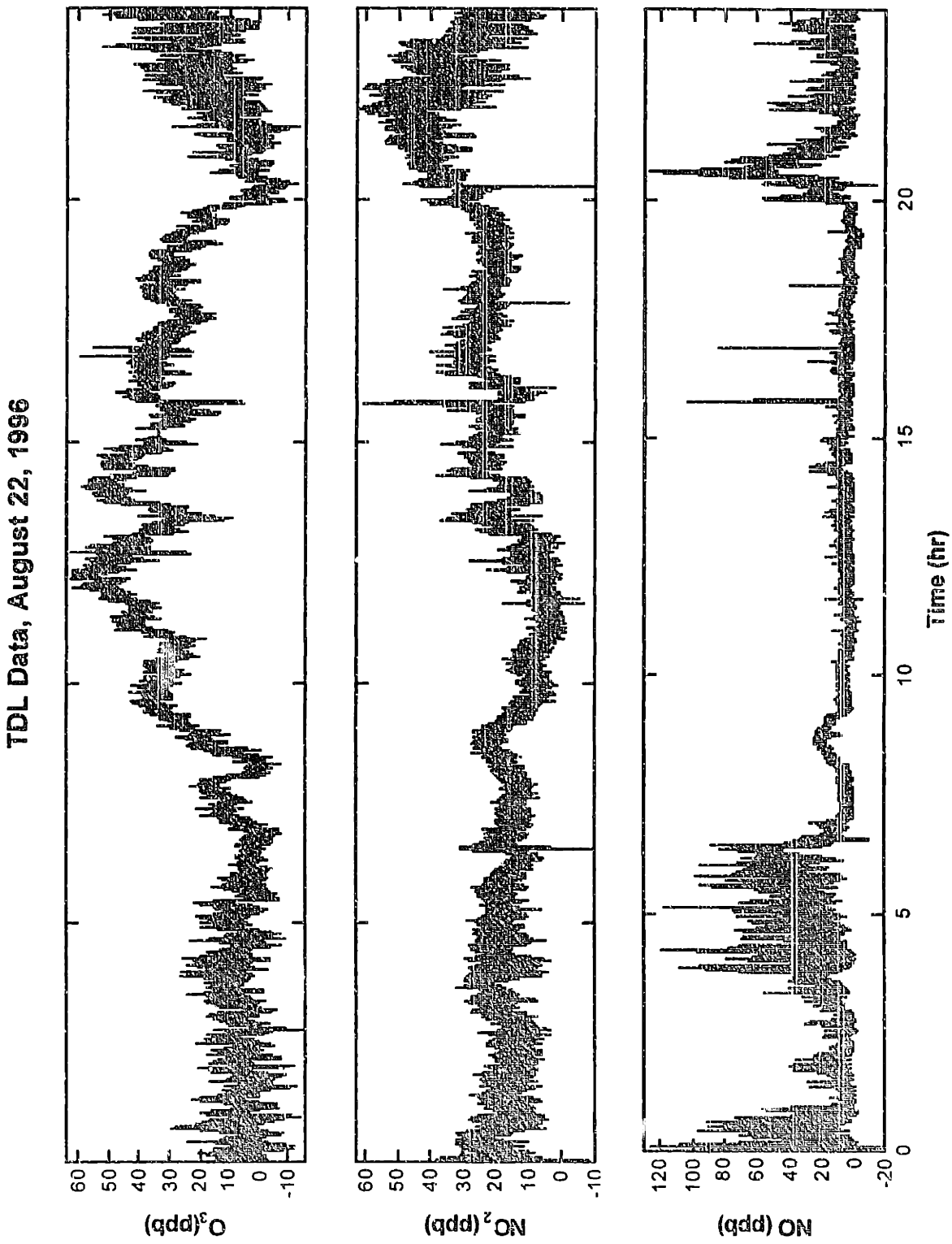
```
# file name: cb4all.dmc
# input date: April 12, 1997
# description: CB4 uncertainty analysis of primary organics + phk + inorganics +
IC

Uncertain Inputs
        jNO2 = lognormal (0.009822, 1.3),
        kNO_O3 = lognormal (1.87066e-14, 1.2),
```

```
        jO3OSD = lognormal (4.55e-05, 1.4),
        rHONO = lognormal (0.1975, 1.6),
        kHO_NO2 = lognormal (1.13245e-11, 1.28),
        jHCHOR = lognormal (3.00667e-05, 1.4),
        jHCHOM = lognormal (4.305e-05, 1.4),
        jALD = lognormal (5.68e-06, 1.4),
        kC2O3_NO = lognormal (1.2333e-11, 2.0),
        kC2O3_NO2 = lognormal (7.3125e-12, 2.0
        kPAN = lognormal (0.00050606, 2.0),
        kPAR_HO = lognormal(8.14E-13 , 1.3),
        kROR = lognormal (2728.0175, 2.0),
        kRORHO2 = lognormal (1.591E+03, 2.0),
        kROR_NO2 = lognormal (1.488E-11, 2.0),
        kOH_OLE = lognormal (2.81155e-11, 1.2),
        kOH_ETH = lognormal(7.98876e-12, 1.15),
        kOH_TOL = lognormal (6.142[illegible]e-12, 1.2),
        kTO2_NO =  lognormal (8.12E-12, 2.0),
        kTO2 = lognormal(4.17, 2.0),
        kHO_XYL = lognormal (2.44363e-11, 1.3),
        rMGLY = lognormal (8.96, 1.6),              # INPUT 22

# PAR + HO
        ror_52 = uniform (0.434, 1.33),          #0.76
#ROR
        ald_54 = uniform (0.629, 1.925),         #1.1
        ho2_54 = uniform (0.537, 1.645),         #0.94
#OH + OLE
        ho2_58 = uniform (0.667, 1.5),           #1.0
        for_58 = uniform (0.667, 1.5),           #1.0
        ald_58 = uniform (0.667, 1.5),           #1.0
#OH + XYL
        ho2_73 = uniform (0.467, 1.05),     #0.7
        mg_73 = uniform (0.53, 1.20),               #0.8
        to2_73 = uniform (0.2, 0.45),               #0.3    INPUT 31

        IC_OLE = lognormal  (2.0986e11, 1.2),
        IC_PAR = lognormal  (6.9439e12, 1.2),
        IC_TOL = lognormal  (1.4785e11, 1.2),
        IC_XYL = lognormal  (1.9913e11, 1.2),
        IC_FORM = lognormal (2.6817e11, 1.2),
        IC_ALD2 = lognormal (3.2993e11, 1.2),
        IC_ETH = lognormal  (2.4842e11, 1.2),
        IC_ISOP = lognormal (8.7557e9 , 1.2),
        IC_NO = lognormal (2.64814e12, 1.2),
        IC_NO2 = lognormal (8.9052e11, 1.2);              #INPUT 41

Uncertain Outputs
        O3, NO, NO2, FORM;

Do Collocation with

  model: CB4,
  dynamic: starting = 0.0, step = 1.0, interval = 10,
  approximation: linear,
  sampling file : samp.pnt,
  number of points : 5000,
  options: looping, no header;

# end allcb4.dmc
```

20.5 Appendix E: TDL Concentration Plots (August 20, 1996 to August 23, 1996)




TDL Data, August 21, 1996
O3 (ppb)
NO2 (ppb)
NO (ppb)
Time (hr)


TDL Data, August 22, 1996
O3 (ppb)
NO2 (ppb)
NO (ppb)
Time (hr)

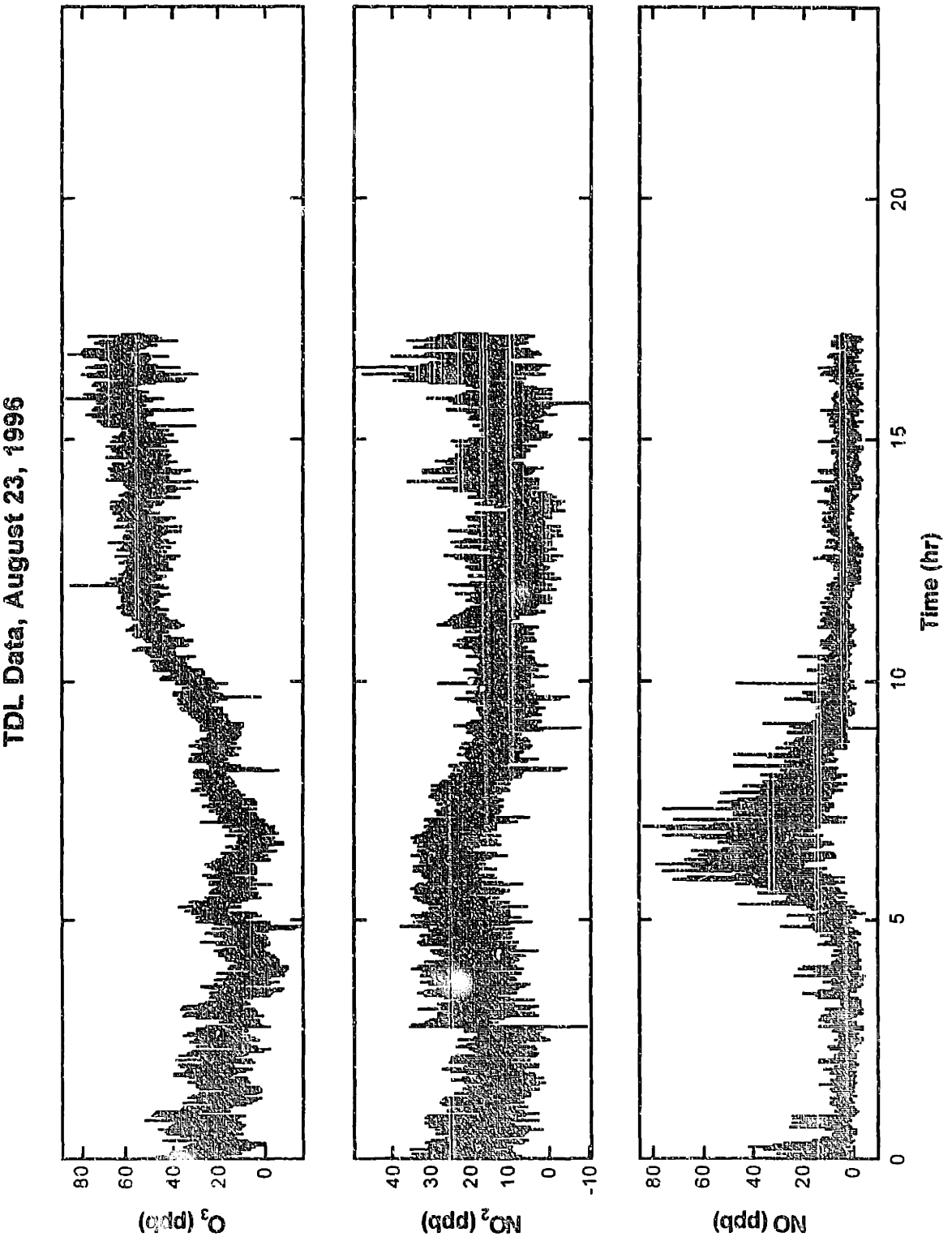


Figure 20-15. Concentrations of Ozone, Nitric Oxide and Nitrogen Dioxide, August 20 to 23, 96.

20.6 Appendix F: Codes for Analysis of Experimental Data

F.1 "Glitch" Removal

```
/*
9/14/96 glitchrm.c

- Each concentration input file start with a start time
- The next line has the headings, which will be ignored.
- The data follows, but only the 3rd and 4th columns are useful.

- Removes glitches in ozone data caused by peak locking at wrong position
  becuase of ramp glitches on 486 computer.

- Assumes that glitches only lasts one correction period
  (consecutive glitches confuse program, fool program into thinking it is one
  peaklocking period by filling in data if consecutive glitches total < Ndata pts
  e.g. I used 400 to fill conc for time 17338, 17340, 17342 on 8/7)
- Assumes that wrong concentration readings are off by a contant factor
- Assumes that the offset factor can be determined by the average of
  the factors at the beginning and end of the peak lock-induced problem period
  (Second reading of the period used if spike due to flow peak lock)
  (First point not printed to output if ozone spike due to flow peak lock)

- If peaklock is not turned on, codes to get around that is in the prog.
- Ozone injection peaks (ppm) removed
- One point glitch removed
  (first point after peak lock removed if spike due to flow peak lock)
- Other glitches that occur within the peaklock correction period not addressed:
  manual adjustment of current (step change in current)
  dewar fill (ramp change in temperature)
*/

#include <stdio.h>
#include <stdlib.h>     /* to use calloc, atof */
#include <string.h>
#include <math.h>

#define Ploc_dur 6.
#define Ndata 150
#define Hspike 8.5
#define Frac_detg 0.15
#define Lgap 20.
#define Ozburst 300.
#define OPG 35.
#define Corr_threshold 30.

/* Ploc_dur = minimum duration of peaklock */
/* Ndata = max number of data point between peak locks */
/* Hspike = absolute height of spike after flow peak lock */
/* Frac_detg = fractional difference used to determine ramp glitch during plock */
/* Lgap = time gap length larger than which real gap and not peak lock */
/* Ozburst = conc above which assume valve toggled to inject burst of ozone */
/* OPG = one point glitch determination.  if readings before and after are
         higher by OPG that point is removed                              */
/* Corr_threshold = concentration below which ramp glitch correction not done */

main(int argc, char *argv[])
{

  double t_start;
  double *t1, *t2, *t3,  *c1, *c2, *c3;
  double *ttmp, *ctmp;
  int i, j;
  int N1, N2, N3;
  float jnk1, jnk2;
  double time, conc;
```

```
  char *filein;
  char fileintmp[ 20], fileout[ 20], filelog[ 20];
  char col_title[ 40], command[ 40];
  FILE *fin, *fout, *flog;

  int index_rgc;
  int ramp_glitch_corr();

  int index_ccs;
  int corr_current_step();

  if (argc == 1){
    printf("please enter ozone file to correct on command line\n");
    exit(0);
  }
  else{
    filein = argv[ 1];
  }

  sprintf(command, "cat %s | tr -d '\015' > %s.tmp\n", filein, filein);
  system(command);

  sprintf(fileintmp, "%s.tmp", filein);
  fin = fopen(fileintmp, "r");

  sprintf(fileout, "%s.gfix", filein);
  fout = fopen(fileout, "w");

  sprintf(filelog, "%s.log", filein);
  flog = fopen(filelog, "w");

  t1 = calloc(Ndata, sizeof(double));
  t2 = calloc(Ndata, sizeof(double));
  t3 = calloc(Ndata, sizeof(double));
  c1 = calloc(Ndata, sizeof(double));
  c2 = calloc(Ndata, sizeof(double));
  c3 = calloc(Ndata, sizeof(double));

  fgets(col_title, 40 , fin);
  sscanf(col_title, "%le", &t_start);
  while (t_start > (3600*24)) t_start = t_start - 3600*24;
  printf("t_start=%le\n", t_start);

  if (argc == 3) t_start = atof(argv[ 2]);

  /* forward past the column title */

  fgets(col_title, 40 , fin);

  N1 = 0;
  N2 = 0;
  N3 = 0;
  i = 0; /* i is the counter for data in input file                    */
  j = 0; /* j is the counter for data in section between peak locks */

  do {                                  /* first time through the loop */
    fscanf(fin, "%f %f %le %le", &jnk1, &jnk2, &time, &conc);

    if (i == 0) {                               /* very first data point */
      t1[ j] = time;
      c1[ j] = conc;
      j++;
      i++;
      fscanf(fin, "%f %f %le %le", &jnk1, &jnk2, &time, &conc);
    }

    if (N1 == 0) {                              /* filling first section */
      if ( ((time - t1[ j-1]) > Ploc_dur) || (j == Ndata) ) {
       /* peak lock ch % t[ j]  & t[ j-1]  or no peaklock in Ndata points */
       if (j == Ndata)
         fprintf(flog, "No peak lock between %le and %le.\n", t1[ 0], t1[ j-1]);
/* else printf ("Sec 1: Peak lock between %le and %le\n", t1[ j-1], time); */
```

```
        N1 = j;
        for (j = 0; j < N1; j ++) fprintf(fout, "%le\t%le\n", t1[ j] +t_start, c1[ j] );
        j = 0;
        t2[ j] = time;
        c2[ j]  = conc;
        i++;
        j++;
       }
       else {
        t1[ j]  = time;
        c1[ j]  = conc;
        i++;
        j++;
       }
     }
     else {
       if (N2 == 0) {                          /* filling second section */
        if (((time - t2[ j-1] ) > Ploc_dur) || ( j == Ndata)) {
                                               /* move to third section  */
          if (j == Ndata)
             fprintf(flog, "No peak lock between %le and %le.\n", t2[ 0] , t2[ j-1] );
/* else printf ("Sec 2: Peak lock between %le and %le\n", t2[ j-1] , time); */

          N2 = j;
          j = 0;
          t3[ j]  = time;
          c3[ j]  = conc;
          i ++;
          j ++;
        }
        else {
          t2[ j]  = time;
          c2[ j]  = conc;
          i ++;
          j ++;

        }
       }
       else {
        if (((time - t3[ j-1] ) > Ploc_dur) || (j == Ndata)) {

          if (j == Ndata)
             fprintf(flog,"No peak lock between %le and %le.\n", t3[ 0] , t3[ j-1] );
/* else printf ("Sec 3: Peak lock between %le and %le\n", t3[ j-1] , time); */

          N3 = j;
          j = 0; /* time, conc to be transferred to t3[ 0]  and c3[ 0]  */
        }
        else {
          t3[ j]  = time;
          c3[ j]  = conc;
          i ++;
          j ++;
        }
       }
     }
   }
   while (N3 == 0);         /* first time through, all arrays filled */

   do {
     if (N3 != 0) {         /* all 3 arrays filled */

       /* Don't bother with corrections if signal is below threshold */

       index_ccs = 0.;
       /* if ((c1[ N1-1] > Corr_threshold) || (c3[ 1]  > Corr_threshold)) { */

       /* current step and ramp within one peaklock period not corrected by program */
       /* index_ccs = corr_current_step(c2, t2, N2, c3[ 1] ); */
       if (index_ccs > 0)
        fprintf(flog,"Current step corrected bet. %le, %le\n",t2[ index_ccs] ,t2[ N2-1] );
       else if (index_ccs < 0)
```

```
        fprintf(flog,"Current ramp not corrected bet. %le, %le\n", t2[ 0], t2[ N2-1] );
       /* ) */

       index_rgc = 0;
       if ((c1[ N1-1] > Corr_threshold) && (c3[ 1] > Corr_threshold)) {
        index_rgc = ramp_glitch_corr(c1, c2, c3, t1, t2, t3, N1, N2, N3);
        if (index_rgc == 1)
           fprintf(flog, "Corrected data between %le and %le\n", t2[ 0], t2[ N2-1] );
      }

       /* print first point of peak lock period if not spike */
       if (c2[ 1] !=0.) {
        if (((c2[ 0] - c2[ 1] ) < Hspike) || (c2[ 0] - c1[ N1-1] < Hspike))
           fprintf (fout, "%le\t%le\n", t2[ 0] +t_start, c2[ 0] );
        else fprintf(flog, "Deleted point after peak lock @time = %le\n", t2[ 0] );
      }
       else  fprintf (fout, "%le\t%le\n", t2[ 0] +t_start, c2[ 0] );

       /* print time, conc to file but remove ozone spikes > Ozburst (300 ppb)
         and one point glitch > 40 ppb diff from before and after */

       for (j = 1; j < N2; j ++) {
        if (c2[ j] < Ozburst) {
           if ((c2[ j+1] - c2[ j] > OPG) && (c2[ j-1] - c2[ j] > OPG) )
              fprintf(flog, "Deleted one point glitch at %le,c=%le\n", t2[ j], c2[ j] );
           else if (c2[ j] == 0.)
              fprintf(flog, "Deleted one point glitch at %le,c=%le\n", t2[ j], c2[ j] );
           else fprintf (fout, "%le\t%le\n", t2[ j] +t_start, c2[ j] );
        }
        else fprintf (flog, "Deleted ozone burst at %le c=%le\n", t2[ j], c2[ j] );
      }
       /* don't print burst of ozone values */

       printf("t2end=%le; t3begin=%le; musical chair!\n", t2[ N2-1], t3[ 0] );
       ttmp = t1;
       ctmp = c1;
       t1 = t2;
       c1 = c2;
       N1 = N2;
       t2 = t3;
       c2 = c3;
       N2 = N3;
       t3 = ttmp;
       c3 = ctmp;
       N3 = 0;
       j = 0;

       t3[ j] = time;
       c3[ j] = conc;
       i ++;
       j ++;
    }

    else {                                        /* filling third section */
       if ((fscanf(fin, "%f %f %le %le", &jnk1, &jnk2, &time, &conc)) == EOF) {
        printf ("in EOF section\n");
        N3 = j;

        if (c2[ 1] != 0.) {
           if ((c2[ 0] - c2[ 1] ) < Hspike) fprintf (fout, "%le\t%le\n", t2[ 0] +t_start,
c2[ 0] );
           else fprintf(flog, "(EOF) Deleted point after peak lock @time = %le\n",
t2[ 0] );
        }

        for (j = 1; j < N2; j++) fprintf(fout, "%le\t%le\n", t2[ j] +t_start, c2[ j] );
        printf ("N2 = %d; t2[ 0] = %le, t2[ N2-1] = %le\n", N2, t2[ 0], t2[ N2-1] );

        if (c3[ 1] != 0.) {
           if ((c3[ 0] - c3[ 1] ) < Hspike) fprintf (fout, "%le\t%le\n", t3[ 0] +t_start,
c3[ 0] );
           else fprintf(flog, "(EOF) Deleted pt after peak lock @time = %le\n", t3[ 0] );
```

```
        }

        for (j = 1; j < N3; j++) fprintf(fout, "%le\t%le\n", t3[ j] +t_start, c3[ j] );
        printf ("N3 = %d; t3[ 0]  = %le, t3[ N3-1]  = %le\n", N3, t3[ 0] , t3[ N3-1] );

        break;
       }

       /* not EOF */
       if (((time - t3[ j-1] ) > Ploc_dur) || (j == Ndata)) {
        if (j == Ndata) printf (" No peak lock between %le and %le\n", t3[ 0] , t3[ j-1] );

        N3 = j;
        j = 0; /* time, conc to be transferred to t3[ 0]  and c3[ 0]  */
       }
       else {
        t3[ j]  = time;
        c3[ j]  = conc;
        i ++;
        j ++;
       }
     }
   }
   while(1);

   fclose(fin);
   fclose(fout);
   fclose(flog);
}


/*****************************************************************************/
/******************* STEP/RAMP CHANGE IN LASER FREQUENCY *********************/
/*****************************************************************************/

/*
Glitches that happen within one peak lock period but more than a single point.
These changes are characterized by a sharp downward movement of
concentration relative to the fluctuations around that time
within one peak lock correction period.  Peak lock brings concentraiton
back up to the level it is supposed to be at.

1. Step change in current occationally occurs when current of laser is changed
not at peak lock to center the peak on the scan
Step change can be corrected by a one-sided factor (avg of correct chunk)
(in case it also coincides with a ramp glitch,
the one sided factor is based on the left side)

2. Ramp change (assumed) in temperature occurs during a dewar fill.
(Ramp changes correction _not coded_ as of 9/19/96. Need to think about it.)
*/

int corr_current_step(c2, t2, N2, c3)
double *c2, *t2;
int N2;
double c3;
{

   /* returns i, t[ i]  is when glitch first occurs if step change is found   */
   /* returns 0 if no step change is found                                    */
   /* returns -1 if a ramp is likely (not corrected)                          */
   /* downcounter keeps track of how many concentration points are moving     */
   /* down after glitch is found.  If downcounter > 0.75*n2, ramp is likely */

   int i, j, n1, n2, downcounter, testindex;
   double fcorr, sum1, sum2;

   /* testindex for third condition in else if */

   n1 = 1;
   n2 = 0;
   sum2 = 0.;
```

```
    downcounter = 0;

    sum1 = c2[ 1];        /* ignore first point after peak lock */

    for (i = 2; i < N2; i ++){
      if (c2[ i] < Ozburst){
        if ((sum1/n1 - c2[ i] < Hspike) ) {

         sum1 = sum1 + c2[ i];
         n1 ++;
        }
        else {
         /* printf ("Possible current step at %le\n", t2[ i]); */
         for (j = i; j < N2; j ++){
           if (c2[ i] < Ozburst) {
             sum2 = sum2 + c2[ j];
             n2 ++;
             if (c2[ j] < (0.7 * sum2/n2)) downcounter ++;
           }
         }

         if (i < (N2-3))testindex = 2;
         else testindex = N2-1-i;

         if (((float)(downcounter)) > ((float)(0.75*n2))) {
             /* ramp change in conc */
           break;
         }

         else if ( ((sum1/n1 - sum2/n2) > Hspike) && ((c3 - sum2/n2) > Hspike) &&
 ((c2[ i-testindex] - c2[ i+testindex]) > Hspike) ) {

           /* add a third condition is added so that a step in a decay
              (with a higher conc after after peak lock) will not be treated
              as a current change.  Such a phenomena occurred at night 8/7 */


           /* really is a glitch */
           /* fcorr = c2[ i-1] /c2[ i]; */
           /* above not good if not > 1 data point % good and bad data */

           fcorr = (sum1/n1) / (sum2/n2);
           for (j = i; j < N2; j ++) c2[ j] = c2[ j] * fcorr;
                 break;
         }

         else {                                          /* not a glitch */
           sum2 = 0.;
           n2 = 0;
           downcounter = 0;
           sum1 = sum1 + c2[ i];
           n1 ++;
         }
        }
      }

    }

    if (((float)(downcounter)) > ((float)(0.75*n2)) ) return (-1);

    else if (sum2 != 0.) return (i);
    else return (0);
}


/*************************************************************************/
/********************** RAMP GLITCH CORRECTION ROUTINE ********************/
/*************************************************************************/


int ramp_glitch_corr(c1, c2, c3, t1, t2, t3, N1, N2, N3)
double *c1, *c2, *c3, *t1, *t2, *t3;
```

```
int N1, N2, N3;
{
  /* determine if there is a glitch.  If glitch, correct data. */
  /* returns 1 if ramp glitch correction happened, 0 if not */

  int j, n;
  double fd1, fd2;
  double fcorr;
  double sum1, sum2a, sum2b, sum3;
  int i1, i2a, i2b, i3;

  fd1 = 0.;                          /* default, no correction */
  fd2 = 0.;
  fcorr = 1.;
  sum1 = 0.;
  sum2a = 0.;
  sum2b = 0.;
  sum3 = 0.;
  i1 = 0;
  i2b = 0;
  i2a = 0;
  i3 = 0;

/* Use average of one quarter of a peak lock period */

  if (N1 < N2) n = (int)(N1/4);
  else n = (int)(N2/4);

  if (n < 1) n = 1;

  for (j = (N1-1) ; j >= ((N1-1)-n); j --) {
    if ((c1[ j] > 0.0) && (c1[ j] < Ozburst)) {
      sum1 = sum1 + c1[ j];
      i1 ++;
    }
  }
  for (j = 1; j < (n+1); j ++) {
    /* ignore first point after peak lock */
    if ((c2[ j] > 0.0) && (c2[ j] < Ozburst)) {
      sum2a = sum2a + c2[ j];
      i2a ++;
    }
  }


  if (N2 < N3) n = (int)(N2/4);
  else n = (int)(N3/4);
  if (n < 1)  n = 1;

  for ( j = (N2-1); j >= ((N2-1) - n); j --) {
    if ((c2[ j] > 0.0) && (c2[ j] < Ozburst)) {
      sum2b = sum2b + c2[ j];
      i2b ++;
    }
  }

  for (j = 1; j < (n+1); j ++) {
    if ((c3[ j] > 0.0) && (c3[ j] < Ozburst)) {
      sum3 = sum3 + c3[ j];
      i3 ++;
    }
  }

  if ( (i1 !=0) && (i2a !=0) && ((sum2a/i2a) > 0.)  ) {

    if (((sum1/i1) - (sum2a/i2a)) > Hspike)
      fd1 = ((sum1/i1) - (sum2a/i2a)) / (sum2a/i2a);
  }

  if ( (i2b !=0) && (i3!=0) && ((sum2b/i2b) > 0.)) {

    if (((sum3/i3) - (sum2b/i2b)) > Hspike)
```

```
        fd2 = ((sum3/i3) - (sum2b/i2b)) / (sum2b/i2b);
    }

    /* Make sure it is a peak lock and not a gap in the data */
    if ((t2[ 0] - t1[ N1-1] > Lgap) || (t3[ 0] - t2[ N2-1] > Lgap)) {
      fd1 = 0.0;
      fd2 = 0.0;
    }

    /* Make sure peak lock: filling contiues when one not found in 150 points */
    if ((t2[ 0] - t1[ N1-1] < Ploc_dur) || (t3[ 0] - t2[ N2-1] < Ploc_dur)) {
      fd1 = 0.0;
      fd2 = 0.0;
      printf("No peak lock found b/t %le, %le\n", t2[ 0], t2[ N2-1] );
    }

    /* criteria for ramp glitch determination */
    if  ((fd1 > Frac_detg) && (fd2 > Frac_detg)) {
      if ((sum2a != 0) && (sum2b != 0) && (i1 != 0) && (i3 != 0))
        fcorr = (sum1 * i2a /(i1 * sum2a) + sum3 * i2b /(i3 * sum2b)) /2;

      /* if ((c2[ 1] != 0) && (c2[ N2-1] != 0)) */
      /* fcorr = ((c1[ N1-1] /c2[ 1] ) + (c3[ 1] /c2[ N2-1] )) /2; */
      /* the above correction factor is not good if we happen to be right in
         the middle of a slow feature during peak lock */

      for (j = 0; j < N2; j ++) c2[ j] = c2[ j] * fcorr;
    }

    if (fcorr != 1.) return (1);
    else return (0);

}
```

F.2 Interpolate Uneven Data onto Time Grid

F.2.1 Main Matlab Script (Perfect.m)

```
function [ gdtime, gdconc] = perfect(series)

% PERFECT(time_series)
% takes an array of rows time, conc
% returns a stream of evenly sampled data with gaps filled by a 3-point
% predictor. Output in 2 vectors -- usage: [ gridtime ,gridconc] =perfect(data);

% get rid of duplicate samples at the same time (HOW ON EARTH DID THAT HAPPEN?)
% also copied into own subroutine called fixduptime.m

delta_time = diff (series(:,1));
[ mindt, i] = min(delta_time);
while (mindt == 0.)
% i-th and (i+1)-th reading has the same time -- move one of them
        if ((series(i, 1)-series(i-1,1)) > (series(i+2, 1)-series(i+1,1)))
                series(i,1) = (series (i,1)+series(i-1,1))/2;
        else
                series(i+1,1) = (series(i+1,1)+series(i+2,1))/2;
        end

        delta_time = diff (series(:,1));
        [ mindt, i] = min(delta_time);
end

ttime = series(:,1);
cconc = series(:,2);

DeltaT=SampRate(ttime);
% for NOx on 8/20/97 DeltaT=1.2721295415636389;
```

```
% for O3 on 8/20/97 DeltaT = 1.6699999999982538  ;

% gridded time
gdtime = (ttime(1): DeltaT: ttime(length(ttime)))';
disp ('made gdtime')

gr_C = Snap_to_grid (ttime, cconc, DeltaT);
disp ('snapped to grid')
figure;
plot (gr_C);

fgr_C = SingleMiss(gr_C);
disp ('filled in single miss')
figure;
plot (fgr_C);

gdconc = FixGap(fgr_C);
disp ('done fix gap')

figure;
plot (series(:,1), series(:,2), 'y-');
hold on;
plot (gdtime, gdconc, 'c*');
hold off;
```

F.2.2 Samp_Rate.m

```
function r = SampRate(time)
% SAMPRATE determines the dominant sample rate of a data stream
% takes a column of time as input; output sampling rate of continuous data
% by calculating time differences between data points and recording the
% delta_t with at least a certain percentage (dom_frac) appearance.
% if no delta_rate satisfies the dom_frac requirement, a mean will be
% taken of the middle range of delta_t data

dom_frac = 0.7;
% fraction about which a sampling rate accepted as dominant

TOL = 0.1001;
% percent accuracy tolerance of sampling rate

delta_time = diff(time);

maxdt = max(delta_time);
r = 0;

while (maxdt >= min(delta_time))
        i = 1;
        count = 0;
        sum = 0;

        mindt = min(delta_time);

        while (i <= length(delta_time))
                if (delta_time(i) <= ((1. + TOL) * mindt))
                        count = count + 1;
                        sum = sum + delta_time(i);
                end
        i = i + 1;
        end

        if (count > (dom_frac * length(delta_time)))
                r = sum/count
%               r = min(delta_time);
        elseif (count > (1-dom_frac)*length(delta_time) )
                disp('no dominant sampling rate found')
                tmp = sort(diff(time));
                index = round((1-dom_frac)*length(delta_time)/2);
                r = mean(tmp(index):tmp(length(delta_time)-index))
```

```
            break;
        else
            i = 1;
            while (i <= length(delta_time))
                if (delta_time(i) <= ((1. + TOL) * mindt))
                    delta_time(i) = 999.;
                end
            i = i + 1;
            end
        end
        if (r > 0), break, end;
end
```

F.2.3 Snap_to_grid.m

```
function [ gc] = Snap_to_grid(time, conc, dt)

% SNAP_TO_GRID gridded data stream based on sample time and precscribed dt
% Usage: Snap_to_grid(sample_time, sample_conc, desired_time_interval)
% Recommend using SampRate to identify suitable dt
% each sample point is matched to gridded time closest to it.
% gaps are labelled 999. for concentration

gridt = time(1): dt: time(length(time));

gi = 1;
si = 1;
ptgap = 0;

while (gi <= length(gridt) )
        if ( ( time(si) > gridt(gi)-dt/2 ) & ( time(si) <= gridt(gi)+dt/2 ) )
        % within grid
            gc(gi) = conc(si);
            if (si < length(time))
                si = si + 1;
            else
                break;
            end
            % in case more than one si pts match to same gi pt
        elseif (time(si) < gridt(gi)-dt/2)
        % belong to last grid -- ignore -- forward to this grid, assign
            while ( time(si) <= gridt(gi) + dt/2)

                gc(gi) = conc(si);

                if (si < length(time))
                    si = si + 1;
                else
                    break;
                end
            end

            % out this loop when time(si) past gridt(gi), gridt assigned

        else
            % time (si) too big already, there is a gap
            gc(gi) = 999.;
        %   disp(gi);
        end

        gi = gi + 1;
end
```

F.2.4 SingleMiss.m

```
function [ fgc] = SingleMiss(gc)

% deal with single misses by interpolation
% since first and last element should never be 999, start with 2nd element

i=2;
while (i <= length(gc)-1)
        if (gc(i) == 999.)
                if(gc(i+1) < 999.)
                        gc(i) = 0.5 * (gc(i+1) + gc(i-1));
                        i = i + 1;
                else
                        while (gc(i) == 999.)
                                i = i + 1;
                        end
                end
        else
                i = i + 1;
        end
end

if (gc(length(gc)) == 999.)
        gc(length(gc)) = gc(length(gc)-1);
end

fgc = gc;
```

F.2.5 FixGap.m

```
function [ cn] = FixGap(c)

% FixGap(gridded time series with gap labeled by 999)
% Uses autoregressive interpolation, from 3 points on
% each side of gap

i=1;
cn = c;

while (i <= length(c) )
        if (c(i) == 999.)
%               disp(i)
% take autocovariance function of corrected data
                ctmp = cn(1:i-1);
                ac = xcov(ctmp,'unbiased');
                ac(1:length(ctmp)-1)=[];

                ngap = 1;
                while (c(i + ngap) == 999.)
                        ngap = ngap +1;
                end

                % cbi * coef(x)  = c^
                cbi=[ cn(i-3) cn(i-2) cn(i-1) c(i+ngap) c(i+ngap+1) c(i+ngap+2)];

                % c(i+ngap) does not equal 999, but problems arise if
                % c(i+ngap+1) or c(i+ngap+2) = 999)  The following treatment
                % is lame, but it might work...

                if (c(i+ngap+1) == 999.)
                        % only c(i+ngap+1) = 999
                        cbi=[ cn(i-3) cn(i-2) cn(i-1) c(i+ngap) c(i+ngap) c(i+ngap+2)];
                end

                if (c(i+ngap+2) == 999.)
                        cbi=[ cn(i-3) cn(i-2) cn(i-1) c(i+ngap) c(i+ngap+1) c(i+ngap)];
                end
```

```
            if ((c(i+ngap+2) == 999.) & (c(i+ngap+2) == 999.))
                % both c(i+ngap+1) and c(i+ngap+2) = 999
                % overrule prev. cases
                cbi=[ cn(i-3) cn(i-2) cn(i-1) c(i+ngap) c(i+ngap) c(i+ngap)];
            end

            A=[ ac(1) ac(2) ac(3) ac(3+ngap+1) ac(3+ngap+2) ac(3+ngap+3);
                ac(2) ac(1) ac(2) ac(2+ngap+1) ac(2+ngap+2) ac(2+ngap+3);
                ac(3) ac(2) ac(1) ac(1+ngap+1) ac(1+ngap+2) ac(1+ngap+3);
                ac(1+ngap+3) ac(1+ngap+2) ac(1+ngap+1) ac(1) ac(2) ac(3);
                ac(2+ngap+3) ac(2+ngap+2) ac(2+ngap+1) ac(2) ac(1) ac(2);
                ac(3+ngap+3) ac(3+ngap+2) ac(3+ngap+1) ac(3) ac(2) ac(1)];
            j = 0;
            while (j < ngap)
                b=[ ac(j+1+3) ac(j+1+2) ac(j+1+1) ac(ngap-j+1) ac(ngap-j+2)
ac(ngap-j+3)]';
                % x=A\b
                % small x's => try constraint mini with sum x = 1
                x=qp(A'*A, (-A'*b), [ 1 1 1 1 1 1], 1, [],[],[],1);
                cn(i+j) = cbi * x;
                j = j + 1;
            end
            i = i + ngap;
        end
        i = i + 1;
end
```

F.3 Concentration Data Analysis (Matlab) Scripts

F.3.1 Hourly Data Main Script

```
function [ hourlydat] = houravg(data)

% houravg (or other denominations) takes one day's worth of data
% input 2-D matrix (time, conc)
% calculating hourly average and other statistical descriptors

fac = 1.;       % default: take hourly averages
%fac = 0.5;     % take half-hour averages


% hour 0 to 23, hourlydat index 1 to 24
% halfhour 0 to 47, hourlydat index 1 to 48
N = 24 / fac - 1;

index1 = 1;
hour = 0;

while (hour <= N)

        index2 = index1;

        if (index2 <= length (data))

                while (data(index2,1) < ((hour+1) * 3600 * fac))

                        if (index2 <= length(data))

                                index2 = index2 + 1;

                        end

                        if (index2 == (length(data)+1) )
                                break;
                        end
```

```
            end
        end

        if (index2 > index1)

            datatmp = data(index1:(index2-1), :);
            % datatmp containts data for the hour

            % orient landscape subplot (4,6,(hour+1));
            subplot (6,4,(hour+1)); %orient tall

            % remove trend
%           smoothd = smooth(datatmp);
%           datatmp(:,2) = datatmp(:,2) - smoothd;

            % for hourly data only -- reduce to min by min data
%           mintmp = mindat(datatmp);

%           hist(datatmp(:,2), 50);
%           prob = myhist(datatmp(:,2));

            freq = 1/(datatmp(2,1) - datatmp(1,1));

            [ f, p] = my_specplot(datatmp(:,2), freq);

%           ac = mycov(datatmp);

            hourlydat(hour+1, 1) = hour * fac;
            hourlydat(hour+1, 2) = mean (datatmp(:,2));
            hourlydat(hour+1, 3) = std (datatmp(:,2));

%           other outputs:
%           hourlydat(hour+1, 4) = mean (mintmp);
%           hourlydat(hour+1, 5) = std (mintmp);
%           hourlydat(hour+1, ?) = length (datatmp);
%           hourlydat(hour+1, ?) = prob;

            clear datatmp;

            index1 = index2;
        else

            hourlydat(hour+1, 1) = hour*fac;
            hourlydat(hour+1, 2) = 0.;
            hourlydat(hour+1, 3) = 0.;
%           hourlydat(hour+1, 4) = 0.;
%           hourlydat(hour+1, 5) = 0.;

        end


        hour = hour + 1;

end
```

F.3.2 Trend Removal

```
function [ smooth] = smooth ( data )

% outputs smoothed data(:,2)
% to remove long term wobbles from one time, conc data stream
% use data(:,2) - smooth
% decimate then spline
% r = 50 point decimation smooth line has too much fluctuations
% r = 100 point decimation smooth line falls behind

r = 300;

dd(:,2) = decimate (data(:,2), r, 'fir');
```

```
dd(:,1) = decimate (data(:,1), r, 'fir');
dd(length(dd), :) = data(length(data),:);
smooth = spline (dd(:,1), dd(:,2), data(:,1));


%plot ((data(:,1)-data(1,1)), data(:,2), 'y', (data(:,1)-data(1,1)), smooth, 'm');
```

F.3.3 MinDat.m

```
function [ outdatnz] = mindat(data)

% takes one hour's worth of data (time, conc) and calculates
% minute by minute averages, outdatnz deletes all zero elements
% function [ outdat] = mindat(data) outputs 60 elements, including 0

hr = fix(data(1,1)/3600);

data(:,1) = data(:,1) - hr*3600;

j = 1;

for i=1:60

        mindat = 0;
        n = 0;

        while(data(j,1) < (i*60))

                mindat = mindat + data(j,2);
                n = n + 1;
                if (j < length(data))
                        j = j + 1;

                else
                        break;
                end
        end

        if (n > 0)
                outdat(i) = mindat / n ;
        end

end

%fix zero dat -- outdat fix may have less than 60 elements

j = 1;
for (i = 1:60)
        if ~(outdat(i) == 0.)
                outdatnz(j) = outdat(i);
                j = j +1;
        end
end
```

F.3.4 Histogram Fitting

F.3.4.1 Myhist.m

```
function [ prob] = myhist (data)

% makes histogram of data
% fit to distributions (normal, lognormal(flip -ve data) )
% output statistics, plot graph of both histogram and fit dist.
```

```
nbin = 50;

% for lognormal, weibull, and gamma fits:
% need to fix data so so that negative data won't cause problem
% data = abs(data);

[ n, hc] = hist (data, nbin);

% fit normal
freqfit = fitnormdist(data, hc);

% fit lognormal
% freqfit = fitlognordist(data, hc);

% fit gamma
% freqfit = fitgamdist(data, hc);

% fit weibull
% freqfit = fitweidist(data, hc);

prob = chisqtest (n, freqfit)

% turn histogram into pdf (normalize n)
% area under curve = data point * bin width
nnor = n / length(data)/ (hc(2) - hc(1));

% turn freq fit back to pdf
pdffit = freqfit / length(data)/ (hc(2) - hc(1));

% plot pdf
bar (hc, nnor);

hold on

% plot distribution fit
plot (hc, pdffit, 'm');
legend ('m', 'normal fit');
% legend ('m', 'lognormal fit');

hold off
```

F.3.4.2 FitNormDist.m

```
function [ freqfit] = fitnormdist(data, hc)

% takes a stream of data and fits a normal distribution
% parameter estimation using matching moment method
% output fitted normal distribution (freq fit)

% fit normal by matching moments
fm = mean (data);
sm = std (data);

% normfit = normpdf(hc, fm, sm);


% Use trapezoid rule to determine fit
hcedge = hc - (hc(2)-hc(1))/2;
hcedge(length(hc)+1)= hcedge(length(hc)) + hc(2) - hc(1);


% By trapezoid rule: calc. pdf at edge of bin
normfitedge = normpdf(hcedge, fm, sm);
normfit = (normfitedge(2:length(normfitedge))+normfitedge(1:length(normfitedge)-1))/2;

% turn normpdf into histogram frequency fit
freqfit = normfit * length(data) * (hc(2)-hc(1));
```

F.3.4.3 FitLognorDist.m

```
function [ freqfit] = fitlognordist(data, hc)

% takes a stream of positive data and fits a lognormal distribution
% parameter estimation using max likelyhood
% output fitted lognormal distribution (freq fit)

nbin = length(hc);

% fit lognormal by max likelihood
lndata = log(data);
fmln = mean (lndata);
smln = std (lndata);
cv = exp(fmln);
uf = exp(smln);

% Use trapezoid rule to determine fit
hcedge = hc - (hc(2)-hc(1))/2;
hcedge(length(hc)+1)= hcedge(length(hc)) + hc(2) - hc(1);

% Trapezoid rule: calc. pdf at edge of bin
for i = 1:(nbin+1)
        if (hcedge(i) > 0)
                lognorfitedge(i)= 1/(smln*hcedge(i)*sqrt(2*pi))*exp(-(log(hcedge(i))-
fmln)^2/(2*smln^2));
        else
                lognorfitedge(i) = 0;
        end
end
lognorfit =
(lognorfitedge(2:length(lognorfitedge))+lognorfitedge(1:length(lognorfitedge)-1))/2;

% turn lognormpdf into histogram frequency fit
freqfit = lognorfit * length(data) * (hc(2)-hc(1));
```

F.3.4.4 FitGammDist.m

```
function [ freqfit] = fitgammdist(data, hc)

% takes a stream of data and fits a gamma distribution
% parameter estimation using matching moment method
% output fitted gamma distribution (freq fit)

% fit normal by matching moments
fm = mean (data);
sm = std (data);
k = fm^2/sm^2;
lambda = fm/sm^2;

% Use trapezoid rule to determine fit
hcedge = hc - (hc(2)-hc(1))/2;
hcedge(length(hc)+1)= hcedge(length(hc)) + hc(2) - hc(1);


% By trapezoid rule: calc. pdf at edge of bin
gammfitedge = gampdf(hcedge, k, 1/lambda);
gammfit = (gammfitedge(2:length(gammfitedge))+gammfitedge(1:length(gammfitedge)-1))/2;

% turn normpdf into histogram frequency fit
freqfit = gammfit * length(data) * (hc(2)-hc(1));
```

F.3.4.5 FitWeiDist.m

```
function [ freqfit] = fitweidist(data, hc)

% takes a stream of data and fits a weibull distribution
% parameter estimation using method outlined in Morgan and Henrion Table 5.6
% Trapezoid rule not used becuase point when cdf = 1 screws up param. est.
% output fitted weibull distribution (freq fit)

% find cdf
sd = sort (data);

% parameter estimation
j = 1;
for i = 1:length(hc)

        while sd(j) < hc(i)
                j = j+1;
        end

        cdf(i) = j / length(data);
end

hcnew = hc;
cdfnew = cdf;

while (cdfnew(length(cdfnew)) > 0.9998)
        hcnew = hc(1:length(hcnew)-1);
        cdfnew = cdf(1:length(cdfnew)-1);
end


y = log(hcnew);
z = log (-log (1-cdfnew));

[ p, S] = polyfit(y,z,1);

khat = p(1)
chat = exp(-p(2)/p(1))

% matlab weipdf is not the same form as Morgan's

for i = 1:length(hc)
        weifit(i) = khat/chat*(hc(i)/chat)^(khat-1)*exp(-(hc(i)/chat)^khat);
end

% turn normpdf into histogram frequency fit
freqfit = weifit * length(data) * (hc(2)-hc(1));
```

F.3.4.6 FitBiMod.m

```
function [ freqfit] = fitbimod(data, chopv)

% takes an hour of bimodal data, fit double normal parameter estimation using
% matching moment method output fitted normal distribution (freq fit)

nbin = 50;
i = 1;

[ n, hc] = hist (data, nbin);
sorteddata = sort(data);

while (sorteddata(i) < chopv)
        i = i +1;
end

data1 = sorteddata(1:i-1);
data2 = sorteddata(i:length(sorteddata));
```

```
% fit normal by using modes as means
fm1 = mean (data1)
sm1 = std (data1);
fm2 = mean (data2)
sm2 = std (data2);

% Use trapezoid rule to determine fit
hcedge = hc - (hc(2)-hc(1))/2;
hcedge(length(hc)+1)= hcedge(length(hc)) + hc(2) - hc(1);

% By trapezoid rule: calc. pdf at edge of bin
normfitedge1 = normpdf(hcedge, fm1, sm1);
normfitedge2 = normpdf(hcedge, fm2, sm2);

normfit1 = (normfitedge1(2:length(normfitedge1))+normfitedge1(1:length(normfitedge1)-
1))/2;
normfit2 = (normfitedge2(2:length(normfitedge2))+normfitedge2(1:length(normfitedge2)-
1))/2;

% turn normpdf into histogram frequency fit
freqfit1 = normfit1 * length(data1) * (hc(2)-hc(1));
freqfit2 = normfit2 * length(data2) * (hc(2)-hc(1));
freqfit = freqfit1 + freqfit2;

prob = chisqtest(n, freqfit)

% turn histogram into pdf (normalize n)
% area under curve = data point * bin width
nnor = n / length(data)/ (hc(2) - hc(1));

% turn freq fit back to pdf
pdffit = freqfit / length(data)/ (hc(2) - hc(1));

% plot pdf
bar (hc, nnor);

hold on

% plot distribution fit
plot (hc, pdffit, 'm');

hold off
```

F.3.4.7 ChiSqTest.m

```
function [ prob] = chisqtest (n, freqfit)

% chi square test
% input: frequency of binned data and fit using known distribution
% output probability Q

chisq = 0;
nbin = length(n);

for i = 1:nbin
       chisq = chisq + (n(i) - freqfit(i))^2 / freqfit(i);
end

degfree = (nbin - 3);
% normfit based on (1) sample mean, (2) std. dev, and (3) normalized to length(data)

%disp ('probability that chi square for correct model less than calc. chi sq')
%disp ('the smaller the probability the better the model')
%prob = chi2cdf (chisq, degfree)

disp ('probability that chi sq for correct model greater than calc. chi sq')
disp ('the larger the probability the better the model')
prob = 1 - chi2cdf (chisq, degfree)
if (prob > 0.01) disp('reasonable model'); end
```

F.3.5 My_Specplot.m

```
function [ F, P] = my_specplot(X, Fs)

% This version takes an input of a uniformly gridded data stream and the sampling
% frequncy (1/s) calles the spectrum function as well as plots the resulting spectrum.
%
% SPECPLOT Plot the output of the SPECTRUM function.
%       SPECPLOT(P,Fs), uses P, the output of SPECTRUM, and Fs, the
%       sample frequency, to successively plot:
%               Pxx - X Power Spectral Density & confidence.
%               Pyy - Y Power Spectral Density & confidence.
%               abs(Txy) - Transfer Function Magnitude.
%               angle(Txy) - Transfer Function Phase.
%               Cxy - Coherence Function.
%       The 95% confidence intervals are displayed on the power
%       spectral density curves.
%
%       SPECPLOT(P) uses normalized frequency, Fs = 2, so that 1.0 on
%       the frequency axis is half the sample rate (the Nyquist frequency).
%
%       Author(s): J.N. Little, 7-9-86
%                  J.N. Little, 11-14-91, revised
%       Copyright (c) 1984-94 by The MathWorks, Inc.
%       $Revision: 1.6 $  $Date: 1994/01/25 17:59:58 $

NFFT = 256;           %default = 256
WINDOW = hanning(256);
NOVERLAP = 0;
P = 0.95;

% [ P,F] = SPECTRUM(X, NFFT, NOVERLAP, WINDOW, Fs)
[ Pxx, Pxxc, F] = psd(X, NFFT, Fs, WINDOW, NOVERLAP, P);

line = 1.* F.^(-2/3);
line1 = 0.05* F.^(-5/3);
newplot;

loglog(F, Pxx, 'r'), ...
%plot(F, Pxx, 'r'), ...
hold on
%loglog(F, [ Pxx Pxx-Pxxc Pxx+Pxxc], 'y'), ...
%plot(F, [ Pxx Pxx-Pxxc Pxx+Pxxc], 'y'), ...
loglog(F, line, 'c--')
loglog(F, line1, 'c--')
hold off

%xlabel('Frequency')
grid on
```

F.3.6 My_Conv.m

```
function [ ac] = mycov ( data )

% calculates unbiased autocorrelation function and plots it
% from input of 2 columns  time, data

ac = xcov(data(:,2),'unbiased');
ac(1:length(data(:,2))-1)=[ ];

tau = data(:,1) - data (1,1);

plot (tau, ac, 'y-');
```

F.4 Calculation of Wind Speed and Direction from Sonic Anemometer Data (Wind.m)

```
function [ speed, dir] = wind (u, v)

%  Modified from wind.c
%  This program takes u and v and returns speed and dir (whole circle bearing)
%  Boom Asmuth Angle (BAA) of 159 degrees added

BAA = 159;

speed = sqrt( u.*u+v.*v );

for (i=1:length(u))

       if (u(i) == 0)
              if (v(i)<0)
                     dir(i)=90;
              else
                     dir(i)=270;
              end


       else
              au = abs(u(i));
              av = abs(v(i));

              beta = av/au;


              if (u(i)>0)
                     if (v(i)>0)
                            dir(i)=360-atan(beta)*57.296 ;
                     else
                            dir(i)=atan(beta)*57.296 ;
                     end
              else
                     if (v(i)>0)
                            dir(i)=180+atan(beta)*57.296 ;
                     else
                            dir(i)= 180-atan(beta)*57.296 ;
                     end %if (v(i) > 0) else
              end % if (u(i) > 0) else
       end % if (u(i) == 0) else
end % for

dir = dir + BAA;
for (i = 1:length(dir))
       if (dir(i)>360)
              dir(i)=dir(i)-360;
       end
end
```

F.5 MactchSCTime.m

```
% matchsctime
% [ newtimestamp] = matchsctime (times, W, conc)

% matches sonic computer time to ozone computer time
% 1a. needs lookup table, file with both ozone and sonic time
%      e.g. 222220C.OUT, cut to 2 columns with corresponding
%      times (for that hour) from 2 computers times = (too, tos) =
%      (time of ozone measurements at oz_comp_time, sonic_time)
%      matched conc stream, (e.g., conc.matched.22H6.22pm.str)
%      and ascii wind file (e.g., W.222202S.asc)
% 1b. The times and met measurements on sonic computer is needed.
```

```
% 1c. lag is also implemented here.
% 1d. Entire sonic data file gets new time stamps
%      that is consistent with ozone time stamp.
% 2.  Since the sonic data file or UV data file
%      will be huge, it is probably best to deal with as small
%      a set of data as possible.  Cut down to the hour.
% 3.  Still too long -- decimate by factor of 9
% 4.  Match to concentration times.
%      (matched conc file need to be for that hour only)

hour = 22;
lag = 4.5;

% make sure 0 < ozone computer time < 24*3600
while (times(1,1) > 24*3600)
       times(:,1) = times(:,1) - 24*3600;
end

% select data for table
basetimes = grephr(times, hour);

[p,S] = polyfit(basetimes(:,2), basetimes(:,1), 1);

% if entered integer values of W, make sure  0 < sonic times < 24*3600
if (W(1:10,5) - round(W(1:10,5))) == zeros(10,1)
              W = W / 100;
end

% Step 1.  W -- newtimestamp
newtimestamp = p(1) * W(:,5) + p(2);
newtimestamp = newtimestamp + lag;



% Step 2.  Select W data for the hour Whr -- ntshr
% index move to next second not of going through all data

index1 = 1;
while W(index1,5) < basetimes(1,2)
        index1 = index1 + 9;
end

index2 = index1;

while (W(index2,5) < basetimes(length(basetimes),2))
        index2 = index2 + 9;
end

% exit when time too big.
Whr = W((index1):(index2-9), :);
ntshr = newtimestamp(index1:(index2-9));


% Step 3.  Decimate Wd -- ntsd
for (i = 1:5)
       Wd(:,i) = decimate(Whr(:,i), 9, 10 , 'fir');
end
ntsd = decimate (ntshr, 9, 10, 'fir');

% plotmet(ntsd, Wd);

% Step 4.  Matchwind to concentration

% remove temperature trend
% [pt,St] = polyfit(Wd(:,5), Wd(:,4), 1);
% Wd(:,4) = Wd(:,4) - (pt(1) * Wd(:,5) + pt(2));

% make sure times are matchable, i.e., ntsd, within conc time range

while (conc(1:1) < ntsd(1))
       conc(1,:) = [];
end
```

```
tic
% Tempdm = interp1(ntsd, Wd(:,4), conc(:,1));
toc

tic
spzdm = interp1(ntsd, Wd(:,3), conc(:,1));
% ozflux = flux(spzdm, conc(:,2));

%spxdm = interp1(ntsd, Wd(:,1), conc(:,1));
%spydm = interp1(ntsd, Wd(:,2), conc(:,1));
%[ speedm, dirm] = wind(spxdm, spydm);
toc

% clean up
clear index1 index2 lag i p S pt St
clear newtimestamp
clear basetimes Whr ntshr
clear ntsd Wd
```

F.6 Flux.m

```
function [ outdat] = flux(w, c)

% takes matched w, c and produce flux vs time plot,
% outputs avg flux, avg flux / mean conc

wmean = mean(w);
cmean = mean(c);

i=1:1:length(c);

cfit = polyfit (i',c,1);
plot (i, c, 'c', i, (cfit(1)*i'+cfit(2)), 'y-');
c = c - (cfit(1)*i'+cfit(2));


wfluct = w-wmean;
cfluct = c-mean(c);
wc = wfluct.*cfluct;

outdat(1) = mean (wc);
outdat(2) = mean (wc) / cmean;

figure
subplot(3,1,1); plot (cfluct); ylabel ('conc fluctuations')
axis([ 0 length(cfluct) min(cfluct) max(cfluct)] );
subplot(3,1,2); plot (wfluct); ylabel ('vert wind fluct')
axis([ 0 length(wfluct) min(wfluct) max(wfluct)] );
subplot(3,1,3); plot (wc); ylabel ('flux')
axis([ 0 length(wc) min(wc) max(wc)] );
xlabel ('point number')
```

21. References

Adamkiewicz, G, Wyckoff, P.S., Tatang, M.A., McRae, G.J., "Inverse Modeling of Urban-Scale Emissions", 22nd NATO/CCMS International Technical Meeting, Clarmont-Ferrad, France, June 2-6, 1997, Preprint Pg. 359, 1997

Aschmann S. M. and Atkinson, R., "Formation and Yields of Methyl Vinyl Ketone and Methacrolein from the Gas-Phase Reaction of O_3 with Isoprene", *Environ. Sci. Technol.* **28**: 1539 – 1542 (1994)

Anderson, S.M., and Zahniser, M.S., "Open-path Tunable Diode Laser Absorption for Eddy Correlation Flux Measurements of Atmospheric Trace Gases", *SPIE* **1433**: 167-178 (1991)

Atkinson, R., Monograph, *J. Phys. Chem. Ref. Data*, 1989

Atkinson, R. *et al.*, "Evaluated Kinetic and Photochemical Data for Atmospheric Chemistry, Supplement III", *J. Phys. Chem. Ref. Data,* **18**: 881–1097 (1989)

Atkinson, R., "Gas-Phase Chemistry of Organic Compounds: A Review", *Atmos. Env.*, **24A**:1 – 41 (1990)

Atkinson, R., "Kinetics and Mechanisms of the Gas-Phase Reactions of the NO_3 Radical with Organic Compounds", *J. Phys. Chem. Ref. Data,* **20**:459-507 (1991)

Atkinson, R., "Gas-Phase Tropospheric Chemistry of Organic Compounds", to be submitted to J. Phys. Chem. Ref. Data in 1991, probably published in 1992

Atkinson, R. *et al.*, "Kinetic and Photochemical Data for Atmospheric Chemistry -- Addendum to IUPAC Review", *J. Phys. Chem. Ref. Data,* **21**: 1204 - 1206 (1992)

Atkinson, R. and Carter, W.P.L., "Kinetics and Mechanisms of the Gas-Phase Reactions of Ozone with Organic Compounds under Atmospheric Conditions", *Chem. Rev.*, **84**:437-470 (1984)

Aumont, B., Jaecker-Voirol, A., Martin, B., Toupance, G., "Tests of Some Reduction Hypothesis Made in Photochemical Mechanims", *Atmos. Envir.*, **30**: 2061-2077 (1996)

G. M. Barrow, "Physical Chemistry", 4th Edition, McGraw Hill , New York, NY, 1979

Baulch, D.L., Cobos, C.J., Cox, R.A., Esser, C., Frank, P. Just, T., Kerr, J.A., Pilling, M.J., Troe, J. Walker, R.W., and Warnatz, J., "Evaluated Kinetic and Photochemical Data for Atmospheric Chemistry: IUPAC Subcommittee on Gas Kinetic Data Evaluation for Atmospheric Chemistry", *J. Phys. Chem. Ref. Data* **21**:411 (1992)

S.W. Benson, "Thermochemical Kinetics, Methods for stimation of Thermochemical Data and Rate Parameters", Second Edition, John Wiley and Sons, New York, N.Y., 1976

J.S. Bendat and A.G. Piersol. "Random Data Measurement and Analysis", John Wiley and Sons, New York, New York, 1986

Blumenthal, D.L., Hering S.V., Roberts, P.T., Lawson, D.R., Watson, J.G., "Sourthern California Air Quality Study Planning Process and Field Measurements", Paper 90.111.1, 82nd AWMA Meeting and Exhibition, Anahein, CA, June 25-30, 1989

Bowman, F.M. and Seinfeld J.H., "Ozone Productivity of Atmospheric Organics", *J. Geophy. Res.* **99**:5309-5324 (1994)

D.W. Byun, "The first Models-3 Design Workshop", EPA Atmospheric Resaerch and Exposure Assessment Laboratory, Research Triangle Park, NC, August, 1994

Carter, W.P.L., "A Detailed Mechanism for the Gas-Phase Atmospheric Reactions of Organic Compounds", *Atmos. Env.* **24A**:481-518 (1990)

Carter, W. P.L., "Development of Ozone Reactivity Scales for Volatile Organic Compounds", Cooperative Agreement No. CR-814396-01-0, US EPA, Office of Research and Development, Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC (1991)

Carter, W.P.L. and Atkinson, R., "Experiemental Investigation of Chamber-dependent Radical Sources", *Int. J. Chem. Kinet.* **14**:1071-1103 (1982)

Carter, W.P.L. and Atkinson, R., "Atmospheric Chemistry of Alkanes", *J. Atmos. Chem.* **3**:377-405 (1985)

Carter, W.P.L., Dodd, M.C., Long, W.D., Atkinson, R., "Outdoor chamber study to test multi-day effects", Report EPA-600/3-84-115, US EPA, Research Triangle Park, N.C., 1985

CFR (Code of Federal Regulations), Protection of the Environment, Title 40, Part 86, US Government Printing Office,Washington, D.C., 1983

Chameides, W.L. and Cowling, E.B., "The State of the Southern Oxidants Study (SOS): Policy-Relevant Findings in Ozone Pollution Research 1988-1994", SOS Science Team and SOS Coordinating Council, 1995

CRC (D.R. Liede, Editor), "Handbook of Chemistry and Physics", 74th Edition, CRC Press, Boca Raton, FL, 1994

Dodge, M.C., "Combined Use of Modeling Techniques and Smog Chamber Data to Derive Ozone Precursor Relationships", US EPA Report EPA-600/3-77-001a, 88-889, (1977)

Donahue, N.M., Dubey, M.K., Mohrschladt, R., Demerjian, K.L., Anderson, J.G., "High Pressure Flow Study of the Reactions HO + NO_x —> $HONO_x$: Errors in the Falloff Region", *J. Geophy. Res.*, **102**:6153 (1997)

EPA, "Air Quality Criteria for Ozone and Related Photochemical Oxidants", EPA/600/AP-93/004, US EPA, Office of Research and Development, Washington D.C., 1993

EPA, "Fact Sheet -- EPA's Revised Ozone Standard", US EPA, Office of Air and Radiation, Office of Air Quality Planning and Standards, Research Triagle Park, NC, July 17, 1997

EPA AIR CHIEF software, Version 4.0, User's Manual, EPA-454/C-95-001 EPA Office of Air Quality Planning and Standards, Emissions, Monitoring and Analysis Division, Emission Factor and Inventory Gorup, Research Triangle Park, NC (1995)

EPA AIRS database, User's Guide, EPA Office of Air Quality Planning and Standards, Technical Support Division, National Air Data Branch, Research Triangle Park, NC (1995)

EPA CAO (Center on Airborne Organics), "1996 Summer Symposium Report: Advanced Instrumentation for Air Quality Measurements", MIT Endicott House, Dedham, MA, July, 1996 (for copy of report, please contact Prof. Jack B. Howard, MIT Department of Chemical Engineering)

Falls, A.H., McRae, G.J., Seinfeld, J.H., "Sensitivity and Uncertainty of Reaction Mechanisms for Photochemical Air Pollution", *Int. J. Chem. Kin.* **XI**:1137-1162 (1979)

B.J. Finlayson-Pitts and Pitts, "Atmospheric Chemistry: Fundamentals and Experimental Techniques", John Wiley and Sons, New York , N.Y., 1986

Gao, D., Stockwell, W.R., Milford, J.B., "First Order Sensitivity and Uncertainty Analysis for a Regional-Scale Gas-Phase Chemical Mechanism", *J. Geophy. Res.* **100**:23153-23166 (1995)

Gao, D., Stockwell, W.R., Milford, J.B., "Global Uncertainty Analysis of a Regional-Scale Gas-Phase Chemical Mechanism", *J. Geophy. Res.* **101**:9107-9119 (1996)

Gautschi, W. Algorithm 726:ORTHPOL — A Package of Routines for Generating Orthogonal Polynomials and Gauss-type Quadrature Rules. ACM Transactions on Mathematical Software, **20:(1)**:21-62 (1994).

Georgopoulos, P. and Seinfeld, J.H., "Mathematical Modeling of Tubulent Reacting Plumes 1: Theory and Model Formulation", *Atmos. Envir.***20**:1791-1807 (1986a)

Georgopoulos, P. and Seinfeld, J.H., "Mathematical Modeling of Tubulent Reacting Plumes 2: Application to the NO-NO_2-O_3 System", *Atmos. Envir.***20**:1791-1807 (1986b)

Gery, M.W., Whitten, G.Z., Killus, J.P, "Development and Testing of the CBM-IV for Urban and Regional Modeling", EPA report, EPA 600/3-88-012, US EPA, Atmospheric Science Research Laboratory, Atmospheric Chemistry and Physics Division, Research Triangle Park, NC (1988)

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., "A Photochemical Kinetics Mechanism for Urban and Regional Scale Computer Modeling", *J. Geophy. Res.* **94**:12925-12956 (1989)

R.G. Ghanem and P.D. Spanos, *Stochastic finite elements: A Spectral Approach,* Springer-Verlag, NY, 1991

Grosjean, D., Williams E. L., Grosjean, E., "Atmospheric Chemistry of Isoprene and of Its Carbonyl Products", *Environ. Sci. Technol.* **27**: 830 - 840 (1993)

Harley, R.A., Hanningan, M.P., Cass, G.R., "Respeciation of Organic Gas Emissions and the Detection of Excess Unburned Gasoline in the Atmosphere", *Envir. Sci. and Tech.* **26**:2395 (1992)

Harley, R.A., Russell, A.G., McRae, G.J., Cass, G.R., Seinfeld, J.H., "Photochemical Modeling of the Sourthern Califronia Air Quality Study", *Envir. Sci. and Tech.* 27:378-388 (1993)

Harley, R.A., Sawyer, R.R., Milford, J.B., "Updated Photochemical Modeling for the California's Southcoast Air Basin: Comparison of Mechanisms and Motor Vehicle Emission Inventories", *Environ. Sci. and Tech*, **31**:2829 (1997)

J.B. Heywood, "Internal Combustion Engine Fundamentals", McGraw Hill Inc., New York, New York, 1988

A.C. Hindmarsh, "ODEPACK, A Systematized Collection of ODE Solvers", in Stepleman R.S. *et al.* Eds. Scientific Computing (IMACS Transaction in Scientific Computation v. 1) North Holland, Amsterdam, 1983 (LSODE pp. 55 – 64)

Jeffries, H.E., "A Different View of Ozone Formation: Fuel Reformulation Alone May Not Reduce Urban Ozone", *Fuel Reformulation*, **3**:58-68 (1993)

Jeffries, H.E., Kamens, R.M., Sexton, K.G., Gerhardt, A.A., "Outdoor smog chamber experiments to test photochemical models", Report EPA-600/3-82-016, US. EPA, , Research Triangle Park, N.C., 1982

Jeffries, H.E., Sexton, K.G., Arnold, J.R., Kale T.L., "Validation Testing of New Mechanisms with Outdoor Chamber Data, Volume 2: Analysis of VOC Data for the CB4 and CAL Photochemical Mechanisms", Final Report, EPA 600/3-89-010b, 1989

Jeffries, H.E., Sexton, K.G., Kamens, R.M., Holleman, M.S., "Outdoor smog chamber experiments to test photochemical models: Phase II", Report EPA-600/3-85-029, US. EPA, , Research Triangle Park, N.C., 1985a

Jeffries, H.E., Sexton, K.G., Morris, T.P., Jackson, M., Goodman, R.G., Kamens, R.M., Holleman, M.S., "Outdoor smog chamber experiments using automobile exhaust", Report EPA-600/3-85-032, US. EPA, , Research Triangle Park, N.C., 1985b

Jungkamp, T.P.W. and Seinfeld, J.H., "Prediction of Bond Dissociation energies and Transition State Barriers by a Modified Complete Basis Set Model Chemistry", submitted to *J. Chem. Phys.* (1996)

J.N. Kapur, *Maximum-Entropy Models in Science and Engineering*, Wiley, NY, NY, 1989.

Kolb, C.E., "The Chemistry of Atmospheric Oxidants: Where Do We Go from Here?", 7th BOC Priestley Conference (1995)

Kramm, G., Dlugi, R.., Foken, T, Molder, N, Muller, H. Seiler, W., Sievering, H, "On the Dry Deposition of Ozone and Reactive Nitrogen Species", *Atmos. Env.*, **29**: 3209-3231, 1995

P.A. Leighton, "Photochemitry of Air Pollution", Academic Press, NewYork, NY, 1961

Leone, J.A. and Seinfeld, J.H., "Analysis of the Characteristics of Complex Chemical Reaction Mechanisms: Application to Photochemical Smog Chemistry", *Env. Sci and Tech* **18**:280-287 (1984)

Leone, J.A. and Seinfeld, J.H., "Comparative Analysis of Chemical Reaction Mechanisms for Photochemical Smog", *Atmos. Environ.* **19**:437-464 (1985)

Lumley and Panofsky, "The Structure of Atmospheric Turbulence", Interscience Publications, New York, NY, 1964

Lurmann, F.W., Carter, W.P.L., Coyner, L.A., "A Surrogate Species Chemical Reaction Mechanism for Urban-Scale Air Quality Simulation Models", US EPA Contract No. 68-02-4104 (1987)

Madronich, S. And Calvert, J.G., "Permutation Reactions of Organic Peroxy Radicals in the Troposphere", *J. Geophy. Res.*: **95**:5697 – 5715 (1990)

J. March, "Advanced Organic Chemistry", John Wiley and Sons, New York, NY, 1992

Matlab™ Manuals, Mathworks, Inc., Natick, MA, 1995

McManus, J.B., Kebabian, P.L., and Zahniser, M.S., "Astimatic Mirror Multipass Absorption Cells for Long-Pathlength Spectroscopy", *Applied Optics*, **34**:3336-3348 (1995)

G.J. McRae, "Mathematical Modeling of Photochemical Air Pollution", Ph. D. Dissertation, Califormia Institute of Technology, 1981

Miller, C., "Two Color Transient Absorption Measurment of HO_2 + RO_2 Reaction Rates", Poster Presentation, Gordon Research Conference on Atmospehric Chemistry, Salve Regina University, RI, June 15-20, 1997

Middleton, P., Stockwell, W. R., Carter, W. P. L., "Aggregation and Analysis of Volatile Organic Compound Emissions for Regional Modeling", *Atmos. Env.*, **24A**:1107 – 1133 (1990)

Milford, J.B., Russell, A.G., McRae, G.J., "A New Approach to Photochemical Pollution Control: Implications of Spatial Paterns in Pollutant Response to Reductions in Nitrogen Oxides and Reactive Organic Gas Emissions", *Envir. Sci. and Tech*, **23**:1290-1301 (1989)

J.H. Moore, C.C. Davis, M.A. Coplan, "Building Scientific Apparatus, A Practical Guide to Design and Construction", Addison-Wesley, London, UK, 1983

M.G. Morgan and M. Henrion, *Uncertainty, A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York, 1992.

Morris, R.E. and Myers, T.C., "User's guide for the Urban Airshed Model", US EPA Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-450/4-90-007A, 1990

Munger, J.W., Wofsy, S.C., Bakwin, P.S., Fan, S-M, Goulden, M.L., Daube, B.C., and Goldstein, A.H. "Atmospheric deposition of reactivie mitrogen oxides and ozone in a temperature deciduous forest and a subarctic woodland 1 Measurements and Mechanisms", *J.Geophy. Res.* **101**: 12639-12657 (1996)

NASA Panal: W. DeMore,S.P. Sander, D.M. Golden, R.F. Hampson, M.J. Kurylo, C.J. Howard, A.R. Ravishankara, C.E Kolb, M.J. Molina, "Chemical Kinetics and photochemical Data for use in Stratospheric Modeling" JPL Publication 92-20, Jet Propulsion Laboratory, Pasadena, CA, 1992

National Research Council (NRC), "Rethinking the Ozone Problem in Urban And Regional Air Pollution", National Acadamy Press, 1991.

D. Nelson, TDL Control and Data Aquisition Software Program, Aerodyne Research, Billerica, MA,1996

L. Newman (editor), "Measurement Challenges in Atmospheric Chemistry", Advances in Chemistry Series, Volume 232, American Chemical Society, Washington D.C., 1993

J.O. Nriagu (editor), "Gaseous Pollutants: Characterization and Cycling", Advances in Environmental Science and Technology Series, Volume 24, Wiley and Sons, New York, NY, 1992

Nuffield Advance Science, "Book of Data", Longman, UK, 1986

W. Pan, "The Role of Aerodols in the Troposphere: Radiative Forcing, Model Response, and Uncertainty Analysis", Ph.D. Thesis, Massachusetts Institute of Technology, 1996

A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* 3rd Edition, McGraw Hill, NY, NY, 1991.

Park, J.H., Rothman, L.S., Rinsland, C.P., Pickett, H.M., Richardson, D.J., and Namkung, J.S., "Atlas of Absoprtion Lines from 0 to 17900 cm^{-1}", NASA, 1987 (NTIS # N87-28955)

Parrish, D.D., Trainer, M., Williams, E.J., Fahey, D.W., Hubler, G., Ebank, C.S , Lui, S.C., Murphy, P.C., Albritton, D.L. Fehsenfeld, F.C., "Measurements of the NO_x-O_3 Photostationary State at Niwot Ridge, Colarado", *J. Geophy. Res.* 91:5361-5370 (1986)

F. Pasquill, "Atmospheric Diffusion" Ellis Honwood, 1974

Paulson, S. E. and Seinfeld, J. H., "Development and Evaluation of a Photooxidation Mechanism for Isoprene", *J.G.R.* 97:20703 –20715 (1992)

Paulson, S. E., Flagan, F. C., Seinfeld, J. H., "Atmospheric Photooxidation of Isoprene, Part I: the Hydroxyl Radical and Ground State Atomic Oxygen Reactions", *Int. J. of Chem. Kin.* 24: 79 – 101 (1992a)

Paulson, S. E., Flagan, F. C., Seinfeld, J. H., "Atmospheric Photooxidation of Isoprene, Part II: the Ozone-Isoprene Reaction", *Int. J. of Chem. Kin.* 24: 103 –125 (1992b)

R.H. Perry and D. Green, "Perry's Chemical Engineers' Handbook", 6th Edition, McGraw Hill, Inc. New York, NY, 1994

Peterson, J.T., "Calculated Actinic Fluxes (290 – 700 nm) for Air Pollution Photochemistry Applications", EPA-600/4-76-025 (1976)

W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, "Numerical Recipes in C", Second Edition, Cambridge University Press, London, UK, 1992

Roberts, J.M., "The Atmospheric Chemistry of Organic Nitrates", *Atmos. Env.*, 24A: 243-287 (1990)

Roberts, P. Korc, M., Blumenthal, D., Mueller, P.K., "Description of the NARSTO-Northeast 1995 Summer Ozone Study", North American Research Strategy for Tropospheric Ozone, 1995

Rothman, L.S., Gamache, R.R., Tipping, R.H., Rinsland, C.P., Smith, M.A.H., Bennner, D.C., Devi, V.M., Flaud, J.M., Camy-Peyret, C., Perrin, A., Goldman, A., Massie, S.T., Brown, L.R.,

Toth, R.A., "The HITRAN Molecular Database Editions of 1991 and 1992", *J. Quant. Spectrosc. Radiat. Transfer*, **48**:469-507 (1992)

Russell, A., Milford, J.B., Bergin, M.S., McBride, S., McNair, L., Yang, Y., Stockwell, W.R., Croes, B., "Urban Ozone Control and Atmospheric Reactivity of Organic Gases", *Science*, **269**:491-495 (1995)

J. H. Seinfeld, "Atmospheric Chemistry and Physics of Air Pollution" Wiley and Sons, NY, NY (1986)

Seinfeld, J.H., "Analysis of Data from the Southern California Air Quality Study", Paper 90.111.2, 82nd AWMA Meeting and Exhibition, Anahein, CA, June 25-30, 1989

Sigrist, M.W., Air Monitoring by Spectroscopic Techniques, Wiley, 1994. (Chapter 5 The Use of Tunable Diode Laser Absorption Spectroscopy for Atmospheric Measurements by H.I.Schiff, G.I. Mackay, and J. Bechara)

W.T. Silfvast, "Laser Fundamentals", First Edition, Cambridge University Press, London, UK, 1996

Stockwell, W.R., "A Homogeneous Gas Phase Mechanism for Use in a Regional Acid Deposition Model", *Atmos. Env.*, **20**: 1615 – 1632 (1986)

Stockwell, W.R., Middleton, P, Chang, J.S., Tang, X., "The Second Generation Regional Acid Deposition Chemical Mechanism for Regional Air Quality Modeling", *J. Geophy. Res.* **95**:16343-16367 (1990)

Stockwell, W.R., Yang, Y. Milford, J.B. "A Compilation of Estimated Uncertainty Factors for Rate Constants in W. P. L. Carter's Detailed Mechanism", Report, Automotive Emissions Cooperative Research Council CRC No. AQIRP-19-92 (1994)

A. Streiwieser and C.H. Heathcock, "Introduction to Organic Chemistry", 3rd Edition, Macmillan Publishing Company, New York, N.Y., 1985

Stump, F.D., Knapp, K.T., Ray, W.D., Snow, R., Burton, C., "The composition of Motor Vehicle Organic Emissions Under Elevated Temperature Summer Driving Conditions", *J. Air Waster Manage. Asso.* **43**:1084-1090 (1992a)

Stump, F.D., Knapp, K.T., Ray, W.D., Snow, R., Eudy, L, "The composition of Motor Vehicle Organic Emissions Under Elevated Temperature Summer Driving Conditions (75 to 105°F) - Part II", *J. Air Waster Manage. Asso.* **42**:1328-1335 (1992b)

M.A. Tatang. *Direct Incorporation of Uncertainty in Chemical and Environmental Engineering Systems*. PhD Thesis, Massachusetts Institute of Technology, 1994.

Tatang, M.A. and McRae, G.J., "Direct Treatment of Uncertainty in Models of Reaction and Trasport", unpublished, 1995

Tatang, M.A., Pan, W., Prinn, R.G., McRae, G.J., "An efficient Method for Parametric Uncertainty Analysis of Numerical Geophysical Models", *J. Geophy. Res.* **102**: 21925-21932 (1997)

Tuazon, E. C. and Atkinson, R., "A Product Study of the Gas-Phase Reaction of Methyl Vinyl Ketone with the OH Radical in the Presence of NO_x", *Int. J. of Chem. Kin.* **21**: 1141 – 1152 (1989)

Tuazon, E. C. and Atkinson, R., "A Product Study of the Gas-Phase Reaction of Mathacrolein with the OH Radical in the Presence of NO_x", *Int. J. of Chem. Kin.* **22**: 591 – 602 (1990a)

Tuazon, E. C. and Atkinson, R., "A Product Study of the Gas-Phase Reaction of Isoprene with the OH Radical in the Presence of NO_x", *Int. J. of Chem. Kin.* **22**: 1221 –1236 (1990b)
J. Villadsen and M. L. Michelsen. *Solution of Differential Equation Models by Polynomial Approximation* Prentice-Hall, N.J., 1978

Walters, R.A. and Saeger, M.L., "The 1985 NAPAP Emissions Inventory: Development of Species Allocation Factors", Report EPA-600/7-89-010f, US. EPA, , Research Triangle Park, N.C., 1990

Wang, Z.S.F., Langstaff, J.E. and, Jeffries, H.E., "Application of Integrated Process Rate Analysis Method for investigation of ubran Airshed Model (UAM) Sensitivity to Speciationin VOC Emissions Data", *Proceedings APCA Annual Meeting,* Conf. 88, **3A**:95 (1995)

Werle, P., Mucke, R., and Slemr, F., The Limits of Signal Averaging in Atmospheric Trace-Gas Monitoring by Tunable Diode Laser Absorption Spectroscopy (TDLAS), Appl. Phys. B57:131-139 (1993)

Wormhoudt, J. Zahniser, M.S., Nelson, D.D., McManus, J.B., Miake-Lye, R.C., and Kolb, C.E., "Infrared Tunable Diode Laser Measurements of Nitrogen Oxide Species in an Aircraft Engine Exhaust", in Optical Technique. in Fluid, Thermal, and Combustion Flow, Soyoung S. Cha and James D. Trolinger, Editors, Proc. SPIE, 2546:552-562 (1995)

P.S. Wyckoff, "Numerical Solution of Differential Equations Through Empirical Eigenfunction Expansions" Ph. D. Thesis, MIT, 1995

Yang, Y-J., Stockwell, W.R., Milford, J.B., "Uncertainties in Incremental Reactivities of Volatile Organic Compounds", *Envir. Sci .and Tech.* **29**:1336-1345 (1995)

Zahniser, M.S., Nelson, D.D., McManus, J.B., and Kebabian, P.L., "A Tunable Diode Laser System for Atmospheric Trace Gas Flux Measurements, Aerodyne Research Inc." prepared for Swiss Federal Resarch Station for Agricultural Chemistry and Engironmental Hygiene (1995a)

Zahniser, M.S., Nelson, D.D., McManus, J.B., and Shorter, J.H., "Tropospheric Measurements of Acidic Gases Using Tunable Diode Laser Spectroscopy", Aerodyne Research, Inc., ARI-RP-539.3 (1995b)

Zahniser, M.S., Personal Communication, 1996

Zimmerman, J. and Poppe, D., "A Supplement for the RADM2 Chemical Mechanism: the Phtooxidation of Isoprene", *Atmos. Env.* **30**: 1225 – 1269 (1996)