ENERGY LABORATORY

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

THE DESIGN AND IMPLEMENTATION OF A DEMONSTRATION SUPPLEMENTARY CONTROL SYSTEM

M. Ruane, J. Gruhl, F. Schweppe,
B. Green (ERT) and B. Egan (ERT)

MIT Energy Laboratory Report No. MIT-EL 80-033

December 1976
revised December 1980

FINAL REPORT

THE ENERGY LABORATORY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

77 Massachusetts Avenue
Cambridge, Massachusetts 02139

THE DESIGN AND IMPLEMENTATION OF A DEMONSTRATION SUPPLEMENTARY CONTROL SYSTEM

M. Ruane, J. Gruhl, F. Schweppe,
B. Green (ERT) and B. Egan (ERT)

MIT Energy Laboratory Report No. MIT-EL 80-033

December 1976
revised December 1980

Prepared under the support of
United States Atomic Energy Commission
Contract No. AT(11-1)-2428

TABLE OF CONTENTS

CONTENTS (2)

TABLE OF FIGURES

FIGURES (2)

TABLE OF TABLES

SUMMARY

The overall goal of the Chestnut Ridge Supplemental Control System (SCS) demonstration project was to demonstrate how an existing monitoring network, existing air quality models, and existing meteorological forecasting methods could be combined with a new control strategy to integrate SCS into electric power system operation. This final report covers the period February 1, 1974 to May 31, 1976.

A complete SCS for four power plants in the Chestnut Ridge region of Pennsylvania was implemented. The design is described in Section 2. The demonstration period, discussed in Section 3, showed that it is definitely possible to integrate a sophisticated SCS into electric power systems operation. The basic methods used in this project are felt to be directly extendable to other situations.

The only new technology originally envisioned for this project was the control strategy which decides the power system's response to predicted or potential violations. One of the key problems was the need for the control strategy to ensure that standards are not violated in spite of the presence of uncertainties in predicted ambient concentration levels. As discussed in Section 2.6, the implemented control strategy accounted explicitly for the uncertainties.

The point source air quality model used during the demonstration period was primarily a state-of-the-art model. However, as discussed in Section 2.5, a relatively new innovation involving downwash modeling was critical to the success of the demonstration.

During the course of the project, a large data base of SO_2 concentrations, meteorological measurements, weather forecasts, and power system data was established and stored in a manner which was easy to access and manipulate. Studies were done using these data, both before and after the actual demonstration. Some of the methodologies used and developed are applicable to a variety of problems including many non-SCS types. The results of these studies will now be summarized.

State-of-the-art air quality modeling was not as satisfactory as initially hoped in coping with the rough terrain in the Chestnut Ridge area. Research on improving point source air quality modeling for rough terrain was successfully undertaken using the data base after the demonstration period was over. The results are discussed in Section 5 with details provided in Appendix E.

The Chestnut Ridge area was discovered to have an unexpectedly high background SO_2 level. The data base enabled this background problem to be addressed in the four ways summarized in Section 4: mean concentration analysis, peak concentration analysis, EPA Larson method, and stochastic modeling. All four approaches are felt to be applicable in other situations where it is desired to understand the true nature of a background

concentration. The stochastic modeling appears to be a new methodology with particularly great potential. Details of these four methods are given in Appendixes A, B, C, and D.

Uncertainty arising from air quality modeling errors, weather forecasting errors, fuel sulfur contents, power system economics, and plant availability plays a central role in SCS analysis, design, and implementation. A systematic analysis methodology was applied to the data base to explore how these various uncertainties propagate through the overall SCS and affect its operation. This work is discussed in Section 6.

The control strategy minimizes cost subject to the constraint that ambient standards are not violated. Because of the uncertainties, the control strategy operates in a conservative fashion, that is, it often takes control actions that would not be required if the uncertainties did not exist. The control strategy was applied to the data base to determine how the overall economics behave and how they are affected by the presence of uncertainty. These results are discussed in Section 7.

During the course of the project, opinions were developed on the potential future role of SCS. We feel that SCS provides a viable tool for dealing with the energy, economic, environmental crisis. These opinions are discussed in more detail in Section 8.

DEFINITIONS OF KEY TERMS

The key terms used in this document are defined as follows:

Air Quality - Ground-level pollutant concentrations and their temporal and spatial distributions

Air Quality Monitoring Network - An array of continuous sampling stations that measure air quality

Air Quality Violation - the occurrence* of an ambient SO_2 concentration that exceeds an ambient air quality standard for SO_2 at any point within a designated liability area

Background Study - The collection and analysis of source, meteorological, and air quality data for the purpose of developing the SCS operational procedures.

Closed-Loop Mode of Operation - The SCS operational mode in which emission control decisions are based upon real-time measurements by the air quality monitoring network.

Constant Emission Control (CEC) - The continuous limitation of SO_2 emissions from a stack through techniques such as stack gas cleaning or use of low-sulfur fuel

Control Decision - The decision as to what degree to curtail emissions

Controlled Emissions - The SO_2 emission rate resulting from implementation of the control decision

Control Strategy - The sequence of recommended control decisions produced by the operating model for the period of the current forecast

Data Storage - Synchronous records of meteorological data, emission data, dispersion model predictions, control decisions and measured air quality

Dispersion Model - A mathematical model that relates meteorological data, emission rates, other source data and terrain factors to air quality in the vicinity of the source

*Many of these definitions have been taken from (EPA, 1976). Modifications have been made to correspond with the use of these terms in this project.

Effective Stack Height - The sum of (1) the physical stack height above grade and (2) the rise of the plume centerline after leaving the stack. The latter is due to the plume's initial vertical momentum and buoyancy

Emission Monitors - A system for the sampling and recording of SO_2 emission rates.

Isolated Source - A source sufficiently remote from other significant sources such that it can and will accept responsibility for attaining and maintaining ambient air quality standards for SO_2 within a designated liability area

National Ambient Air Quality Standards (NAAQS) - The primary NAAQS for SO_2 are a 24-hour average concentration of 365 micrograms per cubic meter (not to be exceeded more than once per year) and a maximum annual average concentration of 80 micrograms per cubic meter; the secondary NAAQS for SO_2 is a 3-hour concentration of 1300 micrograms per cubic meter (not to be exceeded more than once a year)

Open-Loop Mode of Operation - The SCS operational mode in which emission control decisions are based on calculations by the operating model in conjunction with either real-time or projected meteorological and emission data

Operating Model - The key element in the open-loop mode of SCS operation consisting of the control model and the dispersion model. The model generates the control strategy using forecasted meteorology and power system data

Supplementary Control System (SCS) - A system by which SO_2 emissions are curtailed during periods when meteorological conditions conducive to ground-level concentrations in excess of ambient standards for SO_2 either exist or are anticipated

System Upgrade - The continuous and systematic evaluation of the SCS elements and their interrelation in order to improve the reliability of the SCS in maintaining ambient standards for SO_2

Threshold Concentration - A measured or predicted short-term SO_2 concentration and/or rate of change of concentration that serves as an indicator of a potential violation of an ambient SO_2 standard

Time Delay - The time elapsed between the emission control decision and the time that the reduction in emissions begins to affect ambient air quality. This delay is equal to the sum of the time required to achieve the reduced emission level and the time required for the atmospheric transport of the reduced emissions to the point of the maximum ground-level concentration

1.0 INTRODUCTION

1.1 Motivation

The control of sulfur dioxide emissions from large stationary sources has been in the forefront of the national debate on environmental quality and energy supply. Congress, the EPA, the courts, industry and researchers searched for an acceptable trade-off between protecting the public from the effects of SO_2 and using fuel resources containing sulfur. This report documents a research effort on one part of the overall trade-off question: using supplementary control systems on electric power plants in complex terrain.

In its entirety, the trade-off question is sufficiently complex to have consumed almost a decade since the 1967 Air Quality Act and 1970 Clean Air Act were passed. For our purposes, the debate can be traced to Congress' 1970 Clean Air Act, which required the EPA to establish standards for the control of air pollution, and required the states to enact State Implementation Plans (SIP's) to achieve or improve upon the federal standards. Ambient sulfur dioxide standards were set so as to protect the public health and welfare (Federal Register, 1971) and the debate was started. Section 110 of the Act requires any SIP for emission control to include emission limitations and schedules for compliance "and such other measures as may be necessary to attain and maintain ambient air quality standards." The electric utility industry interpreted this as allowing any technique which would meet this goal. EPA, however, maintained that it was required by Congress to disapprove any SIP without continuous emission control techniques, unless such techniques were infeasible.

Continuous emission controls are those which operate all the time. The use of conforming fuel (i.e., fuel whose sulfur content will produce acceptable emissions) is the most common continuous emission control in use. Conforming fuel can be purchased, blended by a user from "clean" and "dirty" supplies, or made by desulfurizing "dirty" fuel. For many future fossil power plants, most of which must burn coal, conforming fuel supplies will be inadequate. Flue gas desulfurization is the alternative means of continuous emission control for such plants.

The utilities' reluctance to use continuous controls was based on a number of factors, including insufficient availability, costs, reliability problems, and secondary environmental impacts, such as sludge disposal. Much of the trade-off debate has centered on the availability, costs, and reliability of the continuous controls. This report does not address those issues directly. Rather, we consider one of the alternatives to continuous controls.

Non-continuous or dynamic emission controls are those which operate only part of the time. Non-continuous emission controls which assist some form of continuous controls are referred to in this document as supplementary control systems (SCS's). Non-continuous controls which are the only controls applied to protect air quality are referred to as intermittent control systems (ICS's).

Supplementary and intermittent control systems both are enacted whenever the scheduled source emissions (whether controlled or uncontrolled) and expected meteorology will combine so as to threaten ambient standards. Both systems act to protect the standards by reducing emissions until the atmosphere's dispersive ability increases. For ICS, the potential need for control is apparent, since no continuous control is in effect. For SCS, any use of the non-continuous system implies that the continuous system's emission limitations are inadequate.

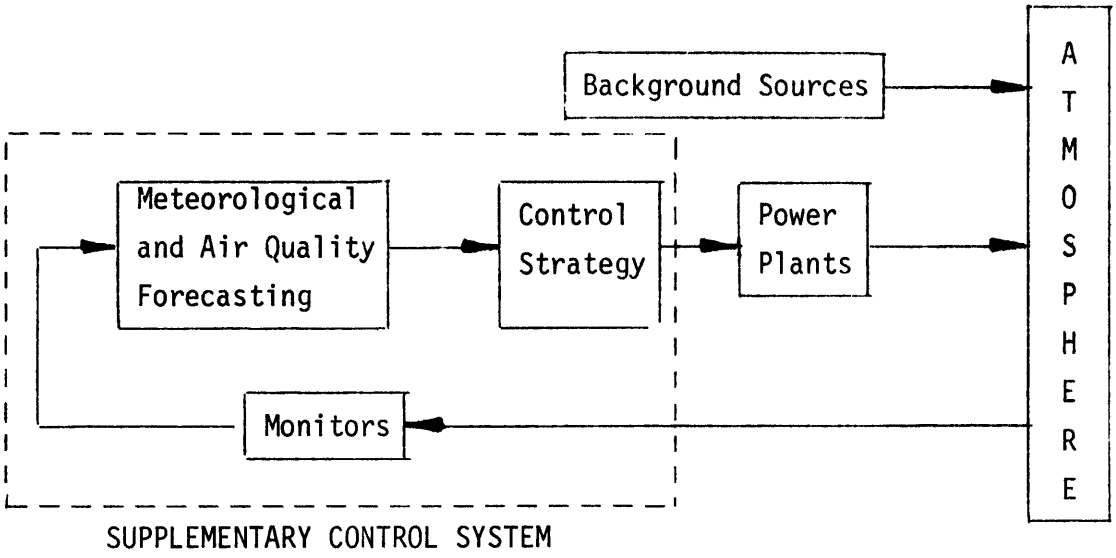
Under what conditions might this inadequacy of continuous emissions control be expected to occur? At present, the most likely case is when the source plume interacts with complex terrain. A second possibility is when other sources or background concentrations combine with the controlled source to cause a threat to the standards. In the future, as the economy grows and more sources compete for the same air resource, the logic which led to the need for continuous controls will lead to the situation where even continuous controls are inadequate to protect the ambient standards at all times. The alternative of tightening allowable emission limits is a possible solution, but one that runs counter to the need for economic, reliable energy and may be limited by the technologies of continuous controls. In the future, and at present in some special cases, even continuous controls may require supplementary help to maintain human health and welfare.

On April 19, 1976, the U.S. Supreme Court declined to hear the Big Rivers case, thereby upholding a lower court's ruling that the EPA is empowered to forbid the use of ICS in any SIP. ICS, as part of the overall trade-off deabte, seems to be a dead issue. The concern over increased atmospheric loadings and suspended sulfate levels has eliminated ICS as a viable compromise for sulfur dioxide control.

We believe that SCS is not a dead issue, nor should it be. This report was being prepared at a time when Congress was attempting to clarify its intentions for the Clean Air Act through the Clean Air Act Amendments of 1976 (U.S. Senate, 1976). Although the 93rd Congress adjourned without enacting these amendments, the Senate Committee Report clearly indicated a preference continuous controls and tended to classify all dynamic controls together as the same technology. We believe Congress should leave open the possibility of using supplementary control systems, and should make clear the distinction between SCS and ICS. Our experience has shown the necessity of SCS to protect air quality standards.

1.2 Project Description

Our experience has been based on a field demonstration of a supplementary control system. Called the Chestnut Ridge SCS after its location in western Pennsylvania, the system supplemented a continuous emission control scheme using a low-sulfur fuel supply which complied with the Pennsylvania SIP. Our SCS was similar to most SCS designs in that it attempted to take advantage of the atmosphere as a time-varying resource and was based on a combination of several existing technologies: monitoring, forecasting, and control, as shown in Figure 1.2.1.



Components of a SCS

Figure 1.2.1

In general terms, the goal of the Chestnut Ridge SCS Demonstration Project was to demonstrate how an existing monitoring network, existing air quality models, and existing meteorological forecasting methods could be combined with a control strategy to integrate SCS into electric power system operation.

The operation of the SCS was intended to provide field information about air quality, reliability, and economics so that the following technical objectives could be satisfied:

1) To demonstrate that an extensive, sophisticated, real-time monitoring system can be operated at a level of data capture and measurement accuracy sufficient for the needs of an SCS.

2) To demonstrate that air quality can be forecast accurately enough and far enough in advance to be useful in an SCS.

3) To demonstrate that a utility suitably prepared in advance can, in fact, switch fuel, increase stack temperature, and/or shift load while under an SCS so that ambient air quality standards in the vicinity are not violated.

4) To demonstrate that the Control Decision Logic can protect air quality standards while meeting various criteria such as cost and reliability of power system operation.

5) To demonstrate that in the presence of uncertainties in monitoring data, forecasts of air quality, emissions and response time, an acceptable reliability can be proven for the operation of an SCS.

6) To establish that a method can be developed to provide sufficient information to the appropriate regulatory agencies so they can ensure satisfactory SCS performance.

7) To define the steps required to transfer SCS technology to other sites and plants.

At the beginning of the project, the control strategy itself was considered to be the only major development required. The development of this control strategy proceeded pretty much as originally expected and did not yield any surprises. However, the overall project itself did not evolve according to the initial plans.

The organization structure involved three organizations: MIT, Environmental Research and Technology (ERT), and Pennsylvania Electric Company (Penelec). While these organizations were completely cooperative, the magnitude of the problems associated with coordinating and authorizing the various tasks was underestimated, causing major deviations from the project schedule. In addition, a major technical problem also arose concering the effect of background concentrations on the SCS operation. Although the plants were chosen under the assumption of their being isolated sources, both Pittsburgh and Johnstown had an effect on the region. This problem gave rise to additional research on background modeling.

For the reader interested in other efforts in this field, several sources are recommended. (Schweppe, 1, 1975) provides a survey of the state of the art of general environmental operating systems for electric power plants, and discusses SCS for SO_2 in detail. (Cadogan, 1, 1975) has a chapter with a comprehensive review of alternative dispatch schemes for SCS which have appeared in the literature. (Noll and Davis, 1, 1976) have several chapters on SCS which provide descriptions of the systems now operating in the U.S. on both utility and industrial boilers. (Montgomery, 1, 1975) provides a description of the most extensive operating SCS of the TVA.

This report is an interim report on the Chestnut Ridge SCS project covering the period from February 1, 1974 to May 31, 1976. It is intended to familiarize the reader with the design, implementation, and testing of the Chestnut Ridge SCS. Because every attempt was made to utilize existing SCS technologies, it effectively is a case study in the problems of technology transfer for SCS. In addition, conclusions on the broader applicability of SCS are drawn.

The work contained in this report was performed under Contract AT(11-1)-2428 of the Energy Research and Development Administration, Division of Biological and Environmental Research.

2.0 SYSTEM DESIGN

The Chestnut Ridge SCS was designed to make maximum use of existing technologies for monitoring and air quality forecasting. It differed from other supplementary control systems in several ways. Power system operating considerations were incorporated directly into the control decisions of the SCS; model predictions and control decisions were formulated probabilistically; the system was designed for use in complex terrain with multiple electrically interconnected and interacting sources. A further, unexpected distinction was the necessity of incorporating background considerations into control decisions.

2.1 Source Characteristics

Three classes of sources affect the Chestnut Ridge air quality. These are local background sources, distant background sources, and the controlled power plants. The Chestnut Ridge site was originally chosen because it appeared to comply with the EPA's criterion for isolated sources, based upon review of the LAPPES data and the relatively large emission rates of the controlled power plants. As described in Section 4, background sources were found to make a significant contribution to Chestnut Ridge concentrations and had to be included in the SCS design.

Resources were unavailable to perform a source inventory study in the Chestnut Ridge and surrounding areas. Emission inventories compiled by the State of Pennsylvania and federal EPA were, therefore, used to describe sources other than the power plants. These inventories were poor, having inaccurate and incomplete data, and contained no data on time scales shorter than yearly average emissions. Most large point sources other than the power plants were in a restricted category and their data could not be released. Area source data were not available. Because of this paucity of information on both local and distant background sources, it was decided to model their aggregate effect on Chestnut Ridge, rather than strive for a detailed source-by-source representation.

Local background sources were assumed to contribute that component of observed concentrations which could not be assigned to the influence of the power plants, or to Pittsburgh or Johnstown. No evidence could be found to justify assigning time or meteorology-dependent variation to local background levels, which were characterized by a constant contribution of 10 ppb to all receptors.

Distant background sources were characterized in two general locations, Pittsburgh and Johnstown. They were represented as effective point sources, although their emissions included both area- and point-source contributions. The entire area of Allegheny County (Pittsburgh and environs) was modeled as three point sources, one on the Monongahela River south of the city, one on the Allegheny River north of the city, and one on the Ohio River west of the city. These sources were located at sites of known, large emitters of SO_2 and were assigned stack heights of 76 m and negligible heat emissions. Their total source strength was chosen to equal the total Allegheny County emission rate of 200,000 tons per year reported by the Allegheny County APCD.

Johnstown, which lies to the east of Laurel Ridge from the Seward and Conemaugh plants, was also modeled as a point source with a 76 m stack and negligible heat emissions. It was somewhat arbitrarily assigned an emission rate of 7,000 tons per year, which is probably a low estimate. No time-varying behavior was assigned to these background sources.

The controlled power plants in the Chestnut Ridge region consisted of four coal-fired, base load plants: Keystone (1,640 MW), Conemaugh (1,700 MW), Homer City (1,200 MW), and Seward (218 MW). Each plant has several independent units consisting of a boiler, generator and stack. Figure 2.1.1 gives the location of the controlled plants in the Chestnut Ridge terrain, while Table 2.1.1 summarizes the characteristics of the power plants. Note that Seward is significantly smaller than the other three controlled plants.

The power plants are interconnected electrically with each other and with the larger, Pennsylvania-New Jersey-Maryland (PJM) power pool. Consequently, the economics of control at the plants must be characterized by two components: an in-plant cost and a system cost. The system cost, or the charges for replacement of lost capacity and energy, is a function of the mix of generation operating and available to the PJM pool. The expected capacity and energy levels of the plants are set each day by the system dispatchers in Johnstown, which has a satellite control center of PJM, as shown in Figure 2.1.2. Although Penelec does not own Keystone and Conemaugh, its Johnstown control center dispatches the plants. Penelec owns 50 percent of Homer City and 100 percent of Seward; these and several other Penelec plants are dispatched from the same control center. This rather special arrangement with Keystone and Conemaugh has been established because of the size and importance of the Chestnut Ridge generating complex.

Any control action must be cleared by the Johnstown dispatchers, who, in turn, inform PJM that the plants will be affected. Other plants in Penelec or in the other utilities shown in the middle of Figure 2.1.2 must be brought on line. If insufficient capacity is available, PJM must purchase energy from the adjacent pools shown at the top of Figure 2.1.2.

The plants are supplied with a combination of mine-mouth coal and contracted deliveries by truck. "As received" coal data on a monthly basis were the best coal sulfur-content information available to the project, but they were not available in real time. For SCS operation, all plants were conservatively assumed to burn 2.4 percent S coal, which is the Pennsylvania limit (4 lb/MBtu). Washing or blending with washed coal was used to guarantee compliance with the 2.5 percent limit when "as received" values were higher.

Although designed for base load operation, the plants had fairly large forced and maintenance outages during the study, as shown by the unit availabilities in Table 2.1.2. These outages were a combination of total and partial outages. During the partial outages, a piecewise, linear load curve for each unit was used to determine emissions of SO_2 and heat.

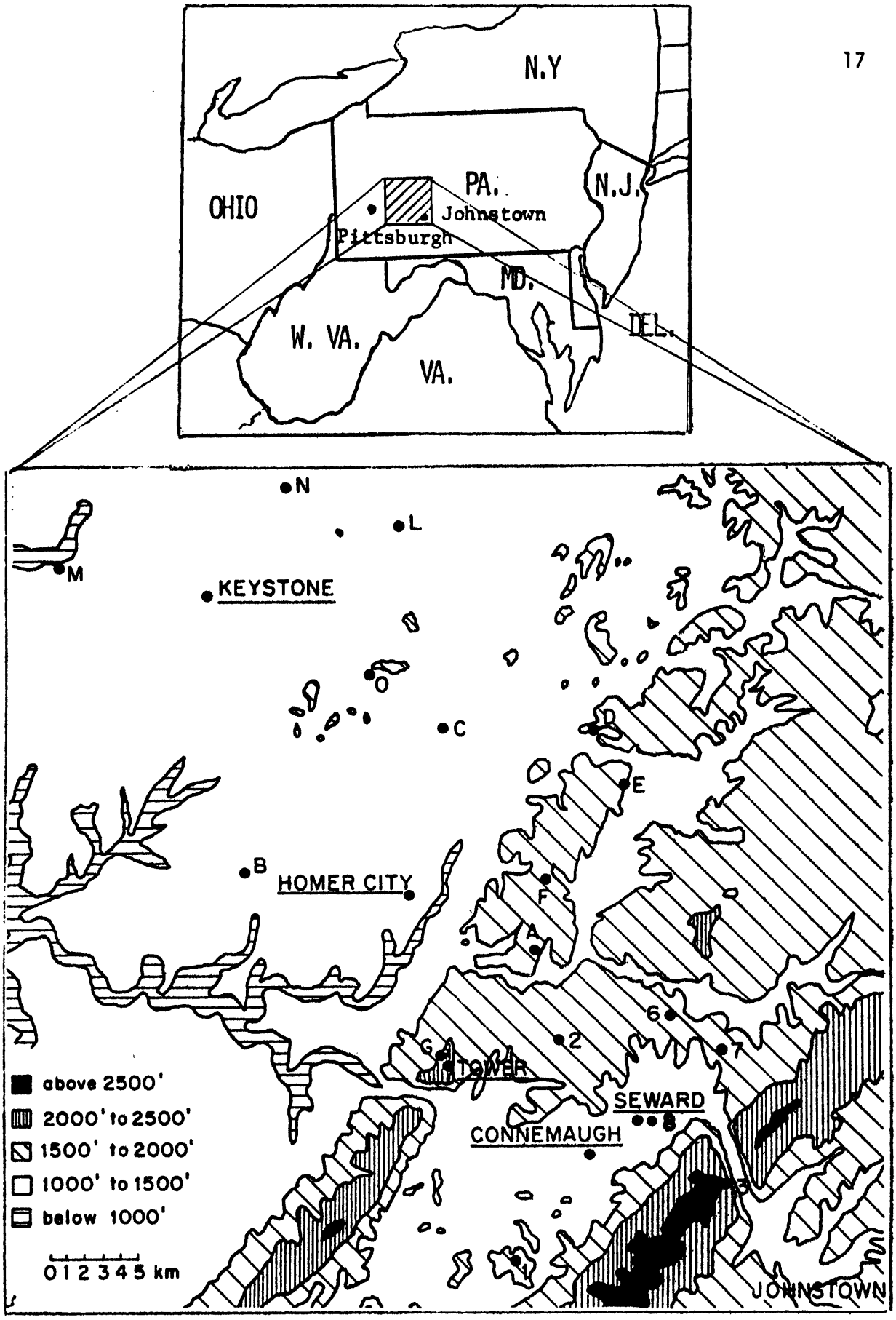


Figure 2.1.1

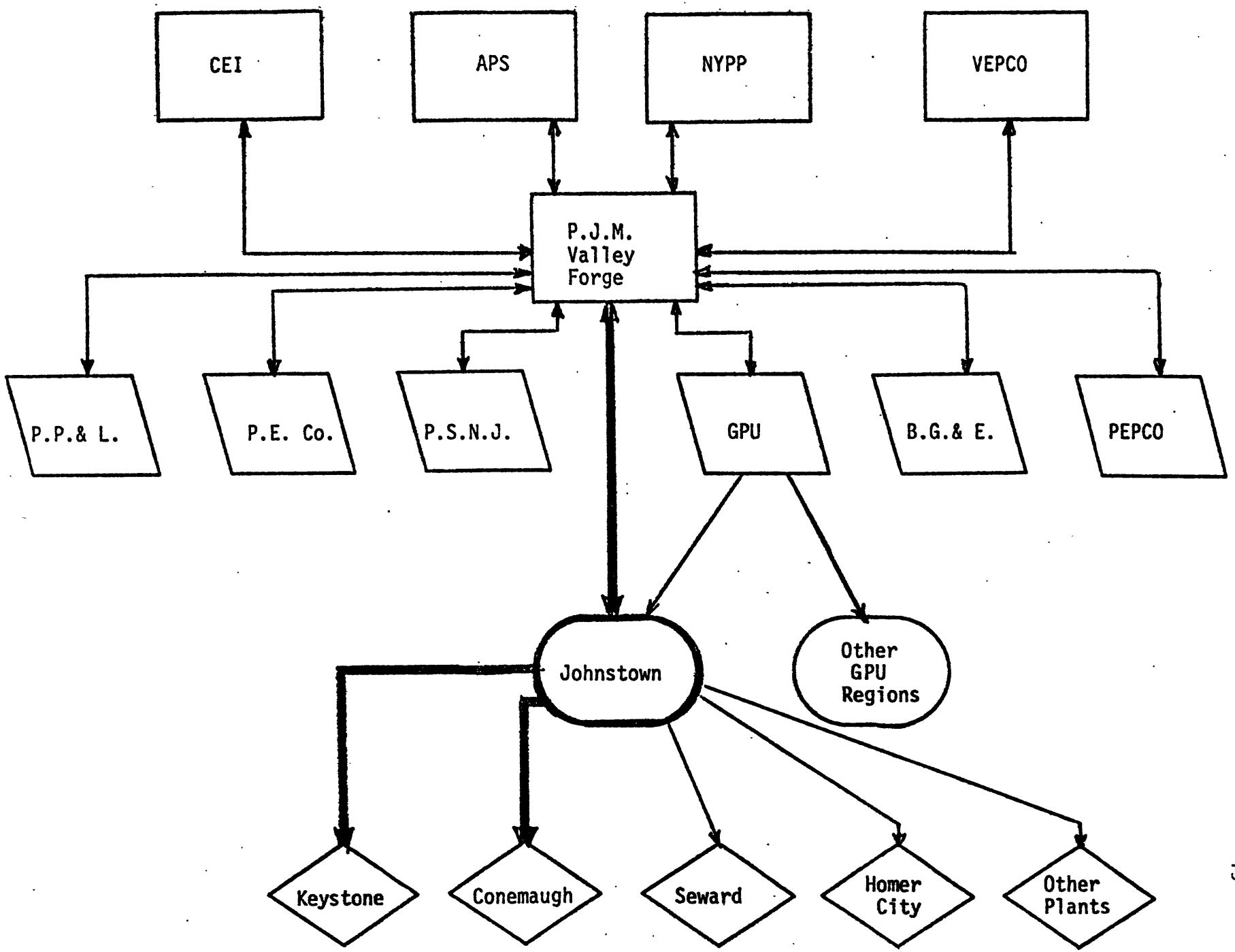
Chestnut Ridge Area

Table 2.1.1

PLANT PARAMETERS

UNIT	CAPACITY	FULL LOAD HEAT RATE	FULL LOAD EMISSIONS	STACK BASE ELEVATION	STACK HEIGHT	STACK U.T.M. COORDINATES	
NAME	MW	BTU/KWH	gm/sec*	ft. MSL	ft	North	East
KEYSTONE 1	820	9939	4111	1020	797	4502.20	640.10
KEYSTONE 2	820	9939	4111	1020	797	4502.20	640.10
HOMER CITY 1	600	9925	3004	1220	796	4486.20	652.73
HOMER CITY 2	600	9925	3004	1220	796	4486.20	652.73
CONEMAUGH 1	850	9754	4182	1080	1000	4472.00	664.25
CONEMAUGH 2	850	9754	4182	1080	1000	4472.00	664.25
SEWARD 5	137	9810	678	1085	232.5	4474.55	666.85
SEWARD 3 & 4	81	10444	427	1085	225	4474.55	666.85

* assuming 2.4% S coal.



Dispatching Hierarchy

Figure 2.1.2

Table 2.1.2

MONTHLY AVAILABILITIES

	HC1	HC2	CN1	CN2	KEY1	KEY2	SEW3	SEW 5
1974 Sept	0.541	0.612	0.496	0.558	0.972	0.294	0.685	0.298
1974 Oct	0.727	0.683	0.109	0.809	0.555	0.0	0.729	0.494
1974 Nov	0.412	0.603	0.196	0.612	0.733	0.0	0.693	0.991
1974 Dec	0.597	0.694	0.624	0.457	0.543	0.0	0.667	0.767
1975 Jan	0.66	0.50	0.507	0.306	0.939	0.362	0.677	0.840
1975 Feb	0.537	0.822	0.593	0.787	0.834	0.900	0.440	0.904
1975 Mar	0.728	0.552	0.577	0.132	0.927	0.460	0.676	0.945
1975 Apr	0.432	0.786	0.728	0.279	0.873	0.816	0.697	0.955
1975 May	0.188	0.541	0.388	0.487	0.539	0.595	0.696	0.917
1975 June	0.817	0.452	0.833	0.732	0.891	0.837	0.608	0.907
1975 July	0.343	0.457	0.593	0.450	0.840	0.760	0.649	0.970
1975 Aug	0.580	0.009	0.663	0.625	0.672	0.637	0.611	0.830
1975 Sept	0.528	0.0	0.589	0.585	0.784	0.744	0.460	0.916
1975 Oct	0.561	0.0	0.880	0.620	0.689	0.468	0.306	0.989
1975 Nov	0.612	0.0	0.417	0.897	0.825	0.007	0.316	0.990
1975 Dec	0.550	0.182	0.670	0.763	0.924	0.613	0.368	0.879
1976 Jan	-MISSING-		0.0	0.491	0.791	0.910	-MISSING-	
1976 Feb	0.684	0.622	0.0	0.744	0.823	0.612	0.460	0.984
1976 Mar	0.547	0.662	0.240	0.459	0.823	0.908	0.575	0.681

Three types of control action were available: load shifting, fuel switching and stack gas temperature modification. Load shifting could be used at all four controlled plants, and involved reducing the output of the units so as to require less fuel and therefore produce lower emissions. Fuel switching involved changing from high- to lower-sulfur coal. This was only of significant interest at Homer City, where a clean fuel reserve was on hand. While it was technically feasible at the other plants, it would have involved a substantial investment to arrange for a clean fuel supply at each plant. Stack gas temperature modification involved adjusting baffles in the air preheaters to increase the heat loss to the stack, thereby raising the effective stack height and reducing concentrations. This is an unusual control method and requires special equipment. Only Keystone and Homer City could effectively reduce concentrations with this method.

2.2 Air Quality Monitoring Network

The air-monitoring network at the Chestnut Ridge power complex comprises 17 stations deployed as illustrated in Figure 2.1.1 and summarized in Table 2.2.1. The instruments at each station are housed in a shelter of approximately 600 cubic feet whose interior temperature is maintained at 70-80°F in all seasons. Each shelter is equipped with several sensors including a computer-controlled SO_2 analyzer whose hourly averaged measurements are printed each hour at a central facility in Concord, Mass., and at a remote read-out teletype at the Homer City Station. The central computer in Concord polls each sensor once a minute, checks the status of the instrument, performs simple validity checks on the measurement reported, and calculates trailing one-, three-, and 24-hour averages on "valid" data. Validity checks are limited to tests against criteria defining excessive variability or excessive steadiness of the instrument's response. The computer also reports changes of the status of instruments -- for example, from proper operation (04 status) to not responding (20) -- and controls the calibration periods of the devices measuring gaseous pollutants. A flameout of the hydrogen flame in the sulfur analyzer is also sensed by the computer, which then tries to reignite the flame automatically by sending a command to the instrument. This feature prevents the loss of SO_2 data that would otherwise occur because of power failures or certain instrument malfunctions.

Strip chart recorders back up the real-time measurements of each SO_2 instrument. These chart records are used to check the operation of the real-time systems, to test the performance of the instruments, to provide more detailed information for case studies, and to fill in the data record on those occasions when the telemetry system fails.

The SO_2 instrument used throughout the network is the Meloy Laboratories Model SA-185R total sulfur analyzer. This flame-photometric detector has a range of 0.005 to 1.0 ppm. It provides excellent sensitivity at the low end of its range because of its logarithmic output. A multipoint calibration is performed every six months. Because the Meloys are not equipped with scrubbers, they measure all sulfurous gases. Consequently, the SO_2

Table 2.2.1

MONITOR LOCATIONS

MONITOR	NAME	U.T.M. COORDINATES NORTH	EAST	ELEVATION ft
1	WEST FAIRFIELD	4466.05	660.30	1500
2	FLORENCE SUB	4478.70	661.50	1740
3	LAUREL RIDGE	4470.65	671.15	2675
6	ARMAGH	4480.30	668.40	1620
7	GAS CENTER	4478.90	670.95	1730
8	SEWARD	4475.42	667.82	1230
A	LUCIUSBORO	4484.10	660.40	1660
B	LEWISVILLE	4486.25	643.15	1330
C	RUSTIC LODGE	4495.60	654.25	1240
D	PENN RUN	4496.70	662.80	1540
E	BRUSH VALLEY	4492.60	664.20	1640
F	LIGGETT	4487.60	660.55	1670
G	PENN VIEW	4477.30	655.90	2040
L	CREEKSIDE	4505.85	650.10	1140
M	GIRDY	4502.60	631.65	1190
N	KEYSTONE DAM	4508.45	644.35	1230
O	PARKWOOD	4498.15	649.45	1250
TOWER	PENN VIEW	4477.30	655.90	2060

measurements reported may be slightly higher than an SO_2-specific device would yield. This is not believed to be a significant factor in this study, because there are no known major sources of reduced sulfur gases in the Chestnut Ridge area. The Meloys used in this program are equipped with automatic reignition by computer command and are calibrated daily and automatically for zero and span by means of a calibrator installed at each station.

Penelec field technicians visit each site in the network every day to check on the operation of the instruments, to maintain them, and to remove and install strip charts.

The data acquired in real time are checked against digitized data from the strip charts for about ten percent of each parameter-month in order to validate the real-time data. Reasons for discrepancies are traced, and the archived data are corrected where possible or assigned a "missing value indicator" if proper values cannot be determined.

2.3 Meteorological Network

The only real-time meteorological data available to the SCS came from the 91.4 m tower at the Penn View site. This location is on top of a rounded knoll on Chestnut Ridge with an elevation approximately 620 m above sea level. The Conemaugh and Seward plants, roughly ten km east of the tower, are built on land almost 300 m lower in elevation in the Conemaugh River Valley between Chestnut and Laurel Ridges. The lack of meteorological data both from this valley and from locations more representative of Homer City and Keystone was early recognized as a potential probem to running and validating the SCS. However, the resources available to the project did not allow any further meteorological instrumentation to be installed.

The Penn View tower is instrumented at three levels -- 12.2 m, 45.7 m, and 91.1 m. The 12.2-m level was given the same station code, G, as the air-quality monitoring station to the north of the tower. The 45.7-m level is designated station S, and the 91.1-m level, station T. The lower and upper levels both have wind speed and direction sensors. Ambient dry bulb temperature and dew point are also measured at the 12.2-m level, and a pyranometer and rain gauge report from the air-quality monitoring shelter. Temperature differences between the middle and lowest levels of the tower and between the top and lowest levels of the tower are measured.

All the meteorological data are reported in real time to the central computer in Concord, Mass., and all the instruments record on strip charts for backup of the real-time reporting system. The meteorological instruments are maintained and serviced by Penelec's field technicians.

Supplementary wind data were available from the National Weather Service stations at Allegheny County Airport (AGC), near Pittsburgh; Blairsville (BSI), about 3 km northeast of the Penn View tower along Chestnut Ridge; Johnstown (JST); and Altoona (AOO), some 50 km east of Johnstown. Blairsville reports only during daylight hours, and Johnstown only from about 0600 to 2000 EST.

None of these stations was more representative than the tower data of the winds experienced by the power plant plumes, but the variation between stations was indicative of the effects of the terrain on flow near the ground. These data were not used in real time but helped the forecasters learn to anticipate the impact of topographic features on the surface winds near Chestnut Ridge.

The nearest upper air station was at the Pittsburgh Airport. The Pittsburgh 00 and 12Z soundings were routinely plotted by the forecasters. Forecasts of the daily maximum surface temperature were made and a dry adiabat extended to the sounding to forecast mixing depth. Verification was done in the same manner.

2.4 Air Quality Model

2.4.1 Deterministic Model

In the choice and development of an air quality model for a supplementary control system, perhaps the most difficult issue that must be addressed is the performance measure or criteria that should serve as the basis for selecting a "best" air quality model. A simple least squares error or a minimum average absolute error may show "best" performance in some sense for all situations but, after a little reflection, it is apparent that it is probably not optimal from an SCS viewpoint. The reason for this is that, in some sense, we would not care about accuracy of predictions for the vast majority of low-concentration situations if we could have a model that could do very well at predicting high levels of pollutant concentration.

The complexity of this problem is better understood with a definition of an ideal measure of desirability for an SCS air quality model:

> An optimum SCS air quality model would be that model which: (1) misses predicting violations with a frequency that is legally acceptable; (2) results in false alarm control actions that are in total less costly than the false alarms of all other models, and (3) is acceptable to the regulatory agencies.

Difficulties arise in trying to reach each of these three objectives. For point (1) there is no precise definition of what is legally acceptable, that is, "not more than once per year" takes no account of the area of the violation and is unclear about the duration of "one" violation. Minimizing the number of false alarms in point (2) is not likely to minimize their cost, and here one could envision varying levels of conservatism with the varying levels of control costs for the different power plants. On point (3), it may be that a model that does well at predicting high-level concentrations is not going to be a mass-conserving dispersion model, and may not even be a physically meaningful model. It is uncertain how such abstract "black box" models would be received by regulatory agencies.

For this project, four air quality models were evaluated. In all cases, there were significant modifications made to the models generally referred to by these generic names; in any event, the four models were essentially:

(1) standard Gaussian,

(2) sector-averaged Gaussian,

(3) the pointwise maximum of predictions from (1) and (2), and

(4) the pointwise sum of the maximum predictions from (1) and (2) on a source-by-source basis.

For example, if m = 1 is standard Gaussian, m = 2 is sector-averaged Gaussian, i are the various sources (including backgrounds), then the concentration C at time t at point x, y, z for models 3 and 4 are:

Model 3:

$$C_3(x,y,z,t) = \max\left[\left(\sum_{\text{all } i} C_{1,i}(x,y,z,t)\right), \left(\sum_{\text{all } i} C_{2,i}(x,y,z,t)\right)\right] \quad (2.1)$$

Model 4:

$$C_4(x,y,z,t) = \sum_{\text{all } i} \max\left[C_{1,i}(x,y,z,t),\ C_{2,i}(x,y,z,t)\right] \quad (2.2)$$

Models 3 and 4 are not mass-conserving models, that is, they showed more total mass of pollutant in the field than is emitted from sources. Although it was clear that these models out-performed models 1 and 2 in predicting high concentrations, they also had considerably higher false alarm rates. Initially, the choice among these four models was at the discretion of the user, but soon, in order to make the analyses comparable across the board, and after some investigation of merits, the sector-averaged Gaussian model was chosen as the primary dispersion model for this project.

The sector-averaged Gaussian formula uses the standard Gaussian dispersion for vertical diffusion and a constant concentration for $\pm$ 11.25° crosswind.

$$C(x,y,z) = \frac{Q}{\sqrt{2\pi}\,\sigma_y\,\sigma_z\,u}\left\{\exp\left[-\frac{1}{2}\left(\frac{z - H_e}{\sigma_z}\right)\right]^2 + \exp\left[-\frac{1}{2}\left(\frac{z + H_e}{\sigma_z}\right)^2\right]\right\} \quad (2.3)$$

where $\sigma_y = .3978x$,

$C(x,y,z)$ = concentration at point x, y, z,

x = downwind distance,

y = crosswind distance,

z = vertical distance,

σ_z = vertical dispersion coefficient,

H_e = effective stack height,

Q = grams per second emission of pollutants, and

u = wind speed, (seven classes),

where downwind direction is computed from the wind direction sector (16 sectors of 22.5°, sector 1 centered at north), and Q is computed from power plant MW levels and % sulfur in the fuel.

ASME (ASME, 1968) dispersion coefficients were used, rather than Pasquill-Gifford curves, because the ASME values are more representative of the elevated releases and averaging periods used at Chestnut Ridge. Pasquill-Turner categories A and B both were assigned to the most unstable ASME class.

The vertical dispersion was calculated by:

$$\sigma_z = Cx^D \quad (2.4)$$

where $C = .4 - .11\,F(S, 2.354) - .05F(S,4.)$

$D = .91 - .08F(S, 2.375) + .01F(S,4.)$

where S = stability class(Pasquill-Turner categories)

$F(\cdot)$ = positive difference (equals zero unless first argument minus second argument is positive).

The effective stack height was computed from the latest empirical plume rise formulas derived by Briggs [a personal communication, later than (Briggs, 1971)]. The buoyancy flux from each stack was estimated by means of linear formulas fit to data on flue gas volumes and temperatures that were provided by Penelec. This procedure may have resulted in an overestimate of the heat emissions because the data on which the formulas were based were taken in the breaching after the air preheaters, and no estimates were incorporated of the losses of heat that may take place in the tall stacks that serve the units at the three larger plants. Plume rise equations included several categories based on wind and stability classes, where

$$F = 37 * 0.07 \times OH \times 10^{-5}$$

and

QH = heat from the source in Btu/hr, then

1. If $S \geq 5$ and $u < 1.37$ m/sec

$$HP = 5\, F^{1/4}\, \alpha^{-3/8} \qquad (2.5)$$

where the stability parameter $\alpha = 4.63\ (10^{-4})$

2. If $S \geq 5$ and $u \geq 1.37$ m/sec

$$HP = 2.4\ (F/\alpha u)^{1/3} \qquad (2.6)$$

3. If $S \leq 5$

$$HP = [1.6(F^{0.3333})(10\ HST)^{0.6667}/u = 1.6\ F^{1/3}(3.5\ x_*)^{2/3}/u \qquad (2.7)$$

$$x_* = 14.43\ F^{2/5}$$

$$H_e = HST + HP \qquad (2.8)$$

HST = height of stack.

The "punch-through" condition incorporated in many EPA models was used in the deterministic operating model. By this convention, if a plume penetrates the top of the mixing layer, its contribution to pollutant levels at the ground is ignored. In the present model, this criterion was modified to the extent that the final rise of the plume had to be at least 50 meters greater than the top of the local mixing layer before punch-through occurred; otherwise, the plume rise was terminated at the top of the mixing layer, if it reached it, and the plume was reflected. Complete multiple reflections of SO_2 from both the ground and mixing layer were used.

The various point sources and background sources were superimposed. Maximum contribution to any point from any background source was limited to 50 ppb. Wind speed was scaled for surface roughness by the scaling factor w where

$$w = (\text{stack height}/150 \text{ meters})\ .06 + .01(1.96^{S-1}). \tag{2.9}$$

The lifting of the plumes over terrain features was modeled by means of a terrain correction factor for each source-receptor pair. This factor is the fraction relating the difference between the elevation of a receptor on the terrain and the elevation of the source stack base and the height that a plume will be lifted as it is transported downwind. A further constraint imposed was that a plume's centerline should not come closer to the terrain than the height given by the product of the terrain correction factor and the height of the plume in flat terrain. Formally, we have then

$$H = H_p - F_t \times \text{Minimum}\ \{t_r - t_s,\ H_p\}, \tag{2.10}$$

in which

H = the height of the plume over the terrain at the receptor,

H_p = the height of the plume over the stack base (i.e., stack height plus plume rise),

$F_t = 1 - T_f$,

T_f being the terrain correction factor,

t_r = the elevation at the receptor, and

t_s = the elevation at the site of the plant.

The top of the mixing layer over the terrain was treated in the same manner, that is, it was moved up and down over the terrain as if it were a non-buoyant plume released above the Pittsburgh Airport (elevation 373 m above mean sea level) from a stack having the height given by the estimate of the mixing depth at the airport. Formally,

$$H_x = \text{DMX} - F_t \times \text{Minimum}\ \{t_r - 373,\ \text{DMX}\}, \tag{2.11}$$

in which

H_x = the mixing depth over the terrain at the receptor, and

DMX = the mixing depth estimated from the Pittsburgh rawinsonde reports.

Minimum mixing depth for stability class 4 was 100 meters, for class 5 was 10,000 meters, regardless of what the calculation of Eq. 2.11 would show for any location.

On the basis of the results of numerical experiments with potential flow models (Egan, 1, 1975), the terrain correction factors given in Table 2.4.1 were assigned to the five stability classes. These values are consistent with the results presented by Egan in his review of plume dispersion in complex terrain.

Table 2.4.1

Terrain Correction Factors for the Stability Classes

Stability	1	2	3	4	5
T_f	.2	.3	.4	.5	.6

Model predictions were made using 3-hour average forecast data for ten 3-hour periods extending 30 hours into the future. Since Pennsylvania adheres to the federal ambient air quality standards for SO_2, this basic three-hour period is appropriate. It is recognized that the federal secondary standard relates to any three consecutive clock hours and that the SCS directly considered only the non-overlapping three-hour periods starting at midnight. The significance of this discrepancy was believed to be insufficient to warrant the effort and complication required to address it directly in this demonstration project. Furthermore, it was felt that forecasters could not make reasonable distinctions in their mid- to far-future predictions for time periods shorter than three hours. Then, too, Penelec's unit commitment schedule, on which emissions forecasts were based, provided estimates of future loads only in six-hour blocks of time.

Daily 24-hour average concentrations were calculated from each set of 8 three-hour projections at all receptors and tested for compliance with the daily SO_2 standards.

The operating model estimated three-hour SO_2 concentrations by sector averaging because the basic time period for forecasts was three hours and the ASME dispersion parameters used are applicable to approximately one-hour concentration estimates in relatively smooth terrain. The sector-averaged three-hour concentration predictions are consistent with the resolution of the wind-direction forecasts for the three-hour periods; they conserve the effluent mass, and they ensure that receptors will be in the predicted path of the plumes. Furthermore, at the distances of maximum anticipated impacts from one to 15 km, the sector-averaged values would give reasonable approximations to the expected, measured concentrations averaged over a three-hour period. This is especially true if one considers the enhanced dispersion encountered in rough terrain and the normal variations of wind speed, direction, and turbulence levels in a three-hour interval. Both these effects should decrease the expected three-hour concentration averages well below the one-hour values calculated at the centerline of a Gaussian plume by a model incorporation σ_y's given by the ASME curves.

2.4.2 Downwash Model for Seward

Special downwash modeling was performed for the area around the Seward plants. The modeling used exactly paralleled that developed in the reference: (Schulman and Egan, 1975). Effective building heights used in the downwash models were developed from physical parameters taken from architectural drawings of the Seward buildings and from on-site inspections, and were computed from equations tested in wind tunnels and verified in actual monitoring (although for a different set of buildings and location). These effective building heights, HBE, were (where wind direction 1 is north):

Wind dir.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
HBE (SEW 34)	36	36	36	36	42	55	42	36	36	36	36	36	36	36	36	36
HBE (SEW 5)	50	50	50	50	50	30	30	36	60	60	60	60	36	30	30	50

(units are meters)

Wind speed classes were as follows:

5	13 - 18 mph
6	18 - 24 mph
7	>24 mph

Formulas used in computing the downwash concentrations consisted of,

HPRIM = buoyancy length of plumes' rise rate

u = wind speed

QH = heat in plume

HST = stack height

F = fraction of plume intercepted by the building wake

GX = vertical mixing parameter

HW = wake height

X = downwind distance

GE = parameter

SCZ = another parameter

$$HPRIM = HST + 300(37 \times QH \times .07/u^3)$$

$$F = \exp[4(L-HPRIM/1.7HBE)].$$

If $F < .03$ there is no downwash, otherwise

$$GX = -.0015(HBE)^{-3}\left(\frac{u}{18}\right)^4$$

$$HW = (HBE^3 + 125X)^{1/3}$$

and $GE = 1$ or

if $GX \cdot X^3 > -7$

$$GE = 1 - \exp(GX \cdot X^3)$$

and

$$SZC = 1.25 * 2.61 * X^{.450} - 25.5$$

and if $SZC > HW$, then set

$$HW = SZC.$$

The concentration in ppb, C, is then

$$C = \frac{Q \cdot F \cdot GE}{\sigma_y \cdot HW \cdot .5 \cdot u} \; 376100 \qquad (2.12)$$

where here again

$$\sigma_y = .3978X.$$

Some very nonlinear effects can be noticed between the maximum downwash concentration and the plant operating levels. Table 2.4.2 shows that, under certain circumstances, an increase in the plant operating level injects enough additional heat into the plume to enable it to better clear the building wakes and decrease maximum concentrations. A series of tables was available to the SCS operators so that they could decide for themselves what would be the best control strategy. With Table 2.4.2 right here, it is instructive to look at an example of one of these difficult situations: suppose Seward units 5 is operating at 125 MW when the wind speed is

Table 2.4.2

Example of Downwash Concentrations* versus Plant Operating Levels

Sew-5	Wind Speed Class 6					
Eff. bldg. ht. = (M)	30.	36.	42.	50.	55.	60.
Max. at Dist. = (M)	540.	650.	760.	910.	1010.	1100.
Megawatts						
170.	0.0	0.0	15.20	33.36	47.37	62.49
165.	0.0	0.0	16.00	34.65	48.91	64.19
160.	0.0	0.0	16.82	35.97	50.45	65.87
155.	0.0	0.0	17.67	37.29	51.99	67.54
150.	0.0	8.18	18.54	39.63	53.53	69.17
145.	0.0	8.69	19.44	39.97	55.04	70.76
140.	0.0	9.22	20.35	41.31	56.53	72.30
135.	0.0	9.78	21.28	42.64	57.99	73.79
130.	0.0	10.35	22.22	43.95	59.40	75.20
125.	0.0	10.93	23.16	45.23	60.76	76.52
120.	0.0	11.54	24.11	46.48	62.05	77.74
115.	0.0	12.15	25.05	47.67	63.26	78.85
110.	0.0	12.77	25.99	48.81	64.37	79.82
105.	0.0	13.40	26.90	49.87	65.36	77.67
100.	0.0	14.03	27.78	50.84	66.21	73.97
95.	0.0	14.65	28.61	51.69	66.92	70.27
90.	0.0	15.25	29.39	52.42	67.44	66.57
85.	6.04	15.83	30.10	52.99	67.75	62.88
80.	6.37	16.37	30.72	53.38	67.83	59.18
75.	6.69	16.87	31.23	53.57	64.47	55.48
70.	6.99	17.30	31.60	53.51	60.17	51.78
65.	7.27	17.66	31.82	53.19	55.88	48.08
60.	7.52	17.92	31.85	52.55	51.58	44.38
55.	7.72	18.05	31.66	51.56	47.28	40.68
50.	7.86	18.03	31.20	50.17	42.98	36.99
45.	7.92	17.84	30.45	45.48	38.68	33.29
40.	7.89	17.43	29.35	40.42	34.38	29.59
35.	7.73	16.76	27.85	35.37	30.09	25.89
30.	7.42	15.79	25.88	30.32	25.79	22.19
25.	6.93	14.46	23.39	25.26	21.49	18.49

*All concentrations in pphm
3 hr. standard = 50., 24-hr standard = 14.

class 6 and the wind direction puts the effective building height at 55 meters. This situation is predicted at about 607.6 ppb maximum concentration, and this could be lowered below 500 ppb if plant load could be increased to greater than 160 MW or decreased to less than 55 MW. A two-way situation such as this did arise during the demonstration where, due to a valve problem, the plant could not be boosted to the higher level, and a "compromise" to drop to about 90 MW was seen as probably being worse than no action at all. The human-to-human communication link was shown tc be vital in a situation such as this.

The first question that comes to mind about this type of downwash model, or the entire air quality model for that matter, is how believable are its predictions? The only validation of the downwash model occurred during the demonstration on the few occasions where monitor 8 (at Seward) was involved in the downwash, and when the mobile monitor was in the area. In these situations, the model predicted the concentrations within 10% of monitored values; such accuracy can only be viewed as largely a matter of luck. The accuracy of the sector-averaged Gaussian portions of the model are dealt with in Chapter 6.

The addition of the downwash model further aggravated an already sparse grid. This original grid covered the 80 km by 80 km area with 1600 points each 2 km apart, see Figure 2.4-1. Although this type of evenly spaced grid may be aesthetically appealing, it is not the most effective manner of covering the likely high-pollutant concentration areas. A grid of about 400 points was eventually used, consisting of about 6 or 7 points approximately downwind in each of the 16 wind directions from each of the four sources. Because of the proximity of the Conemaugh and Seward power plants, these two plants shared some grid points. Individual grid point sites were selected on the basis of their being on elevated terrain in an attempt to find critical sites for the modeled impact of the power plant plumes. An additional 16 points were added that were varied between about .4 and 1.6 km from the Seward plant, being placed exactly at the maximum downwash concentration position as dictated by the forecast meteorologic conditions.

A final comment on the sensitivity of these air quality models is in order. It became necessary for an air quality computer program to be brought up on the MIT computer that was comparable to the ERT model. In doing so, formulas such as those given in this section were used and the air quality model was developed essentially from scratch, due to the great difference in machines and uses for the programs. Initial comparisons of the results of identical situations modeled on the two different computers showed a lack of correlation that was originally thought to have been caused by a gross error somewhere. Instruction-by-instruction comparison of the two programs showed that they were, in fact, the same except for several small differences:

1. Some of the locations of monitors were different by about 100 meters.

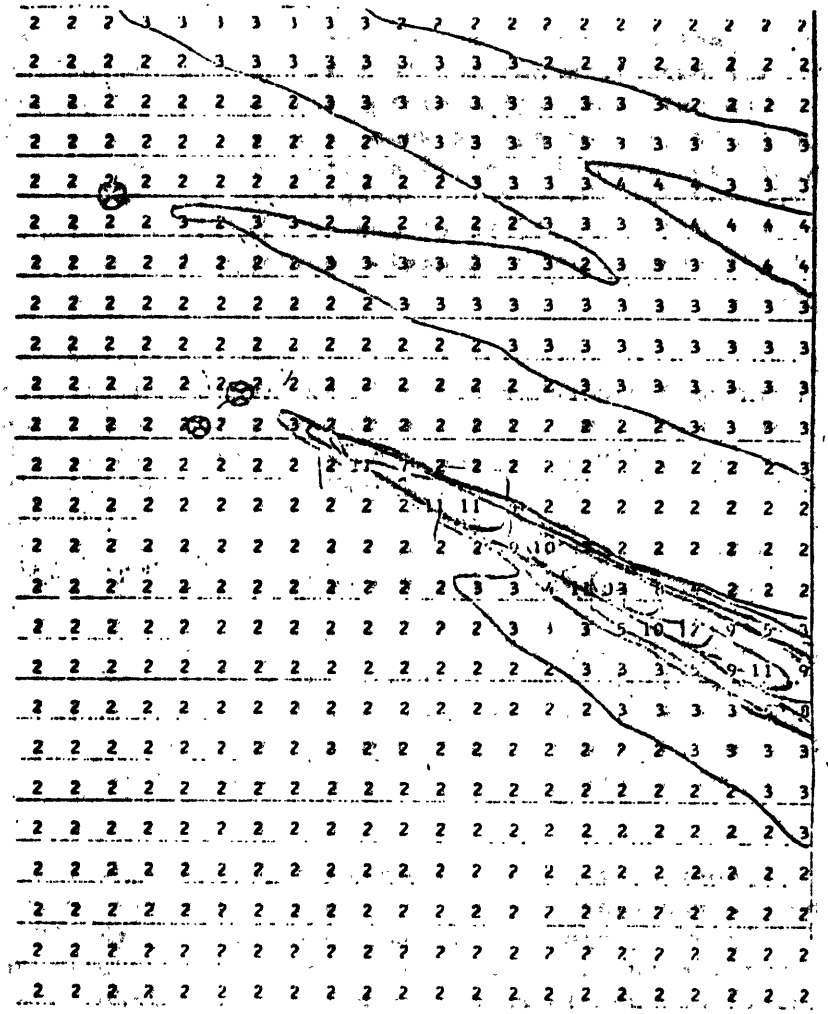


Figure 2.4.1. Portion of the Forty-by-forty Grid Showing Pollutant Concentrations in Parts per Hundred Million

2. Piecewise linear curves were used on the larger MIT machine to better approximate linear curves in the ERT program for heat emissions, diffusion parameters, and other nonlinear relationships.

3. A 1-1/4" difference existed between magnetic north of the wind directions (at ERT) and the universal transverse Mercator north (at MIT).

4. Numerous roundoff differences existed in terms of constants used in the programs, library functions for exponentials and trigonometric functions, and in the number of significant digits carried in the different computers' logics.

All of these differences were small, fractions of a percent, but in total, and through the air quality model, they did appear very large -- often 50% or greater. Eventually, there was no problem making the two programs equivalent $\pm$ 2 ppb, however, this experience cultivated a very healthy respect for the great sensitivities of air quality models to input data variations or errors. For this project, this sensitivity was modeled in a log normal fashion based upon statistics that had been developed, but the subject deserves more research and a more complete treatment. For example, in a sector-averaged Gaussian model, a point predicted at 0 ppb could be 100 meters from a point at 1000 ppb just by being out of the sector of the predicted wind direction. A binary type of probabilistic approach might be an appropriate path for future research.

2.5 Control Strategy

2.5.1 Deterministic Control Strategy

The control strategy of an SCS can operate to produce minimum costs, minimum environmental impact, or to make the most efficient use of limited supplies of clean fuel. In the present effort at Chestnut Ridge, a minimum cost criterion is used.

The control strategy development was dictated by three assumptions. First, it was assumed that the air quality standards should be dealt with probabilistically in response to the wording "not to be exceeded more than once per year" in the standards. Second, the control strategy had to consider the operation and economics of the interconnected power system. The optimum control would maintain standards while minimizing total system costs and disrupting normal operations as little as possible. Third, the control strategy development had to be sufficiently general to allow its application to other plants and systems outside of the Chestnut Ridge power complex. In this way, technology transfer would be facilitated.

The dominant factor affecting any particular control decision is the time scale over which the strategy must be developed. The critical time scales in SCS are those of atmospheric transport, forecasting, plant and system dynamics and the standards themselves. The minimum delay between control and reduced concentrations at the ground-level maximum is between 10 minutes and an hour or more, depending on the meteorology and location

of the maximum. The typical meteorological forecasting period is between three and six hours. Control action implementation requires from minutes (load shifting) to several hours (fuel switching with coal). The standards, being defined as time averages, make it necessary to consider three to 24 hours of concentration data and forecasts before control decisions are possible. The time scales of decisions for the operation of an electric power system are shown in Table 2.5.1.

There can be no single answer to the time scale that should best introduce SCS-type decisions; the level of control will depend both upon circumstances in the power system configuration and upon the environmental problem to be addressed. For example, if a power system does not include any storage facilities and the pollutant that is to be controlled can only be dealt with by load shifting, then the economic dispatch may be a good control level for an SCS. For a system that has a great deal of hydro storage capacity, for example, the SCS may be best operated by shifting maintenance and hydro production schedules to avoid climatologically predictable times of frequent conditions that yield high concentrations.

For the vast majority of systems, the atmospheric processes, forecasting capabilities, and SCS control measures will have time scales that are most appropriately handled at a unit commitment level of control. This allows for a consideration of power system economics and operating procedures that cannot be handled at the dispatch level, such as energy storage via pumped hydro, the avoidance of high power replacement costs during system peak, or the use of "slow" controls such as fuel switching.

In addition to the normal control alternatives of load shifting and fuel switching, the three large plants at Chestnut Ridge could alter stack gas temperatures and, consequently, have partial control of plume rise. This capability was relatively unique and inexpensive to operate, but because of the initially large physical and effective stack heights, could produce at most only a 7% reduction in ground-level concentrations under most conditions. Under conditions of near "punch through," when additional buoyancy might enable the plume to pass through an elevated inversion layer, control could be effectively 100%. Plume temperature modification corresponds to changing exit gas temperature from approximately 300°F to 600°F. The cost of plume ΔT control is small compared to up to $9000/hr/plant for load shifting. Added reserve capacity costs are not considered in these figures. Table 2.5.2 shows the comparative costs of the various control actions when a 2.5 mills/Kwh replacement power cost differential, a $12/ton fuel differential and 1%/40°F efficiency loss are assumed.

In each forecast period, the control strategy had four functions to perform with the support of the air quality forecasters, as shown in Figure 2.5.1. These were (1) check for violations; (2) determine required emissions; (3) determine control action, and (4) inform operators.

In the actual operation of this system during the demonstration phase of this project, some variations were made to the flow chart in Figure 2.5.1. First, an INPUT program was used to collect the met and power data, then a supervisor program called SCSOS was used to step through the entire time horizon, the previous 24 hours, and the subsequent 30 hours. For each time period of three hours, the input data was collected and set into

TABLE . 2.5.1
Utility Scheduling Problems

PROBLEM		TIME SCALE
MAINTENANCE SCHEDULING - NUCLEAR REFUELING	Determine 2-week to 2-month outage period for each plant for each year	1) Plan ahead for 1-2 years 2) Redo as conditions change
UNIT COMMITMENT	Determine hour-by-hour strategy for which plants will be committed (at what level) for next week. Constrained by "Maintenance-Scheduling"	1) Overall plan ahead for 1 week 2) Detailed plan for each day 3) Redo as conditions change
ECONOMIC DISPATCH	Determine minute-by-minute scheduling for each plant. Constrained by "Unit Commitment"	Redo every 5 minutes
AUTOMATIC GENERATION CONTROL (LOAD FREQUENCY CONTROL)	Adjust generation level to maintain system frequency and tie line flows at desired levels. Constrained by "Economic Dispatch"	2 - 10 second time constant

Factors which complicate solution:

1) pumped storage
2) fixed nuclear refueling scheduling
3) gas-oil contracts
4) interruptible loads
5) transmission line losses
6) ability to buy from neighbors
7) start-up cost and time-varying costs, maximum rate of change, maximum and minimum up and down times

Outages:

1) forced out
2) rescheduled maintenance
3) generation reserve

Table 2.5.2

Control Action Comparison

Maximum Control	Keystone	Homer City	Conemaugh	Seward
LOAD SHIFTING	0-1640MW	0-1200MW	0-1700MW	0-218MW
SO_2 Reduction	100%	100%	100%	100%
Cost	$45K/hr	$35K/hr	$45K/hr	$6K/hr
FUEL SWITCHING*	2.4-1.0%	2.4-1.0%	2.4-1.0%	2.4-1.0%
SO_2 Reduction	58%	58%	58%	58%
STACK GAS ΔT	50°F	300°F	300°F	50°F
SO_2 Reduction	negligible	7%	7%	negligible
Cost	$.5K/hr	$3K/hr	$3K/hr	$.1K/hr

* percent sulfur in fuel

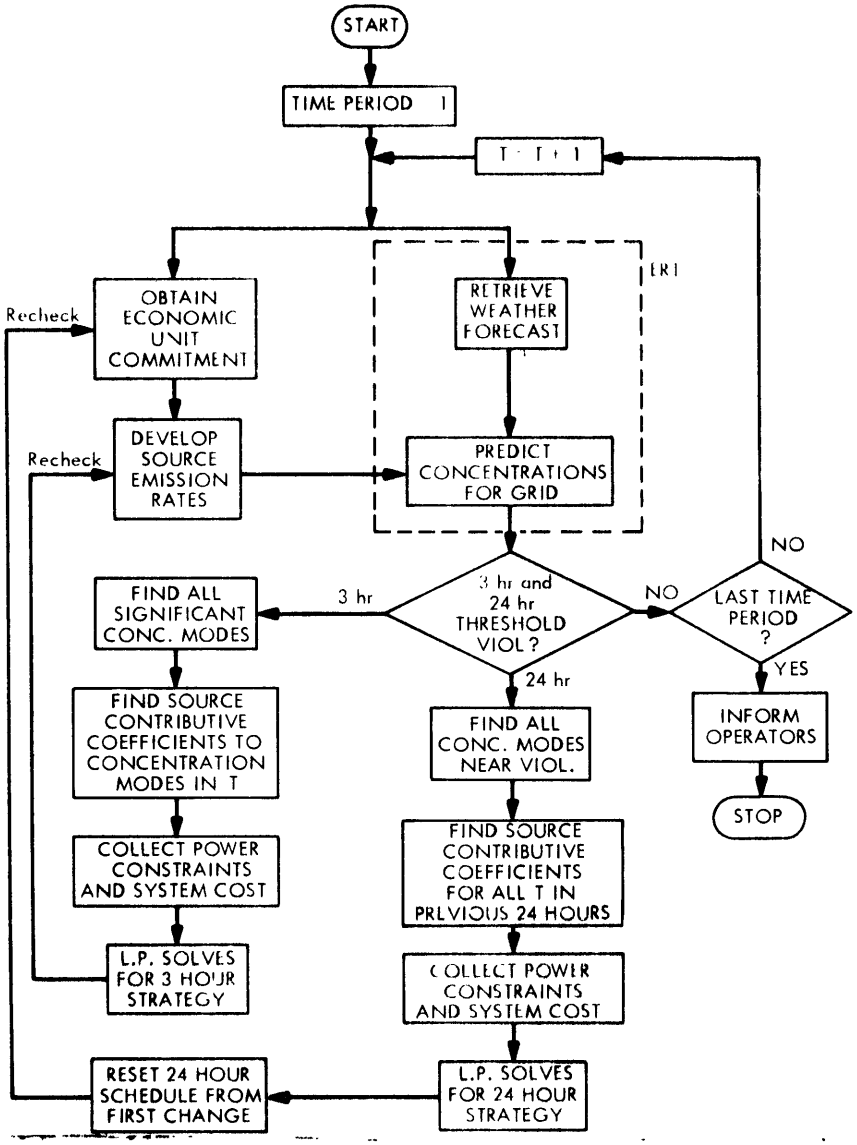


Figure 2.5.1 Block Diagram Flowchart of Control Strategy Logic

proper form by POLPRE. POLFAC then computed the pollution concentrations at each grid point, and ALMODE (later incorporated into SCSOS) sorted out the modes of the violation areas. After the last time period, the contributive factors were collected (in POLFAC) and the linear program control strategy program COSTRA was used just once to solve the entire problem all at once (see Figure 2.5-2). Figure 2.5-1 still represents the most general case, however, for this particular problem it was found that dimensionality would be no barrier to an all-inclusive linear program solution for the entire problem.

The several measures that were used to keep dimensionality in line are described now. The first measure involved looking only at modes of concentration within the violation areas, the contention being that if these highest concentrations can be pushed down below thresholds ,then the surrounding area should also go below the threshold. Even in the case where all plume centerlines are parallel, one can envision hypothetical cases where this contention would not be true. For example, if two sources line up and one plume goes over the other, then dropping some of the load from the upwind source might cause these two plume footprints to add together, resulting in a new mode of pollution concentration. To avoid this problem, no areas of high concentration were assumed to be controllable by a nearby mode if the contributive sources were not (nearly) equally responsible in both situations. Also, after the strategy was developed, checks were made to ensure that this and other nonlinearities hadn't caused incorrect control solutions.

Another dimensionality problem was avoided by exchanging the rectilinear grid for the series of polar grids centered at each source, as described in Section 2.4. One reason for this reduction in dimensionality was that in the rectilinear grid of 2 km by 2 km points, the plumes were able to shoot gaps and show up as apparent new concentration modes at points down three and over one, or down four and over one from other modes. Also, having the six or seven points along each radius available for comparison, it was much easier to sort out multiple downwind concentration peaks that were due only to terrain variations.

Finally, an enormous reduction in dimensionality was effected by following (EPA, 1974) "Guidelines for the Interpretation of Air Quality Standards" which claims that it is acceptable to consider the 24-hour average only for the calendar day. In this way, numerous overlapping and different violations were eliminated that showed up in 24-hour running average periods. In some cases, this reduced the number of decision variables by a factor of 10.

In order to understand the methods used to accomplish the central strategy functions, it is useful to consider first the deterministic control case and then to extend that to a probabilistic formulation. For clarity in this explanation piecewise linear curves are considered to be linear. The mathematical technique used to formulate the control strategy was linear programming. The manner in which this problem was set in a linear programming format is now described.

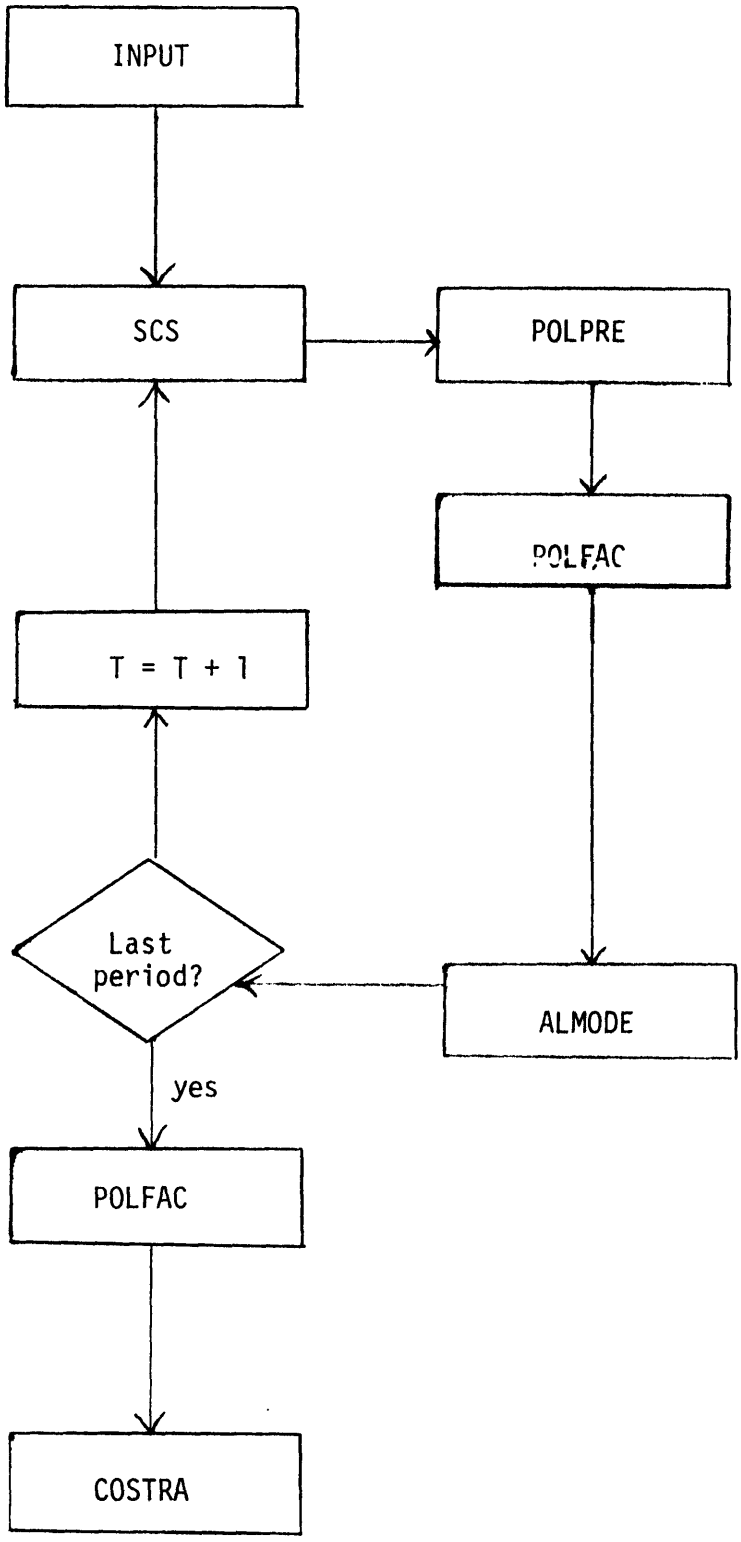


Figure 2.5.2 Flow of Subroutines Used to Determine Optimum SCS Control Strategy

In the course of the air quality check of the power system operation, there was no control strategy formulation if the pollution concentrations did not exceed the control thresholds. That is to say, for

t = time, in discrete hourly steps

x, y = the spatial two-dimensional grid location

$C(x, y, t)$ = SO_2 concentration at x, y and time t

TSO3, TS24 = concentration control thresholds to meet three-hour and 24-hour standards)ppnm)

No control strategy was formulated if:

$$\frac{1}{3} \sum_{t=T-2}^{T} C(x, y, t) < TSO3 \qquad (2.13)$$

and

$$\frac{1}{24} \sum_{t=T-23}^{T} C(x, y, t) < TS24 \qquad (2.14)$$

The control strategy was initiated if these thresholds are violated at any time at any grid point. In this event, the x, y grid was searched for all the significant modes of the averaged concentrations. For

j = index of concentration modes

$CB(j, t)$ = pollution concentration in pphm from background sources at mode j and time t

i = index of emission source (power plant units)

$H(j, i, t)$ = contributive factor of source i to mode j at time t (ppnm/gm/s)

$Q(i, t)$ = emission rate in gm/s of source i at time t

then the concentration at mode j was represented as:

$$C(j, t) = CB(j, t) + \sum_{i=1}^{8} H(j, i, t)\, Q(i, t) \qquad (2.15)$$

The various control modes of each plant unit were represented as:

m = 1 for high sulfur coal, no stack gas ΔT
m = 2 for low sulfur coal, no stack gas ΔT
m = 3 for high sulfur coal, full stack gas ΔT
m = 4 for low sulfur coal, full stack gas ΔT.

Thus one can define the fractional use of each control mode on each unit as:

X(m, i, t) = the 0 to 1 extent use of control mode m at unit i at time t

and a number of relationships follow, such as:

$$\sum_{m=1}^{4} X(m, i, t) \leq 1 \tag{2.16}$$

which constrains each unit to convex (linear) combinations of the control mode . If

QM(m, i, t) = the emission rate from the full use of control mode m at unit i at time t (in the case of stack gas ΔT, this is an "effective emission rate" and varies with meteorological conditions)

then:

$$\sum_{m=1}^{4} QM(m, i, t)\, X(m, i, t) = Q(i, t) \tag{2.17}$$

Also, if

PC(i, t) = maximum power output capability of unit i at time t (MW)

PS(i, t) = power putput strategy for source i at time t (MW) as determined by the control strategy

PM(i, t) = unit commitment strategy for unit i at time t (MW) as determined from economic conditions before addition of air quality standard constraints,

then:

$$\sum_{m=1}^{4} PC(i, t)\, X(m, i, t) = PS(i, t) \tag{2.18}$$

and if there is to be no load shifting whatsoever:

PM(i, t) = PS(i, t)

or if there is to be no load shifting outside the group of controlled sources:

$$\sum_{i=1}^{8} PS(i, t) = \sum_{i=1}^{8} PM(i, t) \quad \text{for all } t. \tag{2.19}$$

If we look now at what is involved in complying with, say, the 24-hour standard (the time previous to T):

$$\frac{1}{24} \sum_{t=T-23}^{T} \left[CB(j, t) + \sum_{i=1}^{8} H(j, i, t) \sum_{m=1}^{4} QM(m, i, t)\, X(m, i, t)\right]$$

$$\leq TS24 \text{ for all modes } j. \tag{2.20}$$

To represent the additional cost implied by air quality control, define

$\lambda S(t)$ = the replacement cost implied by air quality control using power from other sources on the system at time r($/MWhr)

$DM(m, i)$ = the cost of full power from source i in mode m($)

Then the additional cost of control, Z, is defined by

$$Z = \sum_{t=T-23}^{T} \lambda S(t) \sum_{i=1}^{8} (PM(i, t) - PS(i, t)$$

$$+ \sum_{t=T-23}^{T} \sum_{i=1}^{8} \sum_{m=1}^{4} DM(m, i)\, X(M, i, t)$$

$$- \sum_{t=T-23}^{T} \sum_{i=1}^{8} DM(1m\ i) \frac{PM(i, t)}{PC(i,t)} \tag{2.21}$$

where the first term is the replacement cost of power, the next term is the total cost with control and the last term is the total cost without control.

In the linear programming formulation, the cost of control, Z, in dollars, becomes the performance measure to be minimized; the decision variables are the control extents, X(m, i, t); and the constraint equations are those which force the concentration modes to be within the control threshold level (some additional constraint equations are imposed by conditions on the control modes, such as convexity).

2.5.2 Probabilistic Control Strategy

Having examined the deterministic form of the control strategy, we can now consider the concepts which will allow extension to a probabilistic formulation. There are a number of places where uncertainty is introduced into the process of developing an SCS control strategy, see Table 2.5-3. The variations in environmental impact are not considered due to the constant threshold represented by the air quality standards that are the only environmental objectives of this project. The uncertainty in the load and the inexactness of the unit commitment schedule when compared to the actual schedule are both very small, usually within ± 5% of predicted. Large errors result only when generation is forced out of service by breakdowns and this can only decrease pollutant concentrations. Some times this can be very beneficial; for example, in the demonstration phase of this project, just when a load shift was called for at a plant, it experienced a forced outage. The uncertainties of concern thus are pared down to those resulting from unknown sulfur levels in coal being used, those in the air quality model, and those in the meteorological forecast.

Consider the following definitions relevant to understanding the probabilistic model used to evaluate violations and to determine control actions:

$C(t)$ = maximum concentration over all $C(x, y, t)$

$Q(t)$ = emission rate of source without control

$M(t)$ = meteorological function relating $C(t)$ to $Q(t)$ and which includes the effects of effective stack height, wind conditions, mixing depths, and any other pertinent meteorological parameters.

With or without an operating SCS, the observed maximum concentration $\overline{C}(t)$ is related to the actual emissions $Q(t)$ through the meteorological function $M(t)$ as follows (time notation dropped for simplicity):

$$\overline{C} = Q \cdot M \qquad \text{at time } t. \tag{2.22}$$

Assume that there exist probability density functions for M and Q, and we wish to generate a frequency distribution for $\overline{C}$ when no SCS is operating. If A is any concentration value and if Q and M are independent of each other and random variables, then for an increment of emission rate ε:

$$P_C(\overline{C} = A) = P_Q(Q = \varepsilon) \cdot P_M(M = A/\varepsilon)$$

$$+ P_Q(Q = 2\varepsilon) \cdot P_M(M = A/2\varepsilon)$$

$$+ P_Q(Q = 3\varepsilon) \cdot P_M(M = A/3\varepsilon) + \ldots$$

$$+ P_Q(Q = n\varepsilon) \cdot P_M(M = A/n\varepsilon) + \ldots + \Delta\varepsilon$$

or, in the limit as $\Delta\varepsilon \to 0$:

$$P_C(\overline{C} = A) = \int_0^\infty P_Q(Q = \zeta) \cdot P_M(M = A/\zeta)d\zeta \tag{2.23}$$

This equation states that the probability density function for maximum ground-level concentrations can be derived from the convolution of the probability density functions for M and Q. Therefore, the frequency distribution of ground-level concentrations for an uncontrolled plant can be determined from determinations of M and Q.

Consider next the case when the SCS is operating:

$$\overline{C}_c = Q_c \cdot M \quad \text{at time t.} \tag{2.24}$$

In this case, where subscript c denotes the functional value when the SCS is operating, Q_c is no longer independent of meteorology since the operation of the SCS depends on meteorological forecasting.

P_Q will, therefore, also be dependent on P_M and will vary for different control strategies. For computer solutions to the correlated integration, the dependence of these quantities upon each other can be readily simulated. Given P_M and P_Q, it is possible to use control strategy rules for determining Q_c to numerically evaluate $P_{\overline{C}}$ under the SCS control. The value of Q_c in each case is determined from the predicted value of concentration C_{o_c} and from the strategy used. From the resulting distribution of Q_c, the value of P_{Cc} is easily obtained from the equation:

$$P_{C_c}(C_c = A) = \int_0^\infty P_{Q_c}(Q_c = \zeta) \cdot P_M(M = A/\zeta)d\zeta \tag{2.25}$$

Table 2.5-3

KEY FACTORS AFFECTING DESIGN/OPERATION AND ANALYSIS

	DETERMINISTIC	PROBABILISTIC
LOAD	TIME VARYING DAILY, SEASONAL PATTERNS	. WEATHER DEPENDENCE . RANDOM VARIATIONS . MODELING ERRORS
GENERATION	. PLANT TYPES, FUELS, LOCATIONS . TIME VARYING OPERATING SCHEDULES TO FOLLOW LOAD	. FORCED OUTAGES . RESPONSE TO LOAD VARIATIONS
METEOROLOGICAL	. TIME VARYING DAILY, SEASONAL PATTERNS	. RANDOM VARIATIONS . MODELING ERRORS
AMBIENT AIR QUALITY	. INTERACTION OF GENERATION & METEOROLOGY . TIME VARYING BACKGROUND DAILY, SEASONAL PATTERNS	. DISPERSION . RANDOM BACKGROUND EMISSIONS . MODELING ERRORS
ENVIRONMENTAL IMPACT	. POPULATION PATTERN . TIME VARYING SUSCEPTIBILITY	. IMPACT VARIATIONS . RANDOM EXPOSURE PATTERNS

Note the parallel nature of this equation and the equation for P_{C_C} (second previous equation). With the distribution P_C available for the control action Q_c, it is straightforward to calculate the probability of exceeding C_s, the standard, by using the cumulative distribution shown in Figure 2.5-3. In that figure, the function of an SCS is seen to be the shifting of the right hand portion of the distribution so that the probability of exceeding the standard is acceptable (i.e, less than the probability specified by the standard, P_{C_S}).

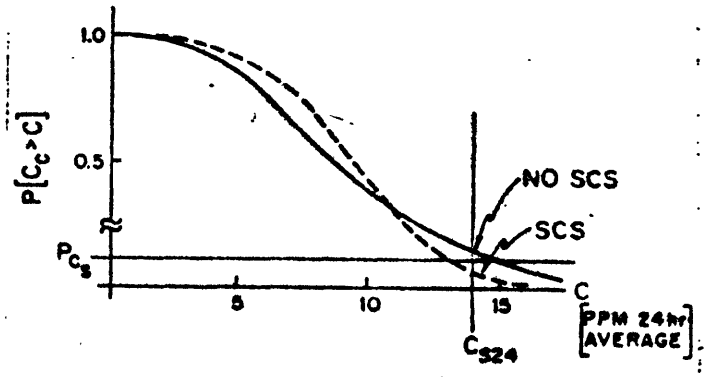


Figure 2.5-3 Cumulative Distribution of P_{C_C}

This probabilistic formulation, as it has been presented, could pose computational problems due to its complexity and often repeated usage. Fortunately, there is a widely used and often substantiated simplifying assumption that can be invoked. That is that the probabilistic distribution of pollutant concentrations at any one point in time and space is log normal. This log normality assumption enables the description of the entire distribution function in terms of two parameters--the geometric mean and geometric deviation. The multiple source modeling requires more consideration. In particular, if each of a number of contributing sources is modeled in a log normal fashion, their summation is not log normal. Thus, an approach that could be taken would involve the log normal modeling of the combination of sources around a midpoint equal to the sum of the means. The variances could then be relegated to the contributing sources by proportional relationship. The simplicity of this mean and deviation modeling makes it easy to introduce it into the deterministic linear program format described previously.

The best way of showing the major features of the control strategy program is with the example in Table 2.5-4. At the date and time shown and with a very conservative threshold on the concentration mode detector(in this case picking up any concentrations greater than one-third of the air quality standards), four areas of high concentrations are picked up.

The conservatism in this threshold must be calculated from considerations of probability of violation being aimed at and the worst case mean and standard deviation of the prediction error.

There are then four very particular situations represented by the lines marked 2. in Table 2.5-4. Historical data can give statistics on what error exists in:

1. knowledge of sulfur emission,

2. the dispersion modeling for these particular meteorological conditions, and

3. the prediction of background concentrations.

In short, the mean and the geometric standard deviation for these situations can be gotten from lookup tables and thus the estimated probability of violation can be computed as well as the thresholds to which each of these situations must be controlled to result in specific probabilities of violation. The probabilistic control problem has thus been reduced to a deterministic problem.

It was found to be crucial to have readily available the unit commitment schedule upon which these predictions have been based. Line 3 in Table 2.5-4 shows this schedule; the minus signs were used to show sources that in no way contributed to the concentration modes (thus facilitating decisions about updating the strategy with new unit commitment information).

Again in Table 2.5-4, the replacement cost of power, "system lambda" was input for the current time period and a computerized predictor displayed the forecasted information based upon historic system lambda curves. The reserve credit was the amount of credit a system was given for having a facility off-line but capable of picking up load at short notice. One final point about Table 2.5-4 was the display of several alternative strategies, thus leaving to the power system operator a choice of what would be best in light of intangible power system considerations. Note that some of these non-optimum strategies were not capable of meeting the desired probability of violation.

The downwash model described in Section 2.4 was included in the control strategy only in a linear approximation. Thus, whenever downwash situations occurred, these were brought to the attention of the operator, complete lists of output versus concentrations were available, and a separate subroutine was used for making on-the-spot checks of downwash strategies.

Further improvements in the control strategy area should be aimed directly at appropriate air quality modeling. Decisions on the eventual desired probability of violation will directly affect the air quality model chosen. For example, in a very conservative model, mixing depths could be systematically reduced from predicted levels and wind directions calculated $\pm$ one or two sectors in a recursive fashion to determine a worst-case situation.

Table 2.5-4

CONTROL STRATEGIES FORM 1-D

SUPPLEMENTARY CONTROL SYSTEM DEMONSTRATION PROJECT

1. DATE: 5/05/75 PREDICTIONS OF SO CONCENTRATIONS AS OF 17*02*50(TIME)
X

PREDICTED PLACES AND PROBABILITIES OF EXCEEDING STANDARDS:

	DATE BEGIN TIME	STAND.	EXP.AVE CONC(PPM)	BAC.GD (PPM)	EST.PROB OF VIOL	LOCATION (UTM) KM EAST	KM NORTH
2.	5/ 5 3:00	3HR.	0.510	0.120	0.5432	664.70	4457.20
2.	5/ 5 4:00	3HR.	0.650	0.100	0.9970	662.30	4459.00
2.	5/ 5 6:00	3HR.	0.470	0.090	0.3672	660.90	4455.10
2.	5/ 5 7:00	1HR.	0.270	0.240	0.0094	658.00	4455.30

BASED ON THE FOLLOWING ASSUMPTIONS OF PLANT OPERATING LEVELS(MW):
(MINUS SIGNS INDICATE SOURCES NOT CONTRIBUTING TO CONC. MODES)

	DATE	TIME	KEY-1	KEY-2	HMC-1	HMC-2	CON-1	CON-2	SEW34	SEW-5
3.	5/ 5	3:00	0.	-80.	0.	-695.	0.	590.	-50.	0.
3.	5/ 5	4:00	0.	-81.	0.	-695.	0.	600.	50.	0.
3.	5/ 5	5:00	0.	-94.	0.	-695.	0.	610.	50.	0.
3.	5/ 5	6:00	0.	-94.	0.	-695.	0.	620.	150.	0.
3.	5/ 5	7:00	0.	100.	0.	-695.	0.	620.	-145.	0.
3.	5/ 5	8:00	0.	-100.	0.	700.	0.	-620.	-145.	0.

ASSUMING(IN MILLS/KWHR OR $/MWHR):
4. SYSTEM LAMBDA 21. AT 3:00; 22. AT 4:00; 24. AT 5:00; 25. AT 6:00;
4. SYSTEM LAMBDA 25. AT 7:00; 27. AT 8:00;
5. RESERVE CREDIT 3. AT 3:00; 3. AT 4:00; 3. AT 5:00; 3. AT 6:00;
5. RESERVE CREDIT 3. AT 7:00; 3. AT 8:00;

1 STRATEGIES 1 AIMING AT PROBABILITY OF VIOLATION OF 10.52001
L=LOAD SHIFTING, F=FUEL SWITCHING, S=STACK GAS TEMP MODIF

	CONTROL TYPE	PLANT	% OF FULL CONTROL	HR.	TOTALS: MAX AVE CONC	STAND.	EST PROB OF VIOL.	CONTROL COST
6. ESTIMATED LEAST COST COMBINATION:								
	F	CON-2	34%	5:00	0.480	3HR.	0.4114	
	L	SEW34	100%	4:00	0.500	3HR.	0.5000	
	L	SEW34	100%	5:00	0.470	3HR.	0.3672	
					0.270	1HR.	0.0094	$ 1680.
7. (L) LOAD SHIFTING ONLY:								
	L	CON-2	21%	4:00	0.480	3HR.	0.4114	
	L	SEW34	100%	4:00	0.500	3HR.	0.5000	
	L	SEW34	100%	5:00	0.470	3HR.	0.3672	
					0.270	1HR.	0.0094	$ 3140.
8. (F) FUEL SWITCHING ONLY:								
	F	CON-2	91%	5:00	0.430	3HR.	0.2041	
	F	SEW34	100%	4:00	0.500	3HR.	0.5000	
	F	SEW34	100%	5:00	0.470	3HR.	0.3672	
					0.270	1HR.	0.0094	$ 1840.

(continued)

Table 2.5-4 (continued)

9. (S) STACK GAS TEMP MODIFICATION ONLY:

S	CON-2	100%	4:00	0.500	3HR.	0.4993	
S	CON-2	100%	5:00	0.631	3HR.	0.9926	
S	CON-2	100%	6:00	0.467	3HR.	0.3549	
S	SEW34	100%	4:00	0.270	1HR.	0.0094	$ 400.
S	SEW34	100%	5:00				

STRATEGIES I AIMING AT PROBABILITY OF VIOLATION OF 10.2000]
L=LOAD SHIFTING, F=FUEL SWITCHING, S=STACK GAS TEMP MODIF

CONTROL TYPE	PLANT	% OF FULL CONTROL	HR.	TOTALS: MAX AVE CONC	STAND.	EST PROB OF VIOL.	CONTROL COST
6. ESTIMATED LEAST COST COMBINATION:							
F	CON-2	93%	5:00	0.429	3HR.	0.2000	
F	CON-2	46%	6:00	0.461	3HR.	0.2000	
F	SEW34	100%	4:00	0.429	3HR.	0.2000	
F	SEW34	100%	5:00	0.270	1HR.	0.0094	$ 2750.
F	SEW34	28%	6:00				
7. (L) LOAD SHIFTING ONLY:							
L	CON-2	58%	4:00	0.429	3HR.	0.2000	
L	SEW34	75%	4:00	0.461	3HR.	0.2000	
L	SEW34	100%	5:00	0.429	3HR.	0.2000	
L	SEW34	34%	6:00	0.270	1HR.	0.0094	$ 7060.
8. (F) FUEL SWITCHING ONLY:							
F	CON-2	93%	5:00	0.429	3HR.	0.2000	
F	CON-2	46%	6:00	0.461	3HR.	0.2000	
F	SEW34	100%	4:00	0.429	3HR.	0.2000	
F	SEW34	100%	5:00	0.270	1HR.	0.0094	$ 2750.
F	SEW34	28%	6:00				
9. (S) STACK GAS TEMP MODIFICATION ONLY:							
S	CON-2	100%	3:00	0.496	3HR.	0.4840	
S	CON-2	100%	4:00	0.631	3HR.	0.9926	
S	CON-2	100%	5:00	0.456	3HR.	0.3082	
S	CON-2	100%	6:00	0.270	1HR.	0.0094	$ 820.
S	SEW34	100%	4:00				
S	SEW34	100%	5:00				
S	CON-2	100%	7:00				
S	HMC-2	100%	8:00				
S	SEW34	100%	6:00				

(continued)

Table 2.5-4 (continued)

STRATEGIES AIMING AT PROBABILITY OF VIOLATION OF .10000
L=LOAD SHIFTING, F=FUEL SWITCHING, S=STACK GAS TEMP MODIF

CONTROL TYPE	PLANT	% OF FULL CONTROL	HR.	TOTALS: MAX AVE CONC	STAND.	EST PROB OF VIOL	CONTROL COST
6. ESTIMATED LEAST COST COMBINATION:							
F	CON-2	33%	4:00	0.396	3HR.	0.1000	
F	CON-2	100%	5:00	0.443	3HR.	0.1000	
F	CON-2	29%	6:00	0.396	3HR.	0.1000	
F	SEW34	100%	4:00	0.270	1HR.	0.0094	$ 3510.
F	SEW34	100%	5:00				
F	SEW34	86%	6:00				
S	SEW34	14%	6:00				
7. (L) LOAD SHIFTING ONLY:							
L	CON-2	82%	4:00	0.396	3HR.	0.1000	
L	SEW34	47%	4:00	0.443	3HR.	0.1000	
L	SEW34	100%	5:00	0.396	3HR.	0.1000	
L	SEW34	62%	6:00	0.270	1HR.	0.0094	$ 9630.
8. (F) FUEL SWITCHING ONLY:							
F	CON-2	33%	4:00	0.396	3HR.	0.1000	
F	CON-2	100%	5:00	0.443	3HR.	0.1000	
F	CON-2	29%	6:00	0.396	3HR.	0.1000	
F	SEW34	100%	4:00	0.270	1HR.	0.0094	$ 3510.
F	SEW34	100%	5:00				
F	SEW34	87%	6:00				
9. (S) STACK GAS TEMP MODIFICATION ONLY:							
S	CON-2	100%	3:00	0.496	3HR.	0.4840	
S	CON-2	100%	4:00	0.631	3HR.	0.8826	
S	CON-2	100%	5:00	0.456	3HR.	0.3082	
S	CON-2	100%	6:00	0.270	1HR.	0.0094	$ 820.
S	SEW34	100%	4:00				
S	SEW34	100%	5:00				
S	CON-2	100%	7:00				
S	HMC-2	100%	8:00				
S	SEW34	100%	6:00				

(continued)

Table 2.5-4 (continued)

STRATEGIES AIMING AT PROBABILITY OF VIOLATION OF 10.0001
L=LOAD SHIFTING, F FUEL SWITCHING, S=STACK GAS TEMP MODIF.

CONTROL TYPE	PLANT	% OF FULL CONTROL	HR.	TOTALS: MAX AVE CONC	STAND.	EST PROB OF VIOL.	CONTROL COST
6. ESTIMATED LEAST COST COMBINATION:							
F	CON-2	65%	3:00	0.303	3HR.	0.0030	
F	CON-2	100%	4:00	0.385	3HR.	0.0030	
F	CON-2	100%	5:00	0.303	3HR.	0.0030	
F	CON-2	68%	6:00	0.243	1HR.	0.0030	$ 8590.
F	SEW34	100%	5:00				
F	HMC-2	80%	8:00				
F	SEW34	100%	6:00				
L	KEY-2	74%	7:00				
F	KEY-2	26%	7:00				
S	KEY-2	26%	7:00				
7. (L) LOAD SHIFTING ONLY:							
L	CON-2	100%	4:00	0.303	3HR.	0.0030	
L	CON-2	45%	5:00	0.385	3HR.	0.0030	
L	SEW34	83%	5:00	0.303	3HR.	0.0030	
L	HMC-2	34%	8:00	0.243	1HR.	0.0030	$ 22810.
L	SEW34	100%	6:00				
L	KEY-2	89%	7:00				
8. (F) FUEL SWITCHING ONLY:							
F	CON-2	65%	3:00	0.303	3HR.	0.0030	
F	CON-2	100%	4:00	0.385	3HR.	0.0030	
F	CON-2	100%	5:00	0.303	3HR.	0.0030	
F	CON-2	68%	6:00	0.252	1HR.	0.0046	$ 7420.
F	SEW34	100%	5:00				
F	HMC-2	80%	8:00				
F	SEW34	100%	6:00				
F	KEY-2	100%	7:00				
9. (S) STACK GAS TEMP MODIFICATION ONLY:							
S	CON-2	100%	3:00	0.496	3HR.	0.4840	
S	CON-2	100%	4:00	0.631	3HR.	0.9926	
S	CON-2	100%	5:00	0.456	3HR.	0.3082	
S	CON-2	100%	6:00	0.269	1HR.	0.0091	$ 840.
S	SEW34	100%	4:00				
S	SEW34	100%	5:00				
S	CON-2	100%	7:00				
S	HMC-2	100%	8:00				
S	SEW34	100%	6:00				
S	KEY-2	100%	7:00				

There are a number of philosophical issues about SCS control strategies that should be raised here. First, SCS is not tied to SO_2 control and, in fact, it offers the only known method of controlling many of the hazardous emissions, including many carcinogens. A common argument against SCS has been the long-range dispersion of sulfates, but, in fact, an SCS can be geared to control sulfates, by, for example, operating coastal units under favorable winds, load-shifting to reduce potential sulfate concentrations, and even fuel switching at times when slow winds are aiming the emissions toward faraway population centers.

Another philosophical issue sometimes raised is that of "air quality equity." In other words, is it fair to pollute one city a little more in order to save another city considerable pollution? How can such a debt be repaid? Questions about methods of handling air quality equity will grow and persist until it is determined how people can be subjected to pollution in direct proportion to their responsibility for it, or until the (impossible) day of perfectly controlled emission sources.

3.0 DEMONSTRATION PERIOD

The Chestnut Ridge SCS was implemented in a demonstration from March 10, 1976 through March 26, 1976 in order to assess its ability to protect ambient air quality standards. The short duration of the demonstration period and several artificial constraints on control action prevented the demonstration from being conclusive, and it cannot be said, on the basis of our observations, whether a permanent implementation of the Chestnut Ridge SCS would maintain ambient air quality at acceptable levels.

The demonstration period began late because of design problems which delayed Penelec's approval of control action. Control during demonstration was limited to the Seward station since it was the only plant which appeared to threaten standards. The demonstration was halted when Seward 5, which accounts for 70% of Seward's capacity, was shut down for conversion from a 232.5-foot to a 600.0-foot stack. The conversion required six weeks and resulted in completely new dispersion patterns in the region.

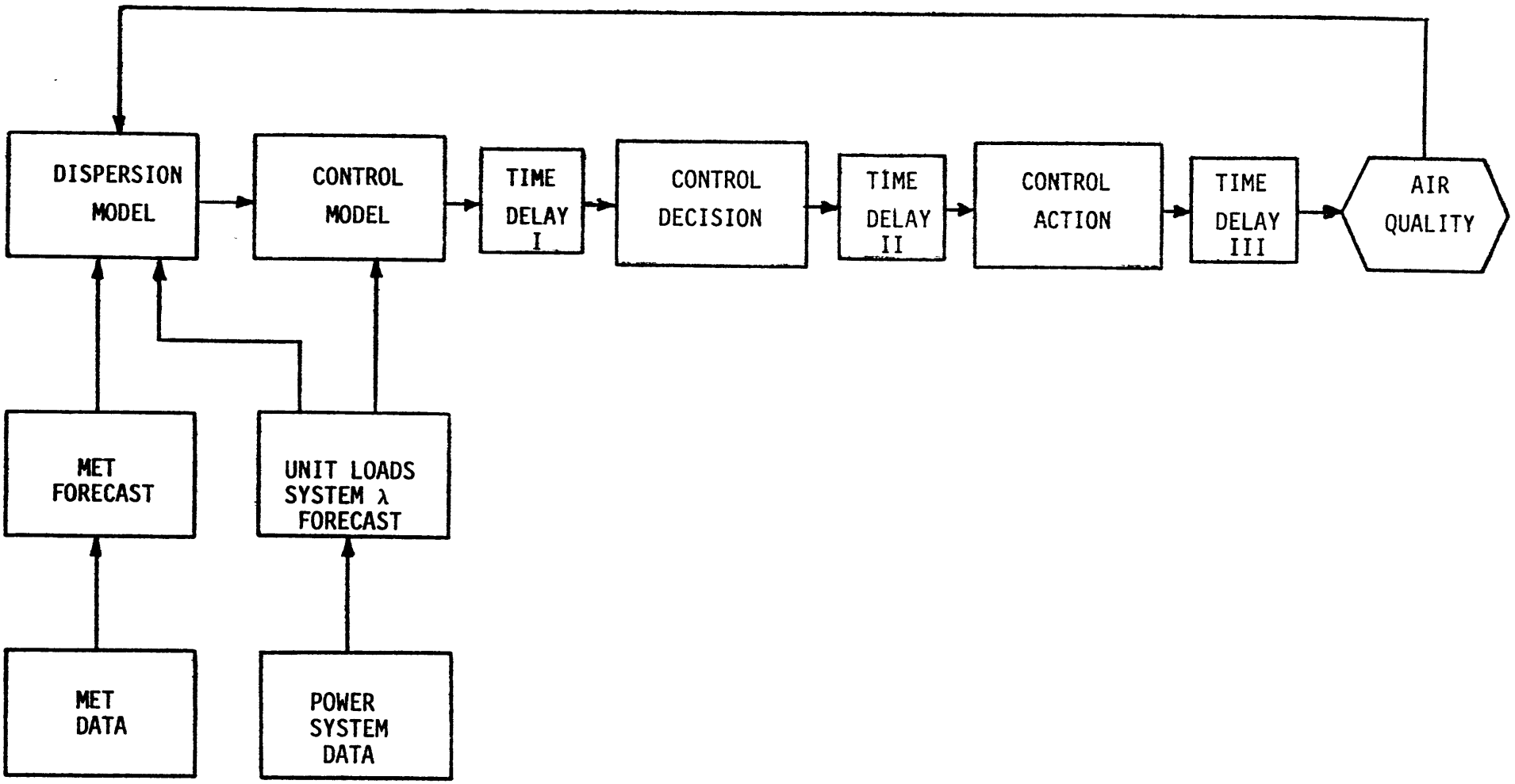
Penelec placed two constraints on control before allowing the demonstration to begin. First, all proposed control actions were subject to approval by Penelec dispatchers before implementation. Second, a ceiling of $10,000 was placed on Penelec's demonstration costs, including replacement power, control, monitoring, and coal analysis. These constraints would not be present in an operating SCS. Although the veto was used only once by Penelec and the cost ceiling was not reached, these constraints directly affected the control actions taken. In order to obtain as many demonstration control actions as possible, without precipitating a veto or exceeding the cost ceiling, the control requested from Penelec was always less than that recommended by the control model.

3.1 Demonstration Approach

The system design described in Section 2 consisted of a dispersion model and a control model. During the demonstration period these components were used in a series of steps which developed recommended control actions for the power plants as functions of meteorological and power system data. The process is summarized in Figure 3.1.1.

Meteorological forecasts for Chestnut Ridge were made twice daily at approximately 0000 and 1200 EST with the aid of National Weather Service synoptic data. Average wind sector, wind speed class, Turner-Pasquill stability class, and mixing depth were forecast for each of the ten successive three-hour periods beginning with 0300-0600 or 1500-1800. Each morning the power system dispatchers would provide the daily schedule of plant outputs and expected system replacement power costs in six-hour blocks beginning at 0600.

The meteorological and power system forecasts were collected at Concord, Massachusetts, where the met forecasters were located. Using an interactive teletype console, the forecasters input the data to the dispersion and control



Demonstration Period Implementation

Figure 3.1.1

models, which were resident on a separate computer system in Cambridge, Massachusetts. After processing the data, a control report was returned via the teletype to Concord. As shown in Figure 3.1.2 the control report gives options for different types of control action. (See also section 2.5.2.)

These options were considered, along with the constraints Penelec had imposed, and a requested control decision was prepared. Full recommended control was never requested. This control decision was telephoned to Johnstown where the dispatchers made the final control decision and requested control action at Seward. If control beyond the next forecast time was indicated, Penelec would be warned of possible control and the situation would be reassessed at the next forecast.

Both three-hour and twenty-four hour standards were considered. The three-hour standards were considered only for the eight daily 3-hour forecast periods while the twenty-four hour standard was based on a daily (0000-2400) average. This three-hour average interpretation differs from the EPA's running three-hour interpretation, and was chosen because of the impossibility of distinguishing meteorological or power system phenomena accurately on a one-hour basis for a 30-hour forecast period. The 24-hour average interpretation was the same as EPA's. Threshold values of 500 ppb/3 hr and 140 ppb/24 hr were chosen and a threshold probability of 0.5 was used.

Several special measures were taken during control action. An instrumented van was on call during the daytime and could "plume chase" or drive to the site of predicted high concentrations. The van recorded wind speed, wind direction, and SO_2 concentration. No emission monitors were available, but special grab samples of coal were taken from the mills during control action and later analyzed for sulfur content. Only load shifting at Seward was used as a control action.

The data from the Chestnut Ridge SCS was collected in several forms, as shown in Figure 3.1.3. The Master Database kept all of the observed monitor, tower, meteorological and plant data, for the duration of the study (September 1974 to March 1976), including the demonstration period. The Operating Database was used only during the demonstration period and automatically kept dated records of the SCS meteorological and power system inputs and the dispersion and control model outputs, including the recommended control actions. Other data were collected on the fuel characteristics, the actual control actions taken, and the observed van results. The total set of data would normally be available to regulatory bodies authorizing permanent SCS operation.

3.2 Demonstration Results

In general terms, the demonstration period included 32 forecasts made over a 16-1/2 day period, with one forecast record missing because of a communications equipment failure. Three forecasts had errors which prevented the preparation of a complete control report; two requested control at plants other than Seward, which violated the demonstration constraints, and three predicted violations or controls only during the next forecast period.

Figure 3.1.2

CONTROL STRATEGIES FORM T-D

SUPPLEMENTARY CONTROL SYSTEM DEMONSTRATION PROJECT

1. DATE: 3/20/76 PREDICTIONS OF SO X CONCENTRATIONS AS OF 12*19*40 (TIME)

PREDICTED PLACES AND PROBABILITIES OF EXCEEDING STANDARDS:

	DATE BEGIN TIME	STAND.	EXP.AVE CONC (PPM)	BACKGD (PPM)	EST.PROB OF VIOL	LOCATION (UTM) KM EAST	KM NORTH
2.	3/20-24:00	24HR.	0.158	0.102	0.5492	667.82	4475.42
2.	3/20-24:00	24HR.	0.154	0.098	0.5376	667.65	4475.47
2.	3/20-24:00	24HR.	0.169	0.108	0.5761	667.56	4475.26

BASED ON THE FOLLOWING ASSUMPTIONS OF PLANT OPERATING LEVELS (MW):

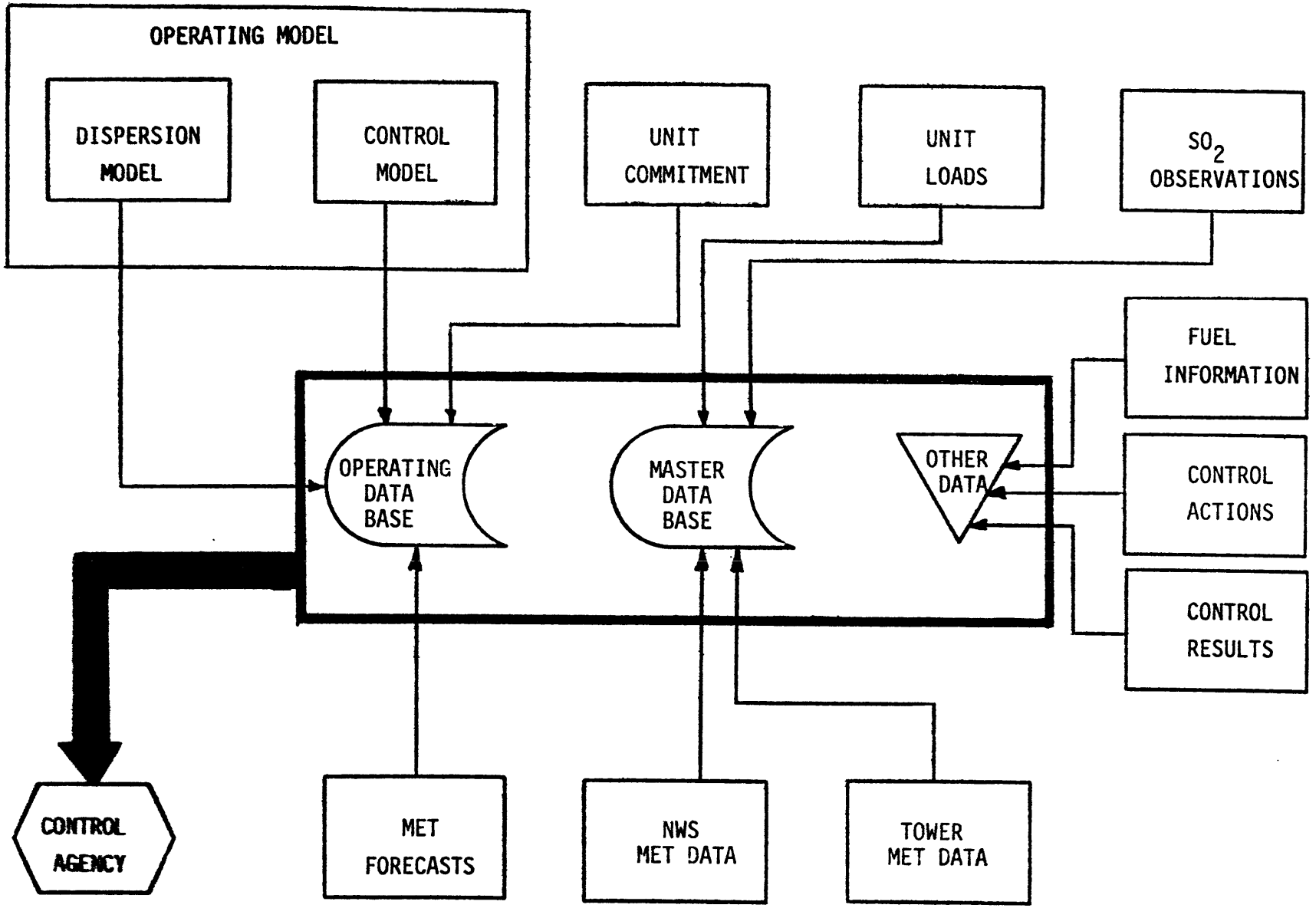
	DATE	TIME	KEY-1	KEY-2	HMC-1	HMC-2	CON-1	CON-2	SEW-4	SEW-5
3.	3/20	12:00	810.	820.	0.	500.	500.	0.	41.	137.

ASSUMING (IN MILLS/KWHR OR $/MWHR):

4. SYSTEM LAMBDA 19. AT 12:00; 19. AT 15:00; 19. AT 18:00; 20. AT 21:00;
4. SYSTEM LAMBDA 13. AT 24:00; 10. AT 27:00; 10. AT 30:00; 17. AT 33:00;
4. SYSTEM LAMBDA 19. AT 36:00; 19. AT 39:00;
5. RESERVE CREDIT 0. AT ALL TIMES

L STRATEGIES 1 AIMING AT PROBABILITY OF VIOLATION OF 10.50001
L=LOAD SHIFTING, F=FUEL SWITCHING, S=STACK GAS TEMP MODIF

	CONTROL TYPE	PLANT	% OF FULL CONTROL	3HRS BEGIN	TOTALS: MAX AVE CONC	STAND.	EST PROB OF VIOL.	CONTROL COST
6.	ESTIMATED LEAST COST COMBINATION:							
	L	SEW-5	54%	12:00	0.132	24HR.	0.4778	
					0.127	24HR.	0.4625	
					0.140	24HR.	0.5000	$ 1680.
7.	(L) LOAD SHIFTING ONLY:							
	L	SEW-5	54%	12:00	0.132	24HR.	0.4778	
					0.127	24HR.	0.4625	
					0.140	24HR.	0.5000	$ 1680.
8.	(F) FUEL SWITCHING ONLY:							
	S	SEW34	100%	12:00	0.158	24HR.	0.5478	
	S	SEW-5	100%	12:00	0.153	24HR.	0.5361	
					0.169	24HR.	0.5746	$ 180.
9.	(S) STACK GAS TEMP MODIFICATION ONLY:							
	S	SEW34	100%	12:00	0.158	24HR.	0.5479	
	S	SEW-5	100%	12:00	0.153	24HR.	0.5362	
					0.169	24HR.	0.5746	$ 180.



Data Collection

Figure 3.1.3

Of the remaining 24 forecasts, 5 predicted control would be required. The 0149-16 March forecast was not implemented because an unexpected rise in wind speed changed the forecast assumptions, eliminating the need for control. The 0003-25 March forecast was not implemented because a forced outage of Seward occurred at 0030 and was projected to last at least until 0600. The three actual control actions are summarized in Table 3.2.1. The control reports for all the forecasts are summarized in Table 3.2.2.

The effectiveness of our demonstration SCS should be considered in terms of four interrelated questions:

- model validity: Do the dispersion and control models accurately reflect the dynamics of the physical and economic systems they represent?
- control strategy: Can an operational system be designed and implemented to generate control recommendations with adequate lead time for SCS?
- control action: Can emission controls be implemented on the basis of control recommendations with sufficient reliability for SCS?
- control impact: Can maintenance of air quality standards be demonstrated to be the result of SCS use?

Section 5 is devoted to the model validity issue while Section 6 is concerned with SCS reliability analysis. This demonstration period analysis can only address the last three questions.

The demonstration experience with control strategy development was favorable. Although there were several hardware, software, and procedural problems when the real-time operation of the SCS began, none were serious. By the end of the demonstration period, a routine for the forecasters and power system dispatchers had evolved which allowed the control strategies to be developed on a regular schedule.

Control actions in the demonstration were limited to load shifting, which is relatively simple to implement. Figure 3.2.1 shows the reduction in Seward station output during the three control actions. There were no problems in obtaining the desired reduction in MW level requested with a 3-hour forecast time period. While the MW level is easy to control, fuel sulfur content is not. Table 3.2.3 shows the variations of fuel sulfur content during the control actions and indicates that small (5-10%) changes in MW levels are probably ineffectual since they lie in the "noise" level of the sulfur contents. The use of emission sensors would help avoid such problems. Fuel Switching was not tried, but would have required 6 to 8 hours lead time; this could have necessitated changes in forecasting procedures.

The small sample size (three) prevents any conclusions from being drawn about the impact of control actions on ambient air quality. Also clouding the issue are the differences between recommended and requested controls and data collection difficulties with the van monitors.

The differences in control requests from model recommendations were always made to reduce costs at the price of increased emissions. It is impossible to correlate such small control actions with air quality changes in the presence of natural concentration variability given three samples. Figure 3.2.2 and Table 3.2.4 show the available van data, and no conclusions can be drawn except that the van did find the plume. Further complicating the van data is the fact that such high concentrations were found that the monitor saturated, distorting the average data. Figure 3.6 shows the locations of the van and Seward station.

Figure 3.2.4 gives the March concentrations at Monitor 8, the Seward monitor, for both 3-hr and 24-hr averages. All other monitor concentrations were below Monitor 8. Even these plots do not clarify the effect of control actions on air quality. One violations is recorded on March 14 (24 hr), so in one sense the SCS failed; given that its control directions were not followed, this is not surprising. Trying to compare the control recommendations at the bottom of Figure 3.2.4 is fruitless; the recommendations consider 396 receptors while Monitor 8 is one sample of these. There is no reason to expect Monitor 8 to correlate with control recommendations; it should be expected that Monitor 8 would not have violations if all control recommendations were followed.

3.3 Conclusions

The demonstration period would support the following statements:

1. Sophisticated dispersion and control models can be incorporated into the time scales of real-time SCS operation.

2. Load shifting is a reliable means for effecting control of emissions in a model-based SCS.

3. No conclusions can be drawn about the effectiveness of the entire model-based SCS as a means of protecting ambient air quality in a reliable manner.

Table 3.2.1

CONTROL ACTION SUMMARY

DATE	MODEL UNIT	MODEL AMOUNT	MODEL REQUEST TIME	ACTUAL UNIT	REQUEST AMOUNT	REQUEST TIME	ESTIMATED COST	PLANT UNIT	ACTION AMOUNT	ACTION TIME	ENERGY LOSS	COST MISC.	COST TOTAL
March 13	SEW 5	42%	0000-0600	SEW 5	10%	Any 1 hr between 0000 - 0600	$160	SEW 5	9%	1411-1520	182.40 (16MWH)	429.08	611.48
	SEW 5	51%	1200-1500										
March 17	SEW 5	46%	0900-1200	SEW 5	15%	0900-1000	$480	SEW 5	13%	0841-1010	484.50 (19MWH)	6.27	490.77
March 20	SEW 5	54%	1200-1500	SEW 5	25%	1300-1400	$570	SEW 5	22%	1215-1420	437.10 (58MWH)	399.37	836.47
										TOTALS	1104.00	833.72	$1938.72

Table 3.2.2

CONTROL REPORT SUMMARY

			FORECAST RECEPTOR VIOLATION						CONTROL STRATEGY			CONTROL IMPACT		
TE	TIME	CONTROL[1] STATUS	BEGIN[2] TIME	STANDARD VIOLATED	CONC. - BG (ppm)	PROB. VIOL.	LOCATION EAST	LOCATION WEST	CONTROL[3] UNIT	DEGREE CONTROL	START[2] TIME	EXPECTED CONC.	PROB. VIOL.	EXPECTED COST
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1631	BG	-	-	-	-	-	-	-	-	-	-	-	-
11	0000	M	-	-	-	-	-	-	-	-	-	-	-	-
	1700	NC	-	-	-	-	-	-	-	-	-	-	-	-
12	0112	C	1500[4]	3 HR.	0.504 - 0.0	0.5034	666.85	4475.55	L-SEW 5	1%	1500	0.500	0.5000	
			1800[4]	3 HR.	0.504 - 0.0	0.5034	666.85	4475.55	L-SEW 5	1%	1800	0.500	0.5000	$ 30
	1440	C-BG	2400[4]	24 HR.	0.186 - 0.150	0.6100	666.25	4475.55	L-SEW 34	100%	2100	0.150	0.5300	
			3600[4]	3 HR.	0.508 - 0.0	0.5060	667.80	4474.20	L-SEW 5	100%	2100	0.500	0.5000	$2680
									L-SEW 5	2%	3600			
13	0006	C	0000	24 HR.	0.167 - 0.166	0.5699	666.15	4475.55	L-SEW 5	42%	0000	0.167	0.5699	
			0000	3 HR.	0.820 - 0.0	0.6906	667.23	4475.47	L-SEW 5	42%	0300	0.500	0.5000	
CONTROL ACTION TAKEN)			0300	3 HR.	0.820 - 0.0	0.6908	666.85	4475.55	L-SEW 5	51%	1200	0.500	0.5000	
			1200	3 HR.	0.505 - 0.0	0.5037	668.73	4473.35				0.283	0.2829	
			1200	3 HR.	0.889 - 0.0	0.7188	667.77	4474.17				0.500	0.5000	$2020
	1547	C-BG-E	1200	3 HR.	0.819 - 0.0	0.6904	667.77	4474.93	E	-	-	E	-	-
			1500	3 HR.	0.625 - 0.004	0.5886	668.58	4474.61						
			1500	3 HR.	0.823 - 0.004	0.6923	667.85	4474.55						
			0000	24 HR.	0.212 - 0.205	0.6623	667.82	4475.42						
			0000	24 HR.	0.215 - 0.207	0.6665	667.65	4475.47						
			0000	24 HR.	0.111 - 0.104	0.4080	667.77	4474.93						
			0000	24 HR.	0.215 - 0.208	0.6678	667.56	4475.26						
			2600[4]	3 HR.	0.934 - 0.0	0.7353	666.53	4464.35						
			2600[4]	3 HR.	0.543 - 0.0	0.5329	661.32	4457.74						
			2600[4]	3 HR.	0.526 - 0.0	0.5203	652.85	4471.65						
			2900[4]	3 HR.	1.357 - 0.0	0.8426	671.15	4470.65						
			2900[4]	3 HR.	0.664 - 0.0	0.6124	674.53	4474.55						
			2900[4]	3 HR.	1.358 - 0.0	0.8427	671.10	4470.07						
			2900[4]	3 HR.	1.357 - 0.0	0.8427	679.32	4469.66						
			2900[4]	3 HR.	1.099 - 0.0	0.7860	667.34	4468.03						
			2900[4]	3 HR.	1.034 - 0.0	0.7670	667.13	4465.35						

(continued)

Table 3.2.2 (cont.)

DATE	TIME	CONTROL[1] STATUS	FORECAST RECEPTOR VIOLATION BEGIN[2] TIME	STANDARD VIOLATED	CONC. ppm	BG	PROB. VIOL.	LOCATION EAST	LOCATION WEST	CONTROL-[3] UNIT	DEGREE CONTROL	START[2] TIME	EXPECTED CONC.	PROB. VIOL.	EXPECTED COST
3/14	0310	C	0600	3 HR.	0.693 -	0.0	0.6088	677.73	4478.40	F KEY-2	69%	0600	0.500	0.5	
										F HC-2	100%	0600			$2580
	1400	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/15	0848	C-E	0600	3 HR.	0.691 -	0.0	0.6273	671.15	4470.65	E	-	-	-	-	-
			0600	3 HR.	0.688 -	0.0	0.6250	671.10	4479.07						
			0600	3 HR.	0.690 -	0.0	0.6271	670.32	4469.66						
			0600	3 HR.	0.530 -	0.0	0.5234	669.69	4466.91						
			0000	24 HR.	0.059 -	0.0	0.1010	671.15	4470.65						
			0000	24 HR.	0.153 -	0.0	0.5366	669.95	4470.94						
			0000	24 HR.	0.058 -	0.0	0.1860	670.32	4469.66						
	1300	C-E	2700[4]	3 HR.	0.849 -	0.0	0.7020	670.95	4478.90	L-SEW 34	100%	2700	0.0	0.0	
			2700[4]	3 HR.	0.706 -	0.0	0.6357	668.82	4476.89	L-SEW 5	100%	2700	0.0	0.0	
			2700[4]	3 HR.	0.815 -	0.0	0.6887	671.19	4470.01	L-SEW 5	39%	4200	0.0	0.0	
			2700[4]	3 HR.	0.522 -	0.0	0.5172	673.26	4481.68				0.0	0.0	
			4200[4]	3 HR.	0.603 -	0.007	0.5746	668.58	4474.51				0.388	0.3993	
			4200[4]	3 HR.	0.783 -	0.007	0.6742	667.85	4474.55				0.500	0.5000	$-150
3/16	0149	C	0600	3 HR.	0.750 -	0.0	0.6585	665.93	4474.17	L-SEW 34	100%	0600	0.0	0.0	
			0900	3 HR.	0.637 -	0.0	0.5965	666.24	4474.06	L-SEW 5	100%	0600	0.414	0.4245	
			0900	3 HR.	0.776 -	0.0	0.6711	666.18	4473.62	L-SEW 5	38%	0900	0.500	0.5000	
			0900	3 HR.	0.750	0.0	0.6584	666.14	4473.84	L-SEW 34	100%	3000	0.481	0.4843	
			3000[4]	3 HR.	0.804	0.0	0.6836	667.77	4474.17	L-SEW 5	100%	3000	0.0	0.4843	$-380
	1516	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/17	0238	C	0900	3 HR.	0.509 -	0.001	0.5071	668.73	4473.35	L-SEW 5	46%	0900	0.304	0.3081	
			0900	3 HR.	0.849 -	0.001	0.7029	667.77	4474.17				0.500	0.5000	
(CONTROL ACTION TAKEN)			0000	24 HR.	0.089 -	0.063	0.3248	668.73	4473.35				0.089	0.3248	
			0000	24 HR.	0.119 -	0.106	0.4349	667.77	4474.17				0.119	0.4349	$2280
	1327	C	0000	24 HR.	0.217 -	0.207	0.6706	668.58	4474.61	F KEY-1	100%	2100	0.212	0.6613	
			0000	24 HR.	0.299 -	0.289	0.7778	667.85	4474.55	F KEY-2	100%	2100	0.294	0.7724	
			2400[4]	3 HR.	0.504 -	0.001	0.5031	677.73	4478.40	F HC-1	100%	2100	0.500	0.5000	
			2700[4]	3 HR.	0.746 -	0.001	0.6563	671.53	4472.25	F HC-2	100%	2100	0.443	0.4512	
			2700[4]	3 HR.	0.843 -	0.001	0.7003	673.22	4471.09	F HC-2	8%	2400	0.500	0.5000	
			2700[4]	3 HR.	0.535 -	0.089	0.5272	678.24	4468.48	L-SEW 5	61%	2700	0.354	0.3636	
			2700[4]	3 HR.	0.523 -	0.111	0.5178	684.01	4465.70	L-SEW 5	61%	3000	0.355	0.3651	
			3000[4]	3 HR.	0.746 -	0.001	0.6563	671.53	4472.25				0.443	0.4512	
			3000[4]	3 HR.	0.843 -	0.001	0.7003	673.22	4471.09				0.500	0.5000	
			3000[4]	3 HR.	0.535 -	0.089	0.5272	678.24	4468.48				0.354	0.3636	
			3000[4]	3 HR.	0.523 -	0.111	0.5178	684.01	4465.70				0.355	0.3651	$2280

(continued)

Table 3.2.2 (cont.)

DATE	TIME	CONTROL[1] STATUS	BEGIN[2] TIME	STANDARD VIOLATED	FORECAST RECEPTION VIOLATION CONC. (ppm)	BG (ppm)	PROB. VIOL.	LOCATION EAST	LOCATION WEST	CONTROL-[3] UNIT	DEGREE CONTROL	START[2] TIME	EXPECTED CONC.	PROB. VIOL.	EXPECTED COST
3/18	1454	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	0215	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/19	0048	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	1652	BG	-	-	-	-	-	-	-	-	-	-	-	-	-
3/20	0006	C	0000	24 HR.	0.186 -	0.001	0.6129	667.23	4475.47	L-SEW 5	85%	1200[4]	0.140	0.5000	$3810
	1133	C	0000	24 HR.	0.158 -	0.102	0.5492	667.82	4475.42	L-SEW 5	54%	1200	0.132	0.4778	
			0000	24 HR.	0.154 -	0.098	0.5376	667.65	4475.47				0.127	0.4625	
(CONTROL ACTION TAKEN)			0000	24 HR.	0.169 -	0.108	0.5761	667.56	4475.26				0.140	0.5000	$1680
3/21	2258*	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	1213	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/22	2252[5]	BG	-	-	-	-	-	-	-	-	-	-	-	-	-
	1426	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/23	2314[5]	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	1334	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/24	2244[5]	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	1432	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/25	0003	C	0000	3 HR.	0.548 -	0.001	0.5371	667.13	4474.83	L-SEW 34	100%	0000	0.001	0.0	
			0300	3 HR.	0.803 -	0.001	0.6833	667.82	4475.42	L-SEW 5	100%	0000	0.290	0.2913	
			0300	3 HR.	0.850 -	0.001	0.7055	667.65	4475.47	L-SEW 5	73%	0300	0.207	0.2994	
			0300	3 HR.	1.384 -	0.001	0.8473	667.13	4474.83	L-SEW 5	10%	0600	0.500	0.5000	
			0600	3 HR.	0.548 -	0.001	0.5371	667.00	4474.92	L-SEW 5	10%	0900	0.500	0.5000	
			0900	3 HR.	0.548 -	0.001	0.5371	667.13	4474.83				0.500	0.5000	
			0000	24 HR.	0.221 -	0.003	0.6777	667.82	4475.42				0.093	0.3397	
			0000	24 HR.	0.227 -	0.003	0.6875	667.65	4475.47				0.094	0.3429	
			0000	24 HR.	0.313 -	0.003	0.7906	667.13	4474.83				0.120	0.4627	$-4200
	1410	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
3/26	2249	NC	-	-	-	-	-	-	-	-	-	-	-	-	-
	1435	NC	-	-	-	-	-	-	-	-	-	-	-	-	-

(continued)

Table 3.2.2 - Footnotes

1 BG: Background-induced violations; M: Forecast missing; NC: no control needed; C: control recommended; E: error occured.

2 3-hr time blocks starting at given time.

3 L: Load shift; F: Fuel switch; E: Error occurred; EW: Seward station; HC: Homer City station; KEY: Keystone station.

4 Violations or controls predicted during next forecast period and can be checked with updated data; no action taken except advanced warning to dispatchers of possible need for control.

5 Forecasts actuallymade before 0000 of given day.

Table 3.2.3

"GRAB SAMPLE" COAL ANALYSIS

SEWARD STATION

MARCH 13, 1976

UNIT-MILL	%S
SEW 34	2.14
SEW 34	1.82
SEW 5 - A	1.80
SEW 5 - A	1.51
SEW 5 - C	1.50

MARCH 17, 1976

UNIT MILL	%S
SEW 34 - A	2.04
SEW 34 - B	1.94
SEW 34 - A	2.04
SEW 34 - B	1.79
SEW 5 - A	1.91
0900	
SEW 5 - A	2.00
1000	
SEW 5 - B	2.07
0900	
SEW 5 - B	2.00
1000	
SEW 5 - C	2.07
0900	
SEW 5 - C	2.01
1000	
SEW 5 - D	2.16
0900	
SEW 5 - D	2.07
1000	

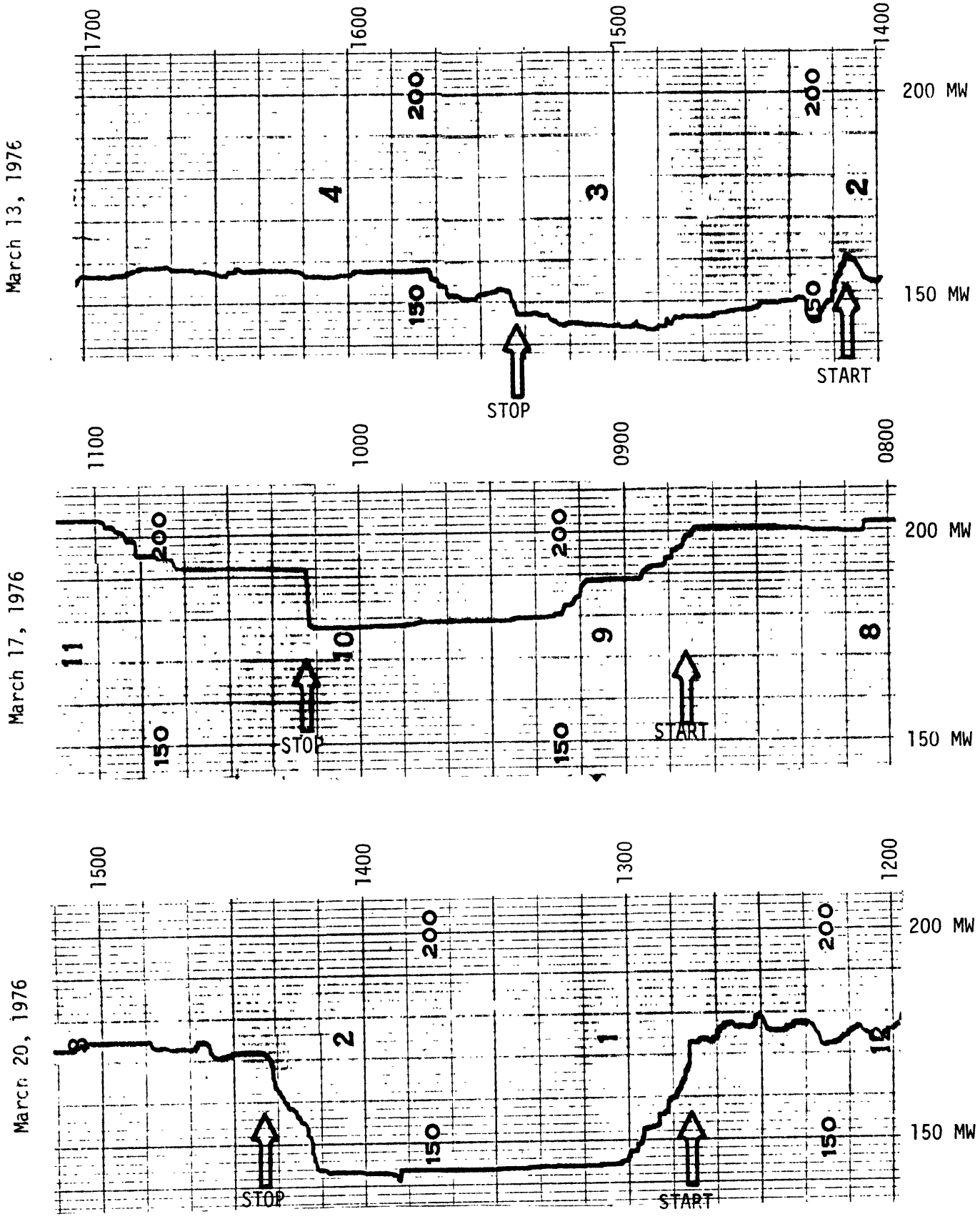
MARCH 20, 1976

UNIT MILL	%S
SEW 34 - B	2.23
SEW 34 - A	2.19
SEW 34 - B	1.99
SEW 5 -A	2.11
SEW 5 - B	2.32
SEW 5 - C	2.04
SEW 5 - D	1.88
During Control:	
SEW 5 - A	2.04
SEW 5 - B	2.10
SEW 5 - C	2.13
SEW 5 - D	2.04

Table 3.2.4

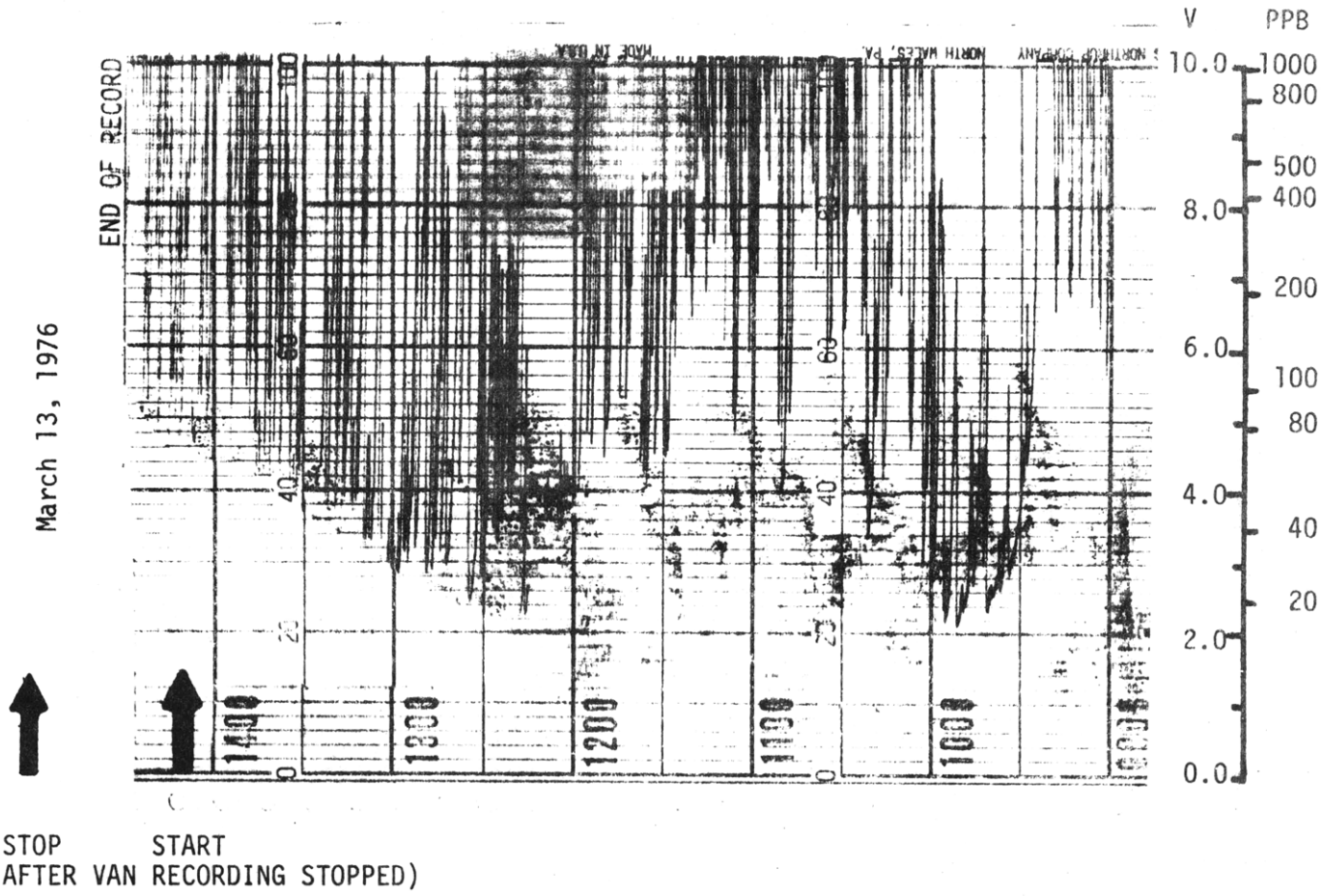
MOBILE VAN SAMPLING DATA

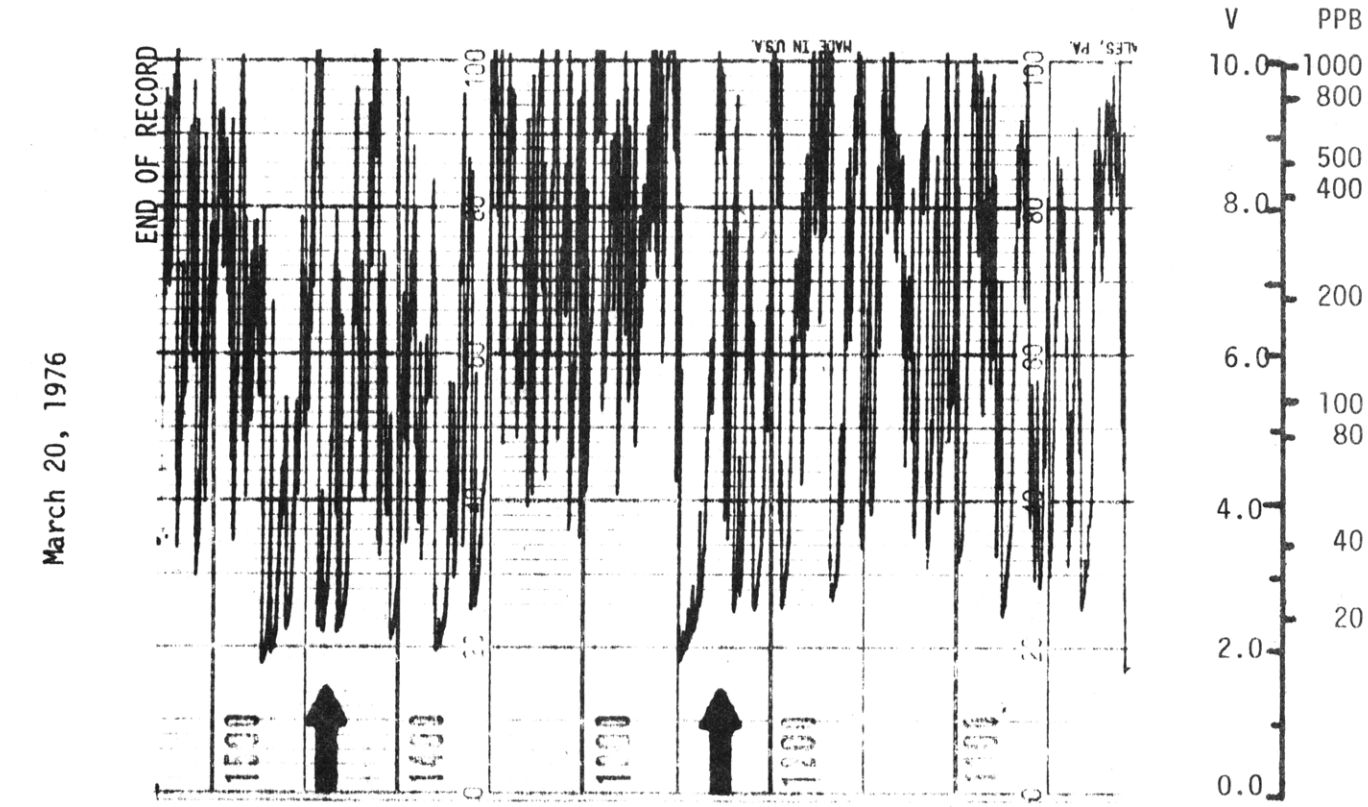
DATE	START TIME	SO_x CONC-PPB AVG.	SO_x CONC-PPB PEAK	WIND SPEED-MPH AVG.	WIND SPEED-MPH PEAK	WIND DIR. -° AVG.	WIND DIR. -° PEAK
March 13	0900	>1000	-	12.0	16.0	255	260
	0930	49	394	14.0	17.0	240	250
	1000	617	1000	14.0	17.0	255	260
	1030	791	>1000	14.0	17.0	270	260
	1100	791	>1000	12.5	17.5	250	260
	1130	481	>1000	12.5	17.0	250	260
	1200	138	1000	12.5	17.0	250	260
	1230	138	1000	12.5	17.0	250	260
	1300	228	>1000	12.5	17.0	250	260
	1330	228	>1000	14.0	17.0	245	260
Start Control	1400	617	>1000	12.0	17.0	255	260
	1430	485	>1000	14.0	18.0	265	260
End Control							
March 20	1040	153	1014	10.0	21.0	216	315
	1110	178	1014	11.0	21.5	198	261
	1140	206	1014	13.3	21.5	202	286
	1210	161	1014	12.5	23.0	202	280
Start Control	1240	138	1014	12.8	23.0	202	360
	1310	198	1014	10.5	22.0	202	252
	1340	138	1014	12.3	22.0	202	261
End Control	1410	138	1014	12.8	21.3	202	243
	1440	135	1014	12.5	21.0	201	250
	1510	178	1014	12.0	21.0	198	216



SEWARD OUTPUT DURING CONTROL ACTIONS

Figure 3.2.1





MOBILE VAN SO_2 READINGS

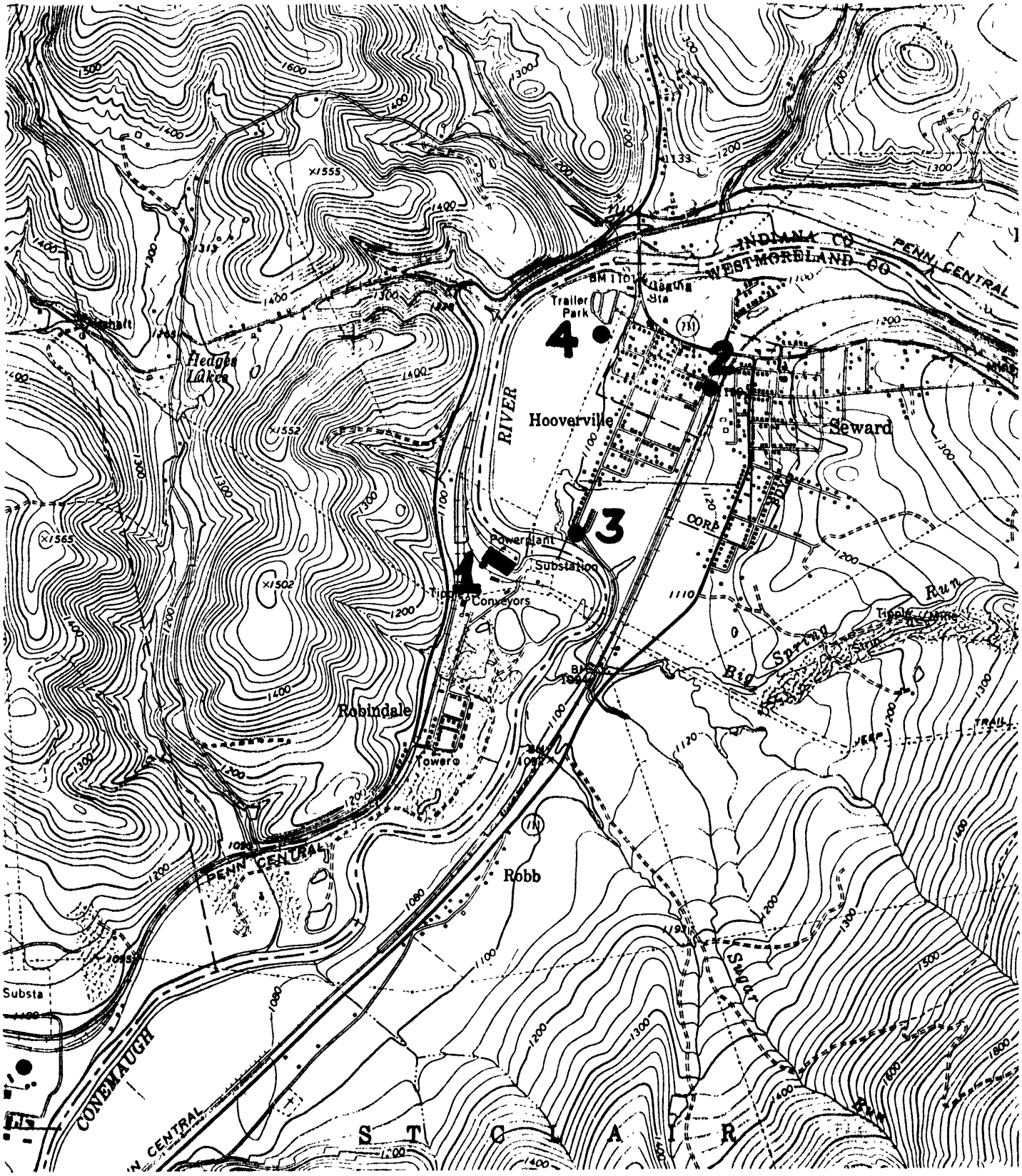
Figure 3.2.2

1 - SEWARD STATION

2 - SEWARD MONITOR - N8H

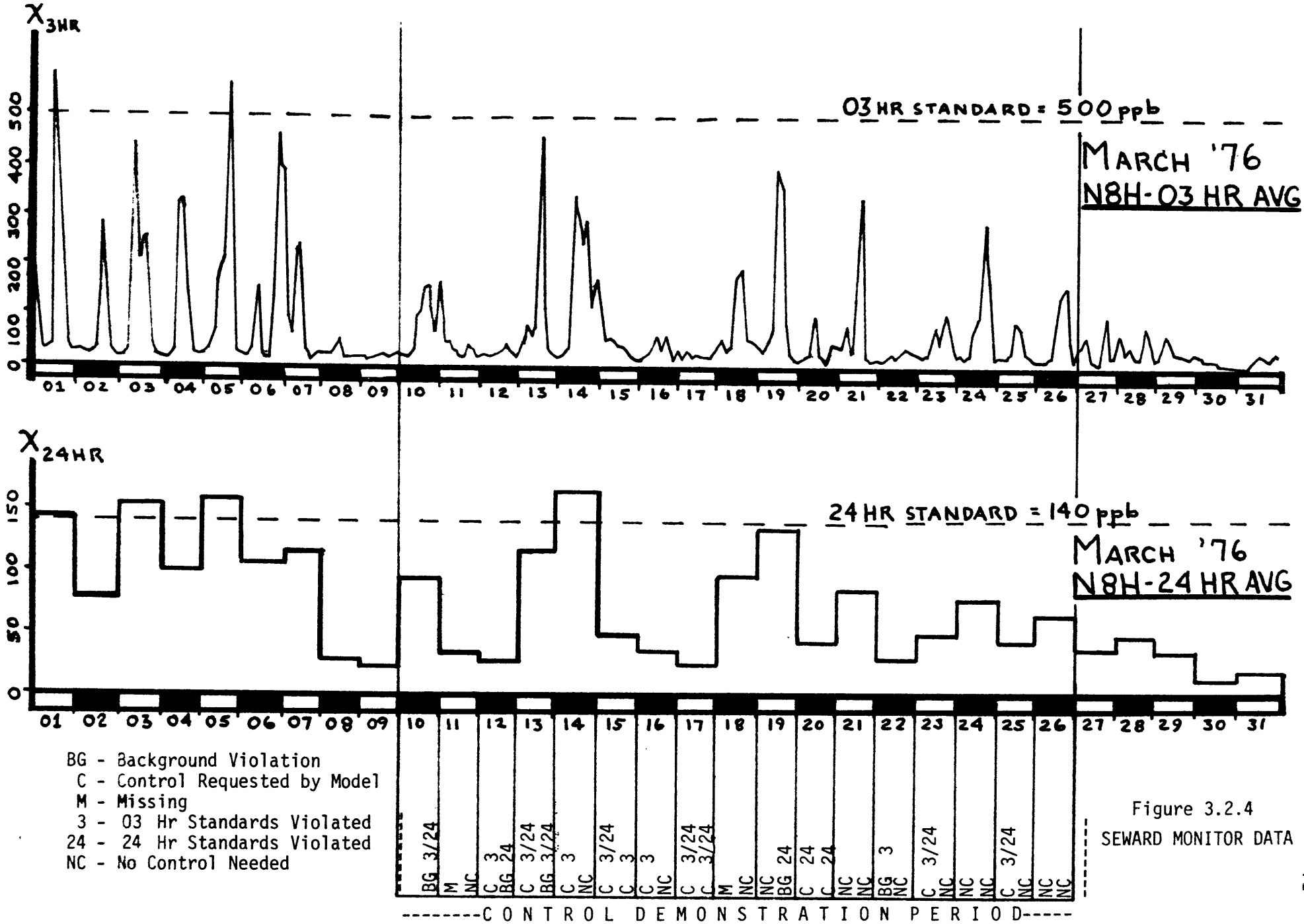
3 - VAN LOCATION March 13, 1976

4 - VAN LOCATION March 20, 1976



CONTROL ACTION MONITORING

Figure 3.2.3



BG - Background Violation
C - Control Requested by Model
M - Missing
3 - 03 Hr Standards Violated
24 - 24 Hr Standards Violated
NC - No Control Needed

Figure 3.2.4
SEWARD MONITOR DATA

4.0 BACKGROUND EFFECTS

Chestnut Ridge was chosen because it appeared to have "isolated" sources. However, analysis of the monitor data showed that this assumption was wrong and that background could contribute to, and at times, cause concentration levels that threatened standards. As would be expected, longer averaging time standards were most threatened. This section describes our analyses of the Chestnut Ridge data. It also includes a summary of a new background-modeling method that was developed for Chestnut Ridge.

4.1 Data

When designing the Chestnut Ridge SCS for the four power plants, three factors combined to make background modeling necessary. First, because of the tall stacks at the three larger plants and the low density of monitors, the observed power plant-caused concentrations were usually on the order of the background concentrations. Thus, attempts to understand the mechanisms of plume interaction with complex terrain required some means of accounting for the background effects. Second, several monitors exhibited high concentrations which correlated with wind directions from Pittsburgh and Johnstown raising the suspicion that background could contribute to or cause potential violations. Finally, the geometry of monitors, plants, terrain, and prevailing winds produced several "ambiguous" wind directions, in which both a power plant and a background source were upwind of a monitor recording high SO_2 values. Without a background model in such cases to apportion some of the concentration to distant sources, control action could not be determined. Our analyses only indicated the need for a background model; the design of a plausible background model is a separate issue. Section 2.1 describes the method actually used in the operating air quality model.

To get some feeling for the range of data values observed at Chestnut Ridge, the monthly maximum 24-hour, three-hour, and one-hour averages for each monitor are presented in Tables 4.1.1 - 4.1.3. A monthly N-hour average is the single highest N-hour (running clock hour) average observed at a monitor in the month. Monitor and meteorological data capture rates are shown in Table 4.1.4. Note that many monitors never exceed 50 percent of the 24-hour and three-hour standards even in their monthly peak averages. With the exception of the Seward monitor (8), which was influenced by downwash, only two monitors ever exceeded the 24-hour standard and only one exceeded the three-hour standard. Over the entire monitoring system, less than five percent of the one-hour averages ever exceeded 200 ppb.

The prevailing winds carry Pittsburgh's emissions directly over the Chestnut Ridge monitoring system and Pittsburgh has been observed to have a strong effect on ambient concentrations. The Pittsburgh urban area is approximately 50-70 km from the monitors and presents an unusual modeling problem both because of its poorly defined emissions and because of the

Table 4.1.1

CHESTNUT RIDGE SCS

SO_2 24-HR RUNNING AVERAGES (MAX) - PPB

MON. #	1974 SEP	OCT	NOV	DEC	1975 JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	1976 JAN	FEB	MAR
1	29	95	24	81	52	61	49	72	45	63	35	37	18	40	83	33	63	44	31
2	58	67	55	117	123	100	58	57	65	41	71	125	(51)	(31)	146 (1)	138	96	66	72
3	58	94	71	99	84	88	52	49	63	55	50	72	48	108	100	55	73	64	63
6	48	117	71	89	84	94	54	72	54	48	60	65	47	67	61	52	55	67	69
7	75	136	101	99	112	114	79	77	99	58	107	139	47	55	219 (23)	81	105	120	59
8	-	-	-	-	279 (126)	311 (110)	142 (2)	157 (4)	80	152 (6)	82	141 (3)	92	236 (32)	144 (1)	(-)	(-)	167	183
A	44	70	46	59	67	70	58	45	54	55	48	54	(47)	62	81	56	75	50	48
B	30	64	56	68	77	86	64	42	53	66	47	73	(50)	48	59	69	89	66	52
C	31	60	53	80	63	69	37	21	44	26	(20)	(10)	(16)	23	55	56	59	37	40
D	62	72	50	57	73	73	46	34	50	45	66	76	53	72	45	43	69	74	52
E	92	114	76	76	103	89	51	(44)	(55)	(57)	40	65	(41)	49	82	64	40	56	56
F	62	109	60	80	87	115	80	49	58	70	55	74	53	64	83	63	73	64	58
G	(25)	(51)	48	115	84	57	52	46	(41)	43	61	74	(42)	(20)	65	58	58	61	64
L	39	78	64	53	85	57	62	55	56	44	50	64	35	77	80	66	60	54	54
M	63	133	89	82	90	82	59	61	(56)	50	48	49	42	68	69	63	67	45	44
N	36	72	40	48	61	49	44	23	31	28	16	(13)	19	61	80	66	127	84	77
O	47	49	33	46	46	37	24	32	25	29	20	33	15	87	60	39	58	15	28
MAX	E	7	7	G	8	8	8	8	7	8	7	8	8	8	7	2	N	8	8
MAX -8	E	7	7	G	2	F	F	7	7	F	7	7	D/F	3	7	2	N	7	N

(·) Number of Ave. > .140 ppm

(·) Low Data Capture

Table 4.1.2

CHESTNUT RIDGE SCS

SO_2 3-HR RUNNING AVERAGES (MAX) - PPB

	1974				1975												1976		
MON. #	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR
1	114	352	42	105	102	173	123	253	118	230	141	84	52	106	201	82	175	118	60
2	165	150	228	152	259	209	128	112	178	99	156	258	(98)	(128)	349	177	186	152	130
3	106	197	111	195	440	336	164	92	135	170	103	145	130	260	290	98	138	119	122
6	249	394	241	124	171	141	125	196	203	151	183	171	152	121	164	149	198	256	171
7	258	300	450	200	270	245	219	172	252	154	270	197	140	183	552 (1)	161	212	308	161
8	-	-	-	-	766 (10)	655 (13)	488	294	480	394	278	488	248	585 (3)	(306)	(-)	(-)	609	609
A	127	122	101	93	79	146	90	94	115	197	117	141	(125)	149	197	75	208	86	158
B	78	79	85	108	92	128	108	79	107	192	98	153	(90)	77	103	86	125	151	113
C	69	157	92	167	80	102	55	42	151	139	(59)	(18)	(38)	52	87	83	102	71	86
D	150	168	109	105	92	114	71	50	126	145	142	138	160	98	67	53	151	115	131
E	209	243	270	152	145	146	111	(66)	(106)	(135)	131	138	85	83	141	92	62	117	98
F	140	240	137	126	220	195	146	89	114	162	135	138	126	153	119	111	206	132	122
G	(60)	(70)	105	192	106	106	112	148	(71)	105	137	153	(89)	(52)	118	120	104	143	91
L	85	111	151	89	72	86	99	103	126	81	135	113	76	143	192	112	97	82	110
M	100	179	130	128	114	121	148	107	(105)	81	87	91	87	100	101	87	146	76	68
N	86	141	74	76	118	104	94	50	83	69	64	(58)	71	137	150	167	178	189	136
O	91	72	45	82	52	64	42	43	59	62	57	52	36	160	94	50	73	35	67
MAX	7	1	7	7	8	8	8	8	8	8	8	8	8	8	7	2	7	8	8
MAX -8		1	7	7	3	3	7	1	7	1	7	2	0	3	7	2	7	7	6

(·) Number of Ave. > .500 ppm.

(·) Low Data Capture.

Table 4.1.3

CHESTNUT RIDGE SCS

SO_2 MONTHLY 1-HR AVERAGES (MAX) - PPB

	1974				1975												1976		
MON.#	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR
1	116	453	62	112	133	225	136	371	205	300	209	127	72	107	217	94	233	123	112
2	206	193	272	164	312	240	158	122	199	193	179	394	(117)	(144)	444	212	217	205	139
3	154	480	228	293	500	563	190	123	183	242	164	198	166	333	308	145	161	147	143
6	538	449	291	176	400	217	179	256	337	177	299	255	208	256	681	193	312	459	240
8	-	-	-	-	917	785	514	528	858	530	421	561	341	916	(398)	(-)	(-)	(724)	857
A	140	142	125	105	115	197	126	133	207	257	137	203	(216)	296	241	82	234	131	180
B	87	113	112	123	102	179	158	100	146	228	137	165	(99)	84	105	101	131	163	133
C	101	193	103	177	84	118	76	54	189	197	(82)	(21)	(55)	85	108	108	121	74	113
D	193	205	176	125	108	121	107	65	204	199	163	158	186	130	76	74	186	142	133
E	287	279	555	203	203	152	125	(91)	(131)	(151)	151	230	(104)	138	162	144	82	148	133
F	165	286	186	138	272	223	169	126	176	202	250	171	204	252	136	120	284	174	147
G	(68)	(90)	165	453	178	142	170	323	(125)	163	235	244	(121)	(54)	202	140	127	185	107
L	102	126	179	102	86	121	163	110	163	107	186	144	108	163	240	140	105	146	119
M	115	193	142	143	129	135	163	113	(191)	109	107	104	94	153	110	94	201	87	78
N	130	186	88	83	136	193	114	80	105	89	118	(83)	115	237	201	315	251	270	256
O	109	98	55	102	54	74	55	60	68	104	71	55	47	201	149	55	85	43	76

(·) Low Data Capture.

Table 4.1.4

CHESTNUT RIDGE SCS

DATA CAPTURE

	1974				1975												1976		
VAR.	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR
N1H	99.31	93.28	99.36	97.85	87.52	99.85	82.86	56.67	91.80	91.67	99.87	83.74	100	88.98	89.59	98.92	98.52	95.11	97.31
N2H	95.83	84.95	100	100	81.05	92.71	94.76	55.14	97185	97.78	99.33	93.68	65.14	60.08	78.33	98.25	98.12	90.95	74.19
N3H	80.42	96.10	99.72	98.52	99.87	99.55	100	55.83	93.28	99.86	86.56	92.47	68.75	88.31	99.17	99.19	77.02	74.43	100
N6H	89.58	98.79	88.47	87.23	95.83	67.41	98.92	36.11	99.33	80.97	100	88.98	95.14	99.06	97.50	100	93.28	73.28	49.33
N7H	100	89.38	98.47	90.86	94.99	96.28	87.11	58.19	91.26	75.69	97.18	68.55	70.42	97.18	84.17	95.96	94.89	95.40	88.58
N8H	-	-	-	-	94.22	93.01	99.60	58.89	91.26	94.86	100	80.24	88.89	93.82	24.31	-	-	15.95	100
NAH	98.75	99.87	99.58	86.16	98.12	97.62	92.88	56.53	98.39	91.25	100	79.84	62.08	85.75	81.94	81.72	93.95	95.83	95.43
NBH	87.36	83.20	84.58	81.45	90.05	95.98	85.22	51.25	87.10	77.08	100	70.30	58.61	81.32	88.75	96.91	96.24	93.82	86.29
NCH	85.77	100	99.58	90.86	96.64	100	99.19	46.53	90.19	94.17	33.74	35.22	10.97	94.35	82.22	91.26	95.16	86.64	98.79
NDH	98.89	100	99.58	77.28	89.25	96.58	96.37	49.86	75.67	74.44	100	69.89	68.47	67.07	96.67	91.26	98.92	99.14	96.77
NEH	98.75	100	97.50	88.84	93.68	99.70	86.83	35.69	58.06	63.47	48.92	88.84	91.25	25.81	72.36	94.22	98.66	100	98.92
NFH	47.36	99.73	97.64	83.06	92.74	96.88	93.41	53.75	80.38	90.00	97.58	67.74	83.89	99.06	78.47	93.28	88.31	78.88	94.89
NGH	44.86	21.24	99.31	93.75	95.03	97.32	97.45	55.28	98.12	52.22	99.87	84.14	81.81	09.41	40.69	94.22	81.72	82.61	77.28
NLH	90.69	83.60	99.44	94.49	98.66	100	93.28	38.33	93.28	86.53	86.69	96.37	97.64	100	95.83	94.62	79.57	94.40	95.97
NMH	88.19	100	97.50	86.42	79.30	98.21	100	51.67	87.50	34.44	93.95	85.89	81.94	83.20	93.47	95.16	85.22	99.86	96.51
NNH	95.97	100	99.86	92.47	99.73	100	97.04	52.92	78.23	94.17	83.71	74.87	38.19	93.95	100	100	91.67	98.13	99.60
NOH	99.58	96.24	99.72	99.73	99.33	94.35	92.34	45.69	83.26	96.11	100	90.99	81.67	81.89	100	100	73.79	5.03	93.95
TOTAL	90.72	90.39	97.58	90.57	93.38	95.61	94.08	50.49	87.88	82.04	90.14	79.51	73.23	77.37	82.56	95.29**	90.32†	81.14	93.75
NG1	100	100	96.25	78.90	97.98	77.23	97.04	100	97.72	57.92	88.04	89.38	92.08	100	52.78	87.77	68.95	83.62	97.58
NT1	100	100	100	78.76	22.04	59.82	54.70	11.25	87.37	15.83	76.34	84.41	92.08	100	99.86	85.08	77.28	88.36	90.99
NG2	100	99.87	99.31	88.58	100	80.80	97.45	99.44	97.72	80.83	97.45	92.07	91.94	90.32	92.36	03.60	68.28	82.33	65.32
NT2	100	100	100	96.10	100	100	97.58	99.44	98.79	80.69	97.31	92.07	92.08	100	99.44	91.26	84.68	99.43	100
NG3	100	99.87	100	97.45	100	100	84.54	100	100	99.86	90.46	97.18	99.58	100	100	78.92	100	100	100
NS8	97.64	100	100	97.45	100	100	85.08	-	85.35	38.89	56.59	99.33	100	100	99.58	90.99	98.92	99.57	100
NT8	97.64	100	100	97.45	100	100	84.54	100	100	72.50	-	-	-	84.95	86.94	96.51	99.60	100	100
NG9	81.53	100	100	97.45	100	100	84.95	99.31	100	100	99.87	99.06	96.67	100	100	99.19	100	100	12.90
NGS*	99.31	100	100	97.45	100	100	100	100	98.39	76.94	87.77	96.64	100	100	100	100	100	99.43	100
NGU	100	100	100	97.45	100	100	85.08	-	84.54	99.72	100	99.60	99.72	100	100	100	97.46	100	100
TOTAL	97.61	99.97	99.54	92.70	92.00	91.79	87.10	79.91	94.99	72.62	79.38	84.97	86.42	97.53	93.10	93.33	89.72	95.27	86.68

*Not in Masterfile.

**89.68 for 17 monitors

†85.00 for 17 monitors

basic uncertainties about the most effective modeling techniques for urban areas over long distances. A number of different methods were tried in order to gain understanding of the background effects of Pittsburgh and other distant sources. Four of the most useful were: (1) analysis of the variation of the mean concentration with meteorological variables, particularly wind direction; (2) analysis of the probability of occurrence of "peak" concentrations; (3) the so-called EPA-Larsen method, and (4) parameter identification of a statistical model representing time and direction dependence of concentrations. All four techniques improve as the data base grows; it is recommended that the mean concentration, peak concentration, and EPA analyses not be performed with less than six months', and preferably, a year's data. The techniques only required monitored data; while emissions data can simplify the validation of these analyses, it is not necessary for model development and validation. These methods were used to circumvent the poor source inventory data problem mentioned in Section 2.1.

4.2 Mean Concentration Analysis

Mean concentration analysis is principally a background modeling technique. The majority of the analyses at Chestnut Ridge were done in what we call the Northern Region (monitors B, C, L, M, N, and O) because of its relatively flat terrain. The analysis can be performed in more complex terrain, such as the Seward region, but the results are much more difficult to interpret. The implicit assumption was made that background behavior in the Northern Region was similar to that elsewhere at Chestnut Ridge.

Briefly, mean concentration analysis entails forming mean concentrations over subsets of the monitor data. The subsets are defined from data having common characteristics, such as time, meteorology, or plant status (referring now to the Chestnut Ridge plants). The following paragraphs describe the most significant results produced by this analysis in the Northern Region. More detail can be found in Appendix A.

All monitors in the Northern Region have the same general mean dependence on wind direction, as measured at the tower located on Chestnut Ridge (monitor G location). Mean concentrations exhibit double peaks at directons which correspond to the directions from the monitor to Pittsburgh and Johnstown. There is a factor of four difference in the highest mean concentration (Pittsburgh direction) and the lowest mean concentration (wind from North). Related to this effect is the fact that the highest daily averages occur when the wind is persistent in the Pittsburgh sector. A seasonal dependence is also observed: during the winter months, all mean concentration values increase. The ratio of the Pittsburgh peak to the north sector minimum decreases in the winter. This effect can be explained by the increased space heating background levels being superimposed on the continuous urban contribution.

Traditionally, concentrations produced by area sources are represented as being proportional to the inverse of wind speed. Efforts to demonstrate this dependence in the northern region produced either a null or negative result. We have not succeeded in identifying the cause of this counter-

intuitive observation. One possible explanation is that the wind speed measured at the tower is a poor indicator of northern region winds and the wind field for the transit from background sources to the monitors.

4.3 Peak Concentration Analysis

Peak concentration analysis examines the highest concentrations rather than the mean values of the concentrations. In the northern region, "peak" concentrations were those in or above the 95th percentile of concentrations at a monitor, during a period of one year. Taking these peaks as our sample, they were categorized as a function of the wind direction and used to estimate the probability of occurrence of peak concentrations. A high probability is interpreted as indicating the direction of a source which has a pronounced effect on peak production at the monitor.

Since SCS operates on peak concentrations, this analysis served two purposes. First, it helped identify the sources, background or otherwise, which contributed to high concentrations. Second, it served to form special data bases which could be used for SCS evaluations. For the present, the discussion will emphasize the background identification problem. The following paragraphs provide the highlights of this technique's application at Chestnut Ridge. Appendix B contains further details.

For a number of monitors, the wind direction with the highest probability of a peak hourly concentration is the wind direction which brings pollutants from Pittsburgh (monitors B, L, M, N, and O). This peak behavior is the basis of our claim that, for SCS purposes, Pittsburgh is as great an influence as are the power plants.

The highest conditional probability of a peak concentration occurring in a single wind direction is 25 percent. Overall, there is approximately a five percent chance of any wind direction seeing a peak concentration. The probability of a peak concentration occurring, and occurring in a given sector is, therefore, very small. However, since the probabilities were calculated using 10° sectors, the conditional probability of 25 percent is nearly nine times larger than the 1/36 which corresponds to a random conditional distribution of peaks.

Because most of the monitors are located to the east of both Pittsburgh and the power plants, it is extremely rare to find a peak concentration which could not be caused by some combination of plant and background. Only a few hundred hourly peak concentrations, out of nearly 13,000 possible hours, are unambiguously due to the power plants.

The probability of an hourly peak occurrence increases during the winter for all wind directions, presumably due to the effect of space heating. In our data base there was no conclusive evidence to indicate that increased instability in the summer months increased the probability of a plant-caused peak concentration.

4.4 EPA-Larsen Method

The Larsen method (Larsen, 1967; Larsen, 1971; Larsen, 1973; Larsen, 1974) has principally been used as an urban analysis tool for determining constant emission reductions required to meet ambient standards. This analysis was performed on a sample of monitors throughout the region (1, 7, 8, B, F, G, L, M, N). The technique assumes the characteristic log-normal distribution of urban concentrations and generates statistical statements of expected concentration levels over various averaging periods. These levels can then be used to generate required emission reductions. It was found that the monitor data did exhibit log-normal distributions, so we felt justified in applying Larsen's concept to rural conditions. The following paragraphs describe the most significant results produced by this analysis at the sample monitors. More detail can be found in Appendix C.

One of the three federal standards will "dominate" at each monitor, i.e., given the distribution of concentrations, one averaging period will be more critical for given numerical standards than others. The annual standard dominated at all but monitors 1, 7, and 8, where the 24-hour standard dominated. Since the dominance function is monotonic, this implies that three-hour standards are least important at the sample monitors, implying either very persistent meteorology, which is not the case at Chestnut Ridge, or background influence.

Larsen's expected maximum concentrations for three-hour, 24-hour, and annual averaging periods were calculated and showed that several monitors threatened the annual standard, including monitor B which is upwind of the plants most of the time. Observed 1975 averages showed a similar problem, implying advection of substantial pollution on the prevailing winds.

Larsen's method, since it is statistical, cannot differentiate between background and power plant concentrations. Using his guidelines, considerable permanent emission reductions, up to 76 percent, would be needed in the region. If background is recognized, if the upwind monitors are used to gauge its impact, and if emission reductions at Chestnut Ridge are used in an attempt to bring total impacts into compliance, some plants would have to be shut down permanently. These conclusions reflect the difficulty of using constant controls to reduce high averages caused by relatively few peak values.

4.5 Stochastic Background Modeling Methodology

The background model used in the control strategy during the demonstration was described in Section 2.1. The "stochastic background model" to be discussed here is the result of a research effort that started too late for results to be incorporated into the operating air quality model.

This stochastic model was developed and verified using only three months of data from four monitors (L, M, N, O) in the "northern sector" (see Table 2.2.1 forcoordinates). Thus, it cannot yet be classified as a valid model for all of Chestnut Ridge. It is discussed here because the modeling methodology appears to be a powerful tool that can be applied to other locations and problems in addition to just Chestnut Ridge.

The modeling methodology is based on the following classical procedures:

- Look for "phenomena" that are present in the data.
- Hypothesize a mathematical structure which might account for the phenomena.
- Estimate the parameters of the mathematical model by "fitting them" to the data.
- Test/verify the validity of the model by seeing how well it fits the data.

This procedure is used repeatedly in an iterative fashion until a "final" model is obtained.

The studies discussed in Section 4.2 - 4.4 and associated appendices uncovered the phenomena that were actually present in the data.

The hypothesized structure of the final model has

- Inputs: Wind direction (hourly)
  Time of day (hour)
  Day of year
- Outputs: Probability distribution of SO_2 background concentration (one-hour average) at all points in region being modeled (two horizontal coordinates).

The structure is dynamic in two ways. The stochastic processes are correlated in time and the output depends on the past history of wind directions (five-ten hours) as well as its present value. The stochastic variation is also correlated in space. The probability distributions are "log-normal" when the wind direction and time inputs are specified.

The parameters of the model were estimated by using a combination of "least squares regression" with "Kalman filtering" techniques.

Model verification testing was done by inspection of the autocorrelation functions and histograms of the "residuals." A residual is the difference between the observed concentration and the value predicted by the model.

Appendix D describes the model and modeling methodology in more detail.

5.0 AIR QUALITY MODELING IN COMPLEX TERRAIN

It was apparent after comparison of calculated and observed maximum concentrations that the operating air-quality model had severe deficiencies in many of its characteristics. A separate study on ways to improve the air-quality model was undertaken after the actual demonstration had taken place. This effort considered only the deterministic portion of the air-quality model as discussed in Section 2.4. Section 4 discussed modeling research that was done on the background. The following discussions are concerned primarily with the power plant effects themselves (i.e., large point-sources).

5.1 Shortcomings of the Original Model

The principal faults of the original operating model are now discussed.

1. Shallow Mixing Layers Associated with Stable Stratifications. The model accepted the predicted low mixing depths associated with stable conditions. These values result from the determination of mixing depths from soundings and surface temperature forecasts (see Section 2.5). They are unrealistic in that the mixing depth under normal stable conditions is essentially either unlimited or non-existent, depending upon how it is used in the model. The result is that plumes from the larger stacks "punch through" the mixing layer, while the background emissions and sometimes the plumes from Seward are "trapped" and produce large, and often unrealistic, concentrations. This particular deficiency is common to many current models, and can be corrected by use of the σ_z curve for stable conditions.

2. Use of Non-Buoyant Plume Dispersion Coefficients for Buoyant Plumes. Buoyant entrainment causes the radius of a buoyant plume to increase linearly with its rise (Briggs, 1971) and can dominate the dispersion of a plume by atmospheric turbulence under many conditions. Commonly used dispersion coefficients, such as the Pasquill-Turner coefficients (Turner, 1, 1969), or the ASME coefficients (ASME, 1968), were based largely on the observation of passive plumes and often underestimate plume dimensions near the source. A review of the Gaussian plume model (Pasquill, 1976) documented this discrepancy. The Chestnut Ridge plants have large heat fluxes (from 1 to 40 times the size of a defined "large" plant - 40 MWe), and are in rough terrain. Both of these factors would tend to make the operating model, which uses the ASME coefficients, liable to serious prediction errors.

3. Lifting of Plumes over Terrain Features. The use of terrain-correction factors to model the lifting of streamlines, and limiting the closeness of approach of plume centerlines to terrain obstacles are improvements over the schemes used by EPA in the Valley models and elsewhere, in which centerlines can get within 10 m of the terrain. As the terrain correction was applied in the original operating model, however, the amount of "lift" experienced by the plume as it passed over the obstacle was independent of the height of the plume as it approached the terrain. A plume much higher than the obstacle will obviously not be lifted as high as a plume approaching from a lower position, as was demonstrated by solutions to the potential flow equations for flow over a semi-circular cylinder and over a hemisphere (Egan, 2, 1975). The result of this deficiency of the

model was to underpredict concentrations due to the power plants at elevated receptors, particularly in situations with light winds and high plume rises.

4. Ignoring Impact of Plumes That Rise through Inversion Base. The operating model ignored the impact of any plume that was modeled to rise into the stable air of an elevated inversion. This "punch-through" feature has a long history extending at least back to the EPA Air Quality Display Model (EPA, 1970). The Chestnut Ridge data base, to the extent that the mixing depths verified by the scheme used in this project are correct, demonstrated that significant concentrations, apparently due to plant impact, occur when the plumes would be modeled to penetrate well into the stable layer. Furthermore, there is no obvious rational basis for assuming that the impact of such plumes is less than their impact in stable stratification.

5.2 Model Improvement Methodology

The problems of the operating model outlined above were addressed by working with a large sample of case studies with one-hour data. The West Fairfield (1), Laurel Ridge (3), and Liggett (F) monitors were selected for this purpose (see Table 2.2.1 for coordinates). The West Fairfield site is in a rather unique location on a moderately elevated knoll of the southwestern end of the valley channel between Chestnut and Laurel Ridges. The plumes from the Seward and Conemaugh plants might be expected to have simultaneous impact at this site because both plants lie roughly to the northeast of the monitoring station.

The Laurel Ridge site is the most elevated monitor in the network and lies only seven km east-southeast of the Conemaugh plant. Of all the relationships between sources and monitors available at Chestnut Ridge, this is the most representative of a large plant in rough terrain.

Liggett was selected because it is in rolling terrain of the sort typical of a large part of the network and in a direction from the Homer City plant that has a high frequency of winds in various meteorological conditions. The Liggett monitor is on moderately elevated terrain that is the northerly remnant of Chestnut Ridge itself.

Cases were selected from the data archived for the period from September 1974 through December 1975.

The initial selection of cases to be examined included all one-hour periods in which the one-hour average wind direction at the 91.1 m level of the meteorological tower fell within the 23° sector centered on the directed trajectory from the plant to the monitor. The sector width for the West Fairfield cases was expanded to 45° because of the relatively low frequency of northeasterly winds and the potential for strong channeling of winds along the valley's axis up to the monitor.

Almost 3,000 hourly cases were identified that fit the wind direction criteria for these three monitors. Of these cases, almost half were for the Homer City to Liggett trajectory. Because this was an unwieldy number of cases, many of which were clearly not indicative of major plume impacts, a further selection process was made to include only cases in which the one-hour SO_2 concentrations exceeded 90 parts per billion (ppb) at Liggett and Laurel Ridge, and 40 ppb at West Fairfield. The Homer City to Liggett set retained 149 cases; the Conemaugh to Laurel Ridge set, 78; and the Conemaugh-Seward to West Fairfield set, only 20.

The model improvement methodology is similar to that discussed in Section 4.5 on background modeling in that a sequence of model changes is hypothesized and then tested to see how well it improves the fit of the model to the data. However, the details are very different. In the present case, a sequence of specific modifications to the model of Section 2.4 was hypothesized. A method of "skill scoring" was developed to provide a way of comparing the relative performance of the modifications.

A major improvement in performance was obtained. The details of the modifications and the skill-scoring procedure are given in Appendix E.

6.0 SCS RELIABILITY ANALYSIS

Section 3 presented an analysis of the demonstration period performance of the Chestnut Ridge SCS design, including issues such as the logistics of developing forecasts and control strategies and the communication and implementation of control decisions within required time limits. The demonstration period experience showed that normal power system operating procedures would allow the execution of load shifting control actions on schedule and in accordance with the degree of control requested by an SCS control model. Because control can be implemented so reliably, the critical elements of SCS reliability must lie in the other parts of the SCS which are concerned with the prediction of air quality in the control region.

One serious drawback to the hindsight analysis of the demonstration period was the small number of actual and near violations that occurred. That small sample of air quality observations prevented any conclusive statements about the overall reliability of the complete system.

Section 5 described an attempt to improve air quality modeling in rough terrain. The modifications were made after the demonstration was concluded, and a simple skill score was used to evaluate the effects of changes in model structure and parameter values. That analysis was restricted to a small (2% of total data) set of 1-hr average data characterized by plume impact at three of the seventeen monitors.

Additional reliability analysis was performed after the demonstration period. The techniques were used (Gaertner, 1974) to examine the question of reliability for the SCS air quality predictions, first with the operating model of the demonstration period and then with the modified design from Section 5. This section presents the results of that additional reliability analysis.

6.1 Approach

Section 2.6 introduced the concept of a meteorological function M(t) which relates emission, Q(t), to the maximum concentration observed, $C_o(t)$, as follows:

$$C_o(t) = M(t) \cdot Q(t) \tag{6.1}$$

For our system assumption of a linear plant heat rate curve, Q(t) can be further expressed as

$$Q(t) = \alpha S(t) \cdot L(t) \tag{6.2}$$

where

L(t) = plant load in MWe

S(t) = fuel sulfur content

α = linear conversion factor

This formulation implies perfect knowledge of these functions which is not true. A better representation is

$$C_p(t) = \alpha M(t) \cdot S(t) \cdot L(t) \cdot R(t) \qquad (6.3)$$

where $C_p(t)$ is the predicted concentration maximum and R(t) is defined as the error ratio at time t:

$$R(t) = C_p(t)/C_o(t) \qquad (6.4)$$

Let us drop the time notation and observe that R contains contributions from all sources of error or uncertainty which prevent a perfect forecast ($C_p = C_o$). R can be factored into the following form:

$$R = R_S \cdot R_L \cdot R_W \cdot R_M \qquad (6.5)$$

where R_S, R_L, R_W, R_M are respectively error ratios for sulfur content errors, plant load errors, weather forecasting errors, and air quality modeling errors.

In a perfect SCS, all factors of R would equal unity at all times. Control strategies could then be implemented with complete assurance that once emission control imits were met, maximum concentrations would meet their targets. In fact, R varies with its components. Uncertainty remains, making a probabilistic strategy necessary. A mean value of R more than 1 represents a generally conservative air quality forecasting system. The smaller the variance of R, the more confidence we have in choosing emission controls.

The structure of this formulation suggests working with ln R instead of R when examining the mean and variance. A linear analysis of R tends to compress the distribution of R values between 0 and 1, and stretch it out between 1 and ∞. Using ln R assures that all ratios of the same relative magnitude are weighted equally. Plots of the samples of R are given on a logarithmic (base e) scale (Figures 6.2.1 to 6.2.11).

Our analysis dealt with R and its components in a form which shows the dependence of C_p on assumptions of sulfur content and forecasts of plant loads and meteorology. Concentration prediction errors are represented by the use of predicted (modeled) or observed (perfect model) concentrations (subscripts p = predicted; o = observed):

$$R = R_S \cdot R_L \cdot R_W \cdot R_M$$

$$= \frac{C_p(S_p, L_o, W_o)}{C_p(S_o, L_o, W_o)} \cdot \frac{C_p(S_p, L_p, W_o)}{C_p(S_p, L_o, W_o)} \cdot \frac{C_p(S_p, L_p, W_p)}{C_p(S_p, L_p, W_o)} \cdot \frac{C_p(S_o, L_o, W_o)}{C_o}$$

$$= \frac{C_p(S_p, L_p, W_p)}{C_o} \qquad (6.6)$$

Rather than analyze the entire 19-month data record, four months in 1975 were chosen to represent seasonal behavior at Chestnut Ridge: May, July, October, and December. In addition, the demonstration period, March 1976, was included to form a total of five "study months." Data were assembled in these periods for predicted and observed fuel sulfur contents, plant load levels, and weather parameters (wind direction, wind speed, mixing height, stability). The data were used in the form of 3 hr averages, corresponding to the 8 daily 3 hr periods used in the demonstration period operations (0-3, 3-6, etc.). These data were used in both air quality models (the oprating model from the demonstration period and the modified model) to develop predictions of ground-level concentrations at Chestnut Ridge for all 3 hr periods in the five study months. Periods of missing data were not included. The eight daily 3 hr averages were combined into daily 24-hr averages as well. From the entire concentration field in each 3 hr period, the maximum monitor (out of 17) and maximum single receptor (out of 394) were available to represent the various C_p's in equation (6.6). The ratios of these C_p's formed sample functions of the distributions of the various R's. Observed monitor data, averaged over the corresponding 3 hr time periods were used for C_o.

6.2 Results for Demonstration Period Model

6.2.1 Two Source vs Eight Source Analysis

6.2.1.1 Choice of Sample Function

The first result obtained allowed a simplification of subsequent calculations. The error ratio of two different models was analyzed for 3 hr and 24 hr average concentrations:

$$R_M^1 = \frac{^2C_p(S_o, L_o, W_o)}{^8C_p(S_o, L_o, W_o)} \tag{6.7}$$

where $^nC_p(.)$ is the 3 hr or 24 hr average concentration predicted by the operating model of the demonstration period with four background sources and with n = 2 or n = 8 power plant sources emitting SO_2. For n = 8, all the Chestnut Ridge power plant units were included as sources. For n = 2, only the two units at Seward were used and the emissions of the other three plants (six sources) were set to zero. The two error ratios (n = 2, n = 8) for 3 hr and 24 hr averages

$$^nR_M = \frac{^nC_p(S_o, L_o, W_o)}{C_o} \tag{6.8}$$

were also computed, where C_o is the observed maximum of the 17 monitors. These three comparisons, R_M^1, ^2R_M and ^8R_M, were motivated by three factors:

- our experience had shown that most observed high concentrations were induced by Seward station
- computation costs could be halved if n = 2 were used instead of n = 8
- the analysis method being used was originally conceived as a one-plant analysis tool. While it can be applied to multiple plants, if they are operated as part of the same SCS strategy, it is most easily interpreted with one plant.

Although both monitor and receptor predictions were available in forming these several ratios, the maximum receptor concentration was used as the sample value for the ${}^{n}C_{p}(S_o, L_o, W_o)$. The rationale for this choice follows.

SCS control should be implemented on the basis of observed or predicted occurrences of high concentrations. As the density of physical monitor locations or mathematical receptor locations increases, the likelihood of actually observing or calculating the true spatially dependent maximum concentration grows. If we were to sample only the maximum monitor concentration predicted by the air quality model, only 17 points would be available to cover the Chestnut Ridge area. A clearly finer grain approximation to the true concentration field is the 394 receptor array. Sampling the maximum receptor as a prediction of the maximum SO_2 level is therefore more consistent with the objectives of SCS operation, although it does introduce a bias on R, discussed below.

6.2.1.2 Model Conservatism

Let us consider a hypothetical, perfect air quality model with perfect inputs and its sample distribution: ${}^{n}R_{M}$.

If the model has k mathematical receptors where concentrations are calculated and the physical monitoring system has ℓ physical monitors, then the following can be said about ${}^{n}R_{M}$:

- $k = \ell$: (same number and same locations): the sample distribution of $\ln {}^{n}R_{M}$ will have mean value = 0 and σ = 0.
- $k > \ell$: the sample distribution of $\ln {}^{n}R_{M}$ will be shifted to have a mean value > 0 and $\sigma = 0$.

The first statement is clear. If it is a perfect model it should predict exactly what is observed at the monitors, and the sampled error ratio will always equal unity. In the second case, the finer grain receptor array, for which perfect concentrations are computed, will more frequently see the actual spatial maximum concentration than will the coarser monitor array. Our error ratio will tend towards values greater than unity. As long as the monitor locations are a subset of the receptor array, underprediction, $\ln {}^{n}R_{M} < 0$, will never occur.

Chestnut Ridge corresponds to the second of the hypothetical cases (394 receptors, 17 monitors, monitors a subset of receptors). If the air quality model were perfect, and perfect input data were available, the limited monitor array would still lead to a distribution of $\ln {}^{n}R_{M}$ which had a mean value > 0, rather than equal to 0. This corresponds to a "conservative" SCS but where does this conservatism arise? Generally this term is used to describe SCS air quality modeling performance with respect to monitored concentrations, where the monitored values are assumed to be perfect. Since no finite monitoring system is perfect, any model with a finer grain receptor array will have some element of conservatism just because it predicts concentrations at more locations (this assumes a reasonable choice of locations).

The question of to what extent $\ln {}^{n}R_{M}$ shows conservatism due to the necessarily limited monitor system or due to true "safety factors" being present in the air quality model structure and parameters cannot be answered completely since no complete air quality data field is available for quantifying the monitor effects.

However, some insight can be gained by considering how "perfectly" an air quality forecast performating at the subset of monitor locations. If it is assumed that that performance can be extrapolated to nonmonitor locations then we can identify a monitor component and prediction component of the total conservatism (Table 6.2.1). If we further assume that the form of the total conservatism is multiplicative:

(Total Conservatism) = (Monitor-Related Conservatism) x (Model-Related Conservatism) (6.9)

then (since the distributions of ${}^{n}C_{p}$ and therefore ${}^{n}R_{M}$ tend to be normal in their logarithms) the monitor-related means and σ's can be subtracted from the receptor-related values to give an estimate of the true "conservatism" statistics.

Table 6.2.1

Model Conservatism Data

$\ln {}^{8}R_{M}$ - 5-month Study Period

	Monitors		Receptors		"Conservatism"	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std.* Dev.
3 hr Average	0.34	0.92	1.10	0.78	0.76	-0.14
24 hr Averag	0.24	0.56	0.71	0.52	0.47	-0.04

$\ln {}^{2}R_{M}$ - 5-month Study Period

	Monitors		Receptors		"Conservatism"	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std.* Dev.
3 hr Average	-0.13	1.12	1.01	0.86	1.14	-0.26
24 hr Average	-0.04	0.71	0.59	0.58	0.63	-0.13

*See text.

The fact that a negative σ value results from this simplified analysis indicates that some correspondence exists between the true conservatism effects in ln $^{n}R_M$ and the monitor effects, i.e., that conditions leading to conservatism (e.g., sector averaging) affect monitor predictions and total receptor predictions similarly. This is understandable since the same air quality model structure is in use. The correspondence tends toward reducing the σ of the receptor maximum ln $^{n}R_M$ as compared to the monitor maximum, and is one indication of the "benefit" of using a finer grain receptor array. Note also that the magnitude of the mean monitor and receptor values is smaller with the longer averaging period. This is simply a reflection of the greater precision expected when smoothing (averaging) occurs over a 24 hr period. In the important receptor case, a gain in precision is obtained at the expense of conservatism. Finally note that ln $^{2}R_M$, as compared to ln $^{8}R_M$, has smaller mean values but larger σ values in the monitor and receptor data. This can be interpreted as saying that the two-source model is a closer mean representation of the observed data (in part because it misses some high predictions of the eight-source model) but that there is less certainty of having values near the mean occur. The "improvement" in true conservatism is ambiguous since it is in part due to the nonconservative monitor performance of the model (although were the two-source model perfect at the monitors it would have slightly more mean "conservatism").

Focusing once again on the receptor data, the two-source model is seen to be over all less conservative than the eight-source model, and has greater variability about its mean error ratio. That the σ value is almost as large as the mean implies underprediction occurs, i.e., the model cannot be perfect. Underprediction is more likely to occur in the two-source than in the eight-source model (Figure 6.2.1 and Figure 6.2.2). The smaller sample in $^{8}R_M$ is due to the necessity of finding periods when valid data for all 8 sources was available.

6.2.1.3 Direct Comparison

The sample distributions of ln R_M^1 (Eq. 6.7) are limited to nonpositive values (Figure 6.3). This results from the fact that the maximums of the two-source model are, by definition, less than or equal to the maximums of the eight-source model. The sample distributions of ln R_M^1 are quite "tight", reflecting good correspondence between the models. Due to the smoothing effects of the longer averaging period, the 24 hr σ was smaller (0.09) than the 3 hr σ (0.14).

The outlying points in the sample distributions of ln R_M^1 represent data from time periods in which the receptor array maximum concentration was caused by a source other than the Seward units. This could occur because of the greater emissions from the other, non-Seward units, or because the meteorology resulted in a plume impact from another plant. Most of the outliers involve ratios of concentrations which are each relatively small numbers and therefore of little interest to SCS control decisions. The outliers were investigated further by looking at the data set which included only those time periods in which either $^{2}C_p$ or $^{8}C_p$ was greater than some cutoff value (200 ppb [3 hr] or 100 ppb [24 hr]). It was found that less than 2% of the sample values both met the cutoff condition and remained outliers, defined as having a ln R_M^1 value less than -0.24. Since this corresponds to an error ratio of 0.78, we can restate our result:

07 1777 15[illegible] EPSIL 5 MONTH OBSVOB/AIRMAP RCP O3H LOG FORCLN EPSIL

MEAN= 1.55586 SIGMA= 0.77576 SKEW= -0.93284 XMIN= -1.07778 XMAX= 2.91906 NV= 272

	RANGE		CUM FREQ	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-3.0000	-2.8800	272	0
2	-2.8400	-2.7200	272	0
3	-2.7200	-2.6400	272	0
4	-2.6400	-2.5200	272	0
5	-2.5200	-2.4000	272	0
6	-2.4000	-2.2800	272	0
7	-2.2800	-2.1600	272	0
8	-2.1600	-2.0400	272	0
9	-2.0400	-1.9200	272	0
10	-1.9200	-1.8000	272	0
11	-1.8000	-1.6800	272	0
12	-1.6800	-1.5600	272	0
13	-1.5600	-1.4400	272	0
14	-1.4400	-1.3200	272	0
15	-1.3200	-1.2000	272	0
16	-1.2000	-1.0800	272	0
17	-1.0800	-0.9600	271	1 *
18	-0.9600	-0.8400	271	0
19	-0.8400	-0.7200	269	2 **
20	-0.7200	-0.6000	266	3 ***
21	-0.6000	-0.4800	263	3 ***
22	-0.4800	-0.3600	260	3 ***
23	-0.3600	-0.2400	258	2 **
24	-0.2400	-0.1200	256	0
25	-0.1200	0.0000	257	1 *
26	0.0000	0.1200	257	0
27	0.1200	0.2400	256	1 *
28	0.2400	0.3600	255	1 *
29	0.3600	0.4800	249	6 ******
30	0.4800	0.6000	246	3 ***
31	0.6000	0.7200	242	4 ****
32	0.7200	0.8400	231	11 ***********
33	0.8400	0.9600	225	6 ******
34	0.9600	1.0800	211	14 **************
35	1.0800	1.2000	200	11 ***********
36	1.2000	1.3200	179	21 **********************
37	1.3200	1.4400	168	11 ***********
38	1.4400	1.5600	153	15 ***************
39	1.5600	1.6800	135	18 ******************
40	1.6800	1.8000	113	22 **********************
41	1.8000	1.9200	90	23 ***********************
42	1.9200	2.0400	71	19 *******************
43	2.0400	2.1600	59	12 ************
44	2.1600	2.2800	48	11 ***********
45	2.2800	2.4000	36	12 ************
46	2.4000	2.5200	21	15 ***************
47	2.5200	2.6400	13	8 ********
48	2.6400	2.7600	6	7 *******
49	2.7600	2.8800	2	4 ****
50	2.8800	3.0000	0	2 **

SCALING FACTOR = 1

Sample Distribution of $\ln {}^{8}R_{M}$ 3 hr. Averages

Figure 6.2.1a

V U= 0.70761 SIGMA= 0.52200 SKEW= 0.25963 VMIN= -0.49936 VMAX= 2.25324 NV= 110

	RANGE		CUM FREQ		0...1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	110	0	
2	-2.8800	-2.7600	110	0	
3	-2.7600	-2.6400	110	0	
4	-2.6400	-2.5200	110	0	
5	-2.5200	-2.4000	110	0	
6	-2.4000	-2.2800	110	0	
7	-2.2800	-2.1600	110	0	
8	-2.1600	-2.0400	110	0	
9	-2.0400	-1.9200	110	0	
10	-1.9200	-1.8000	110	0	
11	-1.8000	-1.6800	110	0	
12	-1.6800	-1.5600	110	0	
13	-1.5600	-1.4400	110	0	
14	-1.4400	-1.3200	110	0	
15	-1.3200	-1.2000	110	0	
16	-1.2000	-1.0800	110	0	
17	-1.0800	-0.9600	110	0	
18	-0.9600	-0.8400	110	0	
19	-0.8400	-0.7200	110	0	
20	-0.7200	-0.6000	110	0	
21	-0.6000	-0.4800	109	1	*
22	-0.4800	-0.3600	107	2	**
23	-0.3600	-0.2400	106	1	*
24	-0.2400	-0.1200	104	2	**
25	-0.1200	0.0000	102	4	****
26	0.0000	0.1200	97	3	***
27	0.1200	0.2400	88	9	*********
28	0.2400	0.3600	84	4	****
29	0.3600	0.4800	73	11	***********
30	0.4800	0.6000	68	5	*****
31	0.6000	0.7200	52	16	****************
32	0.7200	0.8400	39	13	*************
33	0.8400	0.9600	30	9	*********
34	0.9600	1.0800	24	6	******
35	1.0800	1.2000	18	6	******
36	1.2000	1.3200	12	6	******
37	1.3200	1.4400	10	2	**
38	1.4400	1.5600	9	1	*
39	1.5600	1.6800	4	5	*****
40	1.6800	1.8000	3	1	*
41	1.8000	1.9200	2	1	*
42	1.9200	2.0400	1	1	*
43	2.0400	2.1600	1	0	
44	2.1600	2.2800	0	1	*
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of ln $^{8}R_{M}$ 24 hr. Averages

Figure 6.2.1b

VMU= 1.00556 SIGMA= 0.86321 SKEW= -0.26122 VMIN= -1.85895 VMAX= 2.99573 NV= 1039

	RANGE		CUM FREQ.	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	1039	0
2	-2.8800	-2.7600	1039	0
3	-2.7600	-2.6400	1039	0
4	-2.6400	-2.5200	1039	0
5	-2.5200	-2.4000	1039	0
6	-2.4000	-2.2800	1039	0
7	-2.2800	-2.1600	1039	0
8	-2.1600	-2.0400	1039	0
9	-2.0400	-1.9200	1039	0
10	-1.9200	-1.8000	1038	1 *
11	-1.8000	-1.6800	1037	1 *
12	-1.6800	-1.5600	1036	1 *
13	-1.5600	-1.4400	1036	0
14	-1.4400	-1.3200	1035	1 *
15	-1.3200	-1.2000	1033	2 **
16	-1.2000	-1.0800	1030	3 ***
17	-1.0800	-0.9600	1021	9 *********
18	-0.9600	-0.8400	1018	3 ***
19	-0.8400	-0.7200	1010	8 ********
20	-0.7200	-0.6000	996	14 **************
21	-0.6000	-0.4800	986	10 **********
22	-0.4800	-0.3600	965	21 **********************
23	-0.3600	-0.2400	950	15 ***************
24	-0.2400	-0.1200	931	19 *******************
25	-0.1200	0.0000	904	27 ****************************
26	0.0000	0.1200	882	22 **********************
27	0.1200	0.2400	849	33 *********************************
28	0.2400	0.3600	796	53 ***
29	0.3600	0.4800	747	49 ***
30	0.4800	0.6000	706	41 ***
31	0.6000	0.7200	668	38 **************************************
32	0.7200	0.8400	621	47 **
33	0.8400	0.9600	569	52 ***
34	0.9600	1.0800	508	61 **
35	1.0800	1.2000	447	61 **
36	1.2000	1.3200	389	58 ***
37	1.3200	1.4400	337	52 ***
38	1.4400	1.5600	284	53 **
39	1.5600	1.6800	240	44 **
40	1.6800	1.8000	197	43 ***
41	1.8000	1.9200	154	43 ***
42	1.9200	2.0400	115	39 **
43	2.0400	2.1600	97	18 ******************
44	2.1600	2.2800	69	28 ****************************
45	2.2800	2.4000	50	19 *******************
46	2.4000	2.5200	29	21 **********************
47	2.5200	2.6400	20	9 *********
48	2.6400	2.7600	9	11 ***********
49	2.7600	2.8800	3	6 ******
50	2.8800	3.0000	0	3 ***

SCALING FACTOR = 1

Sample Distribution of $\ln {}^{2}R_{M}$ 3 hr. Averages

Figure 6.2.2a

VMU= 0.59553 SIGMA= 0.57894 SKEW= -0.05061 VMIN= -1.11803 VMAX= 2.25324 NV= 134

	RANGE		CUM FREQ.		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	134	0	
2	-2.8800	-2.7600	134	0	
3	-2.7600	-2.6400	134	0	
4	-2.6400	-2.5200	134	0	
5	-2.5200	-2.4000	134	0	
6	-2.4000	-2.2800	134	0	
7	-2.2800	-2.1600	134	0	
8	-2.1600	-2.0400	134	0	
9	-2.0400	-1.9200	134	0	
10	-1.9200	-1.8000	134	0	
11	-1.8000	-1.6800	134	0	
12	-1.6800	-1.5600	134	0	
13	-1.5600	-1.4400	134	0	
14	-1.4400	-1.3200	134	0	
15	-1.3200	-1.2000	134	0	
16	-1.2000	-1.0800	133	1	*
17	-1.0800	-0.9600	133	0	
18	-0.9600	-0.8400	132	1	*
19	-0.8400	-0.7200	132	0	
20	-0.7200	-0.6000	132	0	
21	-0.6000	-0.4800	129	3	***
22	-0.4800	-0.3600	128	1	*
23	-0.3600	-0.2400	123	5	*****
24	-0.2400	-0.1200	119	4	****
25	-0.1200	0.0000	114	5	*****
26	0.0000	0.1200	107	7	*******
27	0.1200	0.2400	100	7	*******
28	0.2400	0.3600	91	9	*********
29	0.3600	0.4800	83	8	********
30	0.4800	0.6000	72	11	***********
31	0.6000	0.7200	57	15	***************
32	0.7200	0.8400	43	14	**************
33	0.8400	0.9600	30	13	*************
34	0.9600	1.0800	25	5	*****
35	1.0800	1.2000	18	7	*******
36	1.2000	1.3200	13	5	*****
37	1.3200	1.4400	10	3	***
38	1.4400	1.5600	8	2	**
39	1.5600	1.6800	3	5	*****
40	1.6800	1.8000	2	1	*
41	1.8000	1.9200	2	0	
42	1.9200	2.0400	2	0	
43	2.0400	2.1600	1	1	*
44	2.1600	2.2800	0	1	*
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of $\ln {}^{2}R_{M}$ 24 hr. Averages

Figure 6.2.2b

V-U= -0.06335 SIGMA= 0.13834 SKEW= -2.89881 VMIN= -0.86786 VMAX= 0.0 NV= 864

	RANGE		CUM FREQ		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.9800	864	0	
2	-2.9800	-2.7600	864	0	
3	-2.7600	-2.6400	864	0	
4	-2.6400	-2.5200	864	0	
5	-2.5200	-2.4000	864	0	
6	-2.4000	-2.2800	864	0	
7	-2.2800	-2.1600	864	0	
8	-2.1600	-2.0400	864	0	
9	-2.0400	-1.9200	864	0	
10	-1.9200	-1.8000	864	0	
11	-1.8000	-1.6800	864	0	
12	-1.6800	-1.5600	864	0	
13	-1.5600	-1.4400	864	0	
14	-1.4400	-1.3200	864	0	
15	-1.3200	-1.2000	864	0	
16	-1.2000	-1.0800	864	0	
17	-1.0800	-0.9600	864	0	
18	-0.9600	-0.8400	863	1	
19	-0.8400	-0.7200	857	6	
20	-0.7200	-0.6000	850	7	
21	-0.6000	-0.4800	836	14	*
22	-0.4800	-0.3600	820	16	**
23	-0.3600	-0.2400	785	35	****
24	-0.2400	-0.1200	709	76	*********
25	-0.1200	0.0000	0	709	**********
26	0.0000	0.1200	0	0	
27	0.1200	0.2400	0	0	
28	0.2400	0.3600	0	0	
29	0.3600	0.4800	0	0	
30	0.4800	0.6000	0	0	
31	0.6000	0.7200	0	0	
32	0.7200	0.8400	0	0	
33	0.8400	0.9600	0	0	
34	0.9600	1.0800	0	0	
35	1.0800	1.2000	0	0	
36	1.2000	1.3200	0	0	
37	1.3200	1.4400	0	0	
38	1.4400	1.5600	0	0	
39	1.5600	1.6800	0	0	
40	1.6800	1.8000	0	0	
41	1.8000	1.9200	0	0	
42	1.9200	2.0400	0	0	
43	2.0400	2.1600	0	0	
44	2.1600	2.2800	0	0	
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 8

Sample Distribution of ln R_M^1 3 hr. Averages

Figure 6.2.3a

V U= -0.06949 SIGMA= 0.09151 SKEW= -1.71339 VMIN= -0.40547 VMAX= 0.0 NV= 110

	RANGE		CUM FRE		.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	110	0	
2	-2.8800	-2.7600	110	0	
3	-2.7600	-2.6400	110	0	
4	-2.6400	-2.5200	110	0	
5	-2.5200	-2.4000	110	0	
6	-2.4000	-2.2800	110	0	
7	-2.2800	-2.1600	110	0	
8	-2.1600	-2.0400	110	0	
9	-2.0400	-1.9200	110	0	
.0	-1.9200	-1.8000	110	0	
.1	-1.8000	-1.6800	110	0	
.2	-1.6800	-1.5600	110	0	
.3	-1.5600	-1.4400	110	0	
.4	-1.4400	-1.3200	110	0	
.5	-1.3200	-1.2000	110	0	
.6	-1.2000	-1.0800	110	0	
.7	-1.0800	-0.9600	110	0	
.8	-0.9600	-0.8400	110	0	
.9	-0.8400	-0.7200	110	0	
20	-0.7200	-0.6000	110	0	
21	-0.6000	-0.4800	110	0	
22	-0.4800	-0.3600	107	3	***
23	-0.3600	-0.2400	103	4	****
24	-0.2400	-0.1200	84	19	*******************
25	-0.1200	0.0000	0	84	**********
26	0.0000	0.1200	0	0	
27	0.1200	0.2400	0	0	
28	0.2400	0.3600	0	0	
29	0.3600	0.4800	0	0	
30	0.4800	0.6000	0	0	
31	0.6000	0.7200	0	0	
32	0.7200	0.8400	0	0	
33	0.8400	0.9600	0	0	
34	0.9600	1.0800	0	0	
35	1.0800	1.2000	0	0	
36	1.2000	1.3200	0	0	
37	1.3200	1.4400	0	0	
38	1.4400	1.5600	0	0	
39	1.5600	1.6800	0	0	
40	1.6800	1.8000	0	0	
41	1.8000	1.9200	0	0	
42	1.9200	2.0400	0	0	
43	2.0400	2.1600	0	0	
44	2.1600	2.2800	0	0	
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of ln R_M^1 24 Hr. Averages

Figure 6.2.3b

Over 98% of the sample error ratios, comparing the two-source and eight-source models, were either formed from low concentrations or were within a range of 0.78 to 1.00. On the basis of this observation and the other reasons given above, it was decided that the two-source model adequately represented the performance of the eight-source model for concentrations at Chestnut Ridge which were significant to SCS control. The two-source model was therefore utilized in all subsequent analysis.

6.2.2. Sulfur Content Analysis

The sample distribution of $\ln R_s$ (Figure 6.2.4) shows that sulfur content errors are not contributing much uncertainty to R. This can be attributed to two things. First, the State of Pennsylvania regulates air quality via emission standards expressed as sulfur content of the fuel being burned. Thus, there is a strong incentive for a source to keep average sulfur levels below standards. At the same time, economics argues for using the highest sulfur content fuel permitted. The power plants therefore tried to maintain an average sulfur content of 2.4% sulfur by weight, the Pennsylvania standard. Second, data on short-term (3 hr basis) variability in observed sulfur content was not available. Although the coal was analyzed using grab samples on a daily basis, the only regularly available data were from shipping purchase records. These tended to involve large quantities of coal and the resulting sulfur statistics were highly aggregated. S_p was the long term average sulfur content, assumed equal to the state standard of 2.4 lb/MBtu. S_o were actual values from shipping purchase forms.

The sulfur content ratios in Figure 6.2.4 were expected to be narrow distributions because of the regulatory/economic incentives and the limited sulfur data. The slightly negative mean reflects the fact that an assumption of exacts 2.4% sulfur is nonconservative when translated into concentration levels, since the regulatory incentives to maintain 2.4% are not reflected in the available data from as-received coal shown on the shipping records. Some coal cleaning is done. Also the averaging scale over which sulfur contents equal 2.4% seems to be longer than the isolated months studied. Both effects produce an optimistic error ratio.

6.2.3 Plant Load Analysis

Variations in plant load are infrequent at Chestnut Ridge because the plants are base-loaded. The plant unit commitment schedule, which was used for the predicted load levels, was updated each day. The only time that the unit commitment schedule would be seriously in error in these conditions would be when a forced outage occurred, or when a plant which had been off line came back into service fastor than expected. The error ratio used here was

$$R_L = \frac{^2C_p(S_p, L_p, W_o)}{^2C_p(S_p, L_o, W_o)} .$$

MEAN= -0.01950 SIGMA= 0.04745 SKEW= -0.8623 MIN= -0.11630 MAX= 0.04546 NV= 1055

	RANGE		CUM FREQ	
1	-3.0000	-2.8800	1055	0
2	-2.8800	-2.7600	1055	0
3	-2.7600	-2.6400	1055	0
4	-2.6400	-2.5200	1055	0
5	-2.5200	-2.4000	1055	0
6	-2.4000	-2.2800	1055	0
7	-2.2800	-2.1600	1055	0
8	-2.1600	-2.0400	1055	0
9	-2.0400	-1.9200	1055	0
10	-1.9200	-1.8000	1055	0
11	-1.8000	-1.6800	1055	0
12	-1.6800	-1.5600	1055	0
13	-1.5600	-1.4400	1055	0
14	-1.4400	-1.3200	1055	0
15	-1.3200	-1.2000	1055	0
16	-1.2000	-1.0800	1055	0
17	-1.0800	-0.9600	1055	0
18	-0.9600	-0.8400	1055	0
19	-0.8400	-0.7200	1055	0
20	-0.7200	-0.6000	1055	0
21	-0.6000	-0.4800	1055	0
22	-0.4800	-0.3600	1055	0
23	-0.3600	-0.2400	1055	0
24	-0.2400	-0.1200	1055	0
25	-0.1200	0.0000	186	869
26	0.0000	0.1200	2	183
27	0.1200	0.2400	1	0
28	0.2400	0.3600	0	0
29	0.3600	0.4800	0	0
30	0.4800	0.6000	0	0
31	0.6000	0.7200	0	0
32	0.7200	0.8400	0	0
33	0.8400	0.9600	0	0
34	0.9600	1.0800	0	0
35	1.0800	1.2000	0	0
36	1.2000	1.3200	0	0
37	1.3200	1.4400	0	0
38	1.4400	1.5600	0	0
39	1.5600	1.6800	0	0
40	1.6800	1.8000	0	0
41	1.8000	1.9200	0	0
42	1.9200	2.0400	0	0
43	2.0400	2.1600	0	0
44	2.1600	2.2800	0	0
45	2.2800	2.4000	0	0
46	2.4000	2.5200	0	0
47	2.5200	2.6400	0	0
48	2.6400	2.7600	0	0
49	2.7600	2.8800	0	0
50	2.8800	3.0000	0	0

SCALING FACTOR = 9

Sample Distribution of ln R_S 3 hr. Averages

Figure 6.2.4a

VMU= -0.01710 SIGMA= 0.04317 SKEW= -0.87232 VMIN= -0.11607 VMAX= 0.04616 NV= 134

	RANGE		CUM FREQ.		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	134	0	
2	-2.8800	-2.7600	134	0	
3	-2.7600	-2.6400	134	0	
4	-2.6400	-2.5200	134	0	
5	-2.5200	-2.4000	134	0	
6	-2.4000	-2.2800	134	0	
7	-2.2800	-2.1600	134	0	
8	-2.1600	-2.0400	134	0	
9	-2.0400	-1.9200	134	0	
10	-1.9200	-1.8000	134	0	
11	-1.8000	-1.6800	134	0	
12	-1.6800	-1.5600	134	0	
13	-1.5600	-1.4400	134	0	
14	-1.4400	-1.3200	134	0	
15	-1.3200	-1.2000	134	0	
16	-1.2000	-1.0800	134	0	
17	-1.0800	-0.9600	134	0	
18	-0.9600	-0.8400	134	0	
19	-0.8400	-0.7200	134	0	
20	-0.7200	-0.6000	134	0	
21	-0.6000	-0.4800	134	0	
22	-0.4800	-0.3600	134	0	
23	-0.3600	-0.2400	134	0	
24	-0.2400	-0.1200	134	0	
25	-0.1200	0.0000	24	110	**********
26	0.0000	0.1200	0	24	**********
27	0.1200	0.2400	0	0	
28	0.2400	0.3600	0	0	
29	0.3600	0.4800	0	0	
30	0.4800	0.6000	0	0	
31	0.6000	0.7200	0	0	
32	0.7200	0.8400	0	0	
33	0.8400	0.9600	0	0	
34	0.9600	1.0800	0	0	
35	1.0800	1.2000	0	0	
36	1.2000	1.3200	0	0	
37	1.3200	1.4400	0	0	
38	1.4400	1.5600	0	0	
39	1.5600	1.6800	0	0	
40	1.6800	1.8000	0	0	
41	1.8000	1.9200	0	0	
42	1.9200	2.0400	0	0	
43	2.0400	2.1600	0	0	
44	2.1600	2.2800	0	0	
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 2

Sample Distribution of ln R_S 24 hr. Averages

Figure 6.2.4b

Note that predicted sulfur values were used in the numerator and denominator. It should be remembered that this analysis seeks to identify the propagation of error through the air quality forecasting. The structure of the denominator with S_p instead of S_o is needed in Equation (6.2.5) to ensure that cancellation among the components of R leads to the correct overall definition of R. For analysis of plant load errors, we must allow only the load parameter to vary; thus S_p must be used in the numerator also.

The narrow distribution of ln R_L (Figure 6.2.5) about a mean of 0 the constancy of the load, a result of the base load operation of the plants. The scattered points above the median are due to full or partial forced outages where the predicted load in the unit commitment exceeded the actual generation. Below the median are points where the plant generation exceeded expectations. These instances occurred during return from outages and also during periods when unexpected demand made it economically attractive to generate more power than had been scheduled.

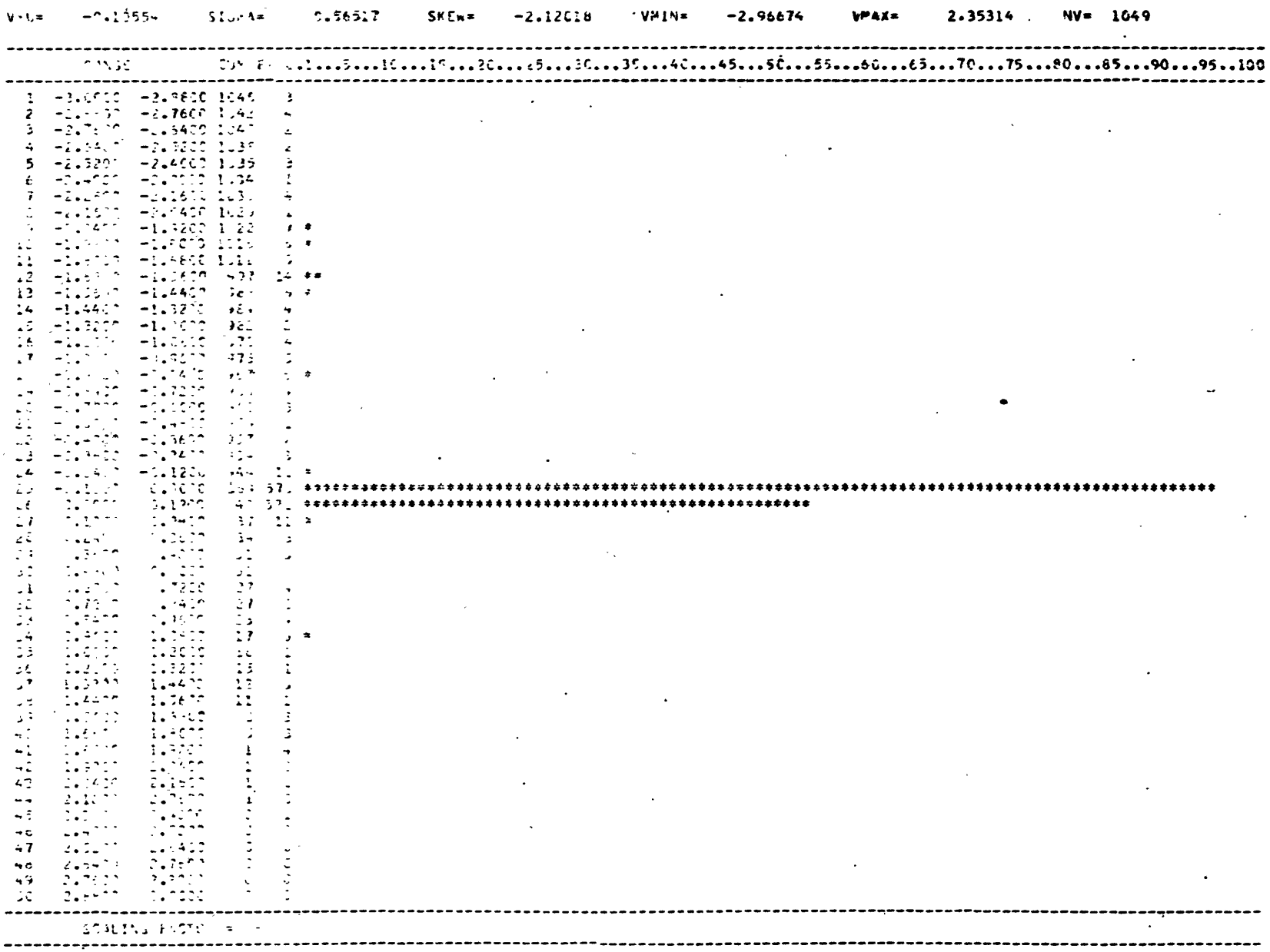
Load errors are not significant to an SCS except for these infrequent cases of over- and underprediction by the unit commitment. These are, in a practical system, easily handled because the system dispatchers have continuous information on each plant's status. If a divergence from the unit commitment schedule is observed, the SCS strategy can be updated with corrected data, effectively reducing load error to zero before a control decision is implemented.

6.2.4 Weather Forecasting Errors

Weather forecasting was done on a regular basis, starting in April 1975, as if the SCS were operational. Wind speed class, wind direction sector, mixing height, and stability class were predicted as 3 hr averages for ten periods or a total of 30 hours. Forecasts occurred at approximately noon (the 1200 forecast) and midnight (the 2400 forecast).

A traditional approach to weather-forecasting errors involves analysis of the observed and predicted parameter values. In keeping with our attempts to establish a statistical basis for SCS decisions, a possible approach is to examine the distribution of forecast parameter differences (i.e., forecasted-observed values). The sample distributions for the forecast differences for the entire 5-month study period, using the forecasts at 2400, tend toward normal distributions (Figure 6.2.6). The forecasts at 1200 show behavior similar to what would be expected from the Law of Large Numbers. Note the discrete valued nature of the distributions arising from the practice of forecasting classes and sectors.

While this type of error analysis may be helpful in educating forecasters about forecast weaknesses, it is difficult to extract SCS reliability information from such distributions, for two reasons. First, the distributions are not independent. There are direct and lagged correlations between parameters such as stability class and wind speed class, or mixing height and stability class. Second, the forecast parameters are merely inputs to a nonlinear system (the air quality model) which produces an output of greater interest — the spatial distribution of concentrations. It is the statistics of the output (or rather of a subset of the output — concentrations which are "high") which determine SCS reliability. The output statistics depend on the input parameters (and their correlations) in a highly complex manner.



Sample Distribution of ln R_L 3 hr. Averages

Figure 6.2.5a

VMU= -0.07704 SIGMA= 0.35809 SKEW= -3.99431 VMIN= -2.30970 VMAX= 0.69315 NV= 133

	RANGE		CUM FREQ.	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	133	0
2	-2.8800	-2.7600	133	0
3	-2.7600	-2.6400	133	0
4	-2.6400	-2.5200	133	0
5	-2.5200	-2.4000	133	0
6	-2.4000	-2.2800	132	1 *
7	-2.2800	-2.1600	132	0
8	-2.1600	-2.0400	132	0
9	-2.0400	-1.9200	132	0
10	-1.9200	-1.8000	132	0
11	-1.8000	-1.6800	130	2 **
12	-1.6800	-1.5600	130	0
13	-1.5600	-1.4400	130	0
14	-1.4400	-1.3200	129	1 *
15	-1.3200	-1.2000	129	0
16	-1.2000	-1.0800	128	1 *
17	-1.0800	-0.9600	127	1 *
18	-0.9600	-0.8400	127	0
19	-0.8400	-0.7200	127	0
20	-0.7200	-0.6000	125	2 **
21	-0.6000	-0.4800	124	1 *
22	-0.4800	-0.3600	124	0
23	-0.3600	-0.2400	123	1 *
24	-0.2400	-0.1200	121	2 **
25	-0.1200	0.0000	43	78 **********
26	0.0000	0.1200	5	38 **********
27	0.1200	0.2400	3	2 **
28	0.2400	0.3600	2	1 *
29	0.3600	0.4800	1	1 *
30	0.4800	0.6000	1	0
31	0.6000	0.7200	0	1 *
32	0.7200	0.8400	0	0
33	0.8400	0.9600	0	0
34	0.9600	1.0800	0	0
35	1.0800	1.2000	0	0
36	1.2000	1.3200	0	0
37	1.3200	1.4400	0	0
38	1.4400	1.5600	0	0
39	1.5600	1.6800	0	0
40	1.6800	1.8000	0	0
41	1.8000	1.9200	0	0
42	1.9200	2.0400	0	0
43	2.0400	2.1600	0	0
44	2.1600	2.2800	0	0
45	2.2800	2.4000	0	0
46	2.4000	2.5200	0	0
47	2.5200	2.6400	0	0
48	2.6400	2.7600	0	0
49	2.7600	2.8800	0	0
50	2.8800	3.0000	0	0

SCALING FACTOR = 1

Sample Distribution of ln R_L 24 hr. Averages

Figure 6.2.5b

MEAN= 14.35593 SIGMA= 59.67953 SKEW= 3.05001 VMIN= -180.00000 VMAX= 180.00000 NV= 967

	RANGE		CUM FREQ		0...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-180.0000	-172.9412	967	7	**
2	-172.9412	-165.8824	960	0	
3	-165.8824	-158.8235	960	0	
4	-158.8235	-151.7647	956	4	*
5	-151.7647	-144.7059	955	3	
6	-144.7059	-137.6471	952	0	
7	-137.6471	-130.5882	945	7	**
8	-130.5882	-123.5294	945	0	
9	-123.5294	-116.4706	945	0	
10	-116.4706	-109.4118	932	13	****
11	-109.4118	-102.3530	932	0	
12	-102.3530	-95.2941	932	0	
13	-95.2941	-88.2353	922	14	****
14	-88.2353	-81.1765	922	0	
15	-81.1765	-74.1177	922	0	
16	-74.1177	-67.0588	886	36	************
17	-67.0588	-60.0000	886	0	
18	-60.0000	-52.9412	886	0	
19	-52.9412	-45.8824	886	0	
20	-45.8824	-38.8235	816	69	************************
21	-38.8235	-31.7647	816	0	
22	-31.7647	-24.7059	816	0	
23	-24.7059	-17.6471	669	146	**
24	-17.6471	-10.5882	669	0	
25	-10.5882	-3.5294	669	0	
26	-3.5294	3.5294	468	201	**
27	3.5294	10.5882	468	0	
28	10.5882	17.6470	468	0	
29	17.6470	24.7059	303	165	**
30	24.7059	31.7647	303	0	
31	31.7647	38.8235	303	0	
32	38.8235	45.8823	201	102	**********************************
33	45.8823	52.9411	201	0	
34	52.9411	60.0000	201	0	
35	60.0000	67.0588	200	1	
36	67.0588	74.1176	127	73	*************************
37	74.1176	81.1764	127	0	
38	81.1764	88.2353	127	0	
39	88.2353	95.2941	81	46	****************
40	95.2941	102.3529	81	0	
41	102.3529	109.4117	81	0	
42	109.4117	116.4706	51	30	**********
43	116.4706	123.5294	51	0	
44	123.5294	130.5882	51	0	
45	130.5882	137.6470	24	27	*********
46	137.6470	144.7058	24	0	
47	144.7058	151.7647	24	0	
48	151.7647	158.8235	8	16	*****
49	158.8235	165.8823	5	3	
50	165.8823	172.9411	3	2	
51	172.9411	180.0000	0	3	**

SCALING FACTOR = 3

Sample Distribution of Wind Direction Errors (Forecasted-Observed) (Degrees) 3 hr. Averages

2400 Forecast

Figure 6.2.6a

```
MEAN=      0.03455      SIGMA=      3.26348      SKEW=      0.25204      VMIN=    -10.56000      VMAX=     13.10000      NV=  1076

        RANGE          CUM FREQ.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
  1 -10.5300 -10.0667 1076   0
  2 -10.0667  -9.6033 1076   0
  3  -9.6033  -9.1400 1076   1
  4  -9.1400  -8.6767 1075  10 **
  5  -8.6767  -8.2133 1065   0
  6  -8.2133  -7.7500 1065   0
  7  -7.7500  -7.2867 1065   0
  8  -7.2867  -6.8233 1058   7 *
  9  -6.8233  -6.3600 1054   2
 10  -6.3600  -5.8967 1029  25 ******
 11  -5.8967  -5.4333 1029   0
 12  -5.4333  -4.9700 1008  21 *****
 13  -4.9700  -4.5067 1008   0
 14  -4.5067  -4.0433  982  25 ******
 15  -4.0433  -3.5800  956   0
 16  -3.5800  -3.1167  963  17 ****
 17  -3.1167  -2.6533  897  66 ****************
 18  -2.6533  -2.1900  780 117 ******************************
 19  -2.1900  -1.7267  727  53 **************
 20  -1.7267  -1.2633  727   0
 21  -1.2633  -0.8000  727   0
 22  -0.8000  -0.3367  727   0
 23  -0.3367   0.1267  345 382 ****************************************************************************************************
 24   0.1267   0.5900  345   0
  5   0.5900   1.0533  345   0
 26   1.0533   1.5167  345   0
 27   1.5167   1.9800  292  53 **************
 28   1.9800   2.4433  292   0
 29   2.4433   2.9067  125 166 *******************************************
 30   2.9067   3.3700  103  23 *****
 31   3.3700   3.8333  103   0
 32   3.8333   4.2967  103   0
 33   4.2967   4.7600   77  26 ******
 34   4.7600   5.2233   77   0
 35   5.2233   5.6867   48  29 *******
 36   5.6867   6.1500   48   0
 37   6.1500   6.6133   33  15 ***
 38   6.6133   7.0767   33   0
 39   7.0767   7.5400   22  11 **
  0   7.5400   8.0033   22   0
 41   8.0033   8.4667   22   0
 42   8.4667   8.9300   11  11 **
 43   8.9300   9.3933   11   0
 44   9.3933   9.8567    9   2
 45   9.8567  10.3200    8   0
 46  10.3200  10.7833    3   5 *
 47  10.7833  11.2467    3   0
 48  11.2467  11.7100    3   0
 49  11.7100  12.1733    2   1
 50  12.1733  12.6367    2   0
 51  12.6367  13.1000    0   2

        SCALING FACTOR =  4
```

Sample Distribution of Wind Speed Errors (Forecasted-Observed) (mph) 3 hr. Averages

2400 Forecast

Figure 6.2.6b

```
VAL=   -52.83752     SIGMA=   433.16137     SKEW=    -0.91176     VMIN= -1700.00000     VMAX=  1300.00000     NV=  1105
------------------------------------------------------------------------------------------------------------------------------------
        RANGE           CUM FREQ.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
------------------------------------------------------------------------------------------------------------------------------------
 1******************* 1103   2 *
 2******************* 1102   1 *
 3******************* 1102   0
 4******************* 1097   5 *
 5******************* 1097   0
 6******************* 1090   7 ***
 7******************* 1088   2 *
 8******************* 1088   0
 9******************* 1083   5 **
10******************* 1083   0
11******************* 1078   5 **
12*********-994.1152 1066  12 ******
13-994.1152-935.2915 1066   0
14-935.2915-876.4678 1052  14 *******
15-876.4678-817.6440 1052   0
16-817.6440-758.8203 1035  17 ********
17-758.8203-699.9965 1010  25 ************
18-699.9965-641.1728 1010   0
19-641.1728-582.3491  977  33 ****************
20-582.3491-523.5254  977   0
21-523.5254-464.7017  948  29 **************
22-464.7017-405.8779  948   0
23-405.8779-347.0542  898  50 *************************
24-347.0542-288.2305  834  64 ********************************
25-288.2305-229.4068  834   0
26-229.4068-170.5834  755  79 ***************************************
27-170.5834-111.7599  755   0
28-111.7599 -52.9364  632 123 *************************************************************
29 -52.9364   5.8871  465 167 ***********************************************************************************
30   5.8871  64.7106  465   0
31  64.7106 123.5341  307 158 *****************************************************************************
32 123.5341 182.3577  307   0
33 182.3577 241.1812  202 105 ****************************************************
34 241.1812 300.0046  116  86 *******************************************
35 300.0046 358.8281  116   0
36 358.8281 417.6516   71  45 **********************
37 417.6516 476.4751   71   0
38 476.4751 535.2986   42  29 **************
39 535.2986 594.1221   42   0
40 594.1221 652.9456   19  23 ***********
41 652.9456 711.7690   11   8 ****
42 711.7690 770.5925   11   0
43 770.5925 829.4160    4   7 ***
44 829.4160 888.2395    4   0
45 888.2395 947.0630    4   0
46 947.06301005.8865    3   1
471005.88651064.7100    3   0
481064.71001123.5334    1   2 *
491123.53341182.3569    1   0
501182.35691241.1804    1   0
511241.18041300.0000    0   1
------------------------------------------------------------------------------------------------------------------------------------
        SCALING FACTOR =  2
```

Sample Distribution of Mixing Depth Errors (Forecasted-Observed) (Meters) 3 hr. Averages

2400 Forecast

Figure 6.2.6c

MEAN= -0.03730 SIGMA= 0.62824 SKEW= -0.31591 XMIN= -3.00000 XMAX= 2.00000 NP= 1126

	RANGE		SUM	FREQ	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-3.0000	-2.6154	1125	1	
2	-2.6154	-2.2308	1125	0	
3	-2.2308	-1.8462	1107	18	**
4	-1.8462	-1.4615	1107	0	
5	-1.4615	-1.0769	1107	0	
6	-1.0769	-0.6923	929	178	**********
7	-0.6923	-0.3077	929	0	
8	-0.3077	0.0769	169	760	**********
9	0.0769	0.4615	169	0	
10	0.4615	0.8462	169	0	
11	0.8462	1.2308	6	163	**********
12	1.2308	1.6154	5	1	
13	1.6154	2.0000	0	5	

SCALING FACTOR = 8

Sample Distribution of Stability Class Errors (Forecasted-Observed) 3 hr. Averages

2400 Forecast

Figure 6.2.6d

From the simple analysis of forecast parameter errors, it was found that input parameter errors were potentially serious in that standard deviations were usually greater than mean values and sufficient to change the parameter class. This was particularly important in the case of wind sectors and stability classes where a change of sector or class meant an abrupt change in the resulting concentration field due to discontinuities in the model.

Rather than pursue the effects of these forecast errors directly, the analysis of weather forecasting errors shifted to the error ratios of the resulting concentration field. The following ratio was used:

$$R_W = \frac{^2C_p(S_p, L_p, W_p)}{^2C_p(S_p, L_p, W_o)} \qquad (6.11)$$

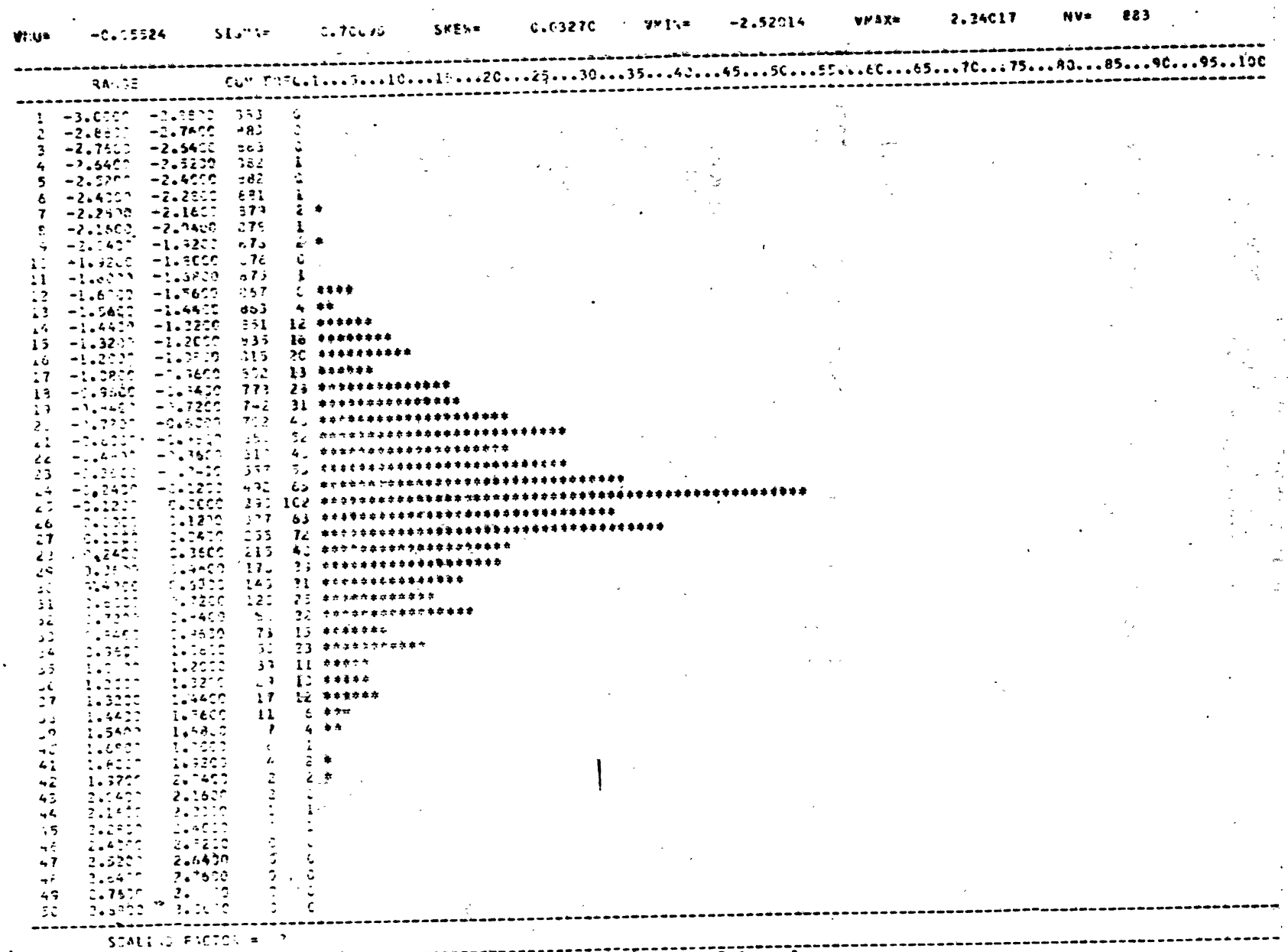
As discussed in Section 6.2.2, the denominator must assume S_p and L_p in order that the correct definition of R result after cancellation among its component ratios. S_p and L_p must be used in the numerator then to isolate the effects of W_p versus W_o.

This error ratio evaluates forecasting accuracy according to the manner in which the set of all meteorological parameters interacts with the air quality concentration model. The effects of, say, wind speed forecast errors alone are not quantified. The distributions of ln R_W (Figure 6.2.7) approach normal distributions, i.e., the errors themselves are log-normally distributed. This simplifies interpretation of the σ values of the distribution. We can see that in both the 3 hr and 24 hr averaging periods there is a nonconservative mean of under 10% (i.e., for the 24 hr geometric mean of -0.098 the linear ratio is $e^{-0.098}$ = .906, and a wide variance compared to the fuel sulfur content and load error ratio sample distributions. Meteorological forecasting errors alone in the 3 hr averaging period will cause error ratios between 0.231 and 3.84, 95% of the time (2 σ levels). For 24 hr averaging periods, the 2 σ levels of the error ratios are .298 and 2.80.

Analyzing individual parameter forecasting errors and their propagation through the concentration equation could be accomplished by repeated application of the analysis techniques used here. Each application would isolate one parameter to decompose R_W into

$$R_W = R_{\substack{\text{wind}\\\text{speed}}} \cdot R_{\substack{\text{wind}\\\text{direction}}} \cdot R_{\text{stability}} \cdot R_{\substack{\text{mixing}\\\text{height}}} \cdot \qquad (6.12)$$

Since the demonstration period was concluded this line of analysis was not pursued.



Sample Distribution of ln R_W 3 hr. Averages

Figure 6.2.7a

VMU= -0.09828 SIGMA= 0.55627 SKEW= -0.06088 VMIN= -1.59423 VMAX= 1.96478 NV= 119

	RANGE		CUM FREQ.		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	119	0	
2	-2.8800	-2.7600	119	0	
3	-2.7600	-2.6400	119	0	
4	-2.6400	-2.5200	119	0	
5	-2.5200	-2.4000	119	0	
6	-2.4000	-2.2800	119	0	
7	-2.2800	-2.1600	119	0	
8	-2.1600	-2.0400	119	0	
9	-2.0400	-1.9200	119	0	
10	-1.9200	-1.8000	119	0	
11	-1.8000	-1.6800	119	0	
12	-1.6800	-1.5600	117	2	**
13	-1.5600	-1.4400	116	1	*
14	-1.4400	-1.3200	115	1	*
15	-1.3200	-1.2000	114	1	*
16	-1.2000	-1.0800	114	0	
17	-1.0800	-0.9600	110	4	****
18	-0.9600	-0.8400	108	2	**
19	-0.8400	-0.7200	106	2	**
20	-0.7200	-0.6000	103	3	***
21	-0.6000	-0.4800	95	8	********
22	-0.4800	-0.3600	86	9	*********
23	-0.3600	-0.2400	75	11	**********
24	-0.2400	-0.1200	59	16	**********
25	-0.1200	0.0000	50	9	*********
26	0.0000	0.1200	45	5	*****
27	0.1200	0.2400	33	12	**********
28	0.2400	0.3600	23	10	**********
29	0.3600	0.4800	14	9	*********
30	0.4800	0.6000	9	5	*****
31	0.6000	0.7200	5	4	****
32	0.7200	0.8400	3	2	**
33	0.8400	0.9600	2	1	*
34	0.9600	1.0800	2	0	
35	1.0800	1.2000	1	1	*
36	1.2000	1.3200	1	0	
37	1.3200	1.4400	1	0	
38	1.4400	1.5600	1	0	
39	1.5600	1.6800	1	0	
40	1.6800	1.8000	1	0	
41	1.8000	1.9200	1	0	
42	1.9200	2.0400	0	1	*
43	2.0400	2.1600	0	0	
44	2.1600	2.2800	0	0	
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of ln R_W 24 hr. Averages

Figure 6.2.7b

6.2.5 Model Errors

Even given perfect knowledge of all input parameters, a concentration model will still contain errors related to incorrect or incomplete model structure. For instance, errors are introduced by the use of wind speed classes instead of continuous wind speeds. In this analysis errors are also introduced by ignoring six of the eight power plant sources and considering only ${}^{2}D(S_o, L_o, W_o)$.

$$R_M = \frac{{}^{2}C(S_o, L_o, W_o)}{C_o} = {}^{2}R_M \qquad (6.13)$$

The distributions for R_M were given in Figure 6.2.2 above. That distribution is repeated in Figure 6.2.8. Recall that even a perfect concentration model would be conservatively biased because of the small probability that the maximum concentration will coincide with a monitor to produce C_o = true maximum. We see then that for both averaging periods we have a conservative model. It is certainly not our elusive perfect model since both distributions, which are approximately normal, have about 15% of their values less than zero, showing underprediction. The $\pm 2\ \sigma$ levels are even broader than the weather-forecasting error ratio, indicating that modifications directed to reducing σ for the modeling errors is the best course of overall model improvement. The smaller geometric mean and that for the 24 hr averaging period are reflections of the smoothing effects of longer averaging times.

6.2.6 Total Error Ratio R

The component error ratio distributions are not independent of each other. In each time period the individual ratios can contribute similar or opposite effects to the total error ratio, so our final analysis looks at their composite behavior:

$$R = \frac{{}^{2}C_p(S_p, L_p, W_p)}{C_o} \qquad (6.14)$$

R is the error ratio of the predicted maximum receptor concentration to the observed maximum monitor concentration. All inputs to the model are predicted values in each time period. The 3 hr average and 24 hr averaged ln R distributions are shown in Figure 6.2.9.

Our overall SCS forecasting system is conservative compared to the available monitor concentrations, as shown by the non-negative geometric means of 0.9 (3 hr) and 0.5 (24 hr). This implies that more conservative control would be applied using the predicted concentrations than if control were implemented only when monitors registered rising values, providing thresholds were the same. The overall system is also not very consistent in its conservatism, since its σ value is as large as the mean for both averaging periods. The distributions appear

VMU= 1.00556 SIGMA= 0.86321 SKEW= -0.26122 VMIN= -1.85895 VMAX= 2.99573 NV= 1039

	RANGE		CUM FREQ		.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.9800	1039	0	
2	-2.9800	-2.9600	1039	0	
3	-2.9600	-2.9400	1039	0	
4	-2.9400	-2.9200	1039	0	
5	-2.9200	-2.9000	1039	0	
6	-2.9000	-2.8800	1039	0	
7	-2.8800	-2.8600	1039	0	
8	-2.8600	-2.8400	1039	0	
9	-2.8400	-1.9200	1039	0	
10	-1.9200	-1.9000	1038	1	*
11	-1.9000	-1.8800	1037	1	*
12	-1.8800	-1.8600	1036	1	*
13	-1.8600	-1.8400	1036	0	
14	-1.8400	-1.8200	1035	1	*
15	-1.8200	-1.8000	1033	2	**
16	-1.8000	-1.7800	1030	3	***
17	-1.7800	-0.9600	1021	9	*********
18	-0.9600	-0.8400	1018	3	***
19	-0.8400	-0.7200	1010	8	********
20	-0.7200	-0.6000	996	14	**************
21	-0.6000	-0.4800	986	10	**********
22	-0.4800	-0.3600	965	21	*********************
23	-0.3600	-0.2400	950	15	***************
24	-0.2400	-0.1200	931	19	*******************
25	-0.1200	0.0000	904	27	***************************
26	0.0000	0.1200	882	22	**********************
27	0.1200	0.2400	849	33	*********************************
28	0.2400	0.3600	796	53	***
29	0.3600	0.4800	747	49	***
30	0.4800	0.6000	706	41	***
31	0.6000	0.7200	668	38	**************************************
32	0.7200	0.8400	621	47	**
33	0.8400	0.9600	569	52	***
34	0.9600	1.0800	508	61	**
35	1.0800	1.2000	447	61	**
36	1.2000	1.3200	389	58	**
37	1.3200	1.4400	337	52	***
38	1.4400	1.5600	284	53	***
39	1.5600	1.6800	240	44	**
40	1.6800	1.8000	197	43	***
41	1.8000	1.9200	154	43	***
42	1.9200	2.0400	115	39	***************************************
43	2.0400	2.1600	97	18	******************
44	2.1600	2.2800	69	28	****************************
45	2.2800	2.4000	50	19	*******************
46	2.4000	2.5200	29	21	*********************
47	2.5200	2.6400	20	9	*********
48	2.6400	2.7600	9	11	***********
49	2.7600	2.8800	3	6	******
50	2.8800	3.0000	0	3	***

SCALING FACTOR = 1

Sample Distribution of ln R_M 3 hr. Averages

Figure 6.2.8a

VMU= 0.59553 SIGMA= 0.57894 SKEW= -0.05061 VMIN= -1.11803 VMAX= 2.25324 NV= 134

	RANGE		CUM FREQ		.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	134	0	
2	-2.8800	-2.7600	134	0	
3	-2.7600	-2.6400	134	0	
4	-2.6400	-2.5200	134	0	
5	-2.5200	-2.4000	134	0	
6	-2.4000	-2.2800	134	0	
7	-2.2800	-2.1600	134	0	
8	-2.1600	-2.0400	134	0	
9	-2.0400	-1.9200	134	0	
10	-1.9200	-1.8000	134	0	
11	-1.8000	-1.6800	134	0	
12	-1.6800	-1.5600	134	0	
13	-1.5600	-1.4400	134	0	
14	-1.4400	-1.3200	134	0	
15	-1.3200	-1.2000	134	0	
16	-1.2000	-1.0800	133	1	*
17	-1.0800	-0.9600	133	0	
18	-0.9600	-0.8400	132	1	*
19	-0.8400	-0.7200	132	0	
20	-0.7200	-0.6000	132	0	
21	-0.6000	-0.4800	129	3	***
22	-0.4800	-0.3600	128	1	*
23	-0.3600	-0.2400	123	5	*****
24	-0.2400	-0.1200	119	4	****
25	-0.1200	0.0000	114	5	*****
26	0.0000	0.1200	107	7	*******
27	0.1200	0.2400	100	7	*******
28	0.2400	0.3600	91	9	*********
29	0.3600	0.4800	83	8	********
30	0.4800	0.6000	72	11	***********
31	0.6000	0.7200	57	15	***************
32	0.7200	0.8400	43	14	**************
33	0.8400	0.9600	30	13	*************
34	0.9600	1.0800	25	5	*****
35	1.0800	1.2000	18	7	*******
36	1.2000	1.3200	13	5	*****
37	1.3200	1.4400	10	3	***
38	1.4400	1.5600	8	2	**
39	1.5600	1.6800	3	5	*****
40	1.6800	1.8000	2	1	*
41	1.8000	1.9200	2	0	
42	1.9200	2.0400	2	0	
43	2.0400	2.1600	1	1	*
44	2.1600	2.2800	0	1	*
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of ln R_M 24 hr. Averages

Figure 6.2.8b

VMU= 0.89934 SIGMA= 0.90249 SKEW= -0.45705 VMIN= -2.43612 VMAX= 2.98904 NV= 982

	RANGE		CUM FREQ		.1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	982	0	
2	-2.8800	-2.7600	982	0	
3	-2.7600	-2.6400	982	0	
4	-2.6400	-2.5200	982	0	
5	-2.5200	-2.4000	981	1	*
6	-2.4000	-2.2800	981	0	
7	-2.2800	-2.1600	981	0	
8	-2.1600	-2.0400	980	1	*
9	-2.0400	-1.9200	978	2	**
10	-1.9200	-1.8000	977	1	*
11	-1.8000	-1.6800	976	1	*
12	-1.6800	-1.5600	973	3	***
13	-1.5600	-1.4400	969	4	****
14	-1.4400	-1.3200	968	1	*
15	-1.3200	-1.2000	961	7	*******
16	-1.2000	-1.0800	953	8	********
17	-1.0800	-0.9600	946	7	*******
18	-0.9600	-0.8400	940	6	******
19	-0.8400	-0.7200	929	11	**********
20	-0.7200	-0.6000	922	7	*******
21	-0.6000	-0.4800	917	5	*****
22	-0.4800	-0.3600	902	15	***************
23	-0.3600	-0.2400	888	14	**************
24	-0.2400	-0.1200	847	41	***
25	-0.1200	0.0000	822	25	*************************
26	0.0000	0.1200	795	27	***************************
27	0.1200	0.2400	768	27	***************************
28	0.2400	0.3600	722	46	**
29	0.3600	0.4800	690	32	********************************
30	0.4800	0.6000	652	38	**************************************
31	0.6000	0.7200	603	49	**
32	0.7200	0.8400	548	55	***
33	0.8400	0.9600	500	48	**
34	0.9600	1.0800	451	49	**
35	1.0800	1.2000	405	46	**
36	1.2000	1.3200	342	63	***
37	1.3200	1.4400	285	57	**
38	1.4400	1.5600	225	60	**
39	1.5600	1.6800	175	50	**
40	1.6800	1.8000	143	32	********************************
41	1.8000	1.9200	110	33	*********************************
42	1.9200	2.0400	89	21	*********************
43	2.0400	2.1600	66	23	************************
44	2.1600	2.2800	49	17	*****************
45	2.2800	2.4000	33	16	****************
46	2.4000	2.5200	24	9	*********
47	2.5200	2.6400	18	6	******
48	2.6400	2.7600	10	8	********
49	2.7600	2.8800	4	6	******
50	2.8800	3.0000	0	4	****

SCALING FACTOR = 1

Sample Distribution of ln R 3 hr. Averages

Figure 6.2.9a

VMU= 0.49885 SIGMA= 0.61450 SKEW= 0.00183 VMIN= -0.95673 VMAX= 2.17617 NV= 132

	RANGE		CUM FREQ.		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	132	0	
2	-2.8800	-2.7600	132	0	
3	-2.7600	-2.6400	132	0	
4	-2.6400	-2.5200	132	0	
5	-2.5200	-2.4000	132	0	
6	-2.4000	-2.2800	132	0	
7	-2.2800	-2.1600	132	0	
8	-2.1600	-2.0400	132	0	
9	-2.0400	-1.9200	132	0	
10	-1.9200	-1.8000	132	0	
11	-1.8000	-1.6800	132	0	
12	-1.6800	-1.5600	132	0	
13	-1.5600	-1.4400	132	0	
14	-1.4400	-1.3200	132	0	
15	-1.3200	-1.2000	132	0	
16	-1.2000	-1.0800	132	0	
17	-1.0800	-0.9600	132	0	
18	-0.9600	-0.8400	131	1	*
19	-0.8400	-0.7200	129	2	**
20	-0.7200	-0.6000	125	4	****
21	-0.6000	-0.4800	121	4	****
22	-0.4800	-0.3600	118	3	***
23	-0.3600	-0.2400	115	3	***
24	-0.2400	-0.1200	110	5	*****
25	-0.1200	0.0000	107	3	***
26	0.0000	0.1200	101	6	******
27	0.1200	0.2400	91	10	**********
28	0.2400	0.3600	79	12	************
29	0.3600	0.4800	68	11	**********
30	0.4800	0.6000	60	8	********
31	0.6000	0.7200	48	12	************
32	0.7200	0.8400	37	11	**********
33	0.8400	0.9600	28	9	*********
34	0.9600	1.0800	22	6	******
35	1.0800	1.2000	15	7	*******
36	1.2000	1.3200	11	4	****
37	1.3200	1.4400	9	2	**
38	1.4400	1.5600	3	6	******
39	1.5600	1.6800	3	0	
40	1.6800	1.8000	3	0	
41	1.8000	1.9200	2	1	*
42	1.9200	2.0400	2	0	
43	2.0400	2.1600	1	1	*
44	2.1600	2.2800	0	1	*
45	2.2800	2.4000	0	0	
46	2.4000	2.5200	0	0	
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of ln R 24 hr. Averages

Figure 6.2.9b

to be approaching normal distributions; the implied log-normality of predicted concentrations agrees with the analysis of observations at monitors (Appendix C), and affirms an important underlying assumption of the probabilistic control model, i.e., that predicted and observed concentrations are log-normally distributed.

6.3 Results - Model Modifications

The modifications described in Section 5 were made to improve the model performance by reflecting the varying terrain conditions with which the plumes interacted. The skill score system (Section 5 - Appendix E) for evaluating the modifications was based on a 1 hr average data set which included only periods with likely plume-terrain interactions. That restricted evaluation gave no indication of performance in the more general forecasting situation at Chestnut Ridge. An error ratio analysis was performed on the modified model but was not able to establish any clear advantages or disadvantages associated with using the modified model instead of the operating period model for the data sample being used. Since insignificant differences were observed between the modified and operating models in the general application to the study period data, it was concluded that the modified model combines the advantages of the operating period model with the improved terrain performance shown in Section 5. The details of the error ratio analysis will now be given.

To evaluate the general performance of the modified model two sets of error ratios were considered. The first compared the two models directly:

$$R''_M = \frac{{}^8C^*_p(S_o, L_o, W_o)}{{}^8C_p(S_o, L_o, W_o)} \qquad (6.15)$$

where ${}^8C^*_p(S_o, L_o, W_o)$ is the maximum receptor concentration predicted by the modified model with background and all 8 power plant sources. The distribution of $\ln R''_M$ is shown in Figure 6.3.1.

In general there is not a significant difference between the two models. This is encouraging since the modified model attempts to perform better during plume-terrain interactions (which are rare), while maintaining its previous performance at other times.

In about 1% of the cases, the modified model predicts 3 times or greater the concentration of the original model. To investigate whether these were indeed critical high concentrations due to rough terrain, the highest five ratios were examined in closer detail. These were found to be cases with stable conditions, wind directions transporting pollutants onto the ridges, and moderate to low wind speed. These are indeed just the situations in which high concentration predictions were expected from the modified model.

VMU= 0.02351 SIGMA= 1.2217-7 SKEW= 1.70153 VMIN= -0.82832 VMAX= 2.54375 NV= 858

	RANGE		SUM	FREQ...1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.9800	550	1
2	-2.9600	-2.7600	650	0
3	-2.7600	-2.6400	750	1
4	-2.6400	-2.5200	850	1
5	-2.5200	-2.4000	850	1
6	-2.4000	-2.2800	950	1
7	-2.2800	-2.1600	950	1
8	-2.1600	-2.0400	950	1
9	-2.0400	-1.9200	950	1
10	-1.9200	-1.8000	950	1
11	-1.8000	-1.6800	950	1
12	-1.6800	-1.5600	950	1
13	-1.5600	-1.4400	950	1
14	-1.4400	-1.3200	950	1
15	-1.3200	-1.2000	950	1
16	-1.2000	-1.0800	950	1
17	-1.0800	-0.9600	950	1
18	-0.9600	-0.8400	950	1
19	-0.8400	-0.7200	771	7 *
20	-0.7200	-0.6000	124	27 ****
21	-0.6000	-0.4800	450	12 **
22	-0.4800	-0.3600	295	35 *****
23	-0.3600	-0.2400	709	34 ******
24	-0.2400	-0.1200	720	11 *
25	-0.1200	0.0000	219	503 **********
26	0.0000	0.1200	173	43 *******
27	0.1200	0.2400	123	51 ********
28	0.2400	0.3600	91	25 ****
29	0.3600	0.4800	51	43 *******
30	0.4800	0.6000	34	22 ***
31	0.6000	0.7200	21	13 **
32	0.7200	0.8400	14	7 *
33	0.8400	0.9600	7	1
34	0.9600	1.0800	6	1
35	1.0800	1.2000	5	1
36	1.2000	1.3200	7	1
37	1.3200	1.4400	6	1
38	1.4400	1.5600	5	1
39	1.5600	1.6800	3	1
40	1.6800	1.8000	2	1
41	1.8000	1.9200	1	1
42	1.9200	2.0400	1	1
43	2.0400	2.1600	2	1
44	2.1600	2.2800	2	1
45	2.2800	2.4000	2	1
46	2.4000	2.5200	2	1
47	2.5200	2.6400	1	1
48	2.6400	2.7600	1	1
49	2.7600	2.8800	1	1
50	2.8800	3.0000	1	1

SCALING FACTOR =

Sample Distribution of ln R''_M 3 hr. Averages

Figure 6.3.1a

VMU= -0.00668 SIGMA= 0.20472 SKEW= 0.34300 VMIN= -0.62012 VMAX= 0.80114 NV= 109

	RANGE		CUM FREQ.	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	109	0
2	-2.8800	-2.7600	109	0
3	-2.7600	-2.6400	109	0
4	-2.6400	-2.5200	109	0
5	-2.5200	-2.4000	109	0
6	-2.4000	-2.2800	109	0
7	-2.2800	-2.1600	109	0
8	-2.1600	-2.0400	109	0
9	-2.0400	-1.9200	109	0
10	-1.9200	-1.8000	109	0
11	-1.8000	-1.6800	109	0
12	-1.6800	-1.5600	109	0
13	-1.5600	-1.4400	109	0
14	-1.4400	-1.3200	109	0
15	-1.3200	-1.2000	109	0
16	-1.2000	-1.0800	109	0
17	-1.0800	-0.9600	109	0
18	-0.9600	-0.8400	109	0
19	-0.8400	-0.7200	109	0
20	-0.7200	-0.6000	108	1 *
21	-0.6000	-0.4800	107	1 *
22	-0.4800	-0.3600	102	5 *****
23	-0.3600	-0.2400	96	6 ******
24	-0.2400	-0.1200	89	7 *******
25	-0.1200	0.0000	25	64 **********
26	0.0000	0.1200	16	9 *********
27	0.1200	0.2400	10	6 ******
28	0.2400	0.3600	5	5 *****
29	0.3600	0.4800	2	3 ***
30	0.4800	0.6000	2	0
31	0.6000	0.7200	1	1 *
32	0.7200	0.8400	0	1 *
33	0.8400	0.9600	0	0
34	0.9600	1.0800	0	0
35	1.0800	1.2000	0	0
36	1.2000	1.3200	0	0
37	1.3200	1.4400	0	0
38	1.4400	1.5600	0	0
39	1.5600	1.6800	0	0
40	1.6800	1.8000	0	0
41	1.8000	1.9200	0	0
42	1.9200	2.0400	0	0
43	2.0400	2.1600	0	0
44	2.1600	2.2800	0	0
45	2.2800	2.4000	0	0
46	2.4000	2.5200	0	0
47	2.5200	2.6400	0	0
48	2.6400	2.7600	0	0
49	2.7600	2.8800	0	0
50	2.8800	3.0000	0	0

SCALING FACTOR = 1

Sample Distribution of ln R''_M 24 hr. Averages

Figure 6.3.1b

A second comparison was made using the observed monitor data:

$$ {}^{8}R^{*}_{M} = \frac{{}^{8}C^{*}_{p}(S_o, L_o, W_o)}{C_o} . \qquad (6.16) $$

The distribution of ln ${}^{8}R^{*}_{M}$ (Figure 6.3.2) should be compared with Figure 6.1 (distribution of ln ${}^{8}R_{M}$). Again, very little noticeable difference is present in terms of the error ratio distributions. The value of σ decreased in the modified model, but only from 0.787 to 0.746. The modified model is a somewhat more conservative mean value of (1.120 vs 1.102) but neither of these changes appears significant.

6.4 Conclusions

The following summarizes the results of the error ratio analysis of the 5-month study period data:

1) Six of the eight sources in the Chestnut Ridge area did not contribute significantly to the analysis of concentrations levels of interest for SCS during the study period. When the full eight-source air quality model was compared to a two-source model (Seward only), 98% of the error ratios comparing the modeled maximum values were between 1.0 and 0.8 except for occasional low concentrations cases of no significant to SCS operation.

2) With the two-source model, the overall SCS error ratio sample distributions are conservative in relation to the monitored data. The geometric means of the distributions correspond to safety factors of 2.5 (3 hr average) and 1.6 (24 hr average) between predicted maxima and observed maxima. An undetermined amount of the conservatism is due to the sparsity of the monitor network which prevents the observation of the true spatial maximum.

3) The error ratio components show that model structure and weather forecasting are the most productive area for efforts at improving overall reliability. Fuel sulfur content and plant load levels offer less opportunity for improvement of reliability (although the study of less aggregated fuel sulfur data might change this conclusion).

4) The modifications made in the operating model of the demonstration period show negligible differences in error ratios comparing the modified model to the operating model and to observed concentrations. This indicates that the modifications of Section 5 did not adversely affect model performance in situations where rough terrain is not a factor.

VMU= 1.12096 SIGMA= 0.74616 SKEW= -0.07698 VMIN= -1.03609 VMAX= 2.94632 NV= 843

	RANGE		CUM FREQ.	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-3.0000	-2.8800	843	0
2	-2.8800	-2.7600	843	0
3	-2.7600	-2.6400	843	0
4	-2.6400	-2.5200	843	0
5	-2.5200	-2.4000	843	0
6	-2.4000	-2.2800	843	0
7	-2.2800	-2.1600	843	0
8	-2.1600	-2.0400	843	0
9	-2.0400	-1.9200	843	0
10	-1.9200	-1.8000	843	0
11	-1.8000	-1.6800	843	0
12	-1.6800	-1.5600	843	0
13	-1.5600	-1.4400	843	0
14	-1.4400	-1.3200	843	0
15	-1.3200	-1.2000	843	0
16	-1.2000	-1.0800	843	0
17	-1.0800	-0.9600	842	1 *
18	-0.9600	-0.8400	840	2 **
19	-0.8400	-0.7200	836	4 ****
20	-0.7200	-0.6000	832	4 ****
21	-0.6000	-0.4800	826	6 ******
22	-0.4800	-0.3600	819	7 *******
23	-0.3600	-0.2400	815	4 ****
24	-0.2400	-0.1200	803	12 ************
25	-0.1200	0.0000	783	20 ********************
26	0.0000	0.1200	762	21 *********************
27	0.1200	0.2400	738	24 ************************
28	0.2400	0.3600	714	24 ************************
29	0.3600	0.4800	677	37 *************************************
30	0.4800	0.6000	639	38 **************************************
31	0.6000	0.7200	589	50 **
32	0.7200	0.8400	541	48 **
33	0.8400	0.9600	503	38 **************************************
34	0.9600	1.0800	444	59 ***
35	1.0800	1.2000	387	57 ***
36	1.2000	1.3200	328	59 ***
37	1.3200	1.4400	282	46 **
38	1.4400	1.5600	241	41 ***
39	1.5600	1.6800	202	39 **************************************
40	1.6800	1.8000	159	43 ***
41	1.8000	1.9200	125	34 **********************************
42	1.9200	2.0400	90	35 ***********************************
43	2.0400	2.1600	74	16 ****************
44	2.1600	2.2800	57	17 *****************
45	2.2800	2.4000	40	17 *****************
46	2.4000	2.5200	25	15 ***************
47	2.5200	2.6400	14	11 ***********
48	2.6400	2.7600	6	8 ********
49	2.7600	2.8800	3	3 ***
50	2.8800	3.0000	0	3 ***

SCALING FACTOR = 1

Sample Distribution of 1a $^{8}R^{*}_{M}$ 3 hr. Average

Figure 6.3.2a

VMU= 0.70294 SIGMA= 0.55541 SKEW= 0.54654 VMIN= -0.49928 VMAX= 2.49281 NV= 109

	RANGE		CUM FREQ.		1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95..100
1	-3.0000	-2.8800	109	0	
2	-2.8800	-2.7600	109	0	
3	-2.7600	-2.6400	109	0	
4	-2.6400	-2.5200	109	0	
5	-2.5200	-2.4000	109	0	
6	-2.4000	-2.2800	109	0	
7	-2.2800	-2.1600	109	0	
8	-2.1600	-2.0400	109	0	
9	-2.0400	-1.9200	109	0	
10	-1.9200	-1.8000	109	0	
11	-1.8000	-1.6800	109	0	
12	-1.6800	-1.5600	109	0	
13	-1.5600	-1.4400	109	0	
14	-1.4400	-1.3200	109	0	
15	-1.3200	-1.2000	109	0	
16	-1.2000	-1.0800	109	0	
17	-1.0800	-0.9600	109	0	
18	-0.9600	-0.8400	109	0	
19	-0.8400	-0.7200	109	0	
20	-0.7200	-0.6000	109	0	
21	-0.6000	-0.4800	108	1	*
22	-0.4800	-0.3600	107	1	*
23	-0.3600	-0.2400	105	2	**
24	-0.2400	-0.1200	102	3	***
25	-0.1200	0.0000	99	3	***
26	0.0000	0.1200	96	3	***
27	0.1200	0.2400	88	8	********
28	0.2400	0.3600	81	7	*******
29	0.3600	0.4800	68	13	*************
30	0.4800	0.6000	61	7	*******
31	0.6000	0.7200	48	13	*************
32	0.7200	0.8400	36	12	************
33	0.8400	0.9600	31	5	*****
34	0.9600	1.0800	22	9	*********
35	1.0800	1.2000	19	3	***
36	1.2000	1.3200	14	5	*****
37	1.3200	1.4400	13	1	*
38	1.4400	1.5600	10	3	***
39	1.5600	1.6800	5	5	*****
40	1.6800	1.8000	4	1	*
41	1.8000	1.9200	3	1	*
42	1.9200	2.0400	2	1	*
43	2.0400	2.1600	2	0	
44	2.1600	2.2800	1	1	*
45	2.2800	2.4000	1	0	
46	2.4000	2.5200	0	1	*
47	2.5200	2.6400	0	0	
48	2.6400	2.7600	0	0	
49	2.7600	2.8800	0	0	
50	2.8800	3.0000	0	0	

SCALING FACTOR = 1

Sample Distribution of $\ln {}^{8}R^{*}_{M}$ 24 hr. Averages

Figure 6.3.2b

7.0 ECONOMIC ANALYSIS

It should immediately be noted that it is difficult to simulate the economics of a situation that contains as many intangibles and human judgments as the operation of an electric power system. It is more difficult to display overall average behavior of such operations, and it is impossible to make these displays perfectly general and accurately transferable to different complex situations. What can be done is a description of a technique which can be used for such analyses and an example of the use of this technique in a particular case.

In Section 7.1 the results of the economic evaluation are presented along with a description of the methodology used for the analysis in this particular situation. The extension of the techniques presented in Section 7.1 to enable transfer of these models to different situations is presented in Section 7.2.

7.1 General Comments and Results

The example represented by the Chestnut Ridge situation is, unfortunately, not the most generalizable type of control situation but it does offer the opportunity for interesting economic analysis. A general rule about the economics of an SCS is that, in most conceivable applications, the costs required to cover uncertainties are between one and two orders of magnitude more than the costs of controlling a system with perfect foresight. Chapter 6 explored the sources and reduction of levels of uncertainty; this section shows the costs implicit in the various levels of uncertainty for the Chestnut Ridge application.

The bottom line as far as the situation that presents itself for economic analysis in the Chestnut Ridge area is the total uncertainty between forecasted and actual ground level concentrations. With the general dominance of the Seward plant in the high concentration episodes, thus eliminating overlaps of various sources, it is a valid simplification to compare maximum observed with maximum predicted concentration at each point in time. Figure 7.1,1 shows the comparison of the 3-hour pollution concentration forecasts from the air quality model using both forecasted and observed meteorological data. Figure 7.1.2 shows the 24-hour comparison. The set of episodes considered were for observations greater than one-half the 24-hr standard and greater than one-third the 3-hour standard. Whether or not it was intentional (and ideally it should be intentional), it can be seen from most of the figures the forecasting was overestimating pollution potential in those situations with high pollution prospects.

Two assumptions were made in developing these scatter diagrams. First, only the midnight, and not the noon, meteorological forecast was used for forecast data. This would tend to show poorer performance in the forecasting than would be the case with noon updated information. However, more than 90% of the episodes that exceeded the study thresholds occurred between 6 a.m. and 6 p.m. thus requiring midnight forecast information for a fuel-switching decision (considering the 6-hour lead time necessary for the decision). A far more important limitation concerning the use of only the midnight forecast data is the

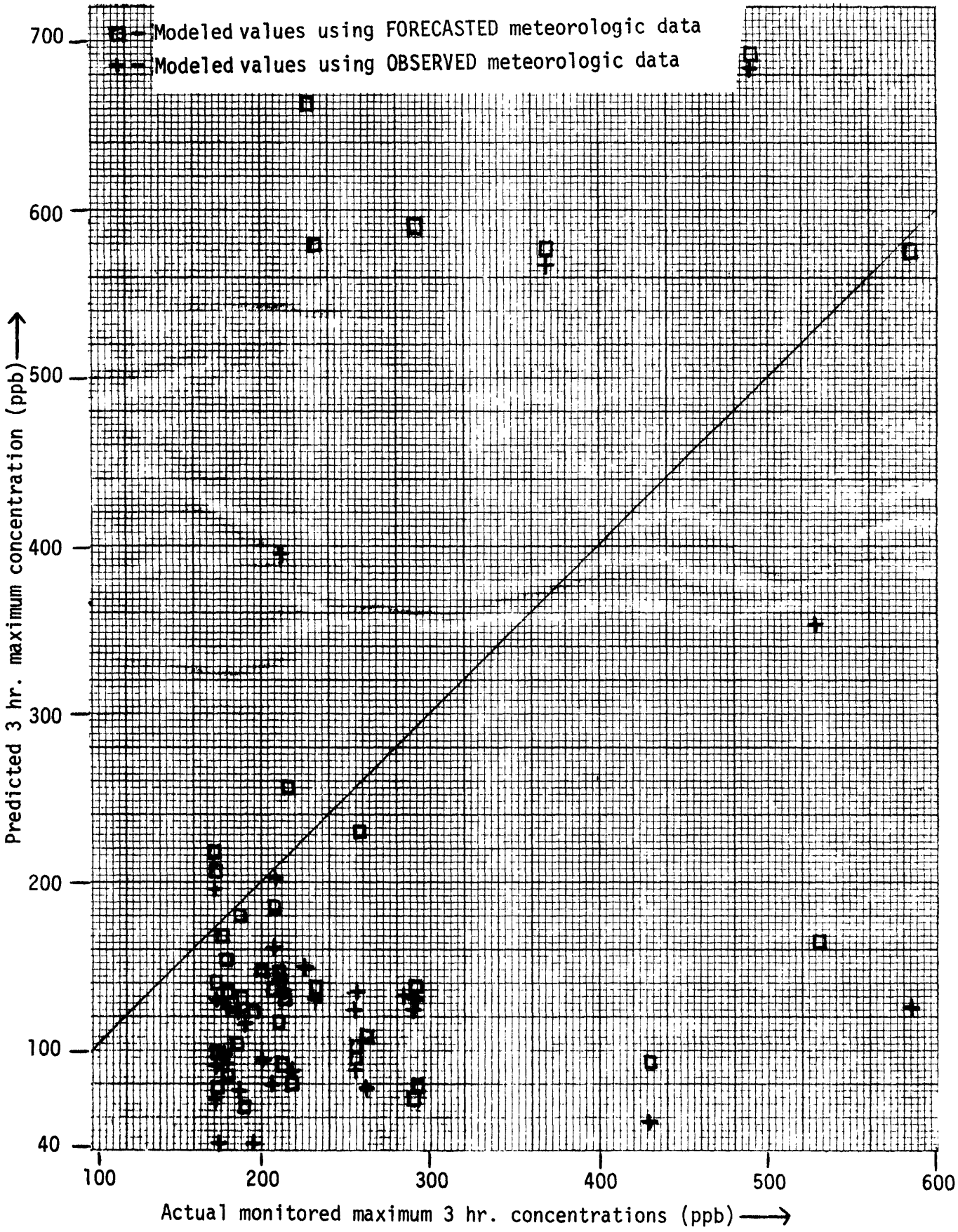


Figure 7.1.1 Comparison of predicted 3 hour pollutant concentrations and monitored levels.

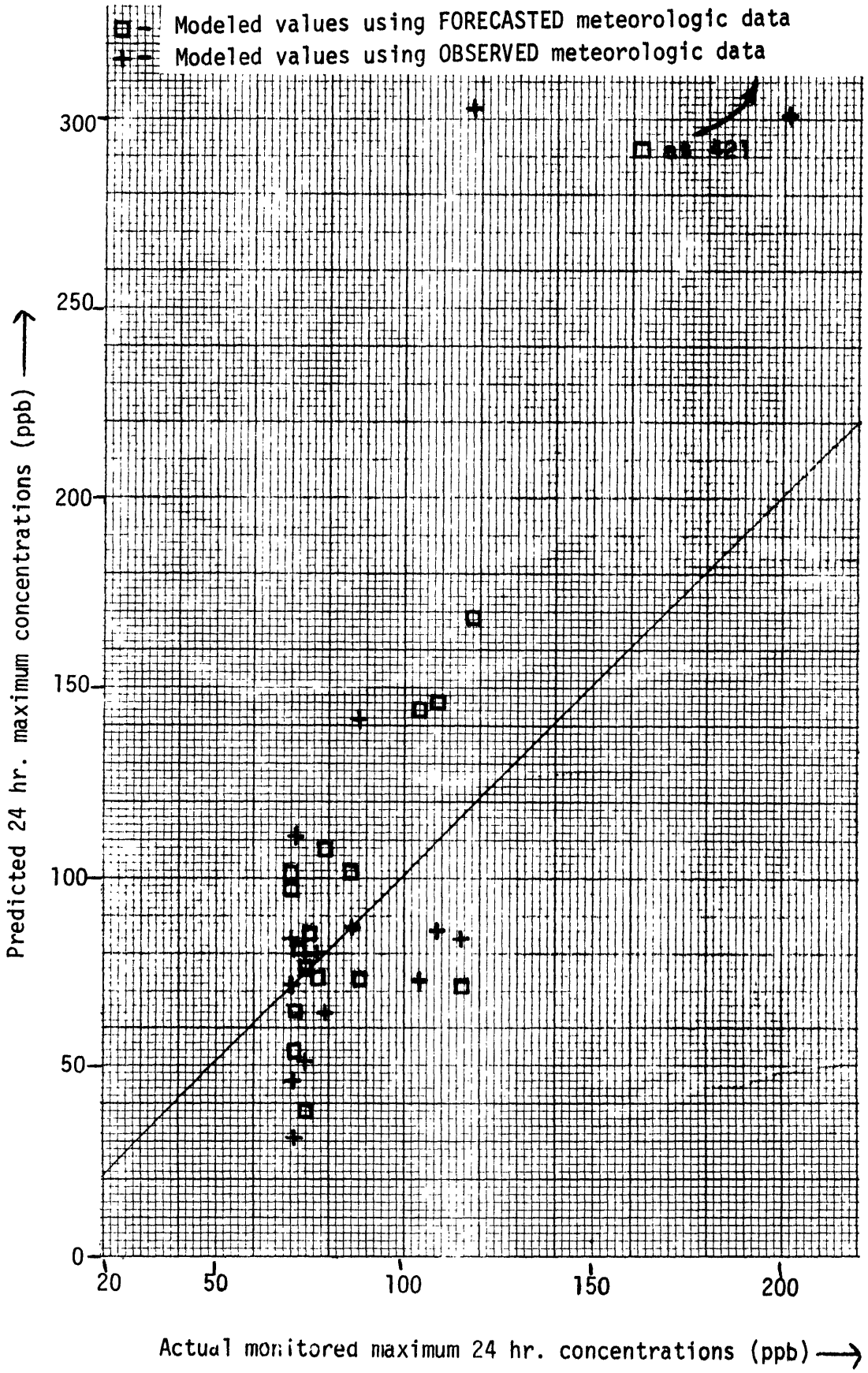


Figure 7.1.2 Comparison of predicted 24 hour pollutant concentrations and monitored levels.

fact that in the actual demonstration period during projected control situations the meteorological forecasters updated their projections every 30 minutes and oftentimes devoted continuous attention to the potential control situation. The use of after-the-fact observed meteorological data in parallel analysis with forecasted data is an effort to reflect the increased accuracy possible with the frequent updates, because with continuous forecaster attention the predicted values closely approached the observed. Of course, this does not accurately reflect the conservatism that could be introduced into the forecasts, but it was virtually impossible due to resource limitations to have forecasters continually committed to this project over large portions of the project period. Also, the direct access to the air quality model in the computer was available for only a short time and in fact the "best" air quality model used in this analysis was not available until after the project had been completed. Fortunately, there was very little variation between economic results from forecasted and observed data.

The second assumption used in developing the scatter diagrams involved the use of actual system operating data for the power plant megawatt levels and the replacement cost of power on the system. Extensive analysis was performed on the effect of using actual versus unit commitment projections of plant operating levels. Figure 7.1.3 shows some typical results. Most projected megawatt levels were within a couple of percentage points of being correct, except in forced outage situations. A forced outage would never cause an otherwise unforeseen pollution episode. There is a chance, however, that the use of observed megawatt levels will result in strategies that will inappropriately count on upcoming forced outages. Only one such simultaneous pollution control and forced outage was observed, with little economic effect. For the baseloaded situation in the Chestnut Ridge area there were very few occasions in which actual power levels were raised significantly higher than forecasted power levels, such as in the case of a plant being brought back from an outage sooner than expected. These are not surprise situations to power system schedulers, who, of course, are aware of the plants coming back on line, so again there is no surprise pollution episode, just a slight economic effect due to the lack of accurate information upon which to formulate the control strategy. Finally, a system lambda, (that is, replacement cost of power) predicter was built for the demonstration phase but not used in this analysis. The Chestnut Ridge plants operate well below system lambda, even with clean fuels, and so, as with the other assumptions, running these simulations using perfect foresight on power system information tends to show predicted costs slightly below what the actual costs would be. Our best guess of the situation is that these very small negative cost effects will counterbalance the very small positive cost effect caused by lack of information about the effect of strategically placed meteorological forecast updating in the control situations.

With the qualifications thus set aside the economic analysis proceeds. Chapter 3 concerning the demonstration period shows the economic results of that period. The objective of this chapter is

(1) to determine what would be typical in terms of economic costs of control in the Chestnut Ridge region, and

RANGE		FREQ.	
-1.2000	-1.1520	3	
-1.1520	-1.1040	0	
-1.1040	-1.0560	0	
-1.0560	-1.0080	0	
-1.0080	-0.9600	33	**
-0.9600	-0.9120	1	
-0.9120	-0.8640	2	
-0.8640	-0.8160	3	
-0.8160	-0.7680	1	
-0.7680	-0.7200	0	
-0.7200	-0.6720	1	
-0.6720	-0.6240	3	
-0.6240	-0.5760	2	
-0.5760	-0.5280	5	
-0.5280	-0.4800	3	
-0.4800	-0.4320	4	
-0.4320	-0.3840	6	
-0.3840	-0.3360	1	
-0.3360	-0.2880	4	
-0.2880	-0.2400	12	
-0.2400	-0.1920	16	*
-0.1920	-0.1440	12	
-0.1440	-0.0960	30	**
-0.0960	-0.0480	86	******
-0.0480	0.0000	239	**********
0.0000	0.0480	1292	********** →
0.0480	0.0960	228	**********
0.0960	0.1440	16	*
0.1440	0.1920	2	
0.1920	0.2400	1	
0.2400	0.2880	0	
0.2880	0.3360	1	
0.3360	0.3840	0	

Figure 7.1.3 Fraction of unit commitment forecasted plant operating level that the actual level differed from forecast -- Seward 5 unit for period April to December 1975 when the plant was between 75% and 100% of rated capacity in the unit commitment forecast.

(2) to develop the methodology and data that would be more or less transferable to the analyses of other potential control situations.

To accomplish these goals there was an exhaustive analysis of one month out of each of the four seasons. Figure 7.1.4 displays the expected values of the costs which would be incurred in "moving" any 3-hr average predicted concentration levels. These optimum strategies show an interesting characteristic shape, that is,

(1) for low predicted concentrations the control costs are low due to a large number of options that can be employed;

(2) at medium high concentrations the costs of control are unexpectedly high; to get to these levels there has had to be a compounding of plumes from different plants and a compounding of two or three high hourly average concentrations, thus reducing the options for control which increase costs;

(3) finally, there is a definite leveling off of costs for high concentrations; this is due to the ease of controlling the downwash situations that predominate in these very high concentration situations.

Figure 7.1.5 shows the monthly costs that would have taken place if the 3-hour control threshold had been set to various levels, and Figure 7.1.6 shows the annual costs for all of Chestnut Ridge, developed from the pattern of the sample months. The geometric growths in these costs as the thresholds decrease are due:

(1) to the greatly increased number of control situations that are accumulated as the thresholds drop, and

(2) to the rapidly increasing expense involved in the necessity for using more and more costly controls to push concentration down to very low levels.

The reason for discussing the 3-hour concentration first is that this is the standard that is the dominant concern in the Chestnut Ridge situation. The 24-hour standard, interpreted as the daily midnight-to-midnight standard, is composed of 8 blocks of 3-hour averages. Figure 7.1.7 shows how the six highest 3-hour concentrations combine to form the 24-hour average. The sharp upswing in the 3-hour concentrations reflects the bimodal nature of their distribution, with the higher mode being caused by the cluster of downwash episodes. It can be seen that there are generally very high 3-hour concentrations responsible for the higher 24-hour levels. Figure 7.1.8 shows how the 24-hour levels are dramatically affected by the 3-hour controls. For example, if the 3-hour threshold is set at 360 ppb then 95% of the 24-hour concentrations would be below 140 ppb. Thus, once the 3-hour control threshold has been set, much of the cost and effort involved in meeting the 24-hour standard has been eliminated, see Figure 7.1.9. As interesting as this figure may appear, it must be warned that the shapes and levels of these curves are not generally transferable to new sites.

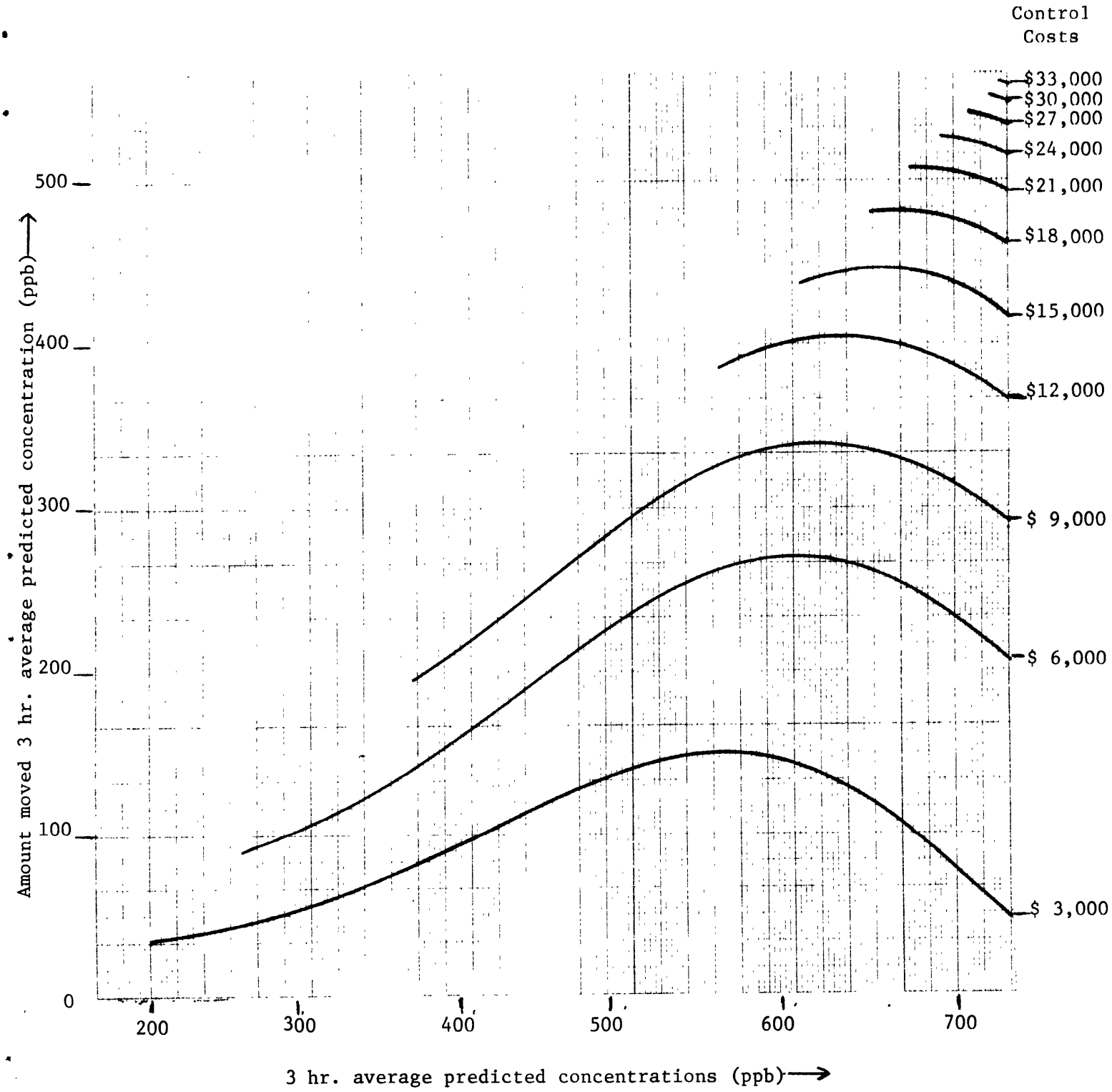


Figure 7.1.4 Total control costs for moving various 3-hr. predicted incidents by various amounts (standard deviations of costs are everywhere about half the displayed mean valves).

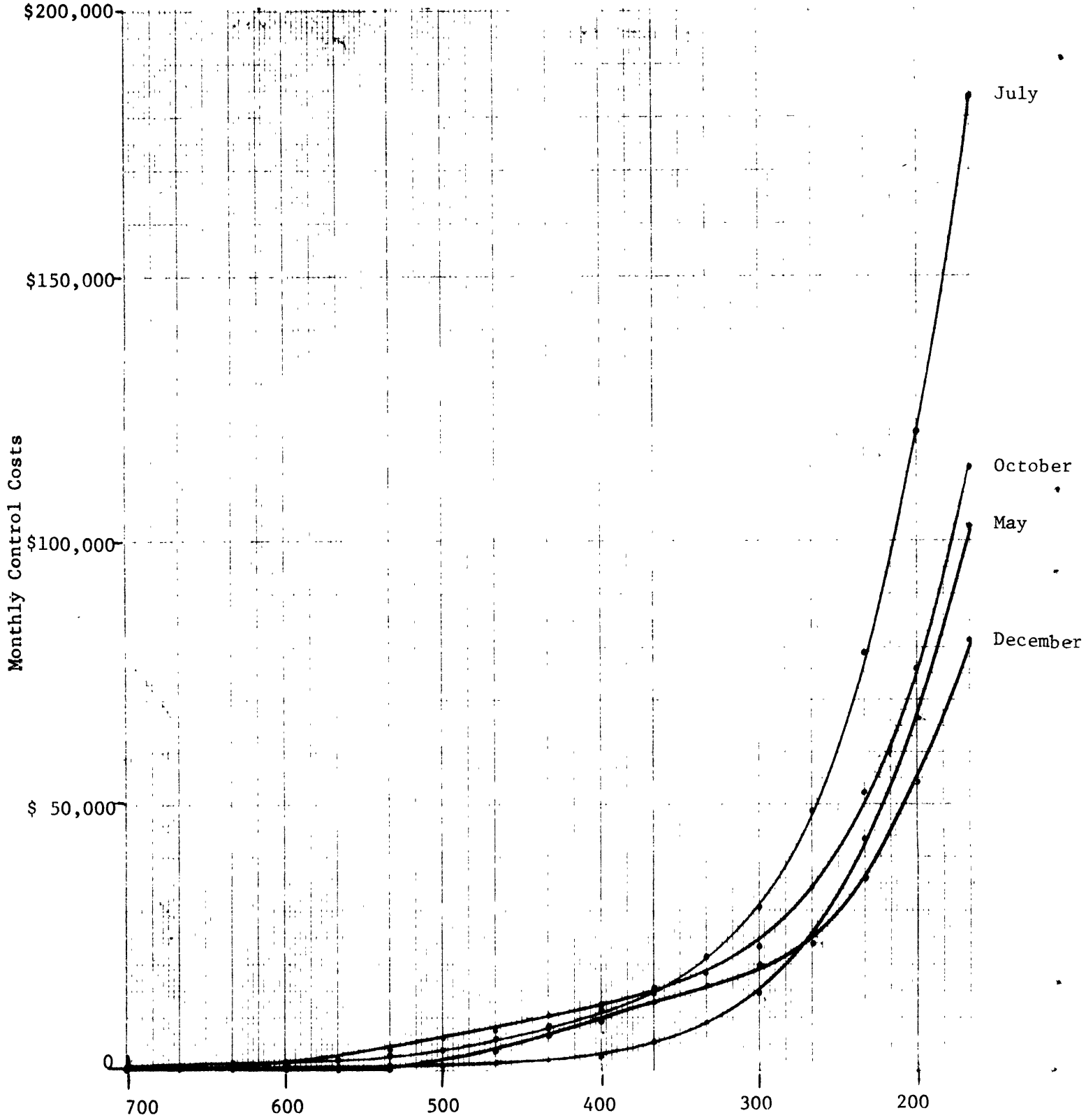


Figure 7.1.5 Total monthly costs for controlling predicted 3 hr. concentrations down to various thresholds (observed meteorological data used here, similar functions for forecasted data).

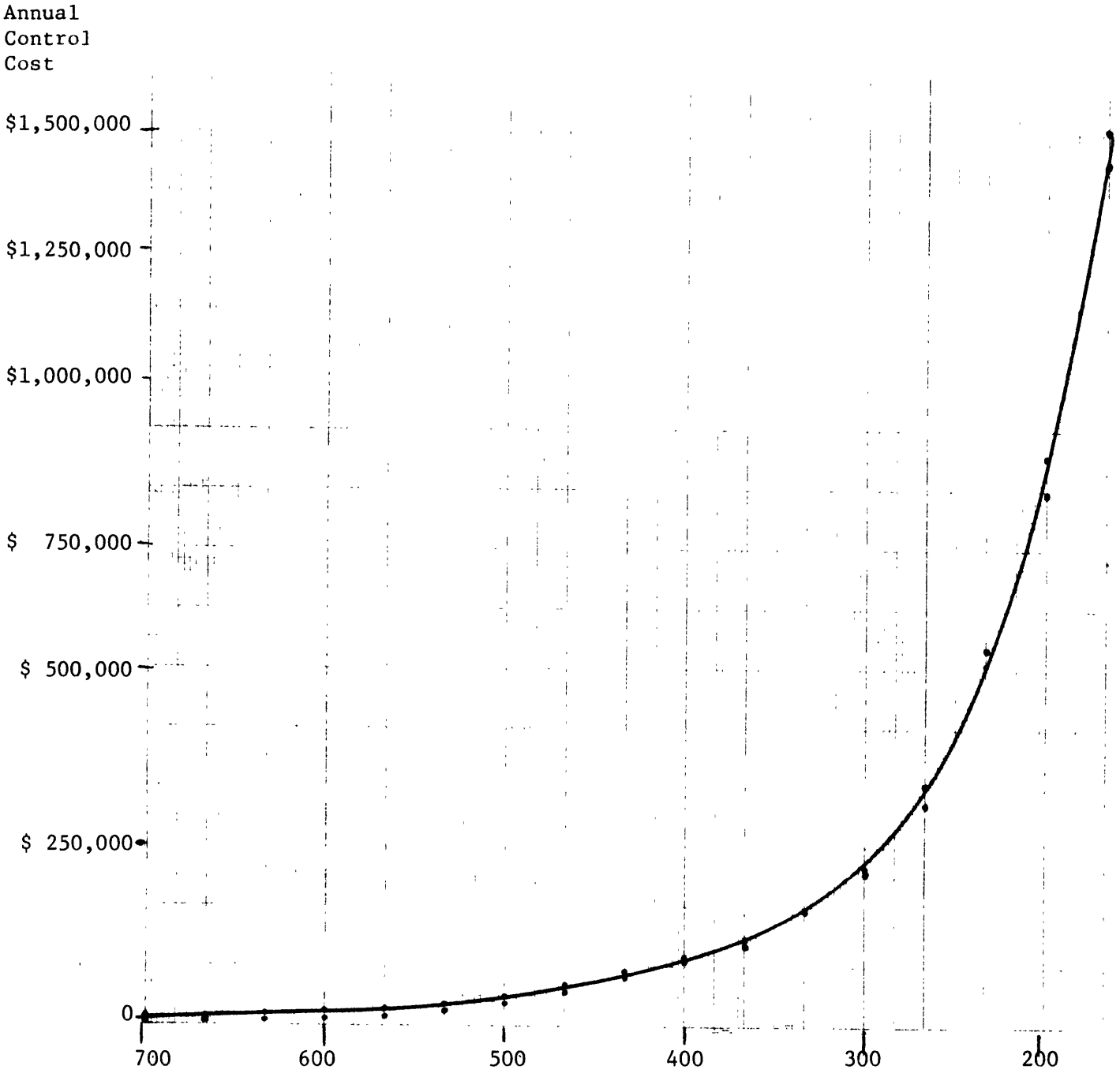
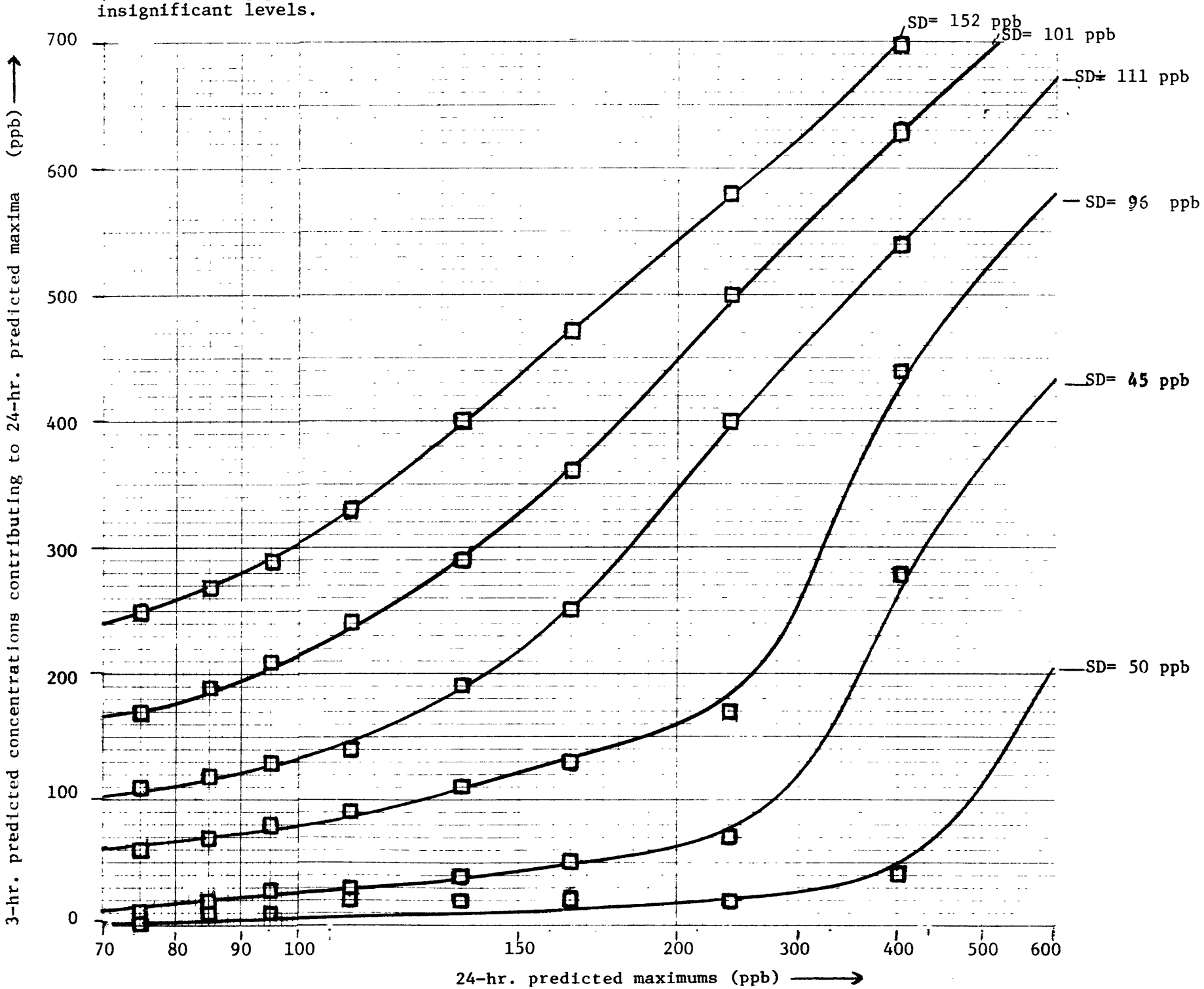


Figure 7.1.6 Annual costs for controlling predicted 3-hr. average concentrations down to various levels (using forecasted or observed data resulted in nearly identical costs as can be seen by the two sets of points).

Figure 7.1.7 Six highest 3-hr. predicted concentration levels contributing to 24-hr predicted maxima (some smoothing performed, SD=standard deviations). The remaining two 3-hr blocks were generally at insignificant levels.



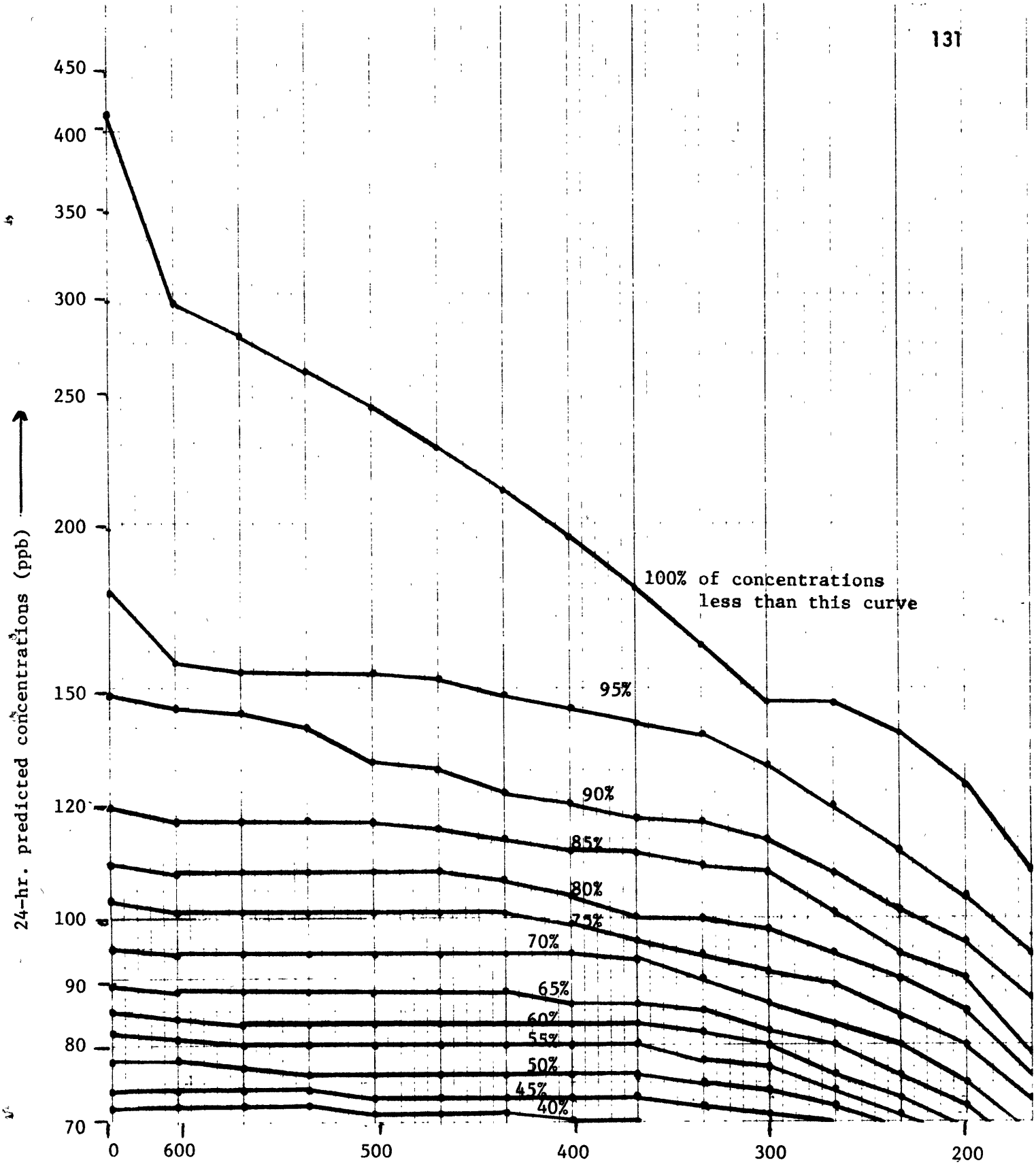


Figure 7.1.8 Distribution of 24-hr. predicted concentrations when the only control is the imposition of various 3-hr. thresholds.

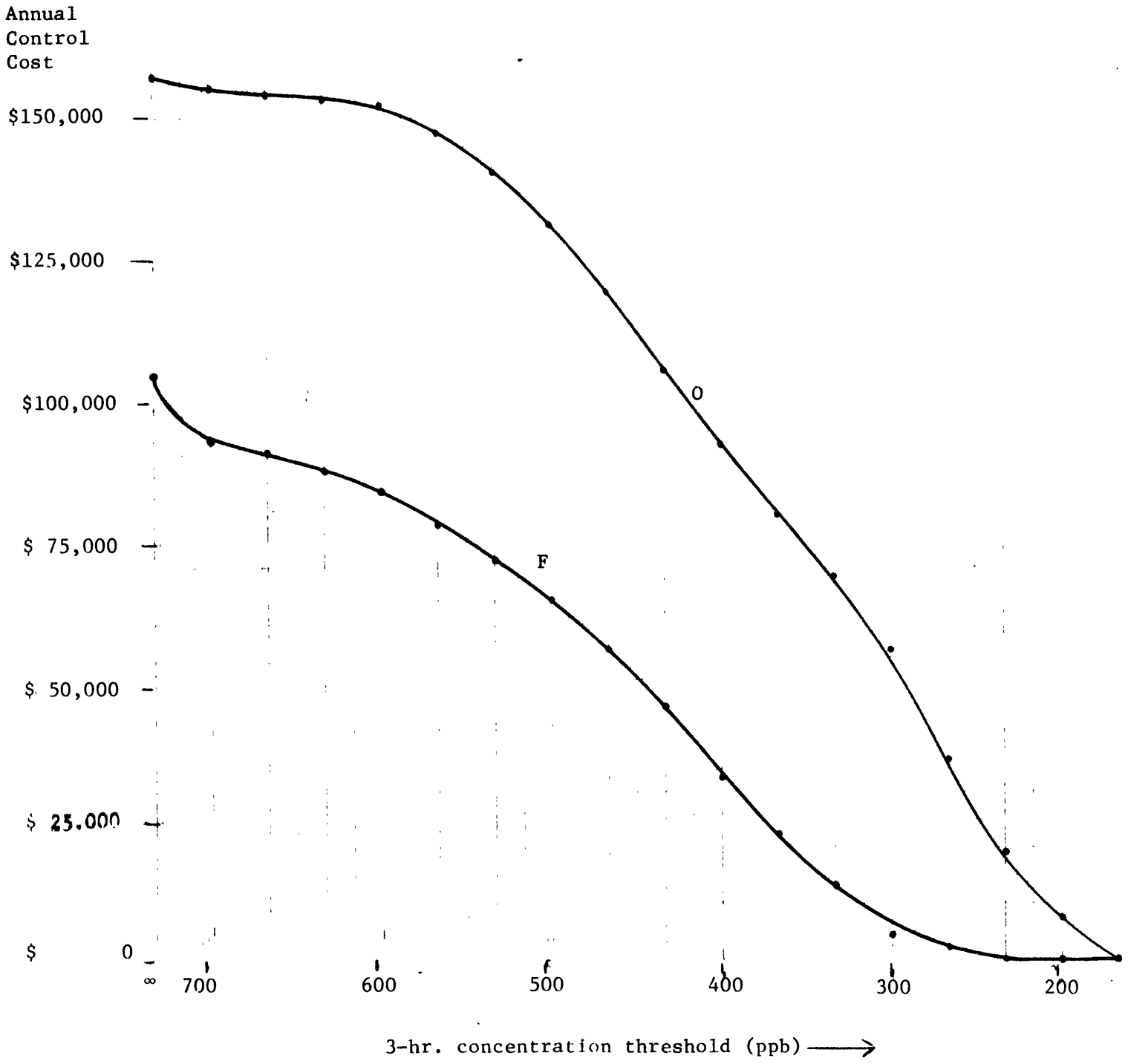


Figure 7.1.9 Cost of controlling predicted 24-hr. concentrations to 140 ppb threshold assuming various 3-hr. control thresholds have already been performed (F = using forecasted meteorological data, O = using observed data). Note that the costly conservatism necessary in meeting 3-hr. thresholds with forecasted data will be somewhat recaptured in controlling 24-hr. situations.

Both the 3-hour and 24-hour control costs have so far been displayed as functions of the control thresholds. The obvious question arises: What should the control thresholds be? To answer that question requires information about the reliability of the air quality predictions, and some review of the situation is presented here.

Table 7.1.1 shows the performance of the final air quality model versus actual concentrations. With the 3-hour standard allowing one violation per year of 500 ppb there is very little information contained in Table 7.1.1 about where the control threshold should be set. Table 7.1.2 shows the relevant control threshold information developed from the various air quality models and experiments that were performed.

For a crude analysis on the basis of this small number of violations the most conservative of the 3-hour thresholds in Table 7.1.2 is used to set the annual control threshold. This level becomes 310 ppb for the 3-hour control threshold because this always leaves at most the one legally allowable violation of the 3- and 24-hr standards at the receptors. This results in $249,500 in annual additional fuel and replacement power costs necessary to meet the ambient threshold (it should be noted that it is very likely that with the new taller stack at the Seward plant this cost would be $0). This amount is probably comparable to the cost of extra power plant and weather forecast personnel necessary to operate such a project. Of course, in this project the higher sulfur fuel used in the fuel-switching scheme was already at the allowable level, and so the $249,500 to meet the ambient standard might be an unnecessary expenditure. It does look like a very small amount, however, if higher-sulfur fuels are allowed as a fuel-switching option, or if the SCS is an alternative to taller stacks, scrubbers, or precleaning the coal, all of which have annualized costs of tens of millions of dollars.

7.2 Technique for Transferability

There is no question about the analytic transferability of the methodologies used in the control strategy and economic analysis. In its most general form the technique that has been used requires the following information:

(1) percentage of predictions above various levels of concentration,

(2) numbers of violations that resulted from situations where the predictions were below various levels of concentration,

(3) statistics on levels of predictions as functions of hours in the day,

(4) statistics on cost of replacement power within and outside system by hours of day,

(5) statistics on sources contributing various fractions to the different predicted levels,

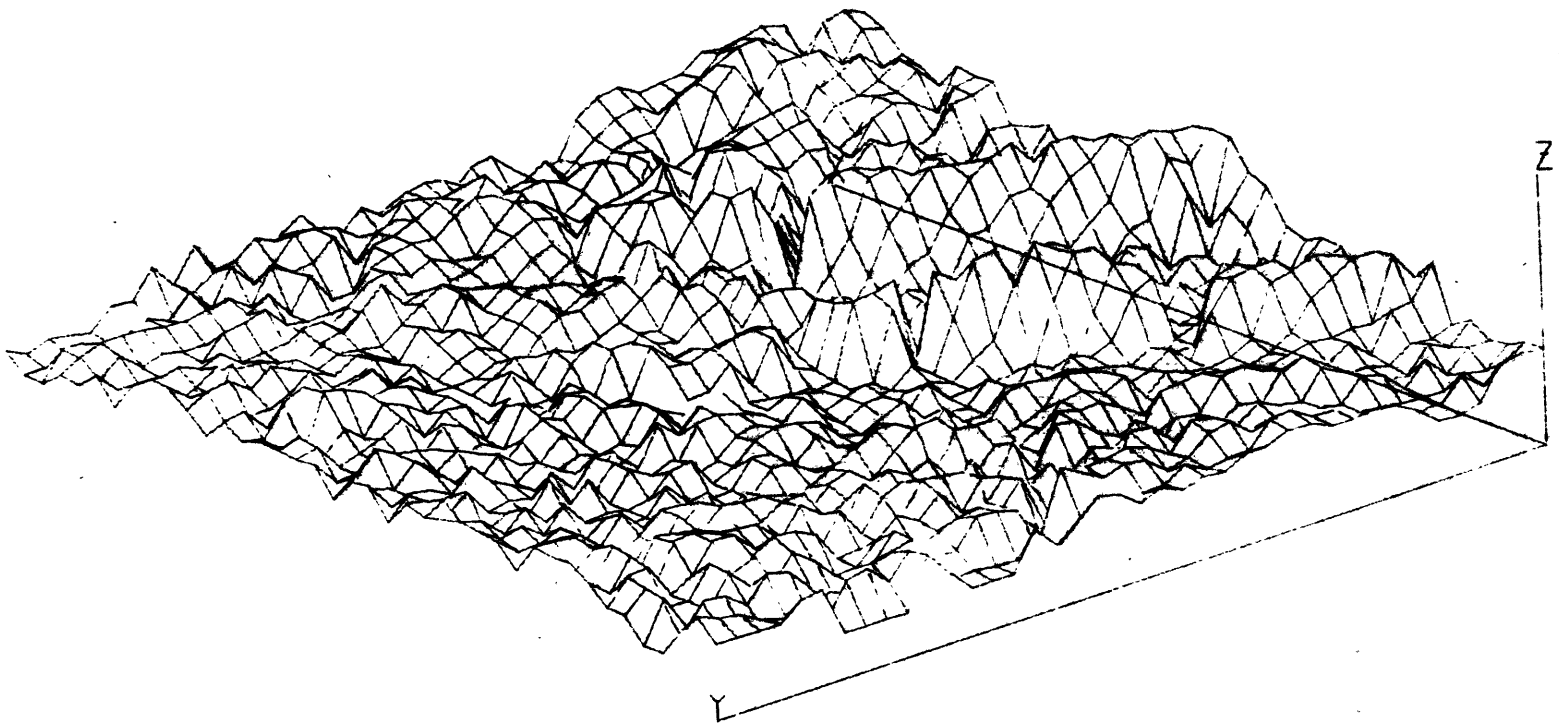


Figure 7.2-1 Midwestern Pennsylvania terrain used in air quality model, variation is from 700 to 2900 feet elevation.

(6) costs of control measures,

(7) variations in forecasting accuracy for one- and eight-hour ahead predictions of concentrations, and

(8) a production costing routine that has a chronological unit commitment scheduler.

This represents a considerable amount of information that is required to make an economic analysis of new weather-dependent control options. Some of this information collection can be avoided by using data from this project. However, a considerable amount of caution should be exercised in transferring graphs presented for the Chestnut Ridge experience to another situation. The principal reasons for this warning include:

(1) the rough terrain of the Chestnut Ridge area (see Figure 7.2.1) is atypical,

(2) the power plants in this area do not generally operate on the economic margin; they are normally baseloaded,

(3) overlapping of plumes is not a dominant effect, and

(4) the air quality model has received an unusual amount of expert fitting to the peculiarities of the Chestnut Ridge environment.

It must be emphasized that the overriding concern in any economic analysis must be the cost penalties associated with the lack of perfectly certain information. The uncertainties in the Chestnut Ridge experiment were about 50 times as costly as the cost of controls if there had been perfect weather and emissions predictions available.

The role of SCS has been the subject of much controversy, aspects of which were summarized in Section 1.1. Much of this controversy arises from a lack of understanding of what SCS can and should do. Some opinions on future roles for SCS will now be summarized.

There is little question that the energy, economic, environmental crisis will last for many years. The pressures to limit oil and natural gas consumption are too strong to wait for solar energy and other new sources to fill the gaps. It is necessary to rely on some combination of coal and nuclear energy sources. The trade-offs between coal and nuclear are too complex to be addressed here, but it is clear that coal will play a major role. It is also clear that the burning of coal will result in some type of atmospheric pollution.

Two major resources needed to deal effectively with the air pollution problems associated with burning coal are knowledge and emission control technology. Knowledge is needed on how the pollutants are formed, how they propagate in the atmosphere, and how they affect flora, fauna, and materials. Emission control technology is needed to reduce the total amount of pollutants that enter the atmosphere. Unfortunately, our resources in both areas are very limited. Our knowledge is rapidly growing, but still pitifully inadequate. Emission control technology is also a rapidly evolving field but it takes many years to develop a viable technology to deal with a particular type of pollutant. One fundamental dilemma is that our knowledge is so inadequate that it is not clear what type of emission control technology should be developed and installed. For example, a lot of work and effort is going into developing technologies to remove sulfur but it is no longer obvious that SO_2 and/or sulfates are the major health problems associated with burning coal.

The long time delays and large costs associated with emission control development and installation translate into a need for some additional way to adapt to our rapidly changing knowledge of environmental impacts. SCS meets this need by exploiting a third resource, the time-varying dispersive potential of the atmosphere. The discussion in Section 1.1 about SCS versus ICS emphasizes that an SCS is not a "competitor" of emission control technology; they are partners.

The Chestnut Ridge project demonstrates one particular type of SCS whose design can be characterized as follows:

- Pollutant: SO_2
- Region: Local area around four power plants
- Time Scale: 3 and 24 hour averages
- Criterion: Minimum cost subject to constraint on not violating standards.

However, the basic SCS concept applies to much more general situations. For example, an SCS is conceptually applicable to sulfate control over multiple state regions for weekly to yearly time scales. Furthermore the criteria of operation are not fixed. It is hoped that in the future our knowledge of health effects will have gotten to the point where we will understand the nonlinear effects of different levels and time durations of pollutant exposure. When this happens, it will be a relatively easy job to modify the control strategy criteria to be directly responsive to the true human impact of coal burning rather than just the fixed standards presently used.

There are of course technical problems associated with many of the more general types of SCS applications discussed above. These problems lie primarily in our limited knowledge of how pollutants form, propagate, and effect human beings. Careful study will usually be needed to determine whether the SCS concept is actually appropriate for any specific problems (pollutants, region, time scale, criteria). Such conditions, however, are not unique to SCS. They are an intrinsic part of the whole energy, economic, environmental crisis.

References

Briggs, G.A. Plume rise: A recent critical review. Nuclear Safety 12(1):15, 1971.

Cadogan, J. A Dynamic Emissions Management System for Electric Utilities. Ph.D. Dissertation in Energy Management and Policy, University of Pennsylvania, 1975.

Egan, B.A. Turbulent diffusion in complex terrain. From lecture notes of the A.M.S. Workshop on Meteorology and Environmental Assessment Boston, Mass., September 29-October 3, 1975.

Environmental Protection Agency. Air Quality Display Model. TRW Systems Group, NAPCA, Washington, D.C., November 1970.

Environmental Protection Agency. Guidelines for the Interpretation of Air Quality Standards. Office of Air Quality Planning and Standards, Monitoring and Data Analysis Division, Research Triangle Park, N.C., August 1974.

Environmental Protection Agency. Guidelines for Evaluating Supplementary Control Systems. EPA-450/2-76-003, Office of Air and Waste Management, Office of Air Quality Planning and Research, USEPA, Research Triangle Park, N.C., February 1976.

Federal Register. National primary and secondary ambient air quality standards. 36(84):8186, 1971.

Gaertner, J. et al. Analysis of the Reliability of a Supplementary Control System for SO2 Emissions from a Point Source. ERT Project 0669, ERT, Lexington, Mass., June 1974.

Larsen, R.I. Determining reduced emission goals needed to achieve air quality goals: A hypothetical case. Journal of the Air Pollution Control Association, 17(12), 1967.

Larsen, R.I. A Mathematical Model for Relating Air Quality Standards. Publication AP-89, USEPA, Research Triangle Park, N.C., 1971...

Larsen, R.I. An air quality data analysis system for interrelating effects, standards and needed source reductions. Journal of the Air Pollution Control Association 23:933, 1973.

Larsen, R.I. An air quality data analysis system for interrelating effects, standards, and needed source reductions: Part 2. Journal of the Air Pollution Control Association 24(6), 1974.

Montgomery, T.L. and Frey, J.W. Tall stacks and intermittent control of SO2 emissions: TVA experience and plans. Mining Congress Journal 61(1):44-51, 1975.

Moses, H. Mathematical Urban Air Pollution Models. Report No. ANL-ES-RPY-001, Argonne National Laboratory, Argonne, Ill., April 1969.

Noll, K.E. and Davis, W.T., eds. Power Generation: Air Pollution Monitoring and Control. Ann Arbor Science, Ann Arbor, Mich., 1976.

Pasquil, F. Atmospheric Dispersion Parameters in Gaussian Plume Modeling Part II: Possible Requirements for Change in the Turner Workbook Values. Report No. ESRL-RTP-058, USEPA, Series No. 4: Environmental Monitoring, 1976.

Peterson, D.W. and Schweppe, F.C. Code for a general purpose system identifier and evaluator (GPSIE). IEEE Transactions on Automatic Control Volume AC-19(6):852, 1974.

Schweppe, F.C. Uncertain Dynamic Systems. Prentice-Hall Electrical Engineering Series, Englewood Cliffs, N.J., 1973.

Schweppe, F., Ruane, M. and Gruhl, J. Economic Environmental Operation of Electric Power Systems. Engineering Foundation Conference -- Systems Engineering for Power: Status and Prospects, CONF 750867, USERDA, Henniker, N.H., August 1975.

Schulman, L. and Egan, B. Development of the Mathematical Downwash Model Used for the Prediction of SO2 Concentration at the Dow Chemical Company, Michigan Division. ERT Document P-1638-B, April 1975.

Turner, D.B. Workbook of Atmospheric Dispersion Estimates (revised 1970) USHEW, Environmental Health Series, 1970.

United States Senate Committee on Public Works. Clean Air Amendments of 1976. Report No. 94-717 Accompanying S.3219, US Government Printing Office, Washington, D.C., 1976.

Woodard, J. Electric Load Modeling. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, Mass., September 1974.

APPENDIX A

Mean Concentration Analysis

Mean concentration analysis is a good starting point to find general trends in a region. The mean concentrations are relatively easy to form and are particularly effective as background indicators. This appendix is divided into three parts: 1) definition of symbols; 2) description of the various averaging processes and their applications; and 3) discussion of some results from this method when applied to Chestnut Ridge.

SYMBOLS

EWV	elevated wind direction at tower location (degrees from North)
EWV	elevated wind speed at tower location (mph)
DMX	mixing depth at Pittsburgh (m)
Wind Sector	one of N divisions of the 360° compass headings; N was either 16 or 36, resulting in sectors of either 22.5° or 10°. Sectors were numbered in ascending clockwise order, starting at north. A 17th and 37th sector was defined to include missing data and variable winds.
Northern Region	monitors B, C, L, M, N, and O
σ	standard deviation of an arithemtic mean
$\% \sigma$	standard deviation as a percentage of mean
C	concentration at specified monitor during a one hour time period, actually composed of the arithmetic mean of 60 samples equally spaced during the hour
hr	hour of the day; hr=1 for the period 0000-0100, etc.
$\lvert\hat{r}\rvert$	distance of resultant vector from monitor to source
EWD^r	angle of resultant wind direction associated with $\hat{r}$
EWV^r	resultant wind velocity associated with $\hat{r}$
t_{ab}	$\frac{\lvert\hat{r}\rvert}{EWV^r}$ transit time for a parcel from source to a to monitor at b, where r is from a to b

Subscripts (all applied to the averaging process)

d,m,s,a	day, month, seasonal, and annual averages (no subscript corresponds to one hour); identifies time period
N	normalized; the mean of the subset of the data divided by the mean of all the data
R	regional or ensemble average of more than one monitor
D	daily average, of hours 1 to 24 on the i th day
H	hourly average, of all the j th hours in the time period
S	sector average of concentration when wind is in sector k

Superscripts

r	indicates the variable is formed from the resultant wind analysis
—	indicates averaging or arithmetic mean

Averages Used

$\overline{C}_m$, $\overline{C}_s$, $\overline{C}_a$	the average concentration at a monitor (missing values ignored) during one month, one season, or annually
$\overline{C}_S(i)$	mean concentration for wind sector i
$\overline{C}_{RS}(i)$	regional average of concentrations at all monitors for the same wind sector i; tends to wipe out point source influence on monitors, but requires similar terrain
$\overline{C}_{RSs}(i)$	regional sector average for sector i in seasonal period
$\overline{C}_R$	regional average over all monitors
$\overline{C}_{NRS}(i)$	the normalized sector average for sector i; formed by averaging the normalized sector averages from each monitor in the region; normalized by $\overline{C}$ (i)
$\overline{C}_{NRS}r(i)$, $\overline{C}_{RS}r(i)$	similar to the above, but using the resultant wind sector
$\overline{EWV}_S$, $\overline{EWV}_S{}^r$	average wind sector speed and average resultant wind sector speed (EWV_S often appears in wind roses)
$\overline{C_S \times EWV_S}$	the average of the product of concentration and sector wind speed (if wind speed and concentration are inversely proportional, the % term will be smaller for this quantity than for $\overline{C}_S(i)$)

$\overline{C}_H(i)$	average of all concentrations occuring during the same hour of the day, i.
$\overline{C}_{RH}(i)$	regional hourly average formed analogously to $\overline{C}_{RS}(i)$
$\overline{C}_D(i)$	the average concentration of hours 1-24 on day i
$C_{NRD}(i)$	normalized regional daily average formed analogously to $\overline{C}_{NRS}(i)$

Application of Averaging Processes

The various averages were applied only to the Northern Region because of the relatively similar terrain found in the area. The monitors are also approximately 30 miles from both of the major regional population centers. Generally a normalized average would be formed so as to eliminate the monitor to monitor differences and highlight the regional phonomena. Three qualities of the data were examined: wind sector dependence, wind speed dependence and time dependence. Seasonal subsets of the data were formed when looking at the various qualities in order to determine seasonal patterns.

Application to Chestnut Ridge

The first work was done with the wind sector average concentrations during the seasons of the year. The seasonal regional sector average, $\overline{C}_{RSs}(i)$, and the seasonal normalized regional sector average, $\overline{C}_{NRSs}(i)$, were both formed. Both $\overline{C}_{RS}$ and $\overline{C}_{NRS}$ show the same general behavior, but the %σ for the $\overline{C}_{NRS}$ is about half of the %σ for $\overline{C}_{RS}$. Consequently, $\overline{C}_{NRS}$ was used for most of the analysis since it reduced the scatter in the data. Table I shows the normalized regional average for the Northern Region for fall and winter, 1974-1975.

The most striking feature is that there are two distinct relative maximums between sectors 7 and 8 and sectors 11 and 12. The first relative maximum comes from a wind direction which corresponds to the following major sources: Seward, Conemaugh, Homer City, Johnstown. The second relative maximum comes from Pittsburgh's direction. The maximum mean concentration at all the monitors is higher in the winter because the seasonal average is higher. It is interesting to note that the difference between the minimum normalized regional average and the maximum is greater in the fall than in the winter. This is a result of increased background, probably due to space heating, which tends to raise the entire concentration distribution in the winter, making normalized differences smaller.

Most normalized sector averages $\overline{C}_{NS}$, at all the monitors in the region are nearly the same, as the Table 1 values of %σ for $\overline{C}_{NRS}$ indicate. This shows that the sector average concentrations are all related to the wind sector in the same manner, but have different magnitudes. Table 3 shows a typical group of sector averages and their %σ values at monitor M.

Notice that the standard deviation in Table 3 is nearly equal to the mean in all sectors. As a predictive tool the sector average method has very wide error brackets.

The above work used 22.5° sectors. Normalized sector averages were also formed using smaller sectors of 10°, and with the resultant wind vector in an attempt to reduce the error brackets and improve the resolution of the relative maximum. Smaller sector size reproduces the 22.5° patterns. Table 2 shows the results of a normalized resultant monthly sector average, $\overline{C}_{NRS}r$, using the 22.5° sectors. The shape of the distributions were the same for $\overline{C}_{NRS}$ and both wind sector sizes with $\overline{C}_{NRS}r$. Conceptually accounting for changes in the wind direction and speed during transit from Pittsburgh and Johnstown with the resultant wind vector model was expected to improve resolution, but no differences were observed. The Northern Region should be a good region for this type of modeling, since it is nearly equidistant from both the urban areas. The null result with the resultant wind vector technique could be because the tower wind data isn't accurate when extrapolated in this manner; it also might be that the urban plume effects actually span 50° or more. Our belief is that both factors contribute to the lack of resolution of the mean values into sharper wind sector peaks.

As a test of the modeling hypothesis that concentration is proportional to 1/EWV, the %σ of C_{Ss} was compared with the %σ $C_{Ss}\hat{r}(hr) \times EWD_{Ss}r(hr)$ in Table 3. The scatter in the data of %σ is nearly the same for both. The conclusion is that if concentration is indeed proportional to 1/EWV, the effect is swamped by other effects and is not significant. An alternate hypothesis is that the tower wind data is not representative of Northern Region wind fields. The average of the product of the concentration and elevated wind speed by wind sector, $\overline{C_S \times EWV_S}$, was also tried using the smaller wind sectors. The average of the product was also examined using the observed EWD at the hour of the concentration less the transit time t_{ab} from the resultant wind vector case. None of these techniques were able to reduce the %σ of the average of the product.

Variations in the daily (24 hour) averages during one-month periods were also considered. $\overline{C}_{Dm}$, $\overline{C}_{RDm}$, and $\overline{C}_{NRDm}$ were all formed. Once again the %σ for the normalized regional average was considerably smaller than the region averages, as shown in Table 4, for two months. There is a large variation in the daily averages, and all the monitors show similar behavior due to the smallness of the %σ. The following meteorologic parameters were postulated as the cause of the variation of the normalized regional daily average: Temperature (through space heating loads), solar insolation, mixing depth, elevated wind speed and elevated wind direction. All but the elevated wind direction were found to have weak or no correlation with high or low daily averages, $\overline{C}_{NRD}$.

It turned out that the highest values of $\overline{C}_{NRD}$ occured when the wind direction was persistently from Pittsburgh, and the lowest values when the wind was from the north. Consequently, the daily averages in the Northern Region are just another reflection of the wind direction dependence described above.

The possiblity of a weekly dependence in the daily average was also investigated. The daily average C_{NRD} occuring on each day of the week, Monday-Sunday, were in turn averaged together to form the average daily average on each day of the week. The result shown in Table 5 is that all days of the week have the same distribution for the daily averages. The seasonal hourly regional average was investigated for the winter and summer months of 1974 and 1975. The results are shown in Figure 1. indicating the trend to a daily maximum occuring in mid morning. The seasonal differences are due to the effects of higher space heating in winter tending to raise the entire curve, and increased turbulent transport in the summer, tending to increase the peak/mean ratio.

TABLE 1

Normalized Regional Sector Averages For Northern Region

Wind Sector	Fall 1974				Winter 1975			
	$C_{NRSs}(i)$	%σ	σ	n	$C_{NRS}(i)$	%σ	σ	n
1	.25	24	.06	41	.67	16	.11	14
2	.29	31	.09	29	.47	11	.05	11
3	.30	30	.09	26	.45	24	.11	6
4	.28	54	.15	32	.60	18	.11	17
5	.44	20	.09	39	.87	10	.09	53
6	.46	24	.11	69	1.13	9	.10	50
7	1.13	25	.28	126	1.05	19	.20	224
8	1.18	23	.27	78	1.12	6	.07	129
9	.88	11	.10	86	1.20	9	.11	92
10	.99	10	.10	152	.96	8	.08	78
11	1.45	14	.20	466	1.25	16	.20	279
12	1.65	8	.13	374	1.20	8	.10	463
13	.94	10	.09	191	.77	17	.13	282
14	.50	16	.08	228	.61	11	.07	80
15	.33	15	.05	118	.57	19	.11	44
16	.21	24	.05	88	.53	26	.14	49
17 CALMS	.88	40	.35	41	.80	18	.14	21
	Mean %σ = 22				Mean %σ = 14			

Dimensionless: $\overline{C}_{NRSs}(i)$ is the normalized regional (B, C, L, M, N, O) Sector (22.5°) seasonal average; σ is the arithmetic standard deviation of the six individual normalized ith sector averages.

Table 2

Normalized Results of Sector Averages for the Northern Region

Wind Sector	November 1974 $\overline{C}_{NRS}\hat{r}$	November 1974 % σ	May 1975 $\overline{C}_{NRS}\hat{r}$	May 1975 % σ
1	.39	27	.43	58
2	.32	49	.35	47
3	.23	44	.33	59
4	.36	93	.31	43
5	.49	32	.65	40
6	.74	29	.93	23
7	1.19	32	1.08	35
8	1.03	20	.90	39
9	.74	15	.72	43
10	1.00	7	1.32	19
11	1.34	13	1.75	15
12	1.36	14	1.61	15
13	.67	11	.79	11
14	.45	13	.55	22
15	.36	18	.55	26
16	.45	16	.78	17
	mean % 27		mean % 32	

$\overline{C}_{NRS}\hat{r}$ (Dimensionless) normalized regional (B, C, L, M, N, O) results of set or average

$\%\sigma = 100 \times \sigma/\overline{C}_{NRS}\hat{r}_{(i)}$ where r (dimensionless) is the arithmetic standard deviation of the six individual i^{th} sector averages.

Table 3

Sample Sector Averages and % σ at Monitor M

22.5° Sectors - November 1974

Sector #	$\bar{C}_{Sm}^{\hat{r}}$	% σ	N	$C_S(hr) \times EWV_S(hr - t_{ab})$	% σ	N
1	15.5	106	11	154.3	112	11
2	8.3	100	4	64.3	101	4
3	8.8	106	5	84.2	103	5
4	7.7	101	3	70.0	100	3
5	23.9	126	7	377.6	199	7
6	32.9	129	20	654.5	128	20
7	22.5	113	51	422.8	119	51
8	25.9	113	23	374.7	115	23
9	28.1	108	34	432.6	108	34
10	32.6	106	58	492.9	106	58
11	48.5	114	129	900.8	117	129
12	41.5	115	177	768.8	113	177
13	21.9	113	71	360.6	123	71
14	13.3	144	79	182.4	125	79
15	9.5	116	15	116.5	113	15
16	12.7	107	15	123.9	107	15

10° Sectors - June - August 1975

Sector #	$C_{Sc}^{\hat{r}}(hr) \times EWD_S^{\hat{r}}(hr)$	% σ	N
11	279.2	140	13
12	182.5	134	11
13	209.4	145	14
14	196.7	142	47
15	154.3	142	53
16	170.8	162	35
17	118.0	140	26
18	381.4	129	44
19	363.5	141	21
20	244.7	131	14
21	220.5	132	28
22	313.3	142	47
23	388.4	132	56
24	507.6	124	95
25	453.4	123	60
26	290.7	130	54

% σ = 100 x σ/mean (dimensionless) where σ is the arithmetic mean of the N occurrence in the sample sector

N is the # of times wind in sector

$C_{Ss}^{\hat{r}}(hr) \times EWV_{Ss}^{\hat{r}}(hr-t_{ab})$ [PPBxMPH]: seasonal average of the products of the concentration in sector i times the wind velocity t_{ab} hours earlier

$C_{Ss}^{\hat{r}}(hr) \times EWV_{Ss}^{\hat{r}}$ [PPBxMPH]: seasonal average of the products of the concentration in sector i times the wind velocity at the same hour

Table 4

Normalized Regional Daily Averages for Northern Region

Day	Jan. 1975 $\bar{C}_{NRDm}(j)$	%σ	σ	March 1975 $\bar{C}_{NRDm}(j)$	%σ	σ
1	1.07	21	23	.75	28	.21
2	.35	40	.14	1.0	20	.20
3	.96	26	.25	.85	52	.44
4	.62	60	.37	1.33	22	.29
5	1.26	50	.63	1.85	39	.73
6	1.12	61	.68	1.81	25	.45
7	.98	55	.54	1.28	26	.33
8	.73	49	.36	.59	27	.16
9	.55	18	.10	1.42	31	.44
10	.79	27	.21	1.46	6	.09
11	.42	26	.11	1.24	17	.21
12	.86	21	.18	.85	60	.51
13	.85	16	.14	.58	47	.27
14	1.02	18	.18	.83	69	.53
15	1.57	13	.21	1.41	10	.14
16	1.27	9	.12	1.88	16	.31
17	1.04	16	.17	1.22	27	.33
18	.61	38	.23	.70	24	.17
19	.72	17	.12	.40	98	.39
20	.48	40	.19	.60	32	.19
21	1.29	11	.14	.91	34	.28
22	2.38	16	.38	.99	92	.42
23	1.81	15	.27	1.23	30	.37
24	1.51	22	.33	.70	10	.07
25	.64	17	.11	.83	19	.16
26	.62	55	.34	.46	59	.27
27	1.20	15	.18	.66	33	.22
28	1.41	9	.13	.58	81	.47
29	.60	20	.12	.75	23	.17
30	.68	41	.28	.41	32	.13
31	.57	112	.64	1.14	17	.19
	Mean %σ-31			Mean %σ-33		

$\bar{C}_{NRDm}$ [Dimensionless] normalized monthly regional (B, C, L, M, N, O) daily average for one month.

$\%\sigma = 100 \times \sigma/\bar{C}_{NRDm}$ [Dimensionless]

σ [Dimensionless] = arithmetic standard deviation of six $\bar{C}_{NDm}$ observed at the monitors on the ith day.

Table 5

Average of $\overline{C}_{NRDs}$ on Days of the Week

	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.
			December 74 - February 1975				
Mean	.77	.95	1.0	1.05	1.19	1.03	.79
σ	.43	.35	.56	.49	.54	.53	.43
			Nov. 1974, March,+May, 1974				
Mean	1.11	.97	.88	.86	.90	1.03	1.05
σ	.43	.43	.60	.41	.36	.62	.42
Combined Mean	.94	.96	.94	.95	1.05	1.03	.92

Table 6

Ratio of Seasonal Average to Regional Seasonal Average C_s/C_{Rs}

Monitor	Fall Sept.- Nov. 74	Winter Dec.- Feb. 75	Spring Mar.- May 75	Summer June - Aug. 75	Fall Sept.- Nov. 75	Mean	%σ
B	1.04	1.42	1.54	1.49	1.10	1.32	13
C	.89	.96	.78	.50	.54	.73	28
L	.91	.82	1.19	1.49	1.13	1.11	26
M	1.53	1.38	1.26	1.44	1.49	1.42	7
N	.83	.78	.63	.42	.93	.72	28
O	.80	.65	.61	.65	.81	.70	9

Monitor	B	C	L	M	N	O
		Winter 74-75				
$\overline{C}_s/\overline{C}_{Rs}$	1.42	.96	.82	1.38	.78	.65
$\frac{\overline{C}_{Ss}/\overline{C}_{Rs}(14\text{-}3)}{\overline{C}_{RSs}(14\text{-}3)}$	1.62	1.01	.68	1.35	.68	.68

$\overline{C}_{NRD_s}$ [Dimensionless] normalized seasonal regional (B, C, L, M, N, O) Daily average

$\overline{C}_s$ [PPB] Seasonal average

$\overline{C}_{Rs}$ [PPB] regional (B, C, L, M, N, O) seasonal average

$\overline{C}_{Ss}(19\text{-}3)$[PPB] Seasonal sector average for sectors (14, 15, 16, 1, 2, 3)

$\overline{C}_{RSs}(19\text{-}3)$ [PPB] regional (B, C, L, M, N, O) seasonal sector average for sectors (14, 15, 16, 1, 2, 3)

Figure 1. Regional Seasonal Hourly Average $\bar{C}_{RHs}$ [PPB]

Average at i^{th} hour for concentration in Northern Region (B, C, L, M, N, O

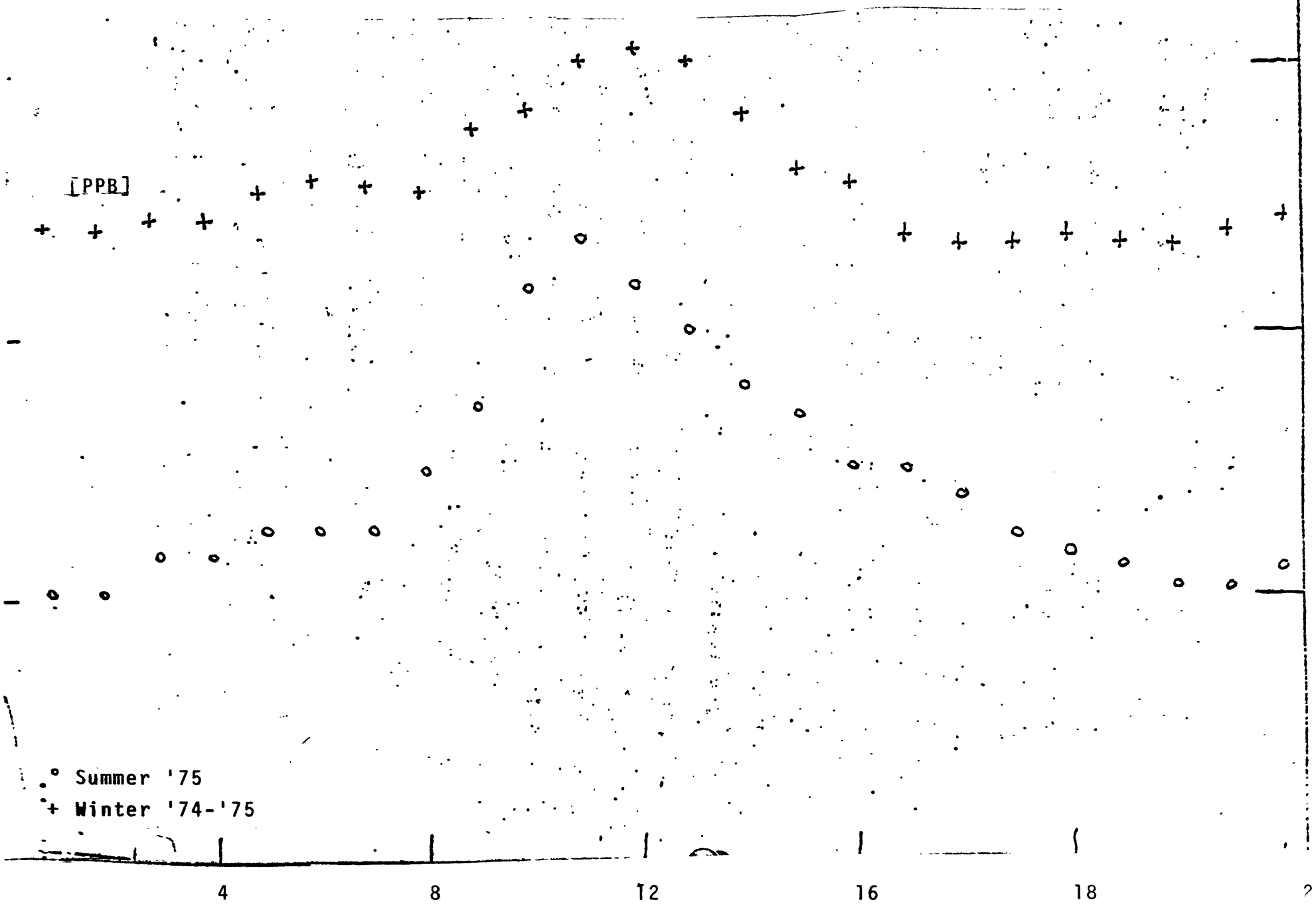
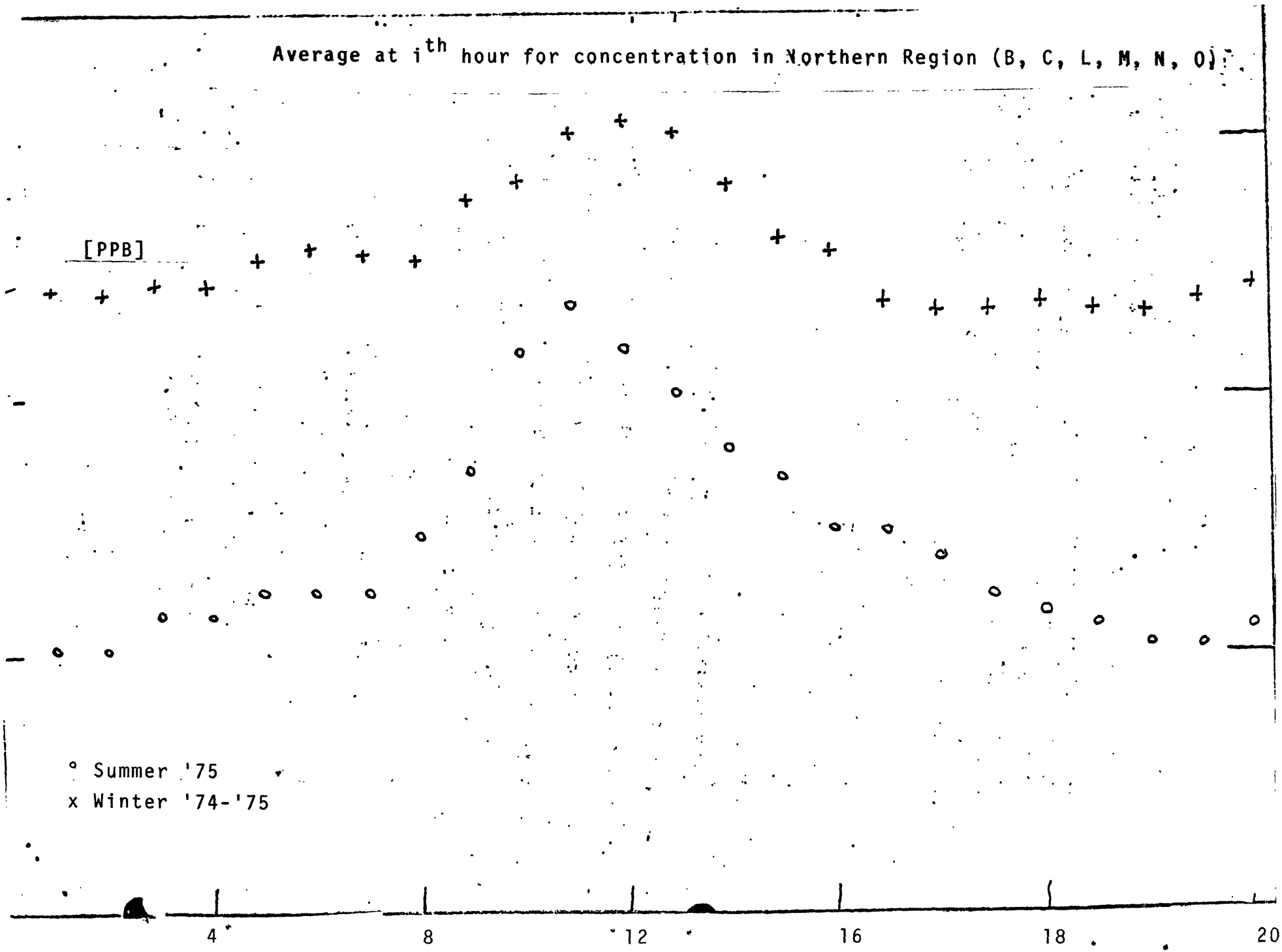


Figure 2 Regional Seasonal ^ Hourly Average $\overline{C}_{RHs}$ [PPB]



APPENDIX B

PEAK CONCENTRATION ANALYSIS

This analysis was directed towards finding the principal sources affecting each monitor, and establishing a database of the highest concentrations from each of the principal sources. The following definitions are used.

Peak concentration, monitor i: a concentration which is in the 95th to 100th percentiles of the concentrations observed at monitor i.

$N_p(\alpha)$; the number of peak concentrations observed in conjunction with the occurrence of meteorological variable α during a specified time period.

$N(\alpha)$: the number of times meteorological variable α occurs during the time period, given that the monitor is working

$P = N_p(\alpha)/N(\alpha)$: the sample probability that a peak concentration will occur when the specified meteoroligcal variable α occurs

EWD: elevated tower wind direction

ENV: elevated tower wind speed.

The first meteorological variable examined was the wind direction EWD. Thirty-six 10° wind sectors were defined, centered on 5°, 15°, etc. A thirty-seventh sector was defined to accept other possible reports (999 - missing, or 888 - variable). September 1974 to August 1975 was analyzed. The space heating season from October to March was analyzed as a subset.

The peak hourly averages are used with this type of analysis because it is the most effective way of identifying the major sources. A longer averaging period requires that the wind direction be persistent over the averaging period. Wind persistence data can be combined with peak hourly concentration data to extrapolate to longer periods if desired.

The general trend is for large value of N_p(EWD) to occur when N(EWD) is large in the same wind sector. P(EWD), the probability of a peak concentration occurring by wind sector, tends to be large when there is a principal source upwind from the monitor. A seasonal trend of higher winter probabilities is also observed, reflecting the space heating component of emissions. Table 1 shows N(EWD), and P(EWD) for monitors B, M, and 8 in the Chestnut Ridge region. Figure 1 shows the wind directions with the highest probability, P(EWD), for all the monitors with the power plants and urban areas shown in their respective geographic locations. For clarity, only wind sectors with two or more adjacent probabilities greater than 0.08 are shown. Due to the scaling problem, monitor 8 is not shown, but Table 1 shows a strong probability peak in the Seward station direction.

The most significant general result is that there is a high probability of peak concentration occurring if the wind direction is from Pittsburgh. Monitors B, C, M, and O have large probabilities of peak concentrations occurring which are almost certainly due to Pittsburgh. Monitors 3, 6, 7, D, E, F, 1, and N also have large probabilities in the Pittsburgh direction. However, these monitors have power plants in the same upwind sector, so it is ambiguous which source is causing the peaks. The existence of high probabilities for the unambiguous monitors was the first indication of the need for explicit consideration of Pittsburgh in the SCS formulation.

This type of analysis is particuarly useful in identifying consistent wind shifts and other complex terrain effects. Initially it was assumed that the tower data were accurately representing the wind field of the entire region, and a search was performed to locate high concentrations downwind of the power plants. In several instances, such as Monitors G, 6, and 7, the majority of the peak concentrations occur offset 10° - 20° from a downwind line to the monitor from the source. Figure 1 also shows that while monitor 1 is strongly affected by Johnstown, monitor 3 is not. Similarly, monitors 1, 2, and G don't seem to be affected by Pittsburgh. In Table 1b below, the mean wind direction for peak concentrations (in this case, greater than 200 ppb, rather than the 95th percentile), the number of peaks, n, and the standard deviation, σ, are shown. It was observed that these data would produce intersections of the wind vectors at the plants and urban sources, if the wind vectors were shifted a fixed amount. If 26° were subtracted from the monitors 1, 2, 6, 7, 8 and 5° is subtracted from monitors E and F, Figure 2 is produced. Table 2b shows data for 95th percentile violations from Northern Region plants. Subtracting 14° from the mean direction would improve the direction indications of large sources. This type of analysis indicates the turning of the wind field over the region and magnitude (about 2° per sector) of the tower error.

Table 3 lists the monitors wnich have annual wind sector probabilities P(EWD) which are greater than 0.20, and the sources which are suspected of causing their peak concentrations. The peak probability observed is 0.259; this is a conditional probability since it is taken over the sample set of data for which peaks have occurred. The total probability of any peaks occurring at all is about 5%, the probability of a peak occurring in a given sector, given that a peak has occurred, is 1/Number of Sectors. The conditional probabilities associated with plant or urban source peaks are therefore three to eight times as great as random peaks.

The next step after determining the wind sectors with large probabilities and their contributing sources, is to examine the variations of the probabilities in these wind sectors with other meteorological variables such as elevated wind speed EWV, and hour of the day, hr. Three groups of wind sectors with large probabilities are discussed below: 1) the wind sectors associated with Pittsburgh at monitors M and N; 2) the wind sectors associated with Seward at monitor 8; and 3) the wind sectors which are unambiguously associated with one of the large power plants.

For each of these three groups, the wind sector was prespecified,

EWD, and the probability that a peak concentration will occur at a specified time of day, or in a specified wind speed range, $V_1 \leq EWD \leq V_2$, was examined. The results are shown in Tables 4 and 5. The probability ($P0° \leq EWD \leq 90°$, $2.5 \leq hr \leq 6.5$) is the probability of a peak concentration occurring when the wind is from the first quadrant and the hour is between 0201 and 0600.

As indicated in the tables, each of the three groups shows different behavior. Peak concentrations are likely to occur:

- in the morning, if the wind is from Pittsburgh
- in the early afternoon, if the wind is from a large plant
- at almost any time of day, when the wind is from Seward.

Both Pittsburgh and Seward are more likely to cause peak concentrations at higher wind speeds. The large plants seem to have a bimodal wind speed distribution with modes at about 10 and 30 mph. Only 97 recorded peak hourly concentrations in the four unambiguous wind directions existed for the large plants.

The Seward case is of special interest in that it indicates a downwash problem as being the cause of the impacts at Seward. This problem was confirmed by a site visit and observations of the plume behavior. Table 6 shows the cumulative percent of concentrations greater than 200 ppb observed at the Seward monitor for wind directions between 230° and 270°, for the period September 1974 to March 1975. The sharp knee in the cumulative curve corresponds to the initiation of downwash.

Table 6

EWV	5	10	15	20	25	30	35	40	45
%	0	0	2	20	52	76	95	98	100

For the month of January 1975, the problem was reversed and the criteria were given as EDW between 230° and 270° and wind speed EWV greater than 20 mph. Fifty-eight of the 76 observed hourly values of concentrations greater than 200 ppb were predicted with the method. A control strategy just based on that simplistic rule would have been nearly 75% effective.

Monitors M and N were both examined with respect to the effect of Pittsburgh because, as indicated in Figure 1, they have very similar probabilities from the direction of Pittsburgh. During the course of one year, monitor M had 162 peak concentrations (95th percentile again) between 230° and 250°, while monitor N had 167 peaks in the same wind sector. The hypothesis was that these concentrations would occur on the same day with a difference of time of occurrence equal to the transit time from M to N. Fifty percent of the peaks were observed to occur simultaneously, which could be due to a simple persistence effect. The remaining 50% occur on different days. This could be the effect of either a closer source than Pittsburgh, local terrain, or possibly erratic plume puffs striking one monitor, but passing over the other.

Table 1

Annual Hourly Peak Data November 1, 1974 - October 31, 1975

Monitor Wind Sector #1	(EWD)	B		M		8	
		N_v(EWD)	P(EWD)	N_v(EWD)	P(EWD)	N_v(EWD)	P(EWD)
1	1-10°	49	.020	53	0	50	.020
2	11-20°	36	.028	42	0	38	.053
3	21-30°	36	0	33	0	36	.056
4	31-40°	30	0	41	0	48	.063
5	41-50°	45	0	44	0	48	.021
6	51-60°	35	0	34	0	41	0
7	61-70°	53	.057	57	0	60	0
8	71-80°	43	.093	43	.093	51	.039
9	81-90P	64	.031	65	.015	61	.016
10	91-100°	75	.120	79	.076	66	.046
11	101-110°	98	.163	99	.131	90	.011
12	111-120°	116	.129	112	.098	99	0
13	121-130°	137	.088	133	.075	130	.023
14	131-140°	261	.088	287	.038	271	.222
15	141-150°	364	.044	371	.040	360	.028
16	151-160°	219	.050	239	.008	228	.013
17	161-170°	159	.113	175	.046	175	.046
18	171-180°	171	.076	167	.030	162	.049
19	181-190°	120	.075	124	.056	118	.034
20	191-200°	125	.032	126	.016	143	.049
21	201-210°	133	.038	127	.016	123	.073
22	211-220°	210	.029	214	.037	195	.031
23	221-230°	331	.067	307	.055	279	.147
24	231-240°	527	.068	513	.156	463	.251
25	241-250°	459	.065	444	.207	358	.187
26	251-260°	511	.115	476	.172	395	.175
27	261-270°	553	.125	501	.086	432	.116
28	271-280°	385	.094	351	.023	302	.036
29	281-290°	413	.056	417	.029	351	.009
30	291-300°	342	.023	368	.016	307	.008
31	201-210°	279	.022	291	.003	265	.004
32	311-320°	167	.072	.84	.005	170	.012
33	321-330°	99	.081	101	0	99	0
34	331-340°	85	.083	85	.024	83	0
35	341-350°	86	.070	86	.012	75	.027
36	351-360°	51	.020	60	.017	46	0
37	999,888	266	.034	307	.036	314	.041

EWD: elevated wind direction

N_v(EWD): the number of times the wind is from the indicated wind sector.

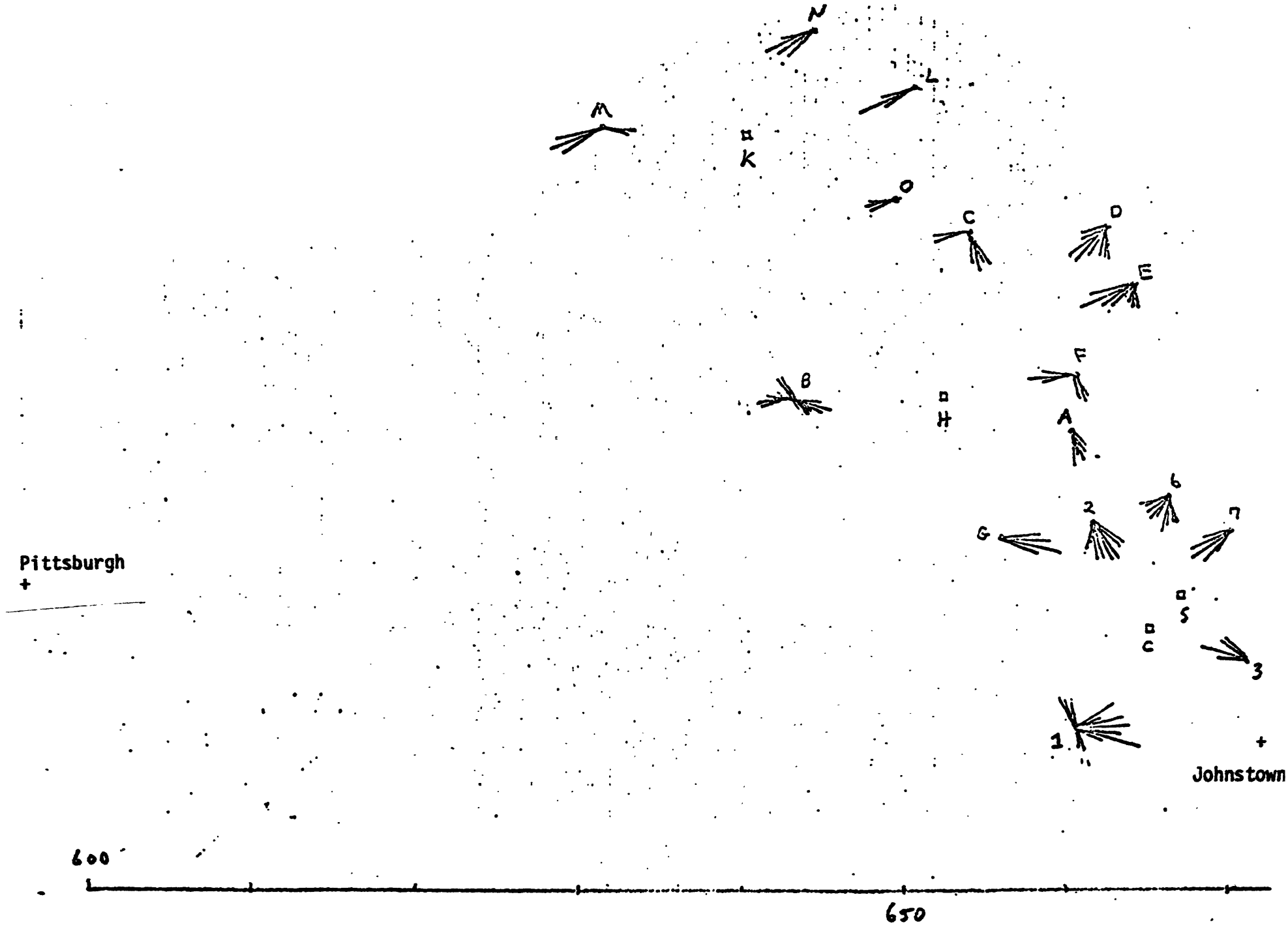
P : the probability of a peak concentration occurring in the indicated wind sector.

Table 1 (continued)

(Winter)

Space Heating Season Peak Hourly Data October 15 - April 15

# Wind Sector	B		M		8	
	N_v(EWD)	P(EWD)	N_v(EWD)	P(EWD)	N_v(EWD)	P(EWD)
1	15	0	18	0	13	0
2	8	0	9	0	4	0
3	11	0	11	0	9	0
4	9	0	11	0	8	0
5	11	0	11	0	8	0
6	7	0	9	0	6	0
7	12	.083	16	0	12	0
8	11	.091	14	.071	9	0
9	24	.042	30	0	16	0
10	43	.070	42	.092	29	0
11	61	.213	65	.154	52	0
12	64	.188	67	.134	47	0
13	74	.122	76	.118	59	.017
14	96	.177	107	.084	76	.013
15	203	.064	193	.067	155	.045
16	124	.048	134	.007	102	.020
17	89	.180	103	.058	81	.049
18	77	.169	76	.053	54	.019
19	60	.117	59	.085	39	.051
20	59	.068	59	.034	40	.100
21	63	.063	63	.032	36	.056
22	91	.044	96	.052	53	.019
23	156	.115	150	.067	88	.264
24	263	.118	263	.232	175	.383
25	251	.092	257	.253	.34	.388
26	317	.126	307	.231	191	.293
27	411	.146	379	.092	272	.162
28	243	.115	230	.026	155	.058
29	287	.066	298	.040	202	.010
30	246	.020	254	.020	179	.006
31	170	.018	189	.005	137	0
32	85	.012	98	.010	64	0
33	39	.103	42	0	30	0
34	39	.103	40	.025	38	0
35	40	.100	43	.023	27	0
36	24	0	28	0	14	0
37	63	.079	80	.063	41	0



+ urban center
° monitor
power plant

Annual Wind Sector Probabilities

Figure 1

Table 1-b

	Johnstown			Plant		
	Mean Direction	σ	n	Mean Direction	σ	n
Monitor 1	149	24	7			
Monitor 2				183	12	11
Monitor 6	165	12	4	230	29	21
Monitor 7	181	20	3	246	15	29
Monitor 8	169	20	8	248	13	117
Monitor E				245	18	12
Monitor F				267	6	13

Table 2-b

	Mean Direction	Direction from Johnstown to Monitor	n	Mean Direction	n	Direction from Pittsburgh to Monitor
Monitor B	175	122	1	249	31	257
Monitor C	159	143	25	259	38	251
Monitor L	168	146	8	259	88	241
Monitor M	143	129	8	250	119	233
Monitor N	175	142	1	246	23	236
Monitor O	175	140	3	254	8	140

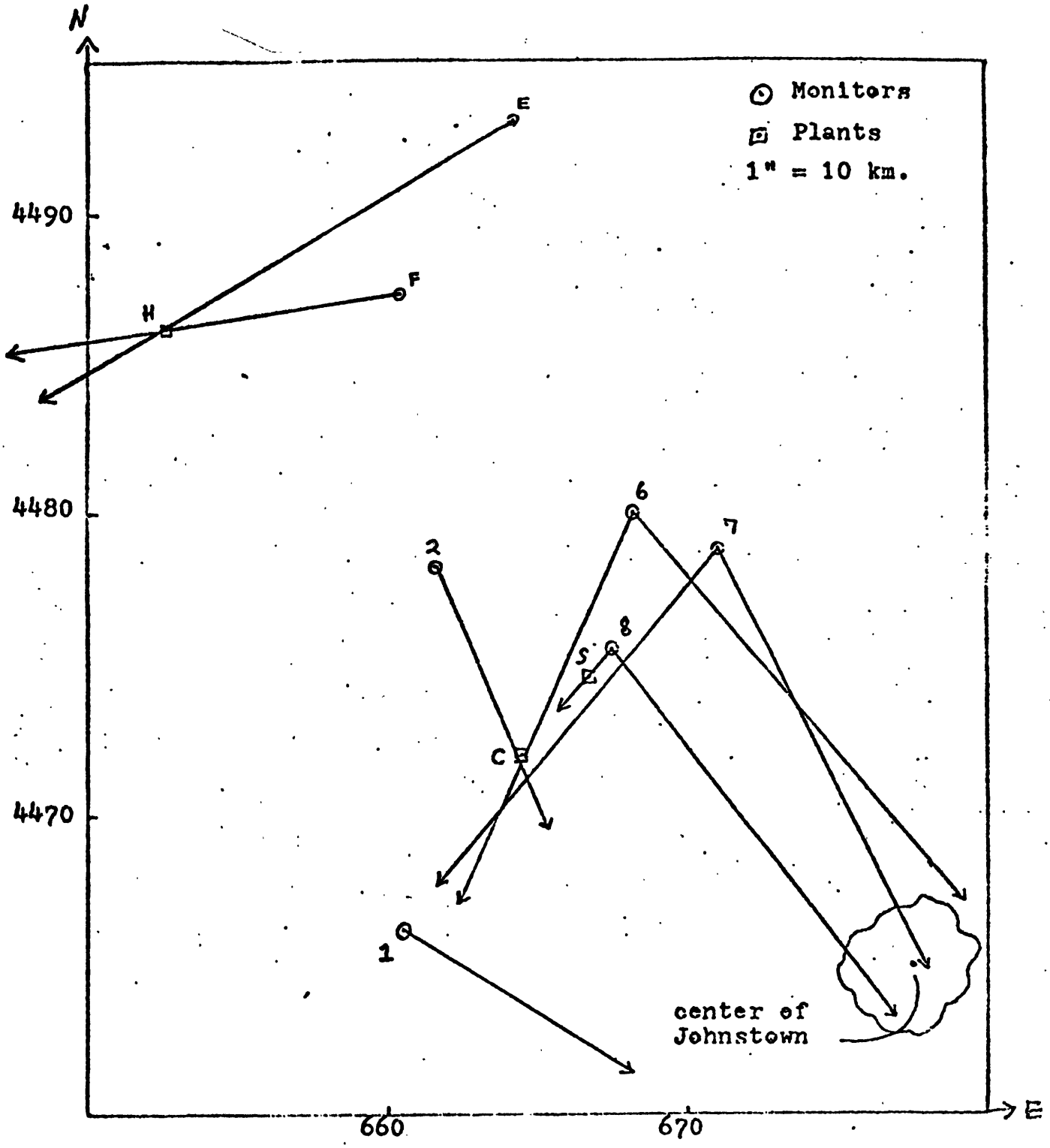


Figure 2

Mean wind direction from Table 1 shifted counterclockwise by:

5° for monitors E and F

26° for monitors 1, 2, 6, 7, and 8

Table 2

November 1974 - October 1975 Annual Data

Monitor	Annual Mean (all data) (ppb)	Mean of Peak Concentrations (ppb)	Maximum 1 hr peak (ppb)	Minimum 1 hr peak (ppb)	Number of 1 hr peaks in 1 yr's data
1	14.36	65	371	39	465
2	26.45	106	394	70	462
3	23.23	102	563	68	462
6	25.09	96	400	63	468
7	27.09	127	606	74	455
8*	48.00	286	917	156	456
A	24.73	83	257	62	461
B	26.00	77	228	61	500
C	15.20	60	197	43	467
D	23.60	80	204	60	500
E	26.40	93	555	65	466
F	30.60	102	272	75	474
G	25.27	82	354	58	468
L	20.91	72	186	54	464
M	25.80	82	191	66	462
N	14.00	57	237	42	455
O	12.73	49	201	34	476

*only 10 months, January - October 1975.

Table 3

P(EWD) > 0.200

Monitor	Wind Sector	P(EWD)	Most Likely Source(s)
1	80-90°	.212	Johnstown
	100-110°	.259	Johnstown
8	230-240°	.251	Seward and Conemaugh Plants
E	240-250°	.222	Pittsburgh and Homer City Plants
G	90-100°	.204	Johnstown, Seward, Conemaugh Plants
	100-110°	.252	
L	250-260°	.227	Pittsburgh and Keystone Plants
M	240-250°	.207	Pittsburgh

Table 4

$P(\theta_1 \leq EWD \leq \theta_2, V_1 \leq EWV \leq V_2)$ Pittsburgh Plume Nov 74-Oct 75

Source Monitor Wind Sector $\theta_1 - \theta_2$ Variable 1 V_1	V_2	M $230 \leq EWD \leq 250$ N_v	P	N $230 \leq EWD \leq 250$ N_v	P	8 $230 \leq EWD \leq 240$ N_v	P	8 $220 \leq EWD \leq 270$ N_v	P
-0.5	2.5	1	0	1	0	0	-	7	.143
2.5	5.5	47	.085	47	.064	20	.050	77	.143
5.5	8.5	95	.095	95	.084	27	0	185	.032
8.5	11.5	124	.170	124	.162	46	.109	202	.064
11.5	14.5	160	.068	160	.137	52	.192	220	.105
14.5	17.5	181	.083	181	.110	64	.297	247	.146
17.5	20.5	154	.189	154	.176	51	.333	210	.238
20.5	23.5	138	.159	138	.080	41	.293	181	.298
23.5	26.5	106	.227	106	.199	43	.349	119	.353
26.5	29.5	78	.205	78	.179	33	.333	82	.329
29.5	32.5	44	.135	44	.290	21	.286	50	.300
32.5	35.5	30	.167	30	.267	13	.462	37	.486
35.5	38.5	6	0	6	0	4	1.000	14	.571
38.5	41.5	2	0	2	0	1	1.000	9	.444
41.5	44.5	1	0	1	0	1	1.000	5	.200
49.5	∞	0	-	0	-	0	-	4	.250

Table 4 (continued)

$P(\theta_1 \leq EWD \leq \theta_2, V_1 \leq EWV \leq V_2)$ Unambiguous wind sectors from large plants Nov 74 - Oct 75

Source Monitor Wind Sector $\theta_1 - \theta_2$	Variable 1	M		N		8		8		Average of all four P's
V_1	V_2	$230 \leq EWD \leq 250$ N_v	P	$230 \leq EWD \leq 250$ N_v	P	$230 \leq EWD \leq 240$ N_v	P	$220 \leq EWD \leq 270$ N_v	P	
-0.5	2.5	2	0	4	0	2	0	1	0	0
2.5	5.5	29	0	15	.133	29	.069	15	.067	.067
5.5	8.5	40	.100	49	.204	40	.100	46	.130	.134
8.5	11.5	46	.217	43	.116	46	.196	35	.171	.174
11.5	14.5	74	.054	26	.038	74	.068	26	.038	.050
14.5	17.5	49	.041	13	.077	49	0	11	0	.030
17.5	20.5	45	.022	13	0	45	.022	8	0	.011
20.5	23.5	44	.068	11	0	44	0	6	0	.017
23.5	26.5	19	.158	3	0	19	0	2	0	.040
26.5	29.5	19	.158	4	0	19	0	1	0	.040
29.5	32.5	4	.250	1	1.000	4	0	0	-	.417
32.5	35.5	6	.333	3	0	6	0	1	0	.083
35.5	38.5	2	0	3	1.000	2	0	3	1.000	.500
38.5	41.5	5	0	1	-	5	0	0	-	0
41.5	44.5	2	.500	1	1.000	2	0	1	0	.375
44.5	∞	1	1.000	0	-	1	0	0	-	.250

N_v is the number of times the wind is in the sector indicated and the speed range indicated.

P is the probability the α peak concentration will occur where the wind is in the indicated sector (sector and the speed range indicated)

Table 5

$P(\theta_1 \leq EWD\ \theta_2,\ t_1 \leq hr \leq t_2)$

Source Monitor	M	N	
Wind sector	230 ≤ EWD ≤ 250	230 ≤ EWD ≤ 250	220 ≤ EWD ≤ 270
1 - 2	P(EWD, hr)	P(EWD, hr)	P(EWD, hr)
Variable N_v	194.5	194.5	87
t_1 t_2			
23 2	.057	.026	.14
3 6	.077	.026	.25
7 10	.129	.057	.30
11 14	.206	.190	.30
15 18	.098	.118	.28
19 22	.098	.036	.22
	-Pittsburgh November 1974 - October 1975-		-Seward, January 1975-

Note C at 87200 (97th, 98th, 99th percentiles)

Source Monitor	A	B	D	M	Avg.
Wind Sector	280 ≤ EWD ≤ 290	90 ≤ EWD ≤ 110	280 ≤ EWD ≤ 290	100 ≤ EWD ≤ 115	of all 4
Variable	P(hr)	P(hr)	P(hr)	P(hr)	P(hr)
N_v	64.5	31.5	64.5	2.6	-
t_1 t_2					
23 2	.031	0	0	0	.008
3 6	.016	.095	0	0	.028
7 10	.155	.286	.031	.154	.157
11 14	.279	.254	.171	.346	.263
15 18	.109	.159	.155	.151	.144
19 22	.031	.032	.031	0	.024

P(EWD, hr) is the probability that a peak concentration will be observed when the wind is in the indicated sector and the time period (hour) indicated.

N_v = 1/6 (# hours wind in sector θ_1 (EWD ≤ θ_2) or one-sixth of the number of hours the wind is from the indicated sector.

APPENDIX C

CALCULATING SOURCE REDUCTION BY LARSEN'S METHOD

Larsen (Larsen, 1967, 1971, 1973, 1974; see references above) recommends a method for determining needed reductions in an area through the calculation of design values of expected concentrations. His procedure requires an assumption of log normality in concentration data. This appendix applies Larsen's method to several of the Chestnut Ridge monitors to determine the need for permanent reductions in SO_2 emissions in the area. Monitors 1, 7, 8, B, F, G, L, M, and N are considered. All concentrations are in ppm.

Log Normality

The distribution of hourly SO_2 data at each monitor was determined for the September 1974 - December 1975 period. The cumulative frequency distributions were then calculated and plotted on log-probability paper. All monitors showed close to straight-line behavior, indicating nearly log normal distributions. At both the high and low tails, the distributions departed slightly for log normality. This is not unusual and the general assumption of log normality will be used.

Dominant Standard

In order to determine the controlling averaging period, Larsen develops breakpoints of the standard geometric deviation for daily averages. Using his method, breakpoints for hourly averages can be found:

Dominant Standard	SGD Day	SGD Hour
1 yr	0 - 2.05	0 - 2.36
24 hr	2.35 - 4.50	2.36 - 6.01
3 hr	4.50 - ∞	6.01 - ∞

Standard geometric deviations (S_g) for each of the monitors can be calculated if two points for the same averaging time are available:

$$S_g = \exp\{\ln(C_a/C_b)/(z_a - z_b)\} \qquad (1)$$

C_a, C_b = concentrations at a and b on the log-probability chart

z_a, z_b = number of deviations from median

Using the 16th (z=1.0) and 50th (z=0) percentile C's, the equation becomes:

$$S_g = C_a/C_b \qquad (2)$$

The values taken from the straight line approximations to the monitor data are as follows:

Table 1

Monitor	C_a	C_b	S_g	Dominant Stnd.
1	.027	.010	2.70	24 hr
7	.048	.018	2.67	24 hr
8	.082	.014	3.42	24 hr
B	.043	.024	1.79	1 yr
F	.052	.024	2.17	1 yr
G	.041	.019	2.16	1 yr
L	.038	.024	1.58	1 yr
M	.047	.024	1.96	1 yr
N	.030	.013	2.31	1 yr

Design Values

In order to determine design values, Larsen next requires a differentiation between standards whose averaging time scales are the same as or different from the averaging time scales of the data. Since Chestnut Ridge data is averaged on an hourly basis, the "different averaging times" method is used. Using Larsen's relation:

$$C_1 = C_2(C_2/C_3)^E \qquad (3)$$

C_1 = concentration at yearly max hour

C_2 = concentration at 10^{th} percentile

C_3 = concentration at 30^{th} percentile

$E = (z_1 - z_2)/(z_2 - z_3) = .2817$

the following table (Table 2) can be developed for expected annual maximum hours at the various monitors:

Table 2

Monitor	$C_2(0.1)$	$C_3(.30)$	C_1	S_g*
1	.20	.017	.40	2.70
7	.59	.031	1.35	2.67
8	1.00	.041	2.46	3.42
B	.16	.033	.25	1.79
F	.24	.036	.41	2.17
G	.19	.029	.32	2.16
L	.16	.028	.26	1.58
M	.19	.034	.31	1.96
N	.18	.021	.33	2.31

*from Table 1 above.

Larsen's Table III (Larsen, 1973) can then be used to find the expected arithmetic mean and the expected maximum for any averagin time (Table 3).

Table 3

Monitor	Expected Arith. Mean	3 hr Max	24 hr Max	5 Yr Max Annual
1	.015	.268	.123	.022
7	.052	.903	.417	.076
8	.048	1.492	.584	.075
B	.032	.193	.119	.041
F	.029	.296	.158	.039
G	.023	.232	.125	.031
L	.050	.211	.142	.061
M	.030	.233	.134	.039
N	.019	.229	.117	.026

These numbers can be compared with those actually observed in 1975 (Table 4), recalling that 1975 represents only a single sample, and had only 83.6% SO_2 data capture.

Table 4

Monitor	C_1	Arith. Mean	3 hr Max.	24 hr. Max.
1	.371	.015	.253	.083
7	.681	.027	.552	.219
8	.917	.048	.766	.279
B	.228	.026	.153	.086
F	.272	.031	.220	.115
G	.323	.024	.137	.084
L	.186	.022	.192	.085
M	.191	.025	.148	.090
N	.315	.016	.167	.080

It should also be noted that the Meloy total sulfur detector used "pins" at +10v (1 ppm) so values such as those predicted by Larsen at Monitors 7 and 8 could not be reocrded even if they occurred.

Larsen presents a method in an earlier paper for finding needed source reduction (R) to meet a standard (q), given present concentrations (c), and a growth rate of concentrations (g).

$$R = \frac{(100\%)\ (gc-q)}{gc} \qquad (4)$$

Assuming g = 1 to examine the present we find the reductions needed in Table 5 (using 5 year max. for annual case):

Table 5

Monitor	3 hr	24 hr	Annual
1	-	-	-
7	44.6%	66.4%	60.5%
8	66.5%	76.0%	60.0%
B	-	-	26.8%
F	-	11.4%	23.1%
G	-	-	3.2%
L	-	1.4%	50.8%
M	-	-	23.1%
N	-	-	-

Of course, equation (4) does not account for background (b) which is an uncontrollable part of the observed concentrations. Assuming that the source alone must account for all of the reduction yields the following equation:

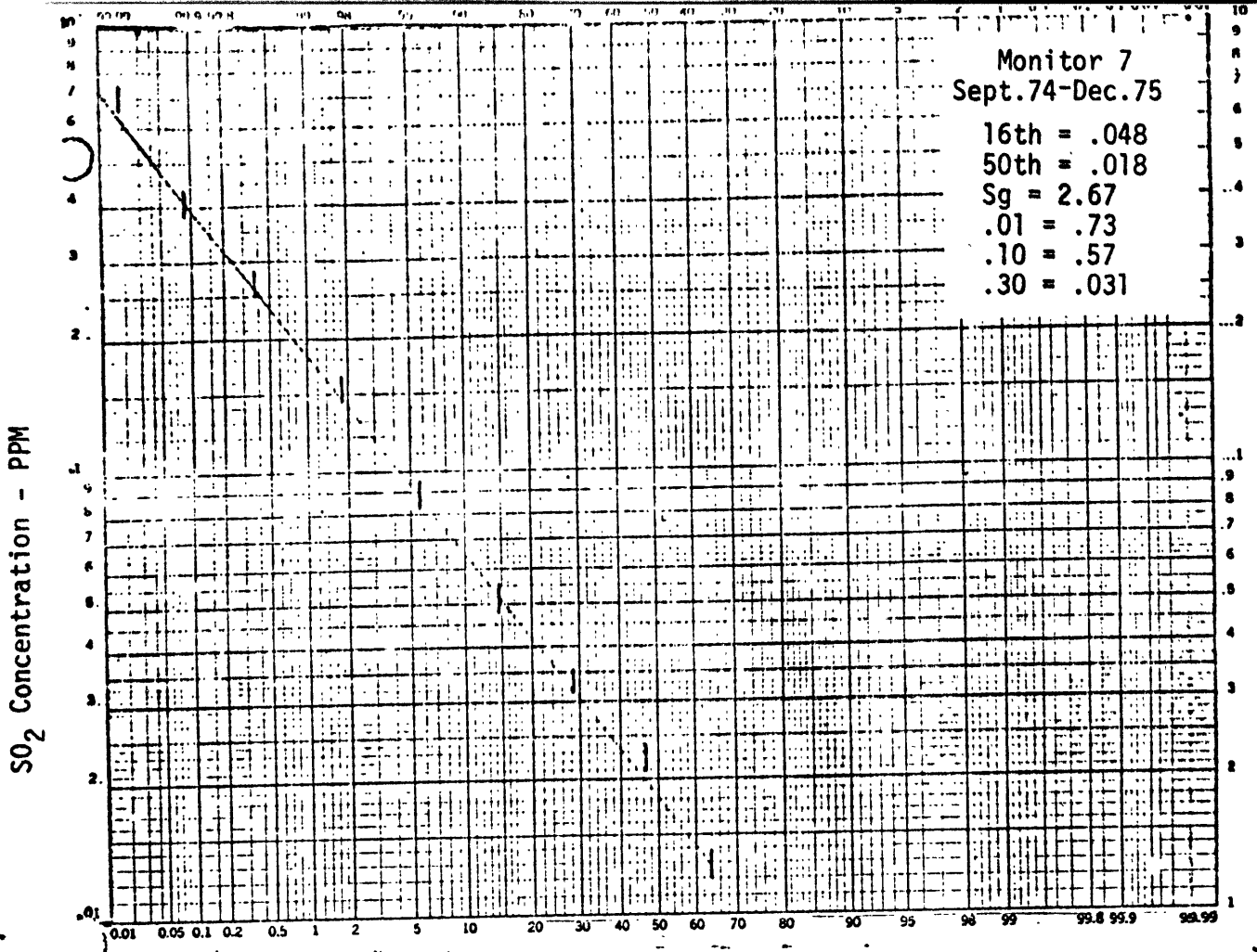
$$R = \frac{(100\%)(gc-q)}{gc-b} \qquad (5)$$

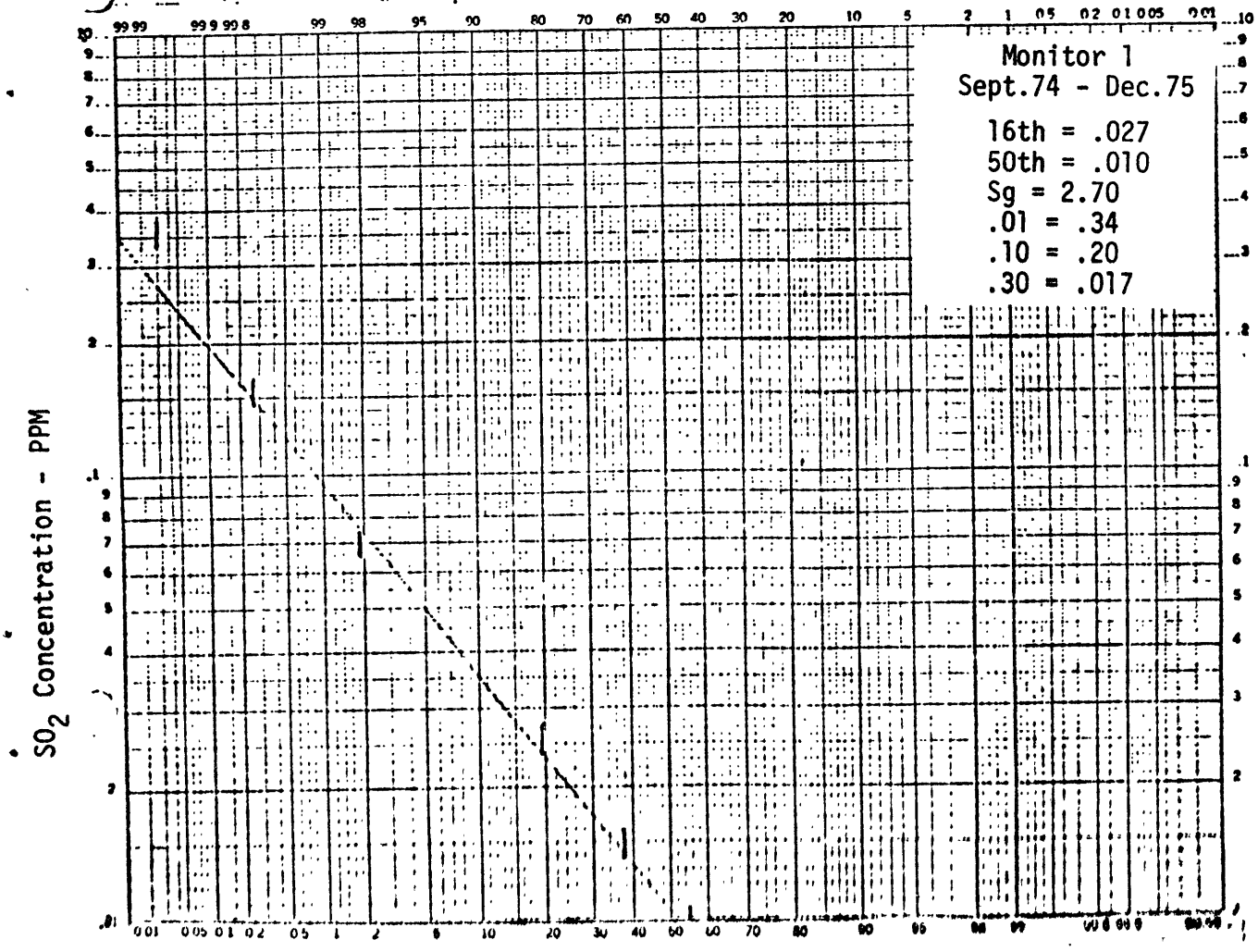
This varies as 1/(1-b) when b is expressed as a percentage of the observed concentration.

Since background levels will contribute the greatest percentage to the lowest (quantitative) standards, background will cause the greatest plant reduction effect when longer-term standards dominate. This is the case at Chestnut Ridge and implies that any permanent emission reduction at the Chestnut Ridge plants would penalize the plants for the background (e.g., if background is 0.015 ppm or 50% of the annual 0.030 ppm allowable, plant reductions would be twice as stringent as when the plant reduces only its proportion of the excess). This line of thinking raises complex legal "chicken before the egg" questions. (e.g., Has the plant "added" to existing background, and is it therefore responsible for reducing all the excess? Or do the plant and the background sources "share" the resource?) The current EPA interpretation clearly puts the control burden on the plant. Looking at monitor B as an upwind "background" monitor, we see that the expected arithmetic mean already exceeds the standard. Even the lowest monitor value, at monitor 1, is 50% of the annual limit.

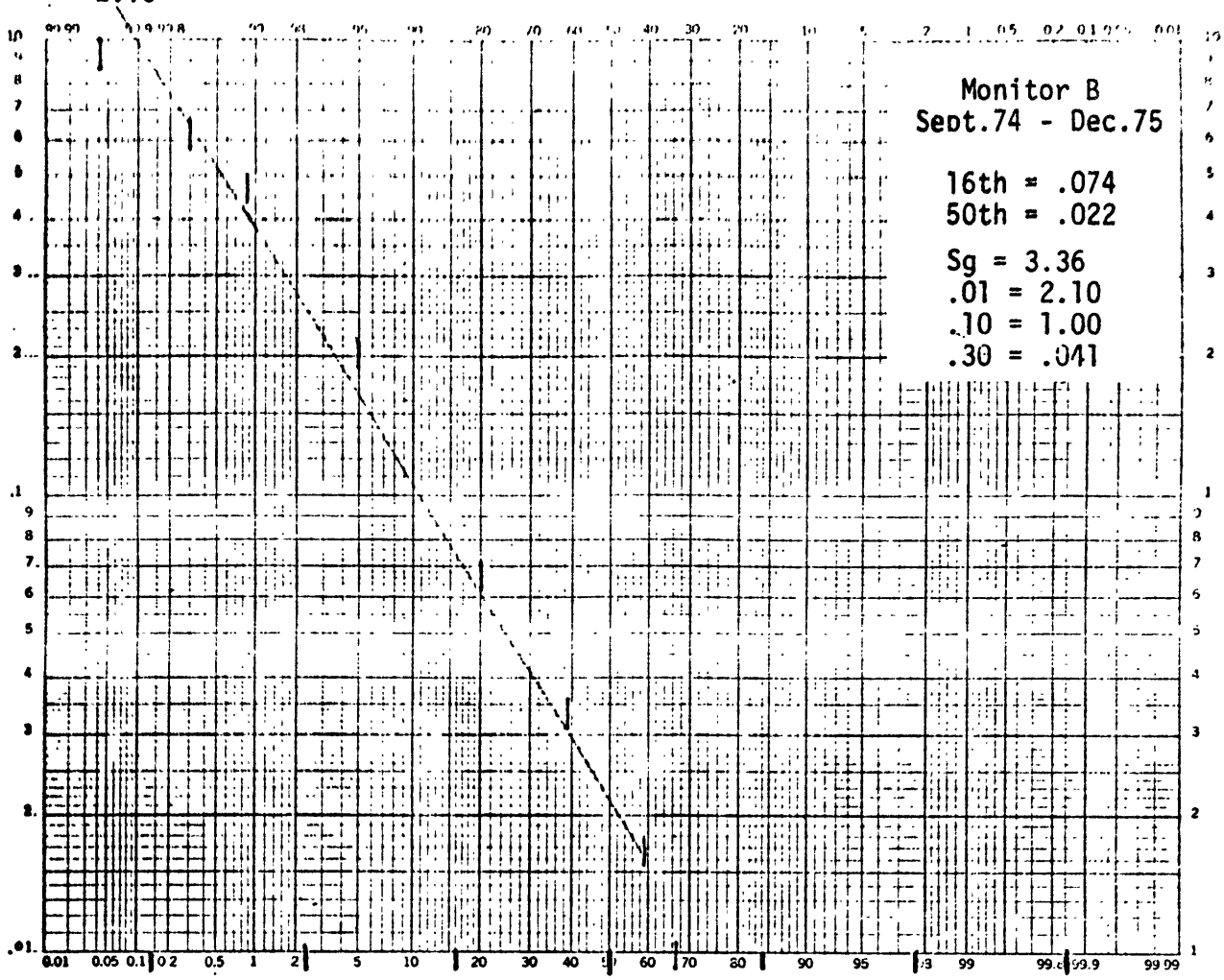
Conclusions

Larsen's method and Chestnut Ridge monitor data show that the 24 hr. and annual standards dominate in determining emissions reductions at Chestnut Ridge. Total emission reductions as high as 76% could be needed. Considering background leads to a prediction that some plants would have to be shut down. 1975 observed Chestnut Ridge data does not agree with Larsen's predictions: violations have occurred, but could have been corrected with less control. Limit problems restricting maximum recorded concentrations by the monitors may have introduced bias in these results.

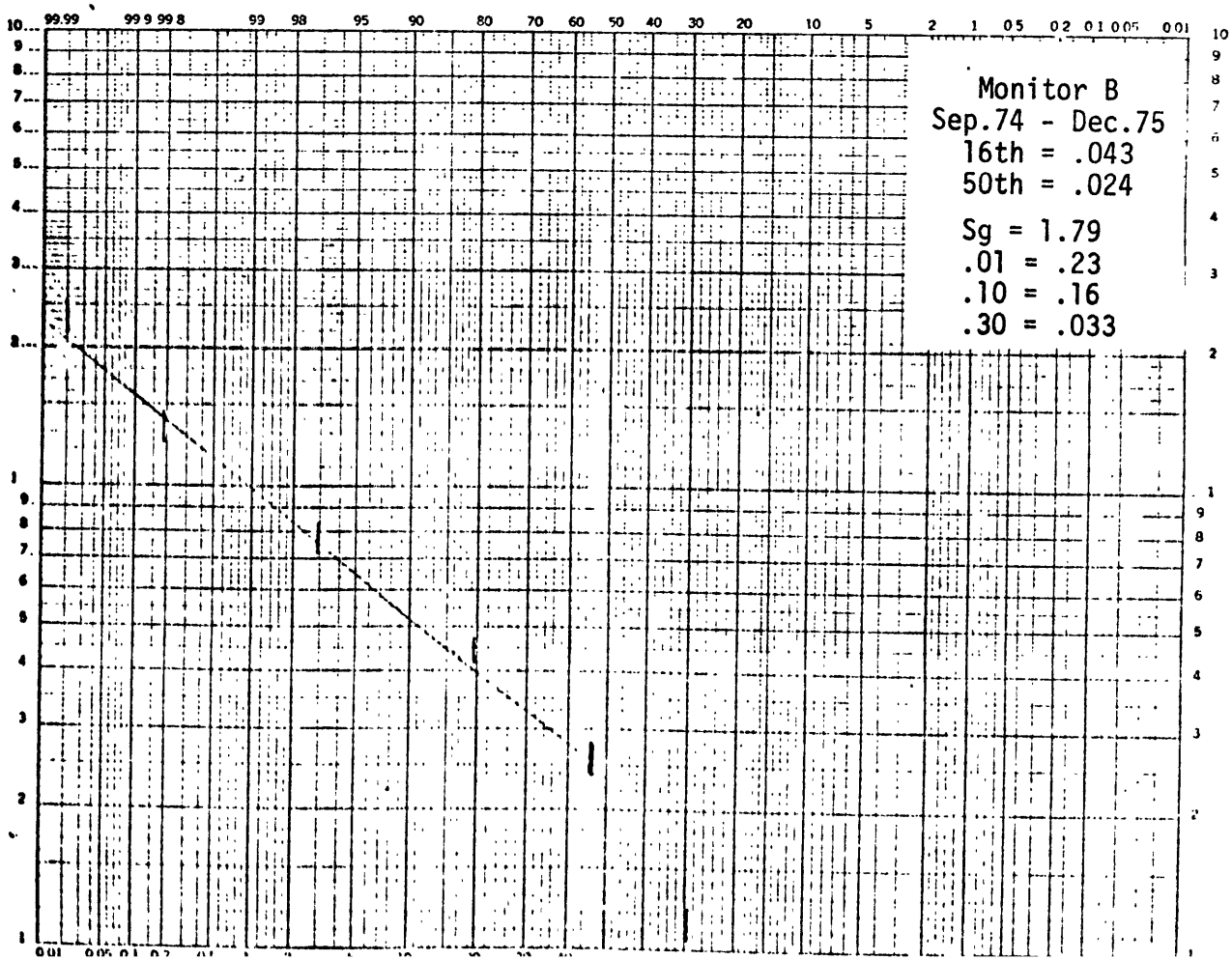

Monitor 7
Sept.74-Dec.75
16th = .048
50th = .018
Sg = 2.67
.01 = .73
.10 = .57
.30 = .031
SO_2 Concentration - PPM


Monitor 1
Sept.74 - Dec.75
16th = .027
50th = .010
Sg = 2.70
.01 = .34
.10 = .20
.30 = .017
SO_2 Concentration - PPM

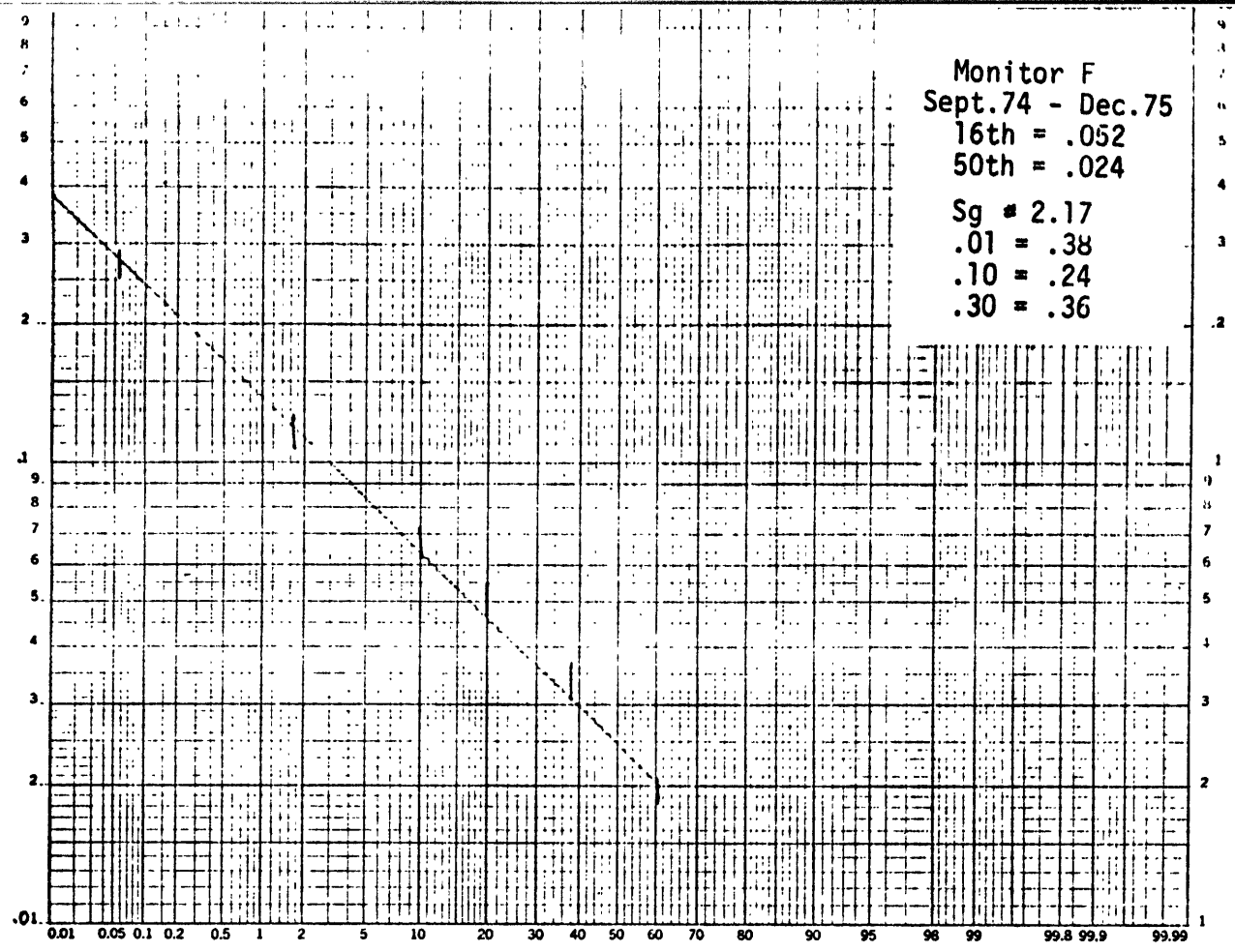
2.10

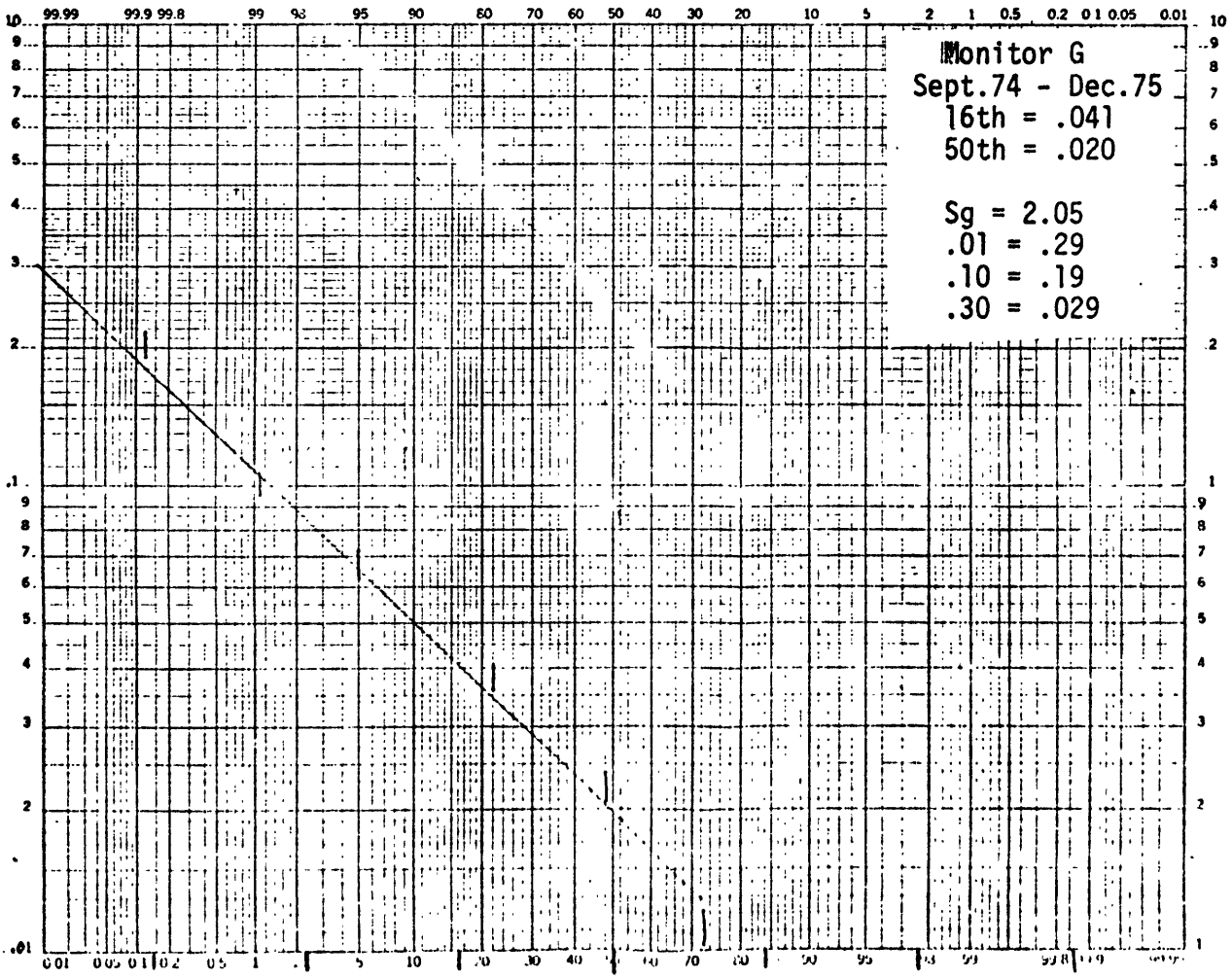


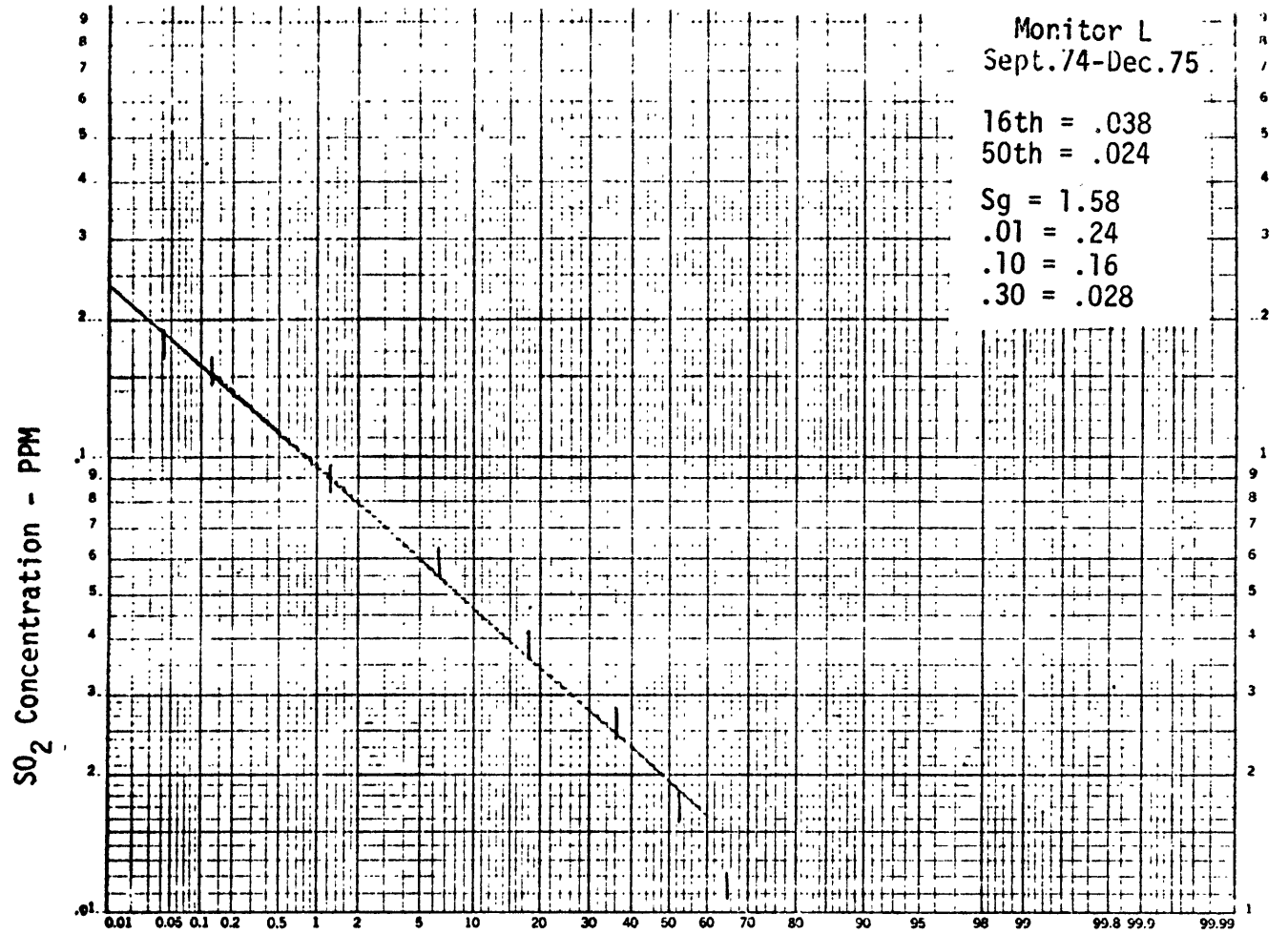
SO_2 Concentration - PPM

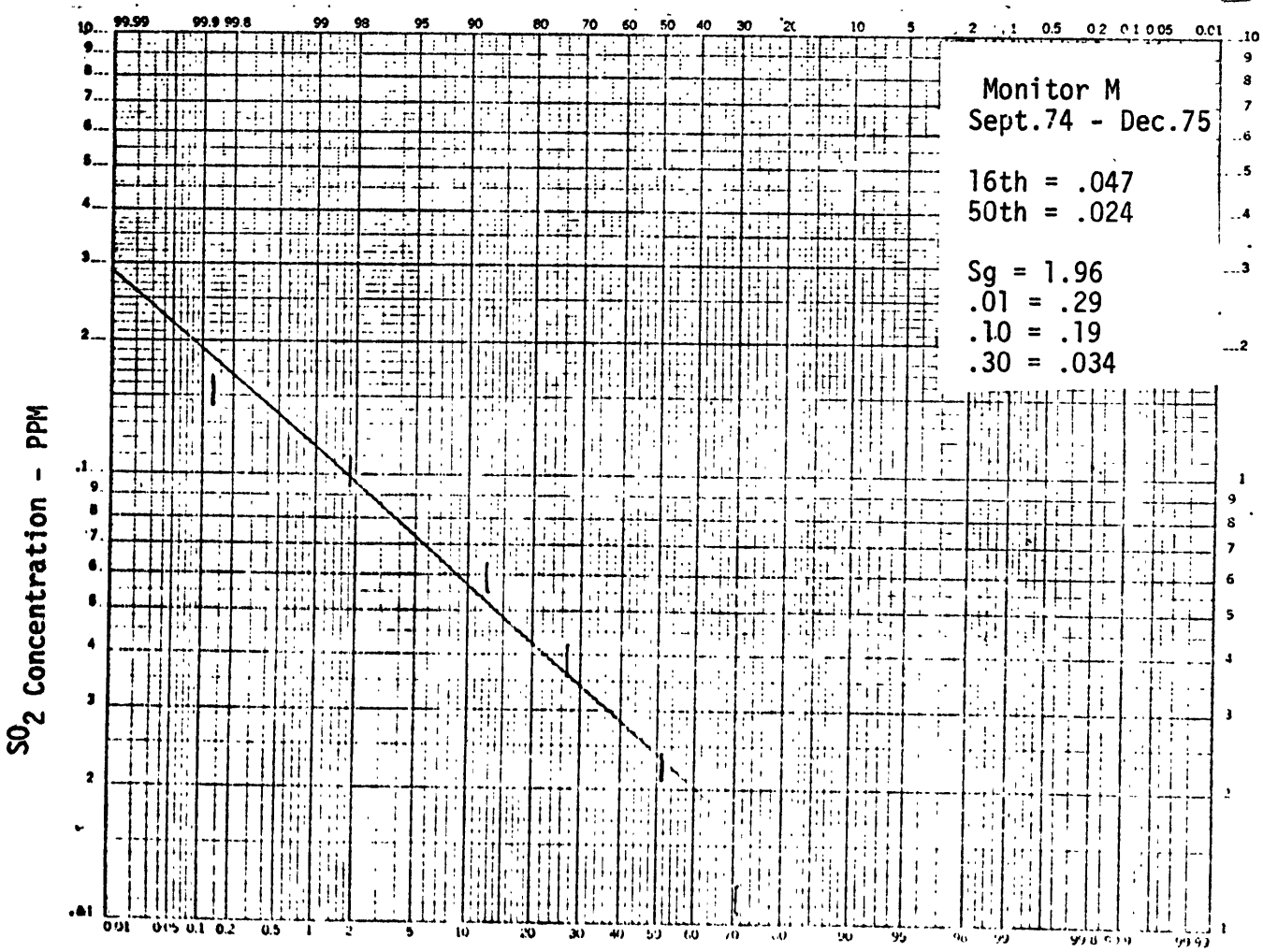


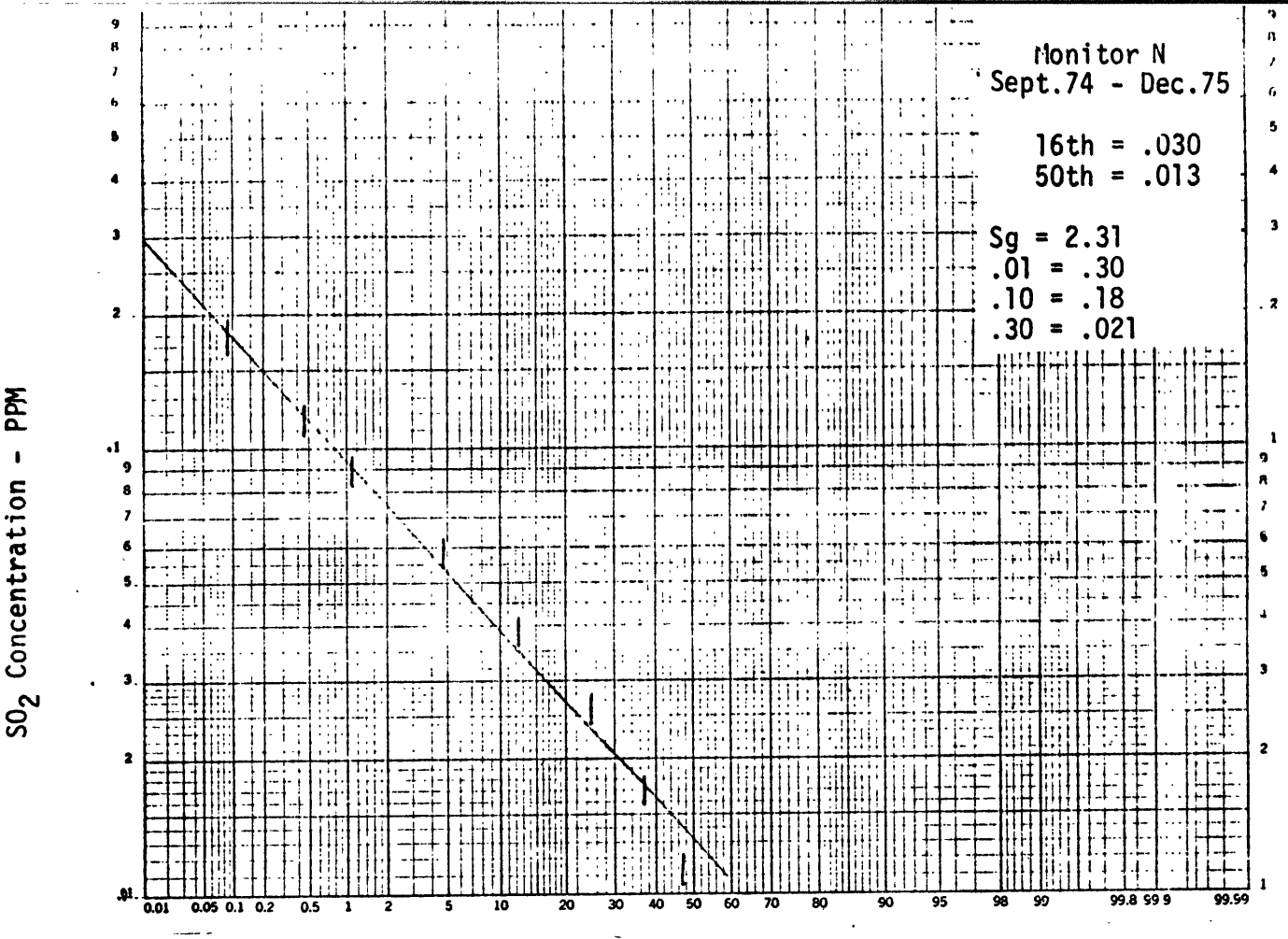
SO_2 Concentration - PPM

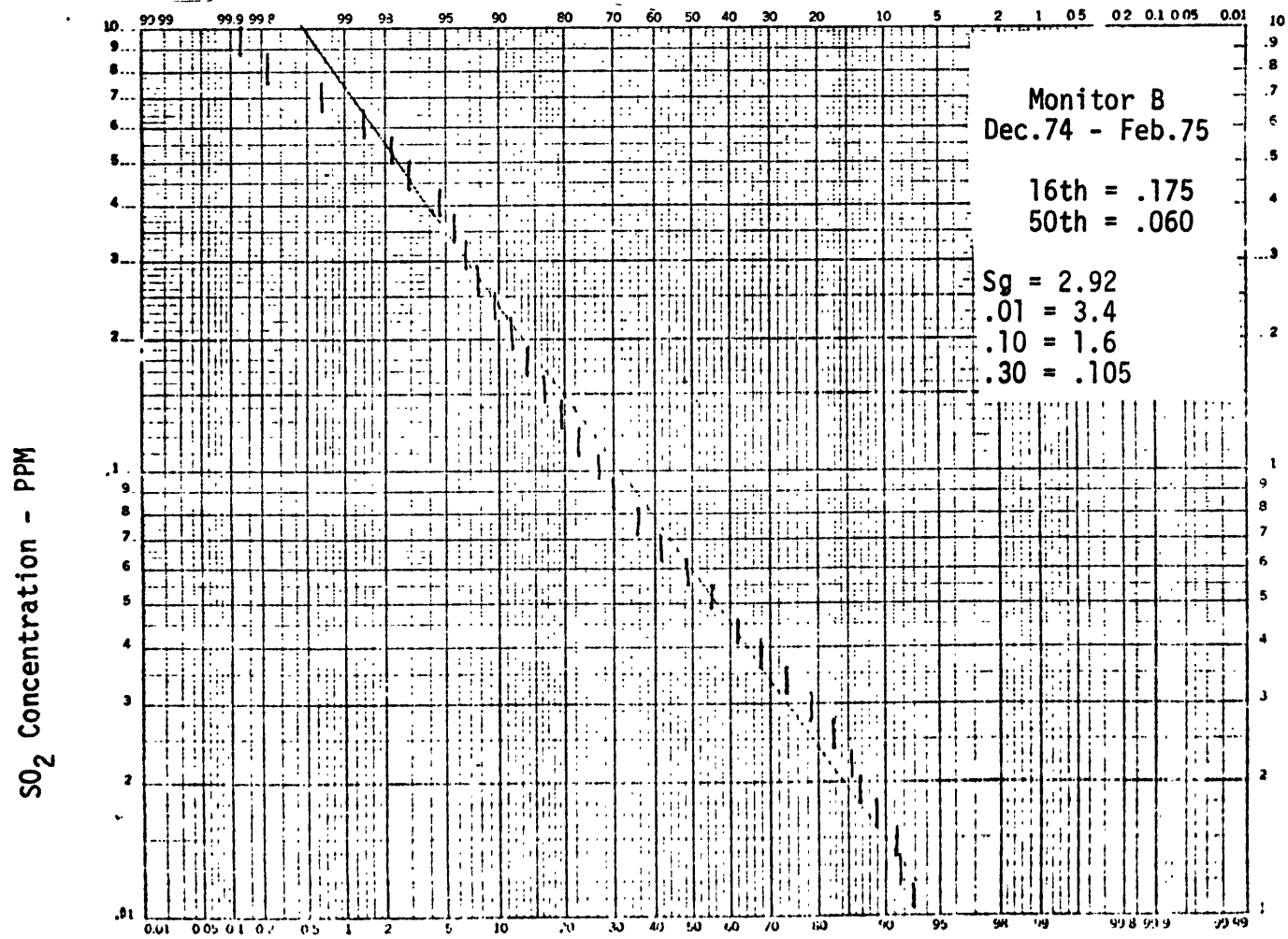

Monitor F
Sept.74 - Dec.75
16th = .052
50th = .024
Sg = 2.17
.01 = .38
.10 = .24
.30 = .36
SO2 Concentration - PPM
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99


Monitor G
Sept.74 - Dec.75
16th = .041
50th = .020
Sg = 2.05
.01 = .29
.10 = .19
.30 = .029
SO2 Concentration - PPM
99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01


Monitor L
Sept.74-Dec.75
16th = .038
50th = .024
Sg = 1.58
.01 = .24
.10 = .16
.30 = .028
SO2 Concentration - PPM


Monitor M
Sept.74 - Dec.75
16th = .047
50th = .024
Sg = 1.96
.01 = .29
.10 = .19
.30 = .034
SO2 Concentration - PPM


Monitor N
Sept.74 - Dec.75
16th = .030
50th = .013
Sg = 2.31
.01 = .30
.10 = .18
.30 = .021
SO_2 Concentration - PPM


Monitor B
Dec.74 - Feb.75
16th = .175
50th = .060
Sg = 2.92
.01 = 3.4
.10 = 1.6
.30 = .105
SO_2 Concentration - PPM

APPENDIX D

MATHEMATICAL BACKGROUND MODEL

Table of Contents

PART II MATHEMATICAL DETAILS

1. Introduction

A mathematical model for the Chestnut RIdge area background SO_2 concentration is stated and its behavior and performance are discussed. The modeling methodology uses measurements made at multiple monitors to develop a weather dependent background model that is applicable over all space and time (hour by hour). Source inventories are not used.

The structure of the model evolved from a long "learning process." The initial starting point was exploratory data analysis which revealed the type of behavior actually present in the data. A series of ever more complex mathematical models were then hypothesized and tested. This evolutionary process has progressed to a point where it is felt to be worthwhile to present the results because:

- they provide insight into the nature of the Chestnut Ridge background;
- the modeling methodology has general value.

However, the "best available" statistical parameter estimation computer codes were not used and only a limited number of model verification tests were actually applied. Thus the model itself is not considered to be "final". It was not explicitly incorporated into the Chestnut Ridge SCS air quality model.

The discussions are divided into two parts. Part I, "General Dicussion," consists of Sections 2 to 7 and discusses the modeling process and the results in general terms. Part II, "Mathematical Details," consists of Sections 8 to 10 and contains some of the more explicit equations needed to completely specify the model and modeling process. References can be found above in the general reference section.

Part I

General Discussion

2) Definitions
3) Nature of Hypothesized Structure
4) Uses of Model
5) Parameter Estimation
6) Model Validation
7) Model Behavior

2. Definitions

Some general terms are defined as a prelude to explicit discussions on background air quality modeling.

It is desired to obtain a mathematical "model" which is considered to consist of two parts:

- Structure: Specific form of the equations
- Parameter values: Values of parameters needed to completely specify the model.

In the case of the SO_2 background model to be discussed the model's input-outputs are:

- Output:
  - Hourly average SO_2 concentrations for any hour and spatial coordinate (2-dimensional horizontal) in region being modeled.
- Input:
  - Time of day, day of year
  - Measured hour by hour elevated tower wind directions.

The process of developing a mathematical model can be viewed as the four step process of Figure 2.1. The fourth step, "Diagnostic Analysis" of Figure 2.1 is needed when the "Hypothesized Structure" (first step) is not valid so that the "validity tests" (third step) are not passed. Many "iterations around the loop" of Figure 2.1 were actually done but only the final hypothesized structure is discussed.

3. Nature of the Hypothesized Structure

Define:

n: time index n=1, 2, 3....hours

$\underline{s}$: two-dimensional spatial coordinate (horizontal)

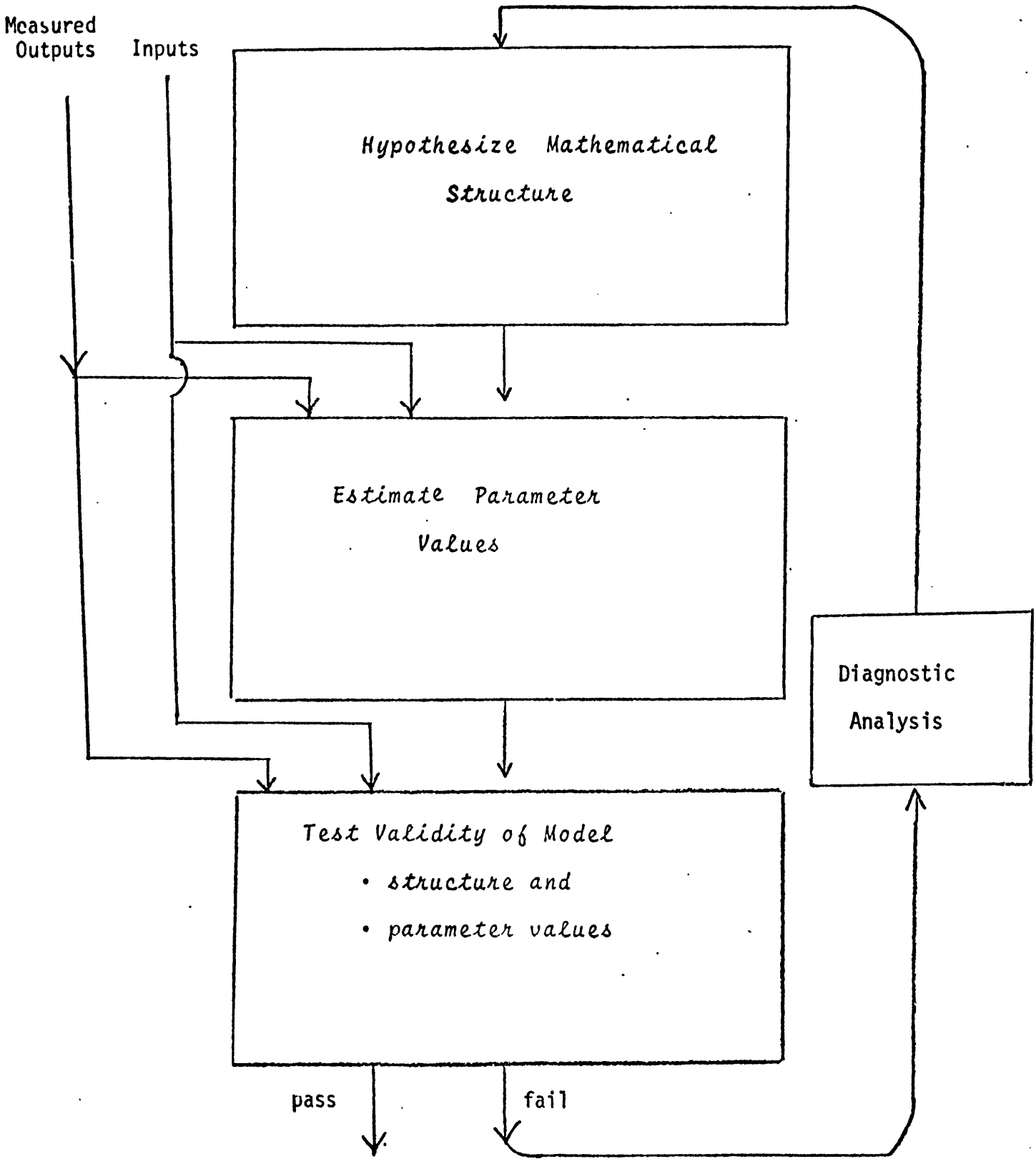
$c(n,\underline{s})$: hourly average concentration at hour n at spatial coordinate $\underline{s}$.

$\theta(n)$: elevated tower wind direction measured at hour n.

The hypothesized structure of the model for $c(n,\underline{s})$ is:

$$c(n,\underline{s}) = m_g(n,\underline{s})\ \Delta_g(n)\ m_\ell(\underline{s})\ \Delta_\ell(n,\underline{s})c_o \qquad (3.1)$$

$m_g(n,\underline{s})$: global concentration field, mean value



Development of Mathematical Models

Figure 2.1

$\Delta_g(n)$: global concentration field: random component

$m_\ell(\underline{s})$: local variation: mean value

$\Delta_\ell(n,\underline{s})$: local variation: random component

c_o: constant

Taking the logarithm (natural) of (3.1) yields

$$\ln c(n,\underline{s}) = \ln m_g(n,\underline{s}) + \ln \Delta_g(n) + \ln m_\ell(\underline{s}) + \ln \Delta_\ell(n,\underline{s}) + \ln c_o \qquad (3.2)$$

The two random components $\Delta_g(n)$ and $\Delta_\ell(n,\underline{s})$ are hypothesized to be log-normal processes. Thus $\ln \Delta_g(n)$ and $\ln \Delta_\ell(n,\underline{s})$ are normal random processes. The "mean values" are defined so that

$$E\{\ln c(n,\underline{s})\} = \ln m_g(n,\underline{s}) + \ln m_\ell(\underline{s}) + \ln c_o \qquad (3.3)$$

when "E" denotes expectation with respect to the random processes $\Delta_g(n)$ and $\Delta_\ell(n,\underline{s})$

The constant $\ln c_o$ can be viewed as the average of $\ln c(n,\underline{s})$ over all time, space, weather conditions, and background source variations. From a dimensional point of view, c_o has the same units as $c(n,\underline{s})$. All the other terms of (3.1) are dimensionless, multiplicative "correction factors" which vary about unity.

3.1 Global Concentration Field: Mean Value

The global concentration field, mean value term, $m_g(n,\underline{s})$ of (3.1) has inputs:

. $\theta(n)$, measured elevated tower wind direction

. time of day, day of year.

It is a dynamic process (has memory) to account for the fact that an instantaneous change in wind direction does not cause a corresponding instantaneous "jump in concentrations" because of the finite "transport time" from the background sources to the geographic region being modeled. The following simple structure is hypothesized to account for this dynamic phenomenon:

$$\ln m_g(n+1,\underline{s}) = \phi_g \ln m_g(n,\underline{s}) + (1-\phi_g) \ln b_g(n+1,\underline{s}) \qquad (3.1.1)$$

$$0 \leq \phi_g < 1$$

The value of the parameter ϕ_g is to be estimated from the observations. If $\phi_g = 0$ or if $b_g(n,\underline{s})$ is constant in time n, then,

$$m_g(n,\underline{s}) = b_g(n,\underline{s})$$

The "input" $b_g(n,\underline{s})$ to (3.1.1) represents the combined multiplicative effects of wind direction and time on mean concentrations, and has the form

$$b_g(n,\underline{s}) = b_\theta[\underline{s},\ \theta(n)]\ b_{h,d}(n) \tag{3.1.2}$$

where

$b_\theta[\underline{s},\theta(n)]$: models dependence on elevated tower wind direction $\theta(n)$

$b_{h,d}(n)$: models dependence on time of day and day of year

$h = 1...24$ (hours of day)

$d = 1...365$ (day of year)

The explicit equations for the structure of $b_\theta[\underline{s},\ \theta(n)]$ and $b_{h,d}(n)$ are quite complex so they are discussed in Part II (Section 8) which contains mathmatical details. As will be described in Section 7.5, attempts to introduce wind speed and mixing height dependence into $b_g(n,\underline{s})$ were not successful.

The equations of Section 8 show that for uniformly distributed wind direction $\theta(n)$, and large N.

$$\frac{1}{N} \sum_{n=1}^{N} \ln b_g(n,\underline{s}) = 0 \tag{3.1.3}$$

$$\frac{1}{N} \sum_{n=1}^{N} \ln m_g(n,\underline{s}) = 0$$

3.2 Global Concentration Field: Random Component

The global concentration field, random component, term $\Delta_g(n)$ of (3.1) is hypothesized to have the following structure:

$$\ln \Delta_g(n+1) = \phi_{g\Delta} \ln \Delta_g(n) + w_g(n) \tag{3.2.1}$$

$w_g(n)$: zero mean, white normal random process

$E\{w_g(n)\} = 0$

$$E\{w_g(n_1)\ w_g(n_2)\} = \begin{cases} R_g & n_1 = n_2 \\ 0 & n_1 \neq n_2 \end{cases}$$

$0 \leq \phi_{g\Delta} < 1$

Thus $\ln \Delta_g(n)$ is a first order, normal Markov process. The values of the parameters $\phi_{g\Delta}$ and R_g are to be estimated from the observations.

Note that $\Delta_g(n)$ does not depend on the spatial coordinate $\underline{s}$. Thus $\Delta_g(n)$ is the same for all $\underline{s}$. Physically $\Delta_g(n)$ can be thought of as modeling a random variation of the overall concentration field "up and down" uniformly over all space. Sources of such overall variations are random background source emission rates and random transport times from distance background sources to the region being modeled.

3.3 Local Variations: Mean Values

The local variations, mean value term, $m_\ell(\underline{s})$ of (3.1) was introduced into the hypothesized structure to account for the observed phenomenon that long term concentration averages (over time, wind direction, etc.) were not equal at different monitor locations. This can correspond to physical phenomena such as local terrain factors (such as height) or, when using the model to account for monitored measurements, it can model the effect of monitor calibration bias. Local sources of background emissions are also represented by this term. For the present discussions, models for $m_\ell(\underline{s})$ only at monitor locations are considered.

Assume there are K monitors at coordinates $\underline{s}_k$ k=1...K. Then the model is simply

$$m_{\ell,k} = m_\ell(\underline{s}) \quad \text{for } \underline{s} = \underline{s}_k \tag{3.3.1}$$

where the values of $m_{\ell,k}$ k=1...K are to be estimated from the observations.

The relationship between the constant c_o of (3.1) and the $m_{\ell,k}$ k=1...K will now be defined. Using (3.2) and (3.1.3) it follows that

$$\frac{1}{N} \sum_{n=1}^{N} \ln c(n,\underline{s}_k) = \ln m_{\ell,k} + \ln c_o \tag{3.3.2}$$
$$k = 1...K$$

Since $\ln c_o$ is, by definition, the long term average of $\ln c(n,\underline{s})$ over all time and space, it follows that the $m_{\ell,k}$ must be constrained so that

$$\sum_{k=1}^{K} \ln m_{\ell,k} = 0 \tag{3.3.3}$$

3.4 Local Variations: Random Component

The local variations, random component term of (3.1), $\Delta_\ell(n,\underline{s})$ is hypothesized to have the following structure:

$$\ln \Delta_\ell(n+1,\underline{s}) = \phi_\ell \ln \Delta_\ell(n,\underline{s}) + w_\ell(n,\underline{s}) \tag{3.4.1}$$

where

$w_\ell(n,\underline{s})$: zero mean, white normal random process

$$E[w_\ell(n,\underline{s})] = 0$$

$$E[w_\ell(n_1,\underline{s}_1)w_\ell(n_2,\underline{s}_2)] = \begin{cases} R_\ell & n_1 = n_2,\ \underline{s}_1 = \underline{s}_2 \\ 0 & n_1 \neq n_2,\ \text{or } \underline{s}_1 \neq \underline{s}_2 \end{cases}$$

$$0 \leq \phi_\ell < 1$$

It is further assumed that

$$E[w_\ell(n_1,\underline{s})\ w_g(n_2)] = 0 \qquad \text{all } n_1,\ n_2$$

where $w_g(n)$ is the white random process associated with the global concentration field, $\Delta_g(n)$. The values of ϕ_ℓ and R_ℓ are to be estimated from the measurements.

From (3.4.1) it follows that $\ln \Delta_\ell(n,\underline{s})$ can be viewed as a "set" of first order, normal, Markov process in time where there is a separate independent process for each spatial coordinate $\underline{s}$. The purpose of $\Delta_\ell(n,\underline{s})$ is to model the local variations (in space) in the concentration field due to the inherently random nature of atmospheric propagation and possibly due to local random sources.

The constant ϕ_ℓ in (3.4.1) accounts for the fact that these local variations have time dynamics (memory) so they are "correlated in time." The hypothesis that $\Delta_\ell(n,\underline{s})$ is "uncorrelated in space" is reasonable for values of $\underline{s}_1$ and $\underline{s}_2$ where $\underline{s}_1-\underline{s}_2$ is sufficiently large. However it is clear that the hypothesized structure becomes invalid as $\underline{s}_1-\underline{s}_2$ goes to zero. This limit on the model's applicability should be remembered. An exact definition for how large $\underline{s}_1-\underline{s}_2$ should be is not available. The actual data processing only showed that the spacing between monitors was sufficient.

3.5 Analysis of Model Structures

A few general results on the behavior of model structures like those of Sections 3.1 - 3.4 are now summarized for later reference.

The dynamic model for $\ln m_g(n,\underline{s})$ of (3.1.1) is of the general form

$$x(n+1) = \phi x(n) + (1-\phi)\, b(n) \tag{3.5.1}$$

If $x(0) = 0$, this equation can be solved to give

$$x(n) = [b(n-1) + \phi\, b(n-2) + \phi^2\, b(n-3) + \ldots]\,(1-\phi)$$

so that ϕ^τ determines how much an input at time $n-\tau-1$ contributes to the out-

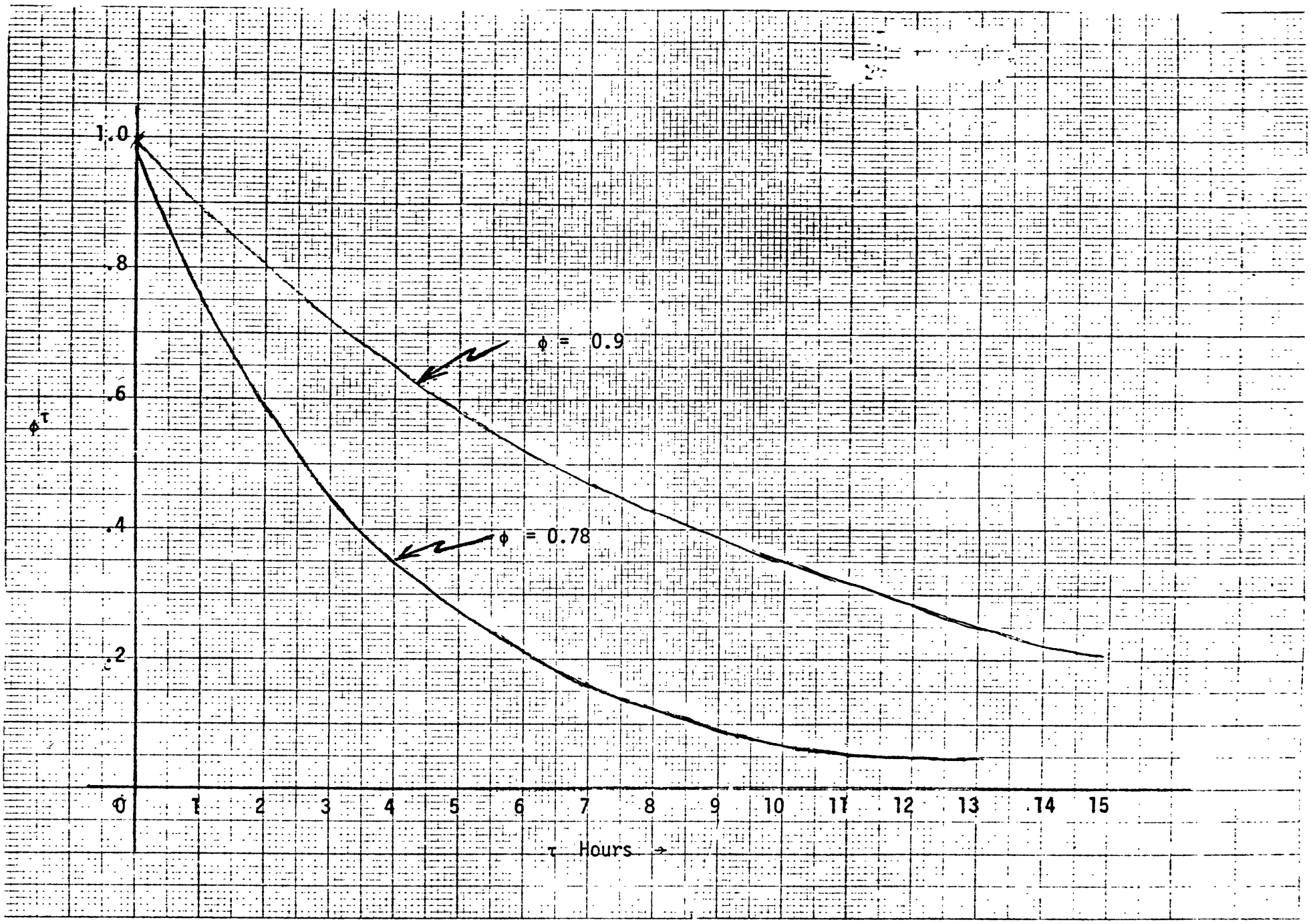


Figure 3.5.1. Behavior of ϕ^τ

put at time n. Figure 3.5.1 plots ϕ^τ for

$$\phi = .9 \text{ and } .78$$

These values correspond to values that were estimated from the observations (see Section 7).

The "effective memory span" of a dynamic system like (3.5.1) can be defined as the time delay required before the effect of an input is "effectively" forgotten. From Figure 3.5.1 it can be concluded that

ϕ	Effective Memory Span
.9	9 to 15 hours
.78	3 to 6 hours

The dynamic models for the random process $\ln \Delta_\ell(n,\underline{s})$ and $\ln \Delta_g(n)$ are of the form

$$x(n+1) = \phi x(n) + w(n) \tag{3.5.2}$$

$$E[w(n)] = 0$$

$$E[w(n_1)w(n_2)] = \begin{cases} R & n_1 = n_2 \\ 0 & n_1 \neq n_2 \end{cases}$$

Assume $\phi < 1$ and $x(n)$ is a stationary process. Define

$$\Gamma = E[x^2(n)]$$

Then

$$\Gamma = \frac{R}{1-\phi 2} \tag{3.5.3}$$

so

$\phi = .9$ implies $\Gamma = 5.2R$

$\phi = .78$ implies $\Gamma = 2.56R$

Consider a lognormal random variable Δ with

$$E\{\ln \Delta\} = \overline{\Delta}$$

$$E\{[\ln \Delta - \overline{\Delta}]^2\} = \Gamma$$

Then

$$E\{\Delta\} = e^{\overline{\Delta} + \Gamma/2} \tag{3.5.4}$$

$$E\{[\Delta - E(\Delta)]^2 = e^{2\overline{\Delta} + \Gamma}(e^{\Gamma}-1)$$

4. Uses of Model

The hypothesized structure of the model was outlined in Section 3. The estimation of the various parameters such as ϕ, etc. will be discussed in Section 5. In the present section, it is assumed that the parameter values are known and the discussion is on the possible uses of the model.

4.1 General

Figures 4.1.1 to 4.1.3 illustrate three general ways the model could be used.

Figure 4.1.1 is intended to be a direct "picture" of (3.1), i.e., a dynamic system with inputs:

- $\theta(n)$: wind direction
- h,d: hour of day, day of year
- $\left.\begin{matrix} w_g(n) \\ w_\ell(n) \end{matrix}\right\}$: white, normal random processes

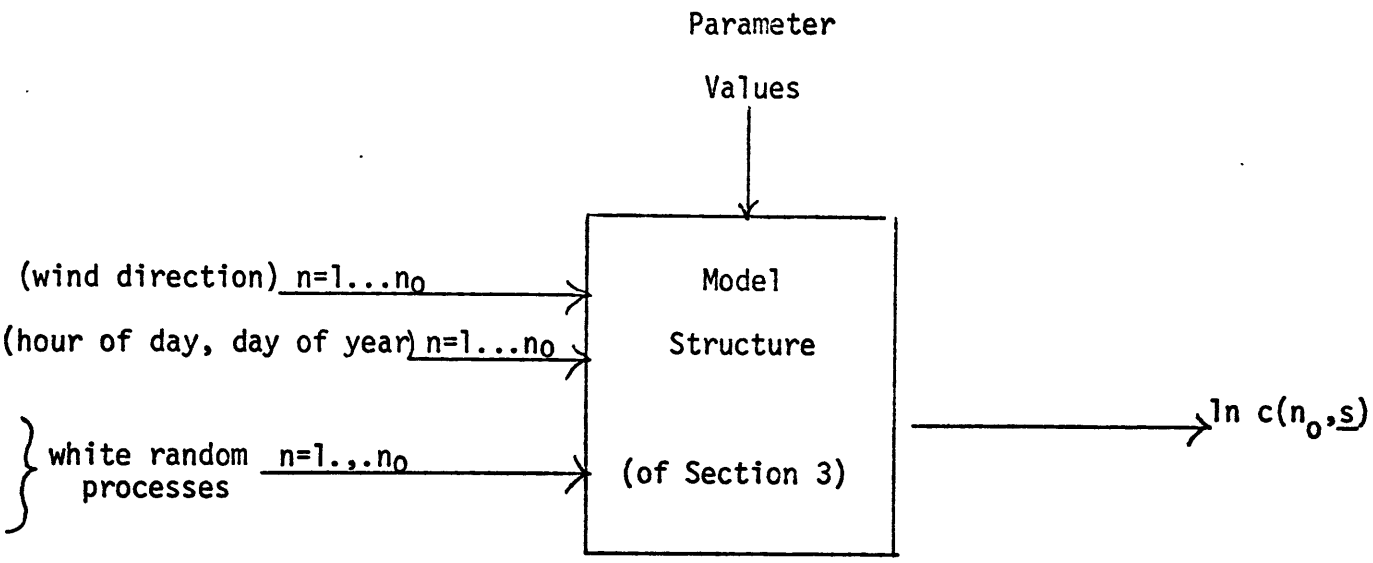
and output $\ln c(n,\underline{s})$ which is a normal random process in time and space. Viewed in this way the model can be used to explore the statistical properties of $c(n,\underline{s})$. It can be used to gain a better understanding of the "physical processes" involved in the real world.

Figure 4.1.2 illustrates one way a model can be used to "predict" concentration values for specified times and spatial coordinates $\underline{s}$. Mathematically

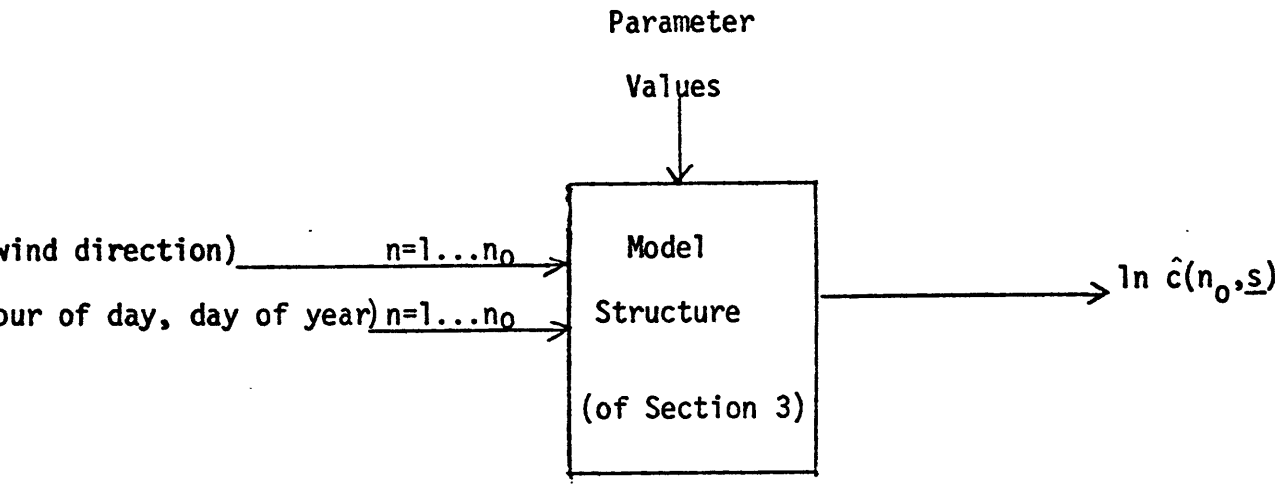
$$\ln \hat{c}(n,\underline{s}) = E\{\ln c(n,\underline{s})\} \qquad (4.1.1)$$

where $\theta(n)$ and h,d are considered to be specified (known, exogenous, etc.) inputs. Note than when "prediction" is done as illustrated in Figure 4.1.2, only the mean value terms of the model (3.1) are used. The values of $\ln \hat{c}(n,\underline{s})$ do not depend on the random terms or their associated parameter values such as ϕ_ℓ, R_g and R_ℓ. Figure 4.1.2 is called open loop prediction to contrast it with the state estimation "closed loop" prediction to be discussed next.

The hypothesized structure of Section 3 is a random process (in time and space) with memory. This "time memory" means a concentration measured at time $n-\tau$ contains "information" which can be used to predict a concentration at time n. Furthermore, the random term, $\Delta_g(n)$, is "correlated in space" so that a concentration measured at spatial coordinate $\underline{s}_k$ contains information which can be used by the model to predict a concentration at coordinate $\underline{s}_j$. Define



Nature of Model of (3.1)
Figure 4.1.1



ln $\hat{c}$ (n,$\underline{s}$): Predicted value of concentration at time n and coordinate $\underline{s}$

Use of Model for "Open Loop" Prediction

Figure 4.1.2

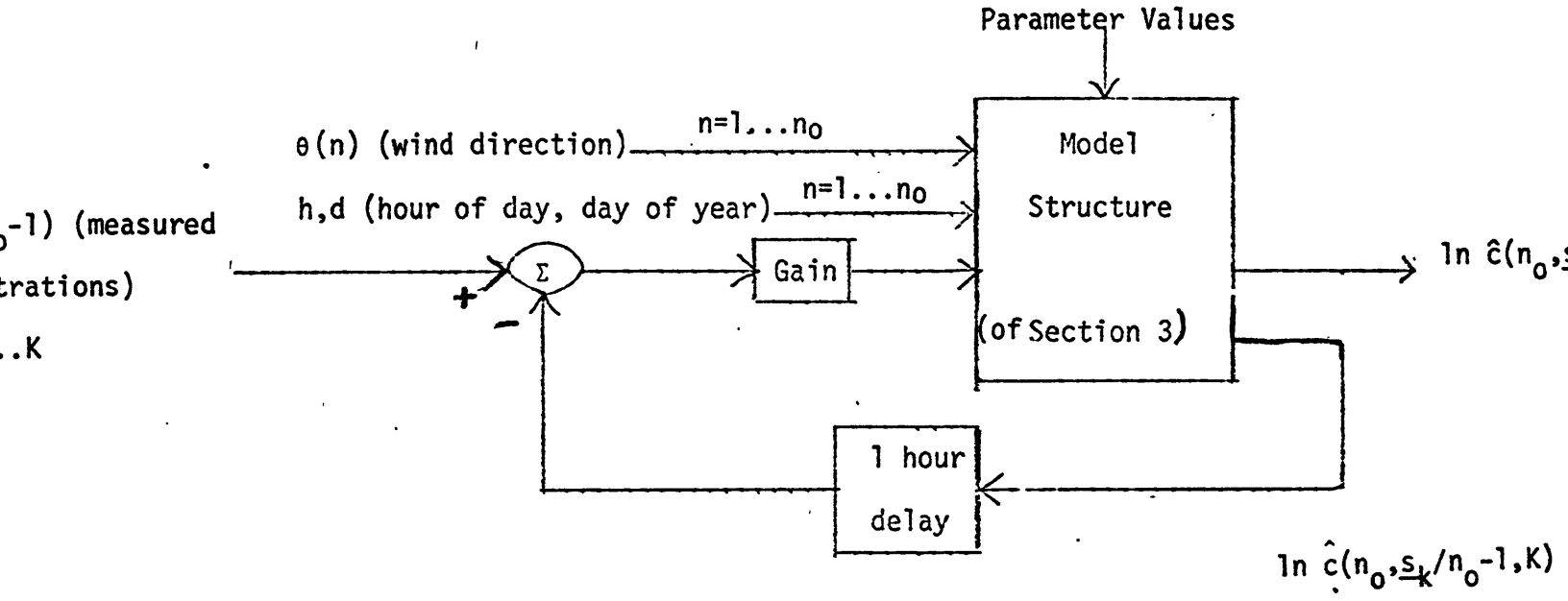


Figure 4.1.3

$\ln \hat{c}(n_0, \underline{s}/n_0-1, K)$: predicted value of concentration at time n_0 and coordinate $\underline{s}$ using measured concentrations $c_k(n)$ $n=1...n_0-1$ $k=1...K$

Figure 4.1.3
Use of Model in "State Estimation"

$c_k(n)$: hourly average concentration measured by monitor at coordinate $\underline{s}_k$ at hour n

$\hat{c}(n_o,\underline{s}/n-\tau,K)$: estimate of concentration at time n_o and coordinate $\underline{s}$ made using the model and the measurements $c_k(n)$, k=1...K n=1...$n_o-\tau$ (4.1.2)

The use of the model to compute $\hat{c}(n,\underline{s}/n-1,K)$ is illustrated in Figure 4.1.3. This type of feedback of information is called state estimation. The theory behind state estimation, discussions on how to compute the gain, etc. are given in Section 9.2. These more detailed discussions show that the $\hat{c}(n,\underline{s}/n-\tau,K)$ do depend on parameters of the random terms of (3.1). The state estimation of Figure 4.1.3 is a closed loop system because the outputs are "feedback" and are compared with the input measurements.

Figure 4.1.3 yields $\hat{c}(n,\underline{s}/n-\tau,K)$ for τ=1. Values for other τ follow in a similar fashion. For example, values for $\tau > 1$ can be obtained by combining Figure 4.1.2 and 4.1.3. As τ gets larger compared to the effective memory span of the model (see Section 3.5) the measured concentrations $c_k(n)$ have less and less effect so that

$$\hat{c}(n, \underline{s}/n-\tau,K) \rightarrow \hat{c}(n,\underline{s}) \quad \text{as } \tau \rightarrow \infty$$

The use of a model to predict concentrations in an open loop fashion as in Figure 4.1.2 is "standard" and will not be discussed further here. The discussion to follow emphasizes the implications the random terms of (3.1) have on other specific uses of the model (the random terms do not affect ln $\hat{c}(n,\underline{s})$ of Figure 4.1.2).

4.2 SCS

As discussed relative to Figure 4.1.3, the model can be used to obtain a state estimator which yields an estimate $c(n,\underline{s}/n-\tau,K)$. This estimate is more accurate than $\hat{c}(n,\underline{s})$ of Figure 4.1.2 and can yield improved SCS performance.

Since the random effects are explicitly modeled, the statistical properties of $c(n,\underline{s})$ of Figure 4.1.1 or $\hat{c}(n,\underline{s})$ of Figure 4.1.2 or $\hat{c}(n,s/n-1,K)$ of Figure 4.1.3 can be calculated. These quantities are needed for the use of a probabilistic control strategy.

The fact that the random terms are explicitly modeled allows a variety of real time "anomaly detection" algorithms to be used in conjunction with the state estimator. This is very important in the development of a reliable SCS.

4.3 Major Point Source Modeling

Many air pollution modeling strategies are concerned primarily with the effects of large point sources. The background is often considered to be a "nuisance". Unfortunately, in the case of the Chestnut Ridge project, the background can be large and random enough to make it very difficult to model large point source effects without first modeling the background.

The state estimation form of Figure 4.1.3 enables one to predict the background concentration at a given monitor using measurements made at other monitors. This is very advantageous for point source modeling studies as a given point source plume almost never affects more than one monitor at a time.

4.4 Dosage-Health Effects Studies

SO_2 ambient standards are presently stated in terms of maximum allowable average values. However health effect impacts of SO_2 (and other pollutants) depend in a complex way on the time duration of exposure as a function of magnitude of concentration. The random terms of (3.1) enable calculation of the actual time-dependent, spatial, probability distributions of the concentrations and hence enable the use of more realistic health effect models.

5. Parameter Estimation

The hypothesized structure of Section 3 contains parameters whose values are to be estimated. Parameter estimation (second step in Figure 2.1) is now discussed.

5.1 Data Base

The hypothesized structure was developed so as to be applicable to the entire Chestnut Ridge data base. However the actual data base used for parameter estimation and model verification was only 3 months - January, February and March 1975 - for the four "northern sector" monitors, L, M, N, O which are located around the Keystone plant.

The actual data used was

$\ln c_k(n)$: logarithm of hourly average SO_2 concentration observed by monitor at location $\underline{s}_k$, k=1...K, n=1...N

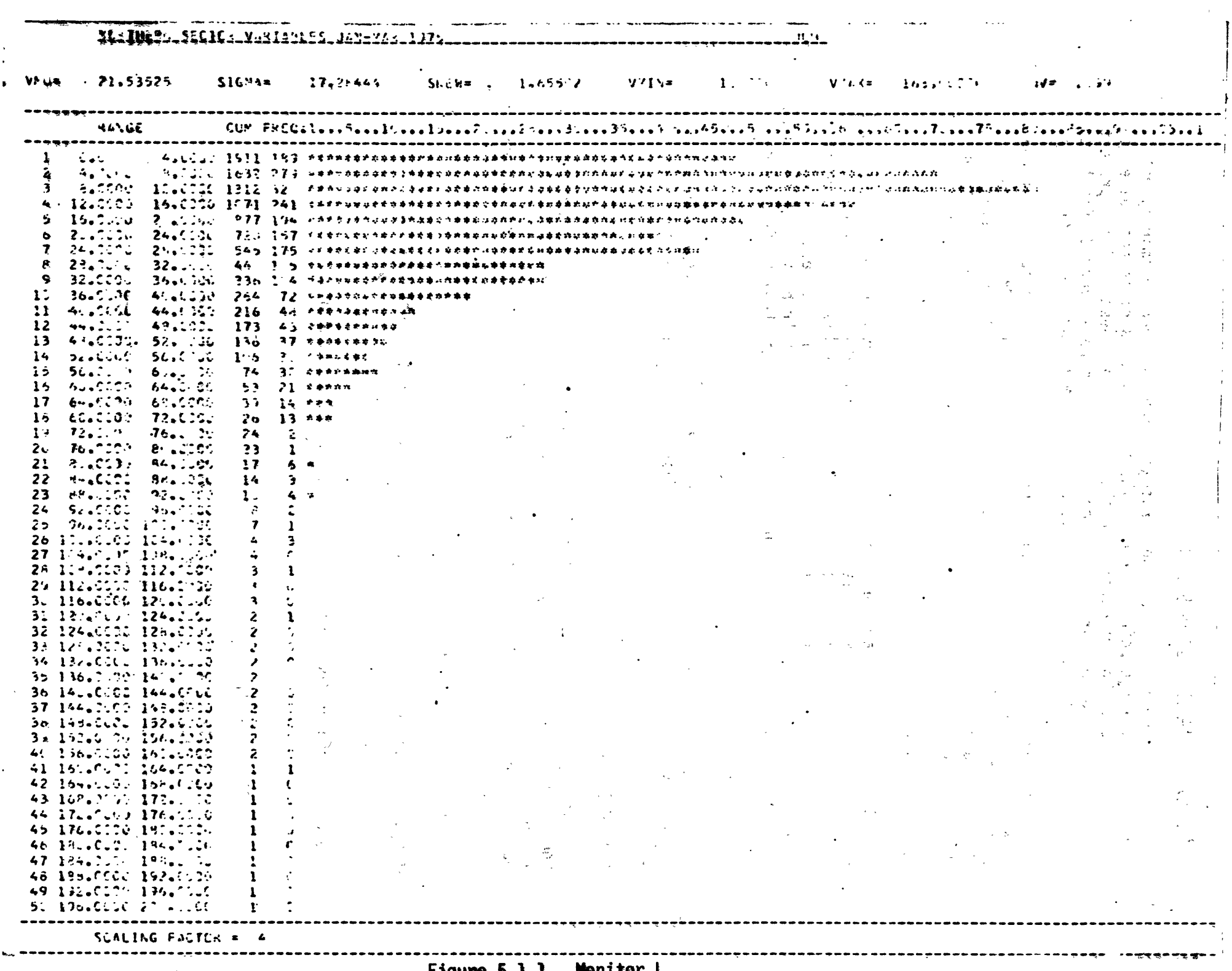
$\theta(n)$: elevated tower wind direction measured at hour n

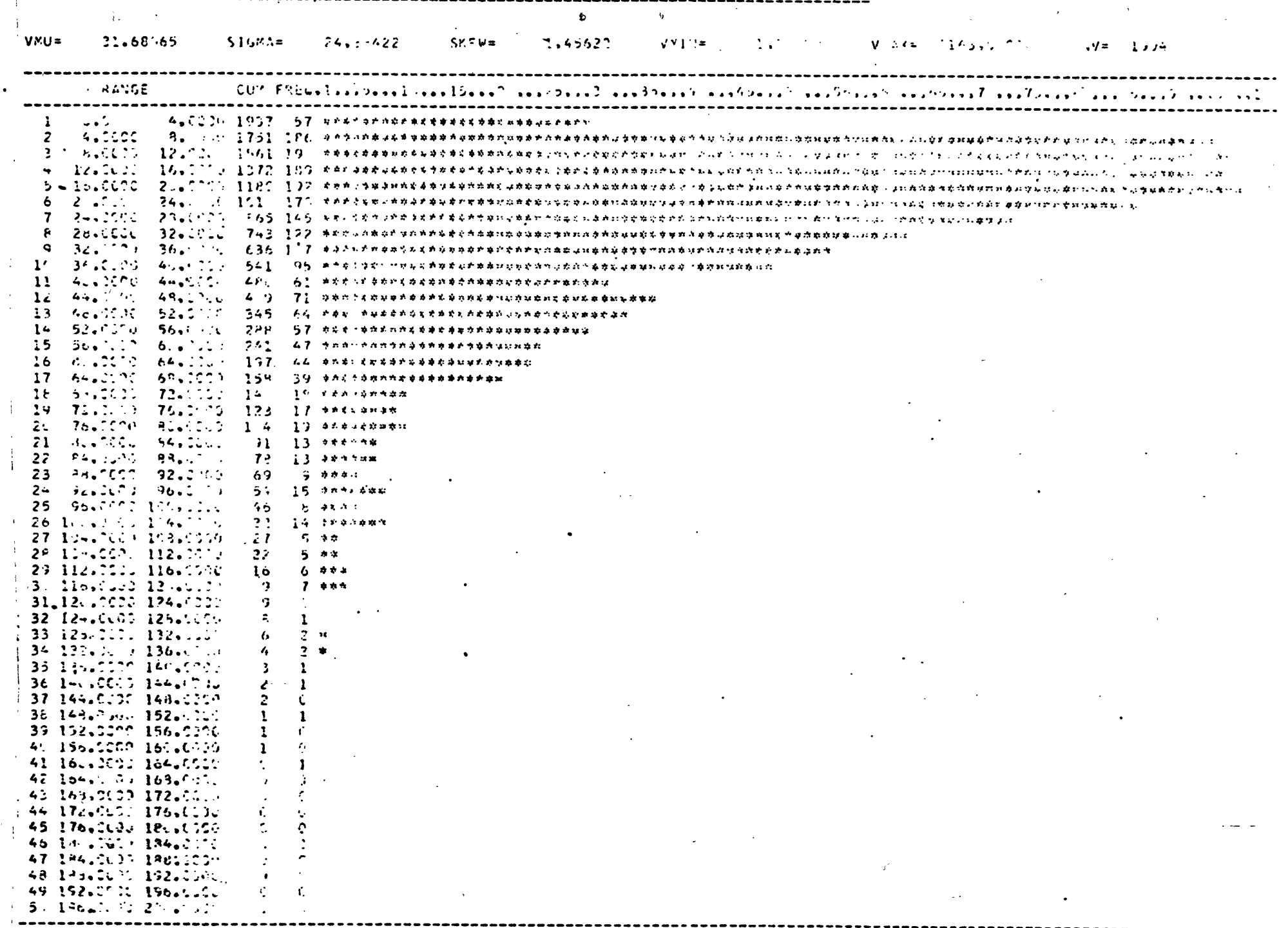
$K = 4$

$N \approx 2000$ (3 months)

Since the monitors did not have 100% data capture, the value of N was actually different for the different monitors.

Figures 5.1.1 to 5.1.4 are histograms of the measured $c_k(n)$ for the four monitors L, M, N, O for the 3 montns January to March 1975.





JURTHERY SECTOR VARIABLES JAN-MAR 1975

VMU= 17.34529 SIGMA= 16.43311 SKEW= 2.50394 VPI= [illegible] VMAX= 193.[illegible] NV= 2156

	RANGE		CUM FREQ	FREQ
1	0.0	4.0000	1795	359
2	4.0000	8.0000	1342	416
3	8.0000	12.0000	1058	224
4	12.0000	16.0000	829	229
5	16.0000	20.0000	668	161
6	20.0000	24.0000	512	156
7	24.0000	28.0000	347	125
8	28.0000	32.0000	311	76
9	32.0000	36.0000	24	71
10	36.0000	40.0000	192	59
11	40.0000	44.0000	141	41
12	44.0000	48.0000	111	3
13	48.0000	52.0000	71	3.
14	52.0000	56.0000	[illegible]	12
15	56.0000	60.0000	53	16
16	60.0000	64.0000	37	21
17	64.0000	68.0000	21	11
18	68.0000	72.0000	16	5
19	72.0000	76.0000	12	4
20	76.0000	80.0000	12	[illegible]
21	[illegible]	84.0000	17	[illegible]
22	84.0000	88.0000	11	1
23	[illegible]	92.0000	3	3
24	[illegible]	96.0000	7	1
25	96.0000	100.0000	6	1
26	100.0000	104.0000	6	[illegible]
27	104.0000	108.0000	6	[illegible]
28	108.0000	112.0000	6	[illegible]
29	112.0000	116.0000	5	1
30	116.0000	120.0000	4	1
31	120.0000	124.0000	4	[illegible]
32	124.0000	128.0000	3	1
33	128.0000	132.0000	3	
34	132.0000	136.0000	2	1
35	136.0000	140.0000	2	[illegible]
36	140.0000	144.0000	2	[illegible]
37	144.0000	148.0000	2	[illegible]
38	148.0000	152.0000	2	[illegible]
39	152.0000	156.0000	1	1
40	156.0000	160.0000	1	[illegible]
41	160.0000	164.0000	1	[illegible]
42	164.0000	168.0000	1	0
43	168.0000	172.0000	1	[illegible]
44	172.0000	176.0000	1	[illegible]
45	176.0000	180.0000	1	[illegible]
46	180.0000	184.0000	1	0
47	184.0000	188.0000	1	[illegible]
48	188.0000	192.0000	[illegible]	[illegible]
49	192.0000	196.0000	[illegible]	1
50	196.0000	200.0000	[illegible]	[illegible]

SCALING FACTOR = 5

Figure 5.1.3. Monitor N

SCRIBERN'S TCP VARIABLES JAN-MAR 1975

VMU= 13.71073 SIGMA= 11.[illegible]574 SKEW= 1.24974 VMIN= 1.[illegible] VMAX= 7[illegible] NV= [illegible]

	RANGE		CUM FREQ		1...5...1 ...1>...2...25...3 ...3>...4 ...45...5 ...5>...6...65...7 ...7>...8 ...85...9 ...9
1	0.0	4.0000	1653	407	**********
2	4.0000	8.0000	1239	415	**********
3	8.0000	12.0000	822	316	**********
4	12.0000	16.0000	491	231	**********
5	16.0000	20.0000	478	213	**********
6	20.0000	24.0000	321	159	**********
7	24.0000	28.0000	211	109	**********
8	28.0000	32.0000	140	71	**********
9	32.0000	36.0000	93	5	**********
10	36.0000	40.0000	57	33	******
11	40.0000	44.0000	44	13	**
12	44.0000	48.0000	2[illegible]	24	****
13	48.0000	52.0000	8	12	**
14	52.0000	56.0000	5	3	
15	56.0000	60.0000	4	1	
16	60.0000	64.0000	2	2	
17	64.0000	68.0000	1	1	
18	68.0000	72.0000	1	0	
19	72.0000	76.0000	[illegible]	1	
20	76.0000	80.0000	[illegible]	[illegible]	
21	80.0000	84.0000	[illegible]	[illegible]	
22	84.0000	88.0000	[illegible]	[illegible]	
23	88.0000	92.0000	[illegible]	[illegible]	
24	92.0000	96.0000	[illegible]	[illegible]	
25	[illegible]	[illegible]	[illegible]	[illegible]	
26	100.0000	104.0000	[illegible]	[illegible]	
27	104.0000	108.0000	[illegible]	[illegible]	
28	108.0000	112.0000	[illegible]	[illegible]	
29	112.0000	116.0000	[illegible]	[illegible]	
30	116.0000	120.0000	[illegible]	[illegible]	
31	120.0000	124.0000	[illegible]	[illegible]	
32	124.0000	128.0000	[illegible]	[illegible]	
33	128.0000	132.0000	[illegible]	[illegible]	
34	132.0000	136.0000	[illegible]	[illegible]	
35	136.0000	140.0000	[illegible]	[illegible]	
36	140.0000	144.0000	[illegible]	[illegible]	
37	144.0000	148.0000	[illegible]	[illegible]	
38	148.0000	152.0000	[illegible]	[illegible]	
39	152.0000	156.0000	[illegible]	[illegible]	
40	156.0000	160.0000	[illegible]	[illegible]	
41	160.0000	164.0000	[illegible]	[illegible]	
42	164.0000	168.0000	[illegible]	[illegible]	
43	168.0000	172.0000	[illegible]	[illegible]	
44	172.0000	176.0000	[illegible]	[illegible]	
45	176.0000	180.0000	[illegible]	[illegible]	
46	180.0000	184.0000	[illegible]	[illegible]	
47	184.0000	188.0000	[illegible]	[illegible]	
48	188.0000	192.0000	[illegible]	[illegible]	
49	192.0000	196.0000	[illegible]	[illegible]	
50	196.0000	200.0000	[illegible]	[illegible]	

SCALING FACTOR = 5

Figure 5.1.4. Monitor 0

Figure 5.1.5 is a histogram of the measured elevated tower wind direction $\theta(n)$ for the 3 months of concern.

Figure 5.1.6 is a histogram of the megawatt output of Keystone for the 3 months of concern. As will be discussed further in Section 7.6, it was very difficult to spot the effects of Keystone in the data.

5.2 Nature of Estimation procedure

The basic idea of parameter estimation is to find the particular set of parameter values which make the model "fit" the observed concentrations $c_k(n)$ "as best as possible." The mathematical theory for doing this is discussed in Section 9 on "maximum likelihood parameter estimation" theory. The basic ideas are discussed here in general terms.

Define the residuals

$$\delta_k(n) = \ln c_k(n) - \ln \hat{c}(n,\underline{s}_k/n-1,K) \quad k=1...K \qquad (5.2.1)$$

to be the difference between the logarithm of the observed concentrations and the values predicted by the state estimator (see Figure 4.1.3). The basic idea underlying the parameter estimation algorithms is to vary the "parameter values" used in the model (see Figure 4.1.3) until the residuals $\delta_k(n)$ $n=1...N$, $k=1...K$ are as small as possible. Because of the normal hypothesized structure of $\ln c(n,\underline{s})$, the maximum likelihood criterion is "effectively" to choose parameter values to minimize the sum of the squared residuals.

A computer program GPSIE (Peterson, 1974) was available to do the parameter estimation using the theory discussed in Section 9. However, a variety of "ad hoc" algorithms were used instead, for the following reasons. The background modeling efforts started out as an exploration of a simple hypothesized structure for which "least squares regression" analysis could be used. As the hypothesized structure grew more complicated (i.e., as iterations around the "loop" of Figure 2.1 were done), it always seemed easier to modify the existing computer codes rather than to transfer the data base to a new computer code (i.e., to GPSIE). The final computer code for parameter estimation was an "engineering approximation" to GPSIE (see Section 9.3) which we feel generated "reasonable" results.

The validity of the resulting model will be discussed in Section 6. However it is first appropriate to list some of the "shortcomings" of the numerical analysis that was done.

First, true maximum likelihood estimates were not obtained.

Second, although the Keystone plant can influence monitors L, M, N, O, the parameters of the hypothetical background model were fit to all of the observations with no attempt to remove or "prefilter out" Keystone effects. (See further discussion in Section 7.6)

[illegible]

[illegible]

	RANGE		[illegible]		
1	.0000	5.0000	2151	1	*
2	5.0000	10.0000	2141	6	***
3	10.0000	15.0000	2132	1	
	15.0000	20.0000	2132		
5	20.0000	25.0000	2127	4	**
6	25.0000	30.0000	2123	[illegible]	*
7	30.0000	35.0000	2122	3	*
8	35.0000	40.0000	2119	1	*
9	40.0000	45.0000	2117	2	*
10	45.0000	50.0000	2114	1	*
11	50.0000	55.0000	2111	1	*
12	55.0000	60.0000	2110	2	*
13	60.0000	65.0000	2104	5	**
14	65.0000	70.0000	2100	4	**
15	70.0000	75.0000	2097	2	*
16	75.0000	80.0000	2094	2	*
17	80.0000	85.0000	2091	1	*
18	85.0000	90.0000	2083	7	****
19	90.0000	95.0000	2073	9	****
20	95.0000	100.0000	2058	15	*******
21	100.0000	105.0000	2035	23	***********
22	105.0000	110.0000	2015	20	**********
23	110.0000	115.0000	1997	19	*********
24	115.0000	120.0000	1974	23	***********
25	120.0000	125.0000	1956	19	*********
26	125.0000	130.0000	1931	25	************
27	130.0000	135.0000	1905	26	*************
28	135.0000	140.0000	1869	36	******************
29	140.0000	145.0000	1817	50	*************************
30	145.0000	150.0000	1747	72	*************************************
31	150.0000	155.0000	1691	56	****************************
32	155.0000	160.0000	1659	32	****************
33	160.0000	165.0000	1625	34	*****************
34	165.0000	170.0000	1584	41	********************
35	170.0000	175.0000	1563	21	**********
36	175.0000	180.0000	1535	28	**************
37	180.0000	185.0000	1521	14	*******
38	185.0000	190.0000	1502	19	*********
39	190.0000	195.0000	1477	25	************
40	195.0000	200.0000	1463	14	*******
41	200.0000	205.0000	1447	16	********
42	205.0000	210.0000	1424	23	***********
43	210.0000	215.0000	1415	9	****
44	215.0000	220.0000	1385	30	***************
	220.0000	225.0000	1366	19	*********
	225.0000	230.0000	1326	40	********************
47	230.0000	235.0000	1272	54	***************************
48	235.0000	240.0000	1202	70	***********************************
49	240.0000	245.0000	1158	44	**********************
50	245.0000	250.0000	1096	62	*******************************
51	250.0000	255.0000	1012	84	**
52	255.0000	260.0000	927	85	**
53	260.0000	265.0000	799	128	**
54	265.0000	270.0000	665	134	***
55	270.0000	275.0000	582	83	**
56	275.0000	280.0000	517	65	*********************************
57	280.0000	285.0000	444	73	*************************************
58	285.0000	290.0000	339	105	***
59	290.0000	295.0000	275	64	********************************
60	295.0000	300.0000	224	51	**************************
61	300.0000	305.0000	180	44	**********************
62	305.0000	310.0000	145	35	*****************
63	310.0000	315.0000	127	18	*********
64	315.0000	320.0000	107	20	**********
65	320.0000	325.0000	94	12	******
66	325.0000	330.0000	86	8	****
67	330.0000	335.0000	68	18	*********
68	335.0000	340.0000	53	15	*******
69	340.0000	345.0000	37	16	********
70	345.0000	350.0000	25	12	******
71	350.0000	355.0000	19	6	***
72	355.0000	360.0000	13	6	***

SCALING FACTOR = 2

41 KYI 50 0.0 1750.000 5 35.00
DATE OF RECORD 44171: 1/ 1/75 HR ENDING 1
DATE OF RECORD 64771: 3/31/75 HR ENDING 2400

Tower Wind Direction

Figure 5.1.5

Figure 5.1.6. Keystone Plant MWe

Third, the state estimator of Figure 4.1.3 is a dynamic system with memory. The associated "startup" problems were ignored and this caused large errors at some times. This startup problem occurred every time there was a time sequence of lost monitor reading (i.e., when data was not captured). This happens for each monitor repeatedly during the 3 months considered. It is conceptually simple to solve the startup problem but the modification was not made.

Fourth, the monitors record "1 part per billion" for any 1-hour average equal to or less than 1 ppb. This phenomenon is not included in the hypothesized model structure so poor performance often resulted in the case of very low concentrations.

6. Model Validation

The third step of the modeling process of Figure 2.1 is to test the validity of the model (structure plus estimated parameter values). Section 10 contains a general discussion on model validity testing. Only two types of tests were actually used:

- autocorrelation functions
- histograms.

Equation (5.2.1) defined the residuals $\delta_k(n)$ as

$$\delta_k(n) = \text{measured} - \text{predicted}$$

$$= \ln c_k(n) - \ln \hat{c}(n, \underline{s}_k/n-1, K)$$

Define

$$\underline{\delta}(n) = \begin{bmatrix} \delta_1(n) \\ \cdot \\ \cdot \\ \cdot \\ \delta_K(n) \end{bmatrix} \qquad (6.1)$$

Define the autocorrelation matrices

$$\underline{P}_{\delta\delta}(\tau) = E[\underline{\delta}(n)\underline{\delta}^T(n-\tau)] \qquad \tau = 0, 1... \qquad (6.2)$$

where "T" denotes transpose.

Then if the hypothesized structure and estimated parameter values correspond exactly to the "real world",

$$\underline{P}_{\delta\delta}(\tau) = 0 \qquad \tau > 0 \qquad (6.3)$$

Figure 6.1 shows $\underline{P}_{\delta\delta}(\tau)$, $\tau = 0, 1, \ldots$ for the structure of Section 3 and with the parameter values estimated as discussed in Section 5. In order to have a standard of comparison, define

	L	M	N	O
	TIME SERIES ANALYSIS		TAU= 0	
L	0.2120E 00	0.2958E-01	0.7523E-01	0.5898E-01
M	0.2958E-01	0.1675E 00	0.5118E-01	0.4268E-01
N	0.7523E-01	0.5118E-01	0.2251E 00	0.6827E-01
O	0.5898E-01	0.4268E-01	0.6827E-01	0.2429E 00
	TIME SERIES ANALYSIS		TAU= 1	
	0.4255E-01	0.2280E-01	0.2515E-01	0.4227E-01
	-0.3265E-02	0.3285E-01	0.8140E-02	0.5291E-02
	0.2364E-01	0.3172E-01	0.5725E-01	0.3950E-01
	0.1299E-01	0.2242E-01	0.2513E-01	0.4009E-01
	TIME SERIES ANALYSIS		TAU= 2	
	-0.5354E-02	0.7635E-03	0.2266E-02	0.1856E-01
	-0.1776E-02	-0.6367E-02	-0.7880E-02	0.5853E-03
	0.1091E-01	0.6889E-02	0.1078E-02	0.2530E-01
	0.1493E-02	0.7738E-02	-0.5461E-02	0.8113E-03
	TIME SERIES ANALYSIS		TAU= 3	
	-0.6303E-02	-0.7417E-02	-0.7471E-02	-0.6898E-02
	0.3033E-02	0.8087E-03	-0.5014E-02	0.5623E-02
	0.5944E-02	-0.1647E-02	-0.5587E-02	0.3486E-02
	-0.5340E-03	-0.1028E-02	-0.5658E-02	-0.3431E-02
	TIME SERIES ANALYSIS		TAU= 4	
	0.7408E-02	0.2983E-02	-0.1211E-02	-0.2758E-02
	0.7943E-03	-0.3342E-02	0.3855E-03	0.7023E-02
	0.1347E-01	0.6463E-02	0.2385E-02	0.7403E-02
	0.3114E-02	0.6752E-02	0.1766E-02	0.7794E-02
	TIME SERIES ANALYSIS		TAU= 5	
	0.2270E-01	0.3792E-02	0.9174E-02	0.8378E-02
	0.1722E-02	0.3086E-02	0.1445E-01	0.3627E-02
	0.1022E-01	0.8300E-02	0.1420E-01	0.9135E-02
	0.6269E-02	0.4747E-02	0.9947E-02	0.8111E-02
	TIME SERIES ANALYSIS		TAU= 6	
	0.1433E-01	-0.4462E-02	0.3538E-03	0.8535E-02
	-0.2098E-02	0.1099E-01	0.2409E-02	0.2230E-02
	0.5217E-02	0.8067E-02	0.1940E-01	0.1990E-01
	0.1025E-01	0.6919E-02	0.9304E-02	0.1193E-01
	TIME SERIES ANALYSIS		TAU= 7	
	0.8047E-02	0.4630E-02	-0.7402E-02	-0.7439E-02
	-0.1491E-02	0.6399E-02	0.4772E-02	0.5346E-02
	0.6456E-02	0.7866E-02	0.1133E-01	0.1360E-01
	0.6195E-02	0.1106E-01	0.2527E-02	0.1638E-01
	TIME SERIES ANALYSIS		TAU= 8	
	0.1390E-01	0.1081E-01	-0.3804E-03	0.8433E-02
	0.1494E-02	-0.7279E-02	0.9830E-03	0.2888E-02
	-0.2918E-02	0.2694E-02	0.2867E-02	0.1373E-01
	0.4142E-02	-0.1502E-02	-0.1418E-01	0.7378E-02
	TIME SERIES ANALYSIS		TAU= 9	
	0.1179E-01	0.2350E-02	0.7493E-02	0.1375E-02
	0.3700E-02	-0.4581E-04	-0.1759E-02	0.1029E-01
	0.4035E-02	0.4754E-02	0.6347E-02	0.6494E-02
	0.7340E-02	0.4043E-02	0.5771E-02	0.8623E-02

NUMBER OF POINTS--N= 2151

Autocorrelation Matrices $P_\delta(\tau)$

Figure 6.1

$\tilde{\underline{\delta}}(n)$: residuals when only the mean value terms of (3.1) are present

$$\tilde{\underline{P}}_{\delta\delta}(\tau) = E[\tilde{\underline{\delta}}(n)\ \tilde{\underline{\delta}}^T(n-\tau)]$$

Figure 6.2 shows $\tilde{\underline{P}}_{\delta\delta}(\tau)$, $\tau = 0,1,\ldots$ The $\underline{P}_{\delta\delta}(\tau)$ of Figure 6.1 can be considered to be a "good approximation" to the condition (6.3) while the $\tilde{\underline{P}}_{\delta\delta}(\tau)$ of Figure 6.2 obviously does not behave like the condition (6.3).

The second type of validity testing involved histograms of the residuals $\delta_k(n)$. Because of the hypothesized structure, these histograms should look like normal distributions. Figures 6.3 - 6.6 show the histograms of $\delta_k(n)$ for the four monitors, L, M, N, O. Figures 6.3 - 6.6 appear to be normally distributed except for the presence of "outliers" (i.e., very large or small values). A study of the actual data showed that almost all of the large (positive) outliers could be associated with the "startup" problem discussed in Section 5.2. Similarly almost all of the small (negative) outliers could be associated with the fact discussed in Section 5.2 that any measurement less than 1 ppb was recorded as 1 ppb. Removal of these outliers makes Figures 6.3 - 6.4 look even more like normal distributions.

Various shortcomings of the data analysis were discussed in Section 5. Another shortcoming is that the above model verification tests are based on "eyeballing" the autocorrelation matrices and histograms. The formal mathematical machinery of statistical hypothesis testing was not used.

```
        L           M           N           O

L   TIME SERIES ANALYSIS        TAU= 0
M     0.7647E 00  0.2708E 00  0.4080E 00  0.4659E 00
N     0.2708E 00  0.5067E 00  0.3954E 00  0.3710E 00
O     0.4080E 00  0.3954E 00  0.9193E 00  0.5522E 00
      0.4659E 00  0.3710E 00  0.5522E 00  0.9008E 00

    TIME SERIES ANALYSIS        TAU= 1
      0.6386E 00  0.2576E 00  0.4222E 00  0.4452E 00
      0.2408E 00  0.4084E 00  0.3526E 00  0.3356E 00
      0.4489E 00  0.3810E 00  0.7907E 00  0.5363E 00
      0.4135E 00  0.3500E 00  0.4950E 00  0.7553E 00

    TIME SERIES ANALYSIS        TAU= 2
      0.5227E 00  0.2230E 00  0.3609E 00  0.3859E 00
      0.2121E 00  0.3789E 00  0.3021E 00  0.3041E 00
      0.4002E 00  0.3376E 00  0.6595E 00  0.4914E 00
      0.3602E 00  0.3143E 00  0.4204E 00  0.6234E 00

    TIME SERIES ANALYSIS        TAU= 3
      0.4491E 00  0.1929E 00  0.3090E 00  0.3246E 00
      0.1957E 00  0.3203E 00  0.2668E 00  0.2758E 00
      0.3023E 00  0.3026E 00  0.5649E 00  0.4414E 00
      0.3188E 00  0.2816E 00  0.3600E 00  0.5459E 00

    TIME SERIES ANALYSIS        TAU= 4
      0.4093E 00  0.1759E 00  0.2626E 00  0.2803E 00
      0.1724E 00  0.2729E 00  0.2449E 00  0.2609E 00
      0.3389E 00  0.2850E 00  0.5036E 00  0.4126E 00
      0.2950E 00  0.2635E 00  0.3333E 00  0.4900E 00

    TIME SERIES ANALYSIS        TAU= 5
      0.3818E 00  0.1607E 00  0.2371E 00  0.2639E 00
      0.1583E 00  0.2430E 00  0.2360E 00  0.2415E 00
      0.3117E 00  0.2707E 00  0.4674E 00  0.3859E 00
      0.2765E 00  0.2467E 00  0.3109E 00  0.4430E 00

    TIME SERIES ANALYSIS        TAU= 6
      0.3440E 00  0.1555E 00  0.2045E 00  0.2274E 00
      0.1479E 00  0.2235E 00  0.2201E 00  0.2287E 00
      0.2807E 00  0.2575E 00  0.4352E 00  0.3759E 00
      0.2580E 00  0.2351E 00  0.2871E 00  0.4074E 00

    TIME SERIES ANALYSIS        TAU= 7
      0.3089E 00  0.1526E 00  0.1766E 00  0.2091E 00
      0.1280E 00  0.1973E 00  0.2081E 00  0.2203E 00
      0.2554E 00  0.2447E 00  0.3987E 00  0.3563E 00
      0.2371E 00  0.2239E 00  0.2609E 00  0.3810E 00

    TIME SERIES ANALYSIS        TAU= 8
      0.2840E 00  0.1486E 00  0.1551E 00  0.1987E 00
      0.1320E 00  0.1646E 00  0.1917E 00  0.2109E 00
      0.2329E 00  0.2297E 00  0.3656E 00  0.3394E 00
      0.2193E 00  0.2052E 00  0.2239E 00  0.3521E 00

    TIME SERIES ANALYSIS        TAU= 9
      0.2570E 00  0.1315E 00  0.1435E 00  0.1787E 00
      0.1209E 00  0.1457E 00  0.1756E 00  0.2053E 00
      0.2165E 00  0.2105E 00  0.3425E 00  0.3205E 00
      0.2025E 00  0.1967E 00  0.2234E 00  0.3350E 00

    NUMBER OF POINTS--N= 2151
```

Autocorrelation Matrices $\tilde{\underline{P}}_{\delta\delta}(\tau)$

Figure 6.2

MEAN= 0.0[illegible] SIGMA= 0.46465 SKEW= -0.35193 VMIN= -2.49900 VMAX= 3.17800 N= 2160

	RANGE		FREQ	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-2.4990	-2.3255	1	*
2	-2.3255	-2.1519	1	*
3	-2.1519	-2.1584	21	
4	-2.1584	-2.0048	0	
5	-2.0048	-1.8313	6	*
6	-1.8313	-1.8178	1	
7	-1.8178	-1.7042	5	*
8	-1.7042	-1.5907	5	*
9	-1.5907	-1.4771	7	*
10	-1.4771	-1.3636	4	*
11	-1.3636	-1.2501	10	**
12	-1.2501	-1.1365	6	*
13	-1.1365	-1.0230	14	***
14	-1.0230	-0.9094	13	***
15	-0.9094	-0.7959	21	*****
16	-0.7959	-0.6824	32	********
17	-0.6824	-0.5688	41	**********
18	-0.5688	-0.4553	71	*****************
19	-0.4553	-0.3417	98	************************
20	-0.3417	-0.2282	122	******************************
21	-0.2282	-0.1147	209	**
22	-0.1147	-0.0011	300	**
23	-0.0011	0.1124	387	**
24	0.1124	0.2260	279	***
25	0.2260	0.3395	206	**
26	0.3395	0.4530	118	****************************
27	0.4530	0.5666	70	*****************
28	0.5666	0.6801	46	**********
29	0.6801	0.7937	31	*******
30	0.7937	0.9072	13	***
31	0.9072	1.0207	20	*****
32	1.0207	1.1343	11	**
33	1.1343	1.2478	9	**
34	1.2478	1.3614	5	*
35	1.3614	1.4749	6	*
36	1.4749	1.5884	2	
37	1.5884	1.7020	0	
38	1.7020	1.8155	1	
39	1.8155	1.9291	0	
40	1.9291	2.0426	1	*
41	2.0426	2.1561	0	
42	2.1561	2.2697	0	
43	2.2697	2.3832	0	
44	2.3832	2.4968	0	
45	2.4968	2.6103	2	*
46	2.6103	2.7238	1	*
47	2.7238	2.8374	0	
48	2.8374	2.9509	0	
49	2.9509	3.0645	0	
50	3.0645	3.1780	1	*

SCALING FACTOR = 4

Figure 6.3. Full Model Residuals for Monitor L from Jan. — Mar. 1975

MEAN= [illegible] SIGMA= 0.81593 SKEW= -0.15655 VMIN= -2.57100 VMAX= 3.62700 NV= 2160

	RANGE		FREQ	1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
1	[illegible]	-2.4870	2	
2	-2.4870	-2.3231	0	
3	-2.3231	-2.1591	1	
4	-2.1991	-2.0752	1	
5	-2.0752	-1.9512	2	
6	-1.9512	-1.8272	1	
7	-1.8272	-1.7033	2	
8	-1.7033	-1.5793	3	
9	-1.5793	-1.4554	5	*
10	-1.4554	-1.3314	6	
11	-1.3314	-1.2074	5	*
12	-1.2074	-1.0835	11	**
13	-1.0835	-0.9595	15	***
14	-0.9595	-0.8356	15	***
15	-0.8356	-0.7116	20	****
16	-0.7116	-0.5876	47	*********
17	-0.5876	-0.4637	74	**************
18	-0.4637	-0.3397	104	********************
19	-0.3397	-0.2158	150	******************************
20	-0.2158	-0.0918	237	**
21	-0.0918	0.0322	282	***
22	0.0322	0.1561	327	**
23	0.1561	0.2801	265	***
24	0.2801	0.4040	157	*******************************
25	0.4040	0.5280	68	**************
26	0.5280	0.6520	48	*********
27	0.6520	0.7759	36	*******
28	0.7759	0.8999	22	****
29	0.8999	1.0238	13	**
30	1.0238	1.1478	7	*
31	1.1478	1.2718	7	*
32	1.2718	1.3957	2	
33	1.3957	1.5197	2	
34	1.5197	1.6436	4	
35	1.6436	1.7676	0	
36	1.7676	1.8916	1	
37	1.8916	2.0155	0	
38	2.0155	2.1395	0	
39	2.1395	2.2634	1	
40	2.2634	2.3874	0	
41	2.3874	2.5114	1	
42	2.5114	2.6353	0	
43	2.6353	2.7593	0	
44	2.7593	2.8832	0	
45	2.8832	3.0072	0	
46	3.0072	3.1312	0	
47	3.1312	3.2551	0	
48	3.2551	3.3791	0	
49	3.3791	3.5030	0	
50	3.5030	3.6270	1	

SCALING FACTOR = 5

Figure 6.4. Full Model Residuals for Monitor M from Jan. - Mar. 1975

MEAN= 0.01949 SIGMA= 0.47911 SKEW= -0.55743 VMIN= -3.46700 VMAX= 0.39500 N= 2.00

	RANGE		FREQ	1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
1	-3.4670	-3.3298	1	*
2	-3.3298	-3.1925	0	
3	-3.1925	-3.0553	0	
4	-3.0553	-2.9180	0	
5	-2.9180	-2.7808	1	*
6	-2.7808	-2.6436	0	
7	-2.6436	-2.5063	0	
8	-2.5063	-2.3691	0	
9	-2.3691	-2.2318	1	*
10	-2.2318	-2.0946	1	*
11	-2.0946	-1.9574	4	****
12	-1.9574	-1.8201	2	**
13	-1.8201	-1.6829	6	******
14	-1.6829	-1.5456	5	*****
15	-1.5456	-1.4084	7	*******
16	-1.4084	-1.2712	5	*****
17	-1.2712	-1.1339	15	***************
18	-1.1339	-0.9967	17	*****************
19	-0.9967	-0.8594	14	**************
20	-0.8594	-0.7222	19	*******************
21	-0.7222	-0.5850	45	***
22	-0.5850	-0.4477	121	**********
23	-0.4477	-0.3105	148	**********
24	-0.3105	-0.1732	164	**********
25	-0.1732	-0.0360	270	**********
26	-0.0360	0.1012	413	**********
27	0.1012	0.2385	329	**********
28	0.2385	0.3757	253	**********
29	0.3757	0.5130	112	**********
30	0.5130	0.6502	76	**********
31	0.6502	0.7874	85	**********
32	0.7874	0.9247	26	**********
33	0.9247	1.0619	21	**********
34	1.0619	1.1992	12	**********
35	1.1992	1.3364	6	**********
36	1.3364	1.4736	7	**********
37	1.4736	1.6109	5	**********
38	1.6109	1.7481	2	**********
39	1.7481	1.8854	2	**********
40	1.8854	2.0226	0	**********
41	2.0226	2.1598	0	**********
42	2.1598	2.2971	0	**********
43	2.2971	2.4343	0	**********
44	2.4343	2.5716	0	**********
45	2.5716	2.7088	0	**********
46	2.7088	2.8460	1	**********
47	2.8460	2.9833	0	**********
48	2.9833	3.1205	0	**********
49	3.1205	3.2578	0	**********
50	3.2578	3.3950	1	**********

SCALING FACTOR = 5

Figure 6.5. Full Model Residuals for Monitor N from Jan. - Mar. 1975

VMU= 0.00563 SIGMA= 0.9605 SKEW= -0.39695 VMIN= -2.40700 VMAX= 2.95300 NV= 2160

	RANGE		FREQ	1...5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100
1	-2.4070	-2.2998	2	
2	-2.2998	-2.1926	7	
3	-2.1926	-2.0854	0	
4	-2.0854	-1.9782	4	*
5	-1.9782	-1.8710	3	
6	-1.8710	-1.7638	10	**
7	-1.7638	-1.6566	4	*
8	-1.6566	-1.5494	6	*
9	-1.5494	-1.4422	7	*
10	-1.4422	-1.3350	7	*
11	-1.3350	-1.2278	4	*
12	-1.2278	-1.1206	15	***
13	-1.1206	-1.0134	10	**
14	-1.0134	-0.9062	12	***
15	-0.9062	-0.7990	15	***
16	-0.7990	-0.6918	21	*****
17	-0.6918	-0.5846	35	********
18	-0.5846	-0.4774	79	********************
19	-0.4774	-0.3702	109	**************************
20	-0.3702	-0.2630	165	**
21	-0.2630	-0.1558	129	********************************
22	-0.1558	-0.0486	208	**
23	-0.0486	0.0586	361	**********
24	0.0586	0.1658	238	**********
25	0.1658	0.2730	213	**********
26	0.2730	0.3802	159	**********
27	0.3802	0.4874	95	**********
28	0.4874	0.5946	70	**********
29	0.5946	0.7018	58	**********
30	0.7018	0.8090	39	*********
31	0.8090	0.9162	30	*******
32	0.9162	1.0234	16	****
33	1.0234	1.1306	14	***
34	1.1306	1.2378	12	***
35	1.2378	1.3450	6	*
36	1.3450	1.4522	1	
37	1.4522	1.5594	4	*
38	1.5594	1.6666	0	
39	1.6666	1.7738	3	
40	1.7738	1.8810	0	
41	1.8810	1.9882	2	
42	1.9882	2.0954	1	
43	2.0954	2.2026	3	
44	2.2026	2.3098	7	
45	2.3098	2.4170	0	
46	2.4170	2.5242	0	
47	2.5242	2.6314	0	
48	2.6314	2.7386	7	
49	2.7386	2.8458	0	
50	2.8458	2.9530	7	

Figure 6.6. Full Model Residuals for Monitor 0 from Jan. - Mar. 1975

7. Model Behavior

The behavior of the model with structure as given in Section 3 and with parameter values estimated as discussed in Section 5 is now summarized.

The estimated value of the constant c_o of (3.1) is

$$\hat{c}_o \simeq 10 \text{ ppb}$$

The fact that $\ln c_o$ is the long term time and space average of $\ln c_k(n)$ does not mean that c_o is the average value of the $c_k(n)$.

7.1 Global Concentration Field: Mean Value

The mean value of the global concentration field $m_g(n,\underline{s})$ is given by (3.1.1).

The estimated value of ϕ_g was

$$\hat{\phi}_g = 0.9$$

As discussed in Section 3.5 this corresponds to an effective memory span of 9 to 15 hours, i.e., the model "remembers past inputs $b_g(n,\underline{s})$ for 9 to 15 hours.

Figure 7.1.1 and 7.1.2 show the wind direction dependence term $b_\theta[\underline{s},\theta(n)]$ of (3.1.2) (as defined in Section 8) for $\underline{s}$ corresponds to the location of monitors M and O. The shape of $b_\theta[\underline{s}, \theta(n)]$ changes with coordinate $\underline{s}$ as the background sources "seen" for a particular wind direction $\theta(n)$ depend on the location of the point $\underline{s}$ within the region being modeled. The main peaks in Figure 7.1.1 and 7.1.2 correspond to Pittsburgh. The secondary peaks at 10-40° are felt to be "fake peaks" that result from the Fourier series nature of the model (see Section 8) and the fact that the data base had very few wind directions in the 0-90° quadrant (see Figure 5.1.5). These fake peaks could probably be removed by using more terms in the Fourier series expansion.

Figure 7.1.3 shows a 24 hour variation of the time of day dependence term $b_{h,d}(n)$ of (3.1.2), i.e., $b_{h,d}(n)$ n=1,,,24, for a specific day d (see Section 8). For the 3 month data base, there was very little change in this shape as a function of day d. Of course for a full year's data base, the shape of $b_{h,d}(n)$ for any 24 hour period will exhibit definite dependence in day d.

7.2 Global Concentration Field: Random Component

The random component of the global concentration field $\Delta_g(n)$, is defined by (3.2.1). The ad hoc parameter estimation procedure (see Section 9.3) imposed the artificial constraint that, [see (3.1.1)]

$$\phi_g = \phi_{g_\Delta}$$

so that from Section 7.1

$$\hat{\phi}_{g\Delta} = 0.9$$

The variance R_g of the white normal random process input $w_g(n)$ was estimated to be

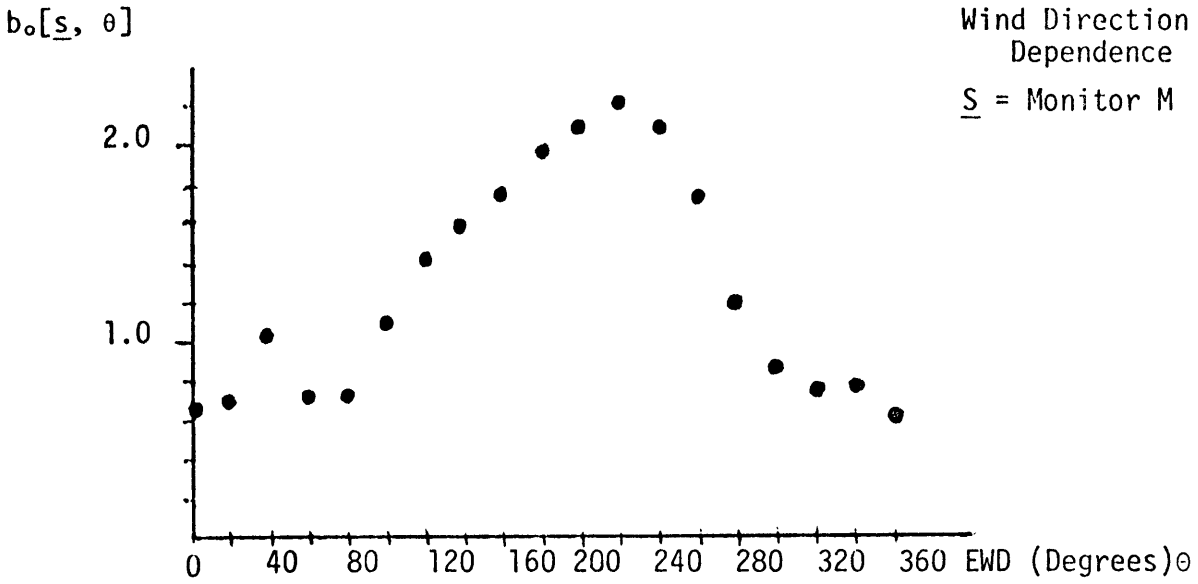


Figure 7.1.1

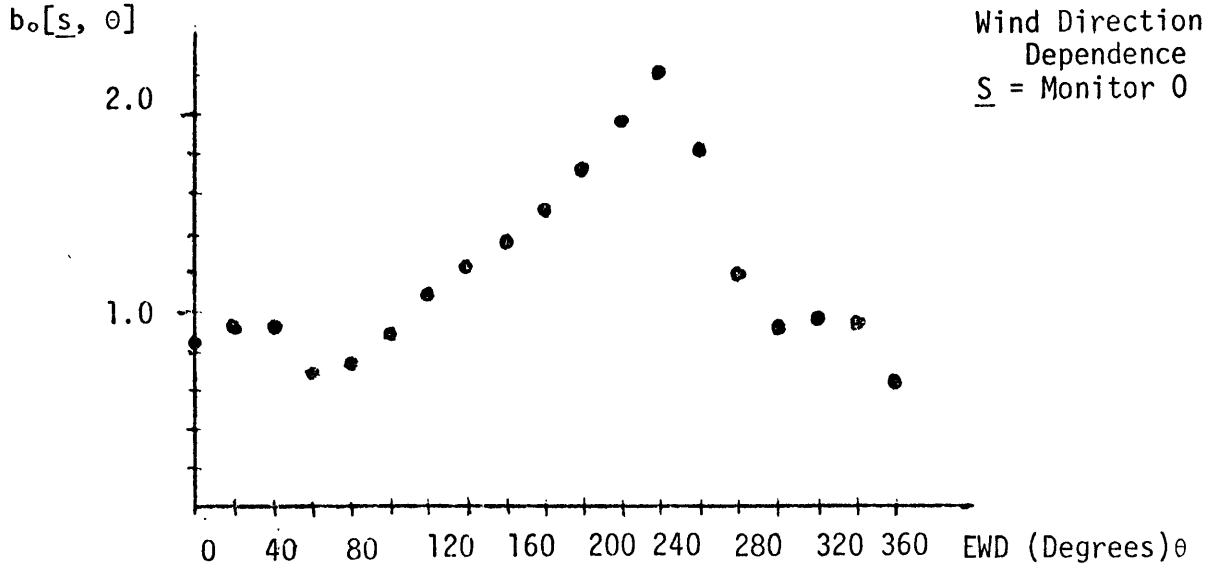


Figure 7.1.2

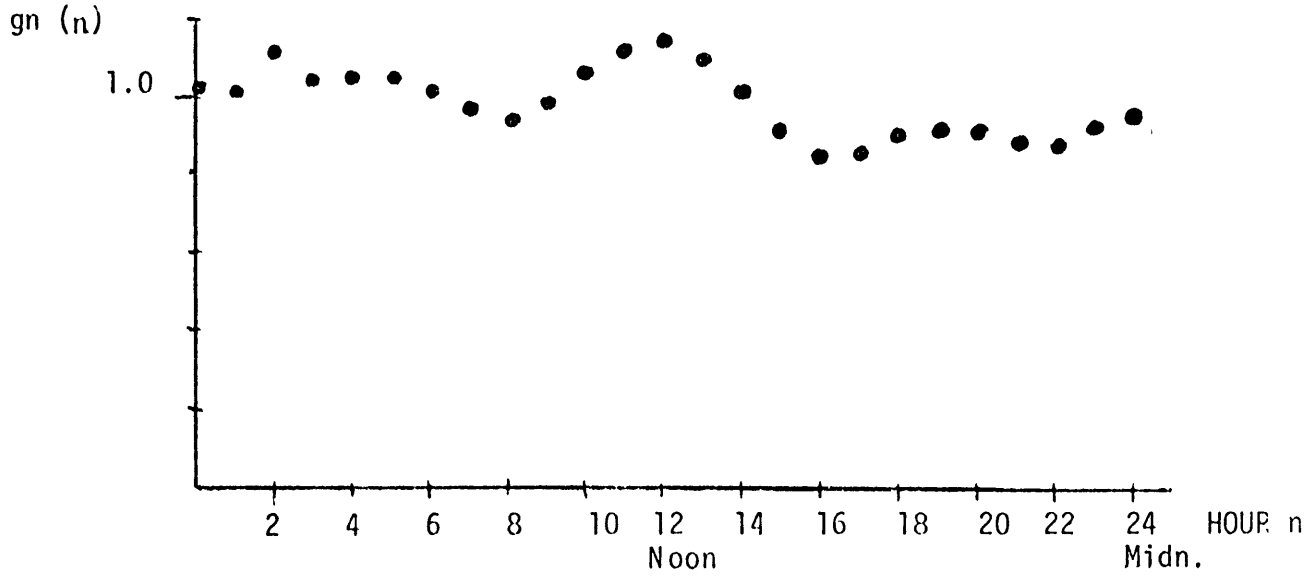


Figure 7.1.3

$$\hat{R}_g = 0.08$$

Thus, using (3.5.3)

$$\Gamma_g = 0.42$$

where

$$\Gamma_g = E\{[\ln \Delta_g(n)]^2\}$$

Thus, using (3.5.4)

$$E\{\Delta_g(n)\} = e^{\Gamma_g/2} = 1.23$$

$$E\{\Delta_g(n) - E[\Delta_g(n)]\}^2 = e^{\Gamma_g}(e^{\Gamma_g}-1)$$
$$= .79$$

7.3 Local Variations: Mean Value

The mean value of the local variation $m_\ell(\underline{s})$ was defined in Section 3.3 only at the monitor locations $\underline{s}_k$, k=1...K by (3.3.1) or

$$m_\ell(\underline{s}) = m_{\ell,k}$$

After satisfying the constraint (3.3.3) the estimated values of $m_{\ell,k}$, k=1...K are as given in Figure 7.3.1.

Monitor	$\hat{m}_{\ell,k}$	$\ln m_{\ell,k}$
L	1.1	0.1
M	1.65	0.5
N	0.75	-0.3
O	0.75	-0.3

Figure 7.3.1

Bias Terms for Monitors L,M,N,O

Figure 7.3.1 labels the $m_{\ell,k}$ as "bias terms" for the respective moni-

tors. The question of whether these biases are due to monitor calibration, local sources, terrain effects, etc. cannot be resolved by modeling of the type being done here. However, Figure 7.3.1 shows that the effect is non-negligible.

7.4 Local Variations: Random Component

The random component of the local variation $\Delta_\ell(n,\underline{s})$ is defined by (3.4.1).

The estimated values of the "time constant" parameter ϕ_ℓ is

$$\phi_\ell = 0.78$$

so that the local variations have an effective memory span (see Section 3.5) of 3 to 6 hours.

The variance Γ_ℓ of the white process input $w_\ell(n)$ was estimated to be

$$\hat{R}_\ell = 0.15$$

Thus using (3.5.3)

$$\Gamma_\ell = 0.384$$

where

$$\Gamma_\ell = E\{[\ln \Delta_\ell(n,\underline{s})]^2\}$$

Thus from (3.5.4)

$$E\{\Delta_\ell(n,\underline{s})\} = e^{\Gamma_\ell/2} = 1.21$$

$$E\{\Delta_\ell(n,\underline{s}) - E[\Delta_\ell(n,\underline{s})]\}^2 = 0.69$$

Comparison of the global concentration field $\Delta_g(n)$ and local variations $\Delta_\ell(n,\underline{s})$ random processes show they have different time dynamics/memory spans (the local variations are faster) but roughly the same magnitude.

7.5 Wind Speed, Mixing Depth Dependence

The hypothesized structure of Section 3 has only one meteorological input, the elevated tower wind direction $\theta(n)$. It was initially expected that meteorological conditions such as

$v(n)$: wind speed

$h(n)$: mixing height

whould also be important. Therefore a variety of parameter estimations (as in Section 5) and model verification (as in Section 6) studies were made where the deterministic function $b_g[n,\underline{s}]$ of (3.1.2) was replaced by

$$\frac{b_g[n,\underline{s}]}{v(n)} \qquad (7.5.1)$$

and also by

$$\frac{b_g[n,\underline{s}]}{h(n)} \qquad (7.5.2)$$

It turned out that the use of (7.5.1) or (7.5.2) did not yield a better model (using the model validity type criterion on Section 6). Therefore wind speed and mixing height were not included in the hypothesized structure (this discussion is an example of an iteration "around the loop" of Figure 2.1).

This "lack of dependence" on wind speed and mixing height has important potential implications relative to the understanding of pollution propagation which should be explored further. However it must be emphasized that care is needed in drawing general conclusions from statistical model developments of the type being discussed here. At the present time, we can only say that inclusion of wind speed and mixing depth as in (7.5.1) and (7.5.2) did not improve the model. This statement does not "prove" that background concentrations are independent of wind speed and mixing depth.

7.6 Effect of Keystone

As discussed in Section 5.2, all the data for the three months was fit to the background model which made no attempt to account for the effect of the Keystone plant.

When the background modeling study was started, it was intended to handle Keystone effects at least partially by the following procedure:

- Estimate background parameters using all the data
- Compute the residuals and study all large (positive) values
- If the wind direction is right and if Keystone was generating, say the large residuals were caused by Keystone (i.e., the monitor was sensing both background and Keystone).
- Remove the "Keystone data" and re-estimate the parameters of the background model.
- Repeat until convergence.

This procedure was not followed because, as discussed in Section 6, almost all of the large residuals were associated with startup problems.

This does not imply that Keystone has no effect on Monitors L, M, N, O.

It merely states that, with a few possible exceptions we were unable to separate Keystone effects from the background using the type of background model being discussed here and using the limited three months data base.

7.7 Summary of Model: January - March 1975

Combination of the parameter values discussed above with the structure of (3.1) yields the overall model which can be summarized as follows:

$c(n,\underline{s})$: SO_2 background concentration at hour n and horizontal coordinate $\underline{s}$

$$c(n,\underline{s}) = m_g(n,\underline{s})\ \Delta_g(n)\ \ m_\ell(s)\ \ \Delta_\ell(n,\underline{s})\ c_o$$

Global Field: Mean Value: $m_g(n,\underline{s})$

. Models effect of wind direction, time of day, day of year

$$\ln m_g(n+1,\underline{s}) = 0.9 \ln m_g(n,\underline{s}) + 0.1 \{\ln b_\theta[\underline{s},\theta(n)] + \ln b_{h,d}(n)\}$$

$b_\theta[\underline{s},\theta(n)]$: effect of wind direction (See Figures 7.1.1, 7.1.2)

$b_{h,d}(n)$: effect of time of day h and day of year d (see Figure 7.1.3).

Global Field: Random Component $\Delta_g(n)$

. Models global random variation in time

$$\ln \Delta_g(n+1) = 0.9 \ln \Delta_g(n) + w_g(n)$$

$$E\{w_g^2(n)\} = R_g = 0.08$$

$$E\{[\ln \Delta_g(n,\underline{s})]^2\} = r_g = 0.42$$

$$E\{\Delta_g(n)\} = 1.23$$

$$E\{[\Delta_g(n) - E[\Delta_g(n)]^2\} = 0.79$$

Local Variations: Mean Values: $m_\ell(\underline{s})$

. Models long term average differences of monitors' readings

$m_\ell(\underline{s}) = m_{\ell,k}$ k=1...K (see Figure 7.3.1)

Local Variations: Random Component $\Delta_\ell(n,\underline{s})$

. Models local random variations in time and space

$$\ln \Delta_\ell(n+1,\underline{s}) = 0.78 \ln \Delta_\ell(n,\underline{s}) + w_\ell(n)$$

$$E\{w_\ell^2(n)\} = R_\ell = 0.15$$

$$E\{[\ln \Delta_\ell(n,\underline{s})]^2\} = r_\ell = 0.38$$

$E\{\Delta_\ell(n,\underline{s})\} = 1.21$

$E\{[\Delta_\ell(n,\underline{s}) - E[\Delta_\ell(n,\underline{s})]^2\} = 0.69$

Time Space Average Concentration

$\ln c_o$: time space average of $\ln c(n,\underline{s})$

$c_o = 10$ ppb

From the above model it is possible to compute the following properties of $c(n,\underline{s})$

$E\{\ln c(n,\underline{s})\} = \ln m_g(n,\underline{s}) + \ln m_\ell(\underline{s}) + \ln 10$

$E\{[\ln c(n,\underline{s}) - E[\ln c(n,\underline{s})]^2\} = \Gamma_g + \Gamma_\ell$

$= 0.8$

$E\{c(n,\underline{s})\} = m_g(n,\underline{s})\ (1.23)\ m_\ell(s)\ (1.21)(10) = 15\ m_g(n,\underline{s})\ m_\ell(\underline{s})$ ppb

$E[c(n,\underline{s}) - E[c(n,\underline{s})]^2 = 2.8$

where "E" denotes expecation over the $\Delta_\ell(n,\underline{s})$ and $\Delta_g(n)$ processes and does not involve averaging over either time or wind direction variations.

Part II

Mathematical Details

8) Model for $b_g(n,\underline{s})$

9) Estimation Theory

10) Model Verification Theory

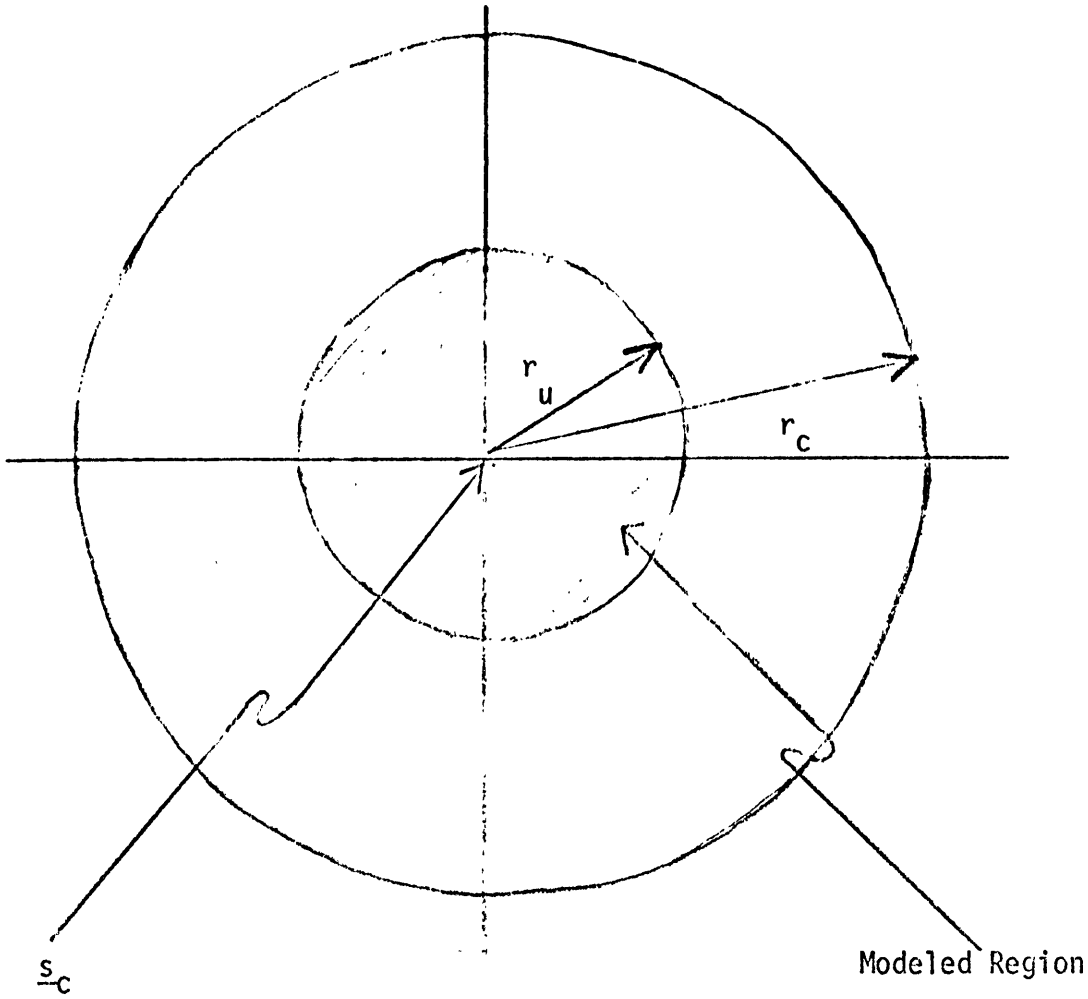
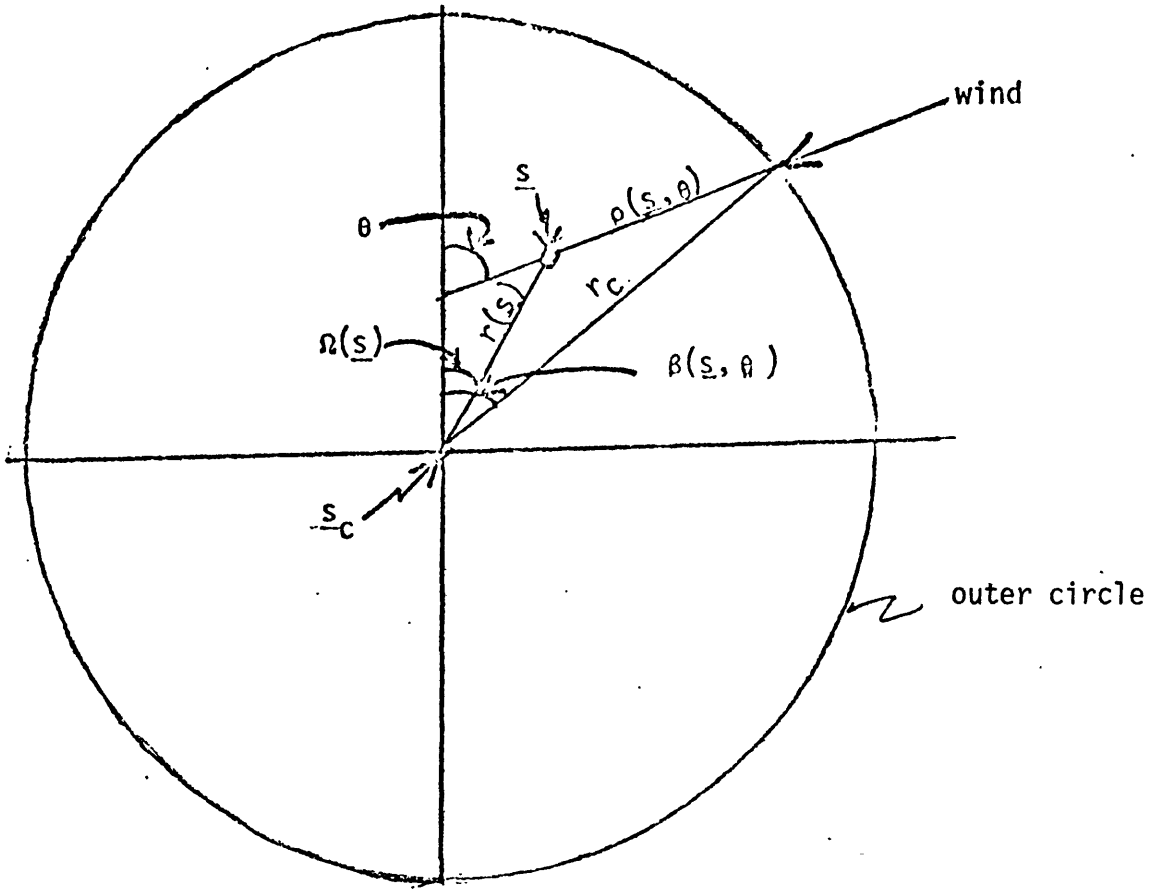


Figure 8.1.1



$\underline{s}_c$: center of outer circle

r_c: radius of outer circle

$r(\underline{s})$, $\Omega(\underline{s})$: polar coordinates of $\underline{s}$

θ: wind direction

$\rho(\underline{s}, \theta)$: distance from $\underline{s}$ to circle in direction θ

$\beta(\underline{s}, \theta)$: angle

Geometry Determing $\beta(\underline{s},\theta)$

Figure 8.1.2

8. Hypothesized Structure of $b_g(n,\underline{s})$

The hypothesized structure of Section 3 contains a term $b_g(n,\underline{s})$ which from (3.1.2)

$$b_g(n,\underline{s}) = b_\theta[\underline{s},\ \theta(n)]\ b_{h,d}(n) \qquad (8.1)$$

Explicit equations for these quantities were not given in Section 3 because they are rather "complicated looking." These are now stated.

8.1 Wind Dependence

The $b_\theta(\underline{s},\ \theta(n)]$ term of (8.1) models dependence on elevated tower wind direction $\theta(n)$.

Consider two concentric circles with radii r_c and r_u where

$$r_c > r_u$$

that are centered at $\underline{s} = \underline{s}_c$ when $\underline{s}_c$ is the "middle of" the geographic area of concern. Define

Modeled region: Area within circle of radius r_u

The background model is to be valid for any location $\underline{s}$ within the modeled region. See Figure 8.1.1. .

Consider a line drawn from some point $\underline{s}$ within the modeled region to the outer circle in the direction of θ, (the wind direction). Define

$\rho(\underline{s},\theta)$: distance from $\underline{s}$ to outer circle

$\beta(\underline{s},\theta)$: value of angle β where line intersects outer circle

Figure 8.1.2 illustrates the necessary geometry.

The hypothesized structure for $b_\theta(s,\underline{\theta})$ is the product of two terms

$$b_\theta(\underline{s},\theta) = b_\beta[\beta(\underline{s},\theta)]\ b_\rho[\rho(\underline{s},\theta)]$$

where

$b_\beta[\beta(\underline{s},\theta)]$: models effect of wind direction

$b_\rho[\rho(\underline{s},\theta)]$: models effect of distance from outer circle.

The function $b_\beta(\beta)$, $0 \leq \beta \leq 360°$ can be viewed as an effective normalized concentration source distributed around the outer circle where $b_\beta(\beta)$is large for β "pointing toward" major background sources such as Pittsburgh. The hypothesized structure is a Fourier series expression

$$\ln b_\beta(\beta) = x_{\beta 1} \sin(\frac{\beta}{360}) + x_{\beta 2} \cos(\frac{\beta}{360}) + x_{\beta 3}\sin(\frac{2\beta}{360}) + \ldots \qquad (8.1.1)$$

where the parameters $x_{\beta j}$ j=1... are to be estimated from the observations.

The distance dependence term is hypothesized to be

$$b_\rho(\rho) = \left[\frac{r_c}{\rho(\underline{s},\theta)}\right]^\gamma \qquad (8.1.2)$$

Thus for $\underline{s} = \underline{s}_c$ (center of the modeled region) $r_c = \rho$ and $b_\rho(\rho) = 1$ while $b_\rho(\rho) > 1$ or < 1 depending on whether $\underline{s}$ is "towards" or "away" from the outer circle. The value of γ is to be estimated. The distance dependence term was not mentioned in Section 7 because, when processing the actual observations, it was found that it did not improve the model for any significant degree. The results of Section 7 are for the case $\gamma = 0$ so $b_\rho(\rho) = 1$ for all θ. The term (8.1.2) is mentioned here because it might prove to be important in other parts of Chestnut Ridge or in other applications.

The value of the outer circle radius r_c and the circle center $\underline{s}_c$ are chosen by engineering judgment. It is reasonable to choose $\underline{s}_c$ and r_c so that the outer circle "passes near" the major background sources. Note that there is no particular reason why the "outer circle" has to be a circle. It could be any closed curve for which it is possible to compute the angle β and the distance ρ for a given coordinate $\underline{s}$ and wind direction θ.

The value of r_u, the radius of the inner circle, does not enter into the model. The inner circle was discussed only to clarify the overall concept.

8.2 Time of Day Dependence

The hypothesized structure of the time of day dependence term $b_{h,d}(n)$ of (8.1) involves two different time scales:

h: hour of day, h=1...24

d: day of year, d=1...365

n: time in one-hour steps, n=1...N

For each value of n, there are associated values of h and d. For the data base actually used (see Section 5.1)

d: 1...90 (3 months)

so

$$N \simeq (24)(90) \simeq 2000$$

The time dependence $b_{h,d}(n)$ is hypothesized to be a Fourier series in hour of day h where each of the daily Fourier coefficients is a polynomial in day of year d.

$$\ln b_{h,d}(n) = a_1(d) \sin\left(\frac{h}{24}\right) + a_2(d) \cos\left(\frac{h}{24}\right) + a_3(d) \sin\left(\frac{2h}{24}\right) + a_4(d) \cos\left(\frac{2h}{24}\right) + \ldots$$

$$a_j(d) = x_{0j} + x_{1j}d + x_{2j}d^2 + \ldots \qquad (8.2.1)$$

Values of the parameters x_{ij}, i=0,1..,j=1... are to be estimated from the measurements. If the 24 hour Fourier series has 6 terms (3 sin, 3 cosine), and if second degree polynomials are used, then there are 3 x 6 = 18 parameters x_{ij}.

As discussed in Section 7.1, the 3 month data base showed no major seasonal dependence (dependence on day d). For processing a year or more of data, the Fourier coefficients $a_j(d)$ might be better modeled as Fourier series in d rather than polynomials in d. The structure (8.2.1) is called a "multiple harmonic expansion" and has been used successfully in other applications such as modeling the hourly dependence of quantities such as weather and electric power demand (see Woodard, 1976).

9. Estimation of Parameters

Some of the mathematical theory underlying the parameter estimation discussions of Section 5.2 is now reviewed. More detailed discussions, derivations, etc. can be found in many places (such as Schweppe, 1974).

Define

$$z_k(n) = \ln c_k(n) \qquad k=1...K;\ n=1...N$$

$$\underline{z}(n) = \begin{bmatrix} z_1(n) \\ \cdot \\ \cdot \\ \cdot \\ z_K(n) \end{bmatrix}$$

where $c_k(n)$ are the measured concentrations at time n at monitor k.

9.1 Maximum Likelihood Probability Estimation

The hypothesized structures of Section 3 have many parameters whose values are to be estimated:

Define:

$\underline{\alpha}$: vector of parameters to be estimated

Define

$$\underline{z}_N = \begin{bmatrix} \underline{z}(1) \\ \cdot \\ \cdot \\ \cdot \\ \underline{z}(N) \end{bmatrix}$$

so $\underline{z}_N$ is a vector with NK elements that represents all of the observed concentrations. For a given value of $\underline{\alpha}$, the hypothesized structure defines the probability distribution of $\underline{z}_N$, i.e.,

$p(\underline{z}_N: \underline{\alpha})$: probability density function of $\underline{z}_N$ given $\underline{\alpha}$

Henceforth $p(\underline{z}_N:\underline{\alpha})$ is called the likelihood function.

Let $\hat{\underline{\alpha}}$ denote the "best estimate" of $\underline{\alpha}$ that can be made from $\underline{z}_N$. The definition of "best" is

$\hat{\underline{\alpha}}$: value of $\underline{\alpha}$ that maximizes $p(\underline{z}_N:\underline{\alpha})$ for given $\underline{z}_N$

$\hat{\underline{\alpha}}$ is called a maximum likelihood estimate. The use of maximum likelihood estimates is reasonable and there is a considerable mathematical theory justifying its use.

For the models of the type hypothesized in Section 3, it can be shown that

$$2 \ln p(\underline{z}_N:\underline{\alpha}) = -N K \ln 2\pi - \sum_{n=1}^{N} \ln|\underline{\Sigma}_z(n/n-1:\underline{\alpha})|$$

$$- \sum_{n=1}^{N} \underline{\delta}_z^T(n:\underline{\alpha})\, \underline{\Sigma}_z^{-1}(n/n-1:\underline{\alpha})\, \underline{\delta}_z(n:\underline{\alpha}) \qquad (9.1.1)$$

where

$\hat{\underline{z}}(n/n-1:\underline{\alpha})$: conditional expectation of $\underline{z}(n)$ given $\underline{z}(1)...\underline{z}(n-1)$ for specified $\underline{\alpha}$

$\underline{\delta}_z(n:\underline{\alpha}) = \underline{z}(n) - \hat{\underline{z}}(n/n-1:\underline{\alpha})$

$\underline{\Sigma}_z(n/n-1:\underline{\alpha}) = E\{\underline{\delta}_z(n:\underline{\alpha})\, \underline{\delta}_z^T(n:\underline{\alpha})\}$

Evaluation of $\hat{\underline{\alpha}}$, the value that maximizes $p(\underline{z}_N:\alpha)$, is done by iterative search techniques (see, for example, GPSIE, Peterson, 1976; Peterson, 1971).

9.2 State Estimation

In order to use (9.1.1) to evaluate $\ln p(\underline{z}_N:\underline{\alpha})$, it is necessary to be able to compute $\hat{\underline{z}}(n/n-1:\underline{\alpha})$ and $\underline{\Sigma}_z(n/n-1:\underline{\alpha})$ for any specified value of $\underline{\alpha}$. For the present discussion, the $\underline{\alpha}$ dependence is dropped from the equations.

Consider a model of the form

$$\underline{x}(n+1) = \underline{\Phi x}(n) + \underline{G}\,\underline{w}(n) + \underline{B}\,\underline{u}(n) \qquad (9.2.1)$$

$$\underline{z}(n) = \underline{H}\,\underline{x}(n)$$

$\underline{w}(n)$: zero mean white, Gaussian process with

$$E\{\underline{w}(n_1)\underline{w}^T(n_2)\} = \begin{cases} \underline{R} & n_1 = n_2 \\ \underline{0} & n_1 \neq n_2 \end{cases}$$

$\underline{u}(n)$: known input

The hypothesized structure of Section 3 is of this form where u(n) corresponds to the input $b_g(\underline{n},s)$, etc. For (9.2.1) it can be shown that

$$\hat{\underline{z}}(n/n-1) = \underline{H}\,\hat{\underline{x}}(n/n-1) \qquad (9.2.2)$$

$$\underline{\Sigma}_z(n/n-1) = \underline{H}\,\underline{\Sigma}_x(n/n-1)\,\underline{H}^T$$

$$\hat{\underline{x}}(n/n-1) = \underline{\Phi}\,\hat{\underline{x}}(n-1/n-1) + \underline{B}\,\underline{u}(n-1)$$

$$\hat{\underline{x}}(n/n) = \hat{\underline{x}}(n/n-1) + \underline{K}(n)\,[\underline{z}(n) - \underline{H}\,\hat{\underline{x}}(n/n-1)]$$

$$\underline{K}(n) = \underline{\Sigma}_x(n/n-1)\,\underline{H}^T\,[\underline{H}\,\underline{\Sigma}_x(n/n-1)\underline{H}^T]^{-1}$$

$$\underline{\Sigma}_x(n/n-1) = \underline{\Phi}\,\underline{\Sigma}_x(n-1/n-1)\,\underline{\Phi}^T + \underline{G}\,\underline{R}\,\underline{G}^T$$

$$\underline{\Sigma}_x(n/n) = \underline{\Sigma}_x(n/n-1) - \underline{\Sigma}_x(n/n-1)\,\underline{H}^T\,[\underline{H}\,\underline{\Sigma}_x(n/n-1)\,\underline{H}^T]^{-1}\,\underline{H}\,\underline{\Sigma}_x(n/n-1)$$

This set of equations is called a state estimator (or a recursive estimator, or in electrical engineering, a Kalman filter.) It is illustrated symbolically in Figure 4.1.3.

The beauty of (9.1.1) and (9.2.2) is that they form a recursive set of equations for evaluating $p(\underline{z}_N:\underline{\alpha})$ for any value of $\underline{\alpha}$ that is relatively easy to implement (for example, GPSIE contains an even more general version of 9.2.2).

9.3 Relation to Algorithms Actually Used

As discussed in Section 5.2, the actual algorithms used were "ad hoc" engineering approximations to (9.1.1) and (9.2.2). To be more precise on the nature of these ad hoc algorithms, define

$$\underline{\alpha} = \begin{bmatrix} \underline{\alpha}_1 \\ \\ \underline{\alpha}_2 \end{bmatrix}$$

$\underline{\alpha}_1$: parameters associated with $b_g(n,\underline{s})$ of Section 8

$\underline{\alpha}_2$: other parameters (ϕ,R)

The estimate $\hat{\underline{\alpha}}_1$ for $\underline{\alpha}_1$ was obtained by linear least squares regression, completely ignoring the dynamics of the hypothesized structure. The estimate $\hat{\underline{\alpha}}_2$ for $\underline{\alpha}_2$ was obtained by fixing $\underline{\alpha}_1 = \hat{\underline{\alpha}}_1$ and then doing brute force search (combined with eyeballing the autocorrelation functions) over $\ln p(\underline{z}_N:\underline{\alpha})$ of (9.1.1) with the approximation that $\underline{\Sigma}_z = \underline{I}$ and the gain $\underline{K}(n)$ of (9.2.2) was a constant matrix calculated in an approximate fashion.

It is felt that $\hat{\underline{\alpha}}$ obtained as discussed above are reasonably close to the maximum likelihood estimates. However this is not proven until a program such as GPSIE is used on the data.

10. Model Verification Theory

The types of model verification tests actually used were discussed in Section 6. The following discusses model verification in more general terms to put what was done in perspective. As in Section 9, detailed discussions, derivations, etc. can be found in various references (such as Schweppe, 1974).

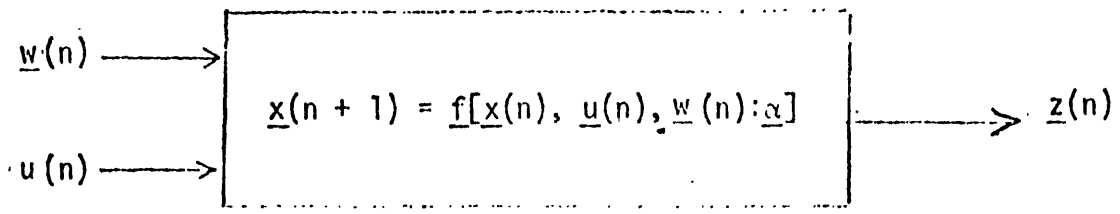
The basic model form to be discussed is summarized in Figure 10.1 where for the background model

$\underline{u}(n)$: measured elevated tower wind direction; hour of day h and day of year d.

$\underline{z}(n)$: ln of observed monitor SO_2 concentrations

$\underline{f}[\ ]$, $\underline{x}(n)$, $\underline{\alpha}$: as in Section 9

The parameter vector $\underline{\alpha}$ includes parameters which define the statistics of the random processes $\underline{w}(n)$. The stochastic input $\underline{w}(n)$ models both random atmospheric phenomena and the very important uncertainties that arise because the measured inputs $\underline{u}(n)$ do not equal "actual" system values. It also models observation errors and local atmospheric variations. For the sake of this



n : n = 1 ... , time

u(n) : known input (measured)

w(n) : random process (not measured directly)

x(n) : state vector (not measured directly)

α : parameters of model (to be estimated)

f[ ] : hypothesized structure of model

z(n) : measured outputs

Basic Model Form

Figure 10.1

discussion, it is arbitrarily assumed that $\underline{w}(n)$ is a normal random process. However, the concepts also apply to other types of probability distributions.

It is assumed that a structure has been hypothesized and an estimate $\hat{\underline{\alpha}}$ of its parameters has been computed. The quantity whose validity is to be tested is the "model" which consists of the hypothesized structure and the estimated parameter values. A precise definition of the boundary between a valid and an invalid model is dependent upon the model's application and depends to a great extent on human judgment.

10.1 Accuracy vs. Validity

There is a basic difference between "accuracy" and "statistical validity". To illustrate this difference, consider the following simple example which defines a mathematical model and the corresponding "real world". First consider the "real world":

$$c = Q + w$$

$$Q = 1$$

$$w : \text{zero mean random variable}$$

$$E\{w^2\} = R_{real} \qquad (10.1.1)$$

Then consider the "model":

$$c = Q + w$$

$$Q = 1$$

$$w : \text{zero mean random variable}$$

$$E\{w^2\} = R_{model}$$

$$\hat{c} = 1 \qquad (10.1.2)$$

Define the model "accuracy" as follows:

$$\text{error} = \delta = \hat{c} - c$$

$$E\{\delta^2\} = R_{real} \qquad (10.1.3)$$

$$(R_{real})^{1/2} : \text{accuracy of model}$$

Consider first the case where

$$R_{model} = 10^4 \qquad (10.1.4)$$

$$R_{real} = 10^{-2}$$

In this case the model is accurate as the standard deviation of the error is 10^{-1}. However, the model is statistically invalid as the model "expects" the error to be around 10^2. The use of this invalid model in an SCS probabilistic control strategy results in much unnecessary expense. Since the controller has to be "sure" standards are not violated, the controller calls for substantial control action to compensate for errors that do not really exist.

In the case where

$$R_{model} = 10^{-2} \qquad (10.1.5)$$

$$R_{real} = 10^4$$

the model is again invalid. This results in an SCS control strategy which does not prevent violation of standards.

In the case where

$$R_{real} = R_{model} = 10^4 \qquad (10.1.6)$$

the model is statistically valid. The control strategy is as "expensive" as in (10.1.4) but at least the costs are actually required because of the uncertainty that exists in the real world.

The final case where

$$R_{real} = R_{model} = 10^{-2} \qquad (10.1.7)$$

is of course the ideal situation.

This difference between accuracy and statistical validity arises when dealing with probabilistic models which provide a measure of the accuracy of the predicted value as well as the predicted value itself.

10.2 Statistical Hypothesis Testing

The principle underlying statistical hypothesis testing (as it is being used here) can be summarized as follows. Assume a model is specified. Then there are two hypotheses H_0 and H_1 where

H_0: Hypothesis is that model is valid

H_1: H_0 is not true.

Testing of these hypotheses is done using "test statistics" where

Test Statistic: Some quantity which can be computed from the available observations and whose statistical properties are known when H_0 is true.

The actual testing proceeds as illustrated in Figure 10.2.1.

It is important to emphasize that statistical hypothesis testing does not prove H_0 true even if all test statistics of interest have yielded consistent results. It only indicates that it was not possible to reject H_0 using the available-observations with the chosen set of test statistics.

It would be ideal to be able to define a "complete set of test statistics" where, by definition, if the testing process of Figure 10.2.1 goes through all the test statistics of a complete set without rejecting H_0, then there exists no test statistic which will reject H_0. Unfortunately a complete set of test statistics does not exist (to the author's knowledge). The test statistics to follow form a "good set" of test statistics for use in Figure 10.2.1.

The statistical properties of the test statistics are discussed only for the case where N is very large and $\underline{\alpha}$ is the true value of $\underline{\alpha}$. More precise equations are available.

Define

$\hat{\underline{z}}(n/n-1)$: best estimate of $\underline{z}(n)$ made from $\underline{z}(m)$, $u(m)$

$m=1...n-1$

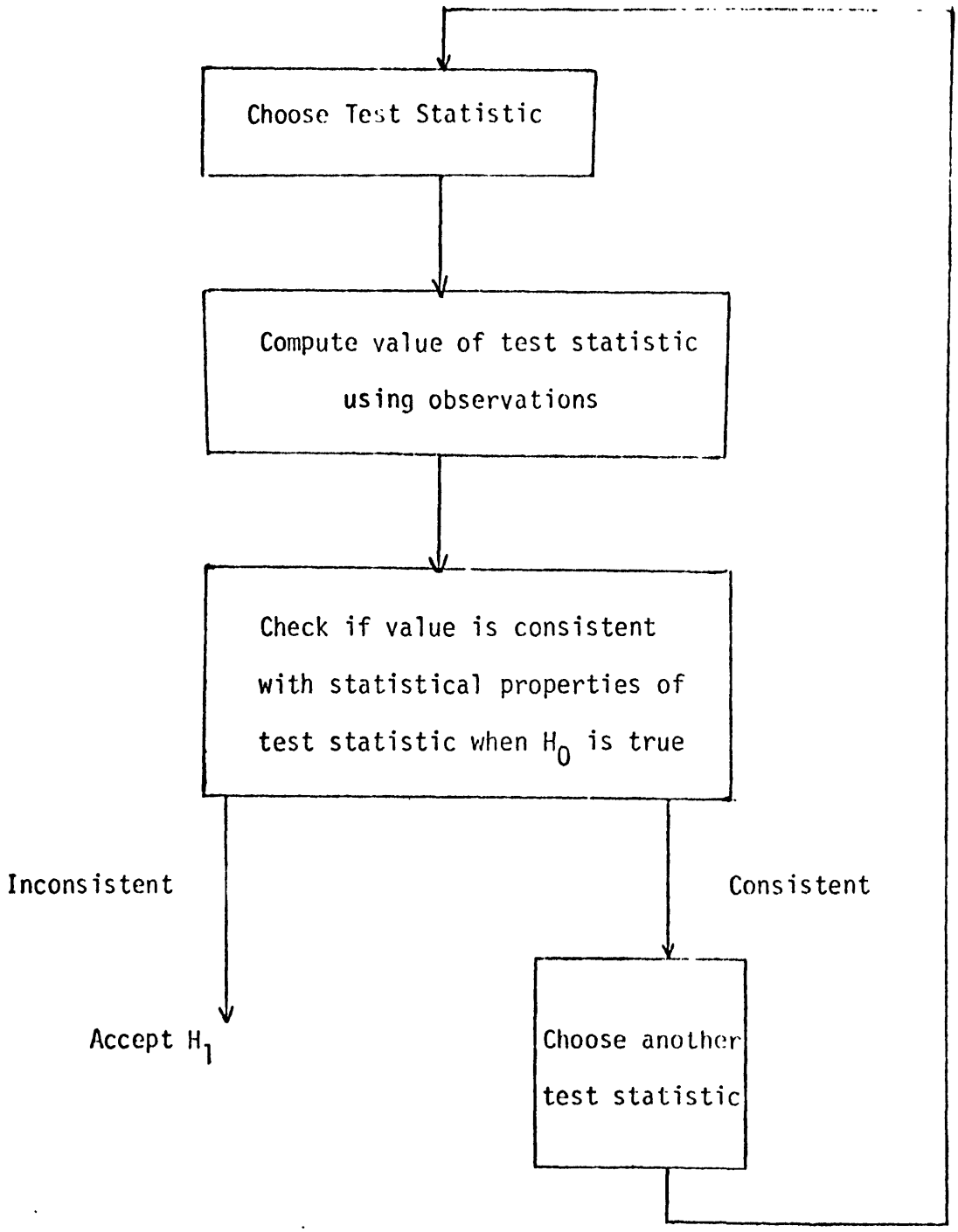
In general $\hat{\underline{z}}(n/n-1)$ is the output of a state estimator which first computes an $\hat{\underline{x}}(n/n-1)$ and then solves for $\hat{\underline{z}}(n/n-1)$ (see Section 9). Define

$$\underline{\delta}_z(n) = \underline{z}(n) - \hat{\underline{z}}(n/n-1) \qquad (10.2.1)$$

$$\underline{\Sigma}_z = E\{\underline{\delta}_z(n)\ \underline{\delta}^T{}_z(n)\}$$

"Individual Residual Magnitude" test statistics are defined by the vector

$$\underline{r}_z(n) = \sqrt{\underline{\Sigma}_z^{-1}}\ \underline{\delta}_z(n) \qquad (10.2.2)$$



Principle of Statistical Hypothesis Testing
Figure 10.2.1

When H_0 is true

$$E\{\underline{r}_z(n)\} = \underline{0}$$

$$E\{\underline{r}_z(n)\ \underline{r}_z^T(n)\} \simeq \underline{I}$$

and the $\underline{r}_z(n)$ have a normal (Gaussian) probability distribution.

One way to test the distribution is to plot histograms of the components of $\underline{r}_z(n)$. If the "main shape" is Gaussian but there are "outliers", the measurement causing the outliers should be studied individually to see if they are bad data, etc. (As discussed in Section 7.6, it was expected that outliers would be caused by the Keystone plume but this rarely if ever happened.) The study and removal of outliers is actually a form of robust estimation theory. Formal hypothesis tests on the normality of a time series can also be applied.

A "Sum of Squared Residuals" test statistic is defined by

the scaler

$$\xi = \sum_{n=1}^{N} \underline{\delta}_z^T\ \underline{\Sigma}_z^{-1}\ \underline{\delta}_z(n) = \sum_{n=1}^{N} \underline{r}_z^T(n)\ \underline{r}_z(n) \qquad (10.2.3)$$

When H_0 is true

$$E\{\zeta\} \simeq N$$

$$E\{\zeta - N^2\} \simeq 2N$$

and for large N, $\zeta(N)$ has a normal (Gaussian) probability distribution.

"Residual Whiteness" test statistics are defined by the matrices

$$\underline{P}_{rr}(\tau) = \frac{1}{N} \sum_{n=1}^{N} \underline{r}_z(n)\ \underline{r}_z^T(n-\tau) \quad \tau = 0,1,\ldots \qquad (10.2.4)$$

When H_0 is true

$$E\{\underline{P}_{rr}(\tau)\} \begin{cases} \simeq \underline{I} & \tau \\ = \underline{0} & \tau > 0 \end{cases}$$

and simple formula for the variances of $\underline{P}_{rr}(\tau)$ are available.

For sufficiently large N, the elements of $\underline{P}_{rr}(\tau)$ can be considered to be

normally distributed.

"Residual-Input Correlation" test statistics and "Residual-Output Correlation" test statistics are defined by the matrices

$$\underline{P}_{ru}(\tau) = \frac{1}{N} \sum_{n=1}^{N} \underline{r}_z(n)\, \underline{u}^T(n-\tau) \qquad \tau = 0,1,\ldots \tag{10.2.5}$$

$$\underline{P}_{rz}(\tau) = \frac{1}{N} \sum_{n=1}^{N} \underline{r}_z(n)\, \underline{z}^T(n-\tau) \qquad \tau = 1,2\ldots \tag{10.2.6}$$

respectively. When H_0 is true

$$E\{\underline{P}_{ru}(\tau)\} = \underline{0}$$

$$E\{\underline{P}_{rz}(\tau)\} = \underline{0}$$

and formula for their variances can be worked out.

An "Information Matrix" test statistic is as follows. Define the estimated information matrix to be $\hat{\underline{I}}_\alpha$ while the actual information matrix is $\underline{I}_\alpha$. When H_0 is true

$$E\{\hat{\underline{I}}_\alpha\} = \underline{I}_\alpha$$

and formula for the variance of the elements of $\hat{\underline{I}}_\alpha$ can probably be developed. There are at least three different ways of computing an $\underline{I}_\alpha$. If H_0 is not true, they can yield three different sets of numbers. Thus all three can be used as separate test statistics.

10.3 Accuracy of Parameter Estimates

Assume the model yields test statistics as in Section 10.2 which are consistent with H_0. Unfortunately this implies little or nothing about the accuracy of the parameters, i.e., about $\hat{\underline{\alpha}} - \underline{\alpha}$. Another class of tests directed toward evaluation of the accuracy of parameter estimates is now discussed.

The key concept here is the so-called "Cramer-Rao" or "Information" inequality which, assuming the hypothesized structure is exact, provides a lower bound on the estimated parameter error covariance matrix for any unbiased estimate, i.e.,

For any unbiased $\hat{\underline{\alpha}}$

$$\underline{\Sigma}_{\alpha} \geq \underline{I}_{\alpha}^{-1}$$

$$\underline{\Sigma}_{\alpha} = E\{(\hat{\underline{\alpha}} - \underline{\alpha}_{true})\ (\hat{\underline{\alpha}} - \underline{\alpha}_{true})^{T}\} \qquad (10.3.1)$$

$\underline{I}_{\alpha}$: information matrix

If $\underline{\Sigma}_{\alpha}$ has some very large main diagonal elements, this inequality states that the corresponding parameter estimates are very inaccurate; i.e., the parameters are effectively "not identifiable." If $\underline{I}_{\alpha}$ is not of full rank, then some of the parameters are completely not identifiable. Thus even if the hypothesized structure is valid, the parameter estimates are very inaccurate. Another way to say this is that the model may be statistically valid but "not valid" for many uses because of inaccuracies in the parameter estimates.

A basic problem in using $\underline{\Sigma}_{\alpha}$ is the question of deciding how large an error variance is important. An error variance that is 10% of the estimate might be critical for some parameters while 100% error variance for a different parameter might be of little concern. One way to evaluate such factors is by simulation to determine how the behavior of the physical system varies when the parameter values are varied around the extremes of the uncertainty ellipsoid defined by $\underline{I}_{\alpha}$. This is a reasonable test provided that only the input-output behavior of the model is of concern. In some cases a singular $\underline{I}_{\alpha}$ can be quite acceptable. To illustrate a different approach, assume one of the unknown parameters, say, α_k, denotes the emission rate of a background point source that exists but may or may not be in actual operation. Assume $\hat{\alpha}_k = 1$. Then if the square root of the corresponding main diagonal element of $\underline{\Sigma}_{\alpha}$ is 10, one can conclude that the source may not actually be operating; i.e., the true value of α_k may be zero.

10.4 Engineering Judgment

Unfortunately it is possible to have situations where a model will pass any possible statistical hypothesis test (as in Section 10.2) and parameter accuracy test (as in Section 10.3) even though the model is wrong (either in its hypothesized structure or its estimated parameter values). Such is life.

A third class of validity test is the use of "engineering judgment." This is not as conceptually elegant as the first two but it is equally (if not more) effective.

Two ways to use engineering judgment are to ask

- Are the estimated parameter values reasonable from an engineering point of view?
- Does the model behave (under simulation) in a way that is consistent with engineering judgment?

A third way to use engineering judgment is to ask the question

- Could there exist a "phenomenon" which the statistical tests of Section 10.2 cannot detect?

If such a phenomenon could be present, nothing can be done (except to "change the experiment" by making more observations.) However one at least learns that there is a "danger" that the phenomenon actually exists.

10.5 Other Validity Tests

The preceding discussion has not explicitly covered many model validation tests that have been applied to air quality models in other studies. Some of these tests are (see Moses, 1969).

- Scatter diagrams, correlation measures
- Contingency tables
- Skill scores
- Effectiveness of model as functions of particular variables
- Spatial temporal variation (Isopleth charts)

Many of these, such as Isopleths, are vehicles for the application of engineering judgment (as in Section 10.4). Others such as "skill scores" are really associated with evaluating the accuracy of the model rather than its validity (as defined in Section 10.1).

The statistical hypothesis testing and parameter accuracy techniques discussed in Sections 10.2 and 10.3 are not intended to replace the use of these other proven tests. They merely provide another set of tools that have the advantage of providing a self-consistent approach that is directly applicable to probablistic air quality models. As discussed in Section 10.2 a complete set of all possible statistical hypothesis tests was not given.

10.6 Relation to Tests Actually Used

The model validation tests actually used and discussed in Section 6 involved

Autocorrelation function:

$$P_{\delta\delta}(\tau) \qquad \tau = 0,1\ldots$$

Histogram of $\underline{\delta}_z(n)$

These tests are the same as those of (10.2.2) and (10.2.4) ecept that $\underline{\delta}_z(n)$ is not normalized to get $\underline{r}_z(n)$.

. Another type of test used (not discussed in Section 6) involved partial checks on the nonsingularity of the information matrix. It was well behaved.

APPENDIX E

AIR QUALITY MODEL DEVELOPMENT FOR ROUGH TERRAIN

Table of Contents

1. Introduction

The deterministic portion of the operating air quality model used in the demonstration control strategy is described in Section 2.4 of the main text. Section 5 of the main text discusses some of the shortcomings of this model and briefly summarizes subsequent research on point source modeling in rough terrain. This appendix contains the details of this research.

The basic appraoch is to prepare a sequence of explicit model modifications and then to compare their performance on the available measurements. Section 2 describes the objective scheme used to evaluate each model modification. Nine model modifications are discussed in Section 3. A "skill scoring" scheme and its results are discussed in Section 4. A final discussion is given in Section 5. References are found in the above general reference section.

2. Model Evaluations

An objective scheme was used to evaluate each model modification. The scheme involved assigning each case-hour for a given modification to one of five categories of success or failure by means of the ratios of calculated to observed concentrations. The five categories are gross underprediction (<<), moderate underprediction (<), success (✓), moderate overprediction (>), and gross overprediction (>>). Because the operating model is used to predict concentrations averaged over three hours and these are compared to one-hour averaged measurements, and because it is never certain that the measured concentration is the maximum that occurred at the dominant arc on which the monitor resides, some latitude must be allowed in classifying a modeled case-hour as a success. The criteria used are defined in Figure 2.1.

Section 5 of the main text discussed how the measurements used for the model evaluation were chosen. Measurements from Monitors 1 (West Fairfield), 3 (Laurel Ridge) and F (Ligget) were used.

3. Nine Model Modifications

The basic model discussed in Sections 2.4 of the main text was changed in a sequence of nine modifications. The changes from one modification to the next sometimes related to more than one problem area so that the individual effects of each change are not always susceptible to separate evaluation. There were six general areas of change that were made. The characteristics of each modification with respect to these six areas are listed in Table 3.1.

A feature common to all modifications is that the stability class for each one-hour case was determined by the one-hour averaged temperature difference measured between 45.7m, and 12.2 m on the Penn View tower. This ΔT was used rather than the ΔT through the thicker layers to 91.1 m for two reasons: first, the data capture rate for the lower ΔT was higher, and second, the measurement of lapse rate nearer the ground should be more closely associated with turbulence levels aloft because it is more representative of the heat flux that drives convective turbulence. The lapse rate through the thicker layer tends toward the adiabatic in response to the vertical motions associated with convective overturning in unstable conditions. In stable conditions, the strongest temperature gradients are also closer to the ground. The preference for the use of the lower ΔT as a stability determinant was supported by F. Pasquill (personal communication). The temperature gradients associated with each stability class were common to all nine modifications; they are listed in Table 3.2.

Table 3.2

Stability Classification* by Temperature Difference Measured between 45.7 m and 12.2 m at Meteorological Tower

Stability Class	ΔT (NS8) °F
1	<-1.2
2	-1.2
3	-1.1 to -1.0
4	-0.9 to -0.4
5	≥ -0.3

* Corresponding to Nuclear Regulatory Guide 23.

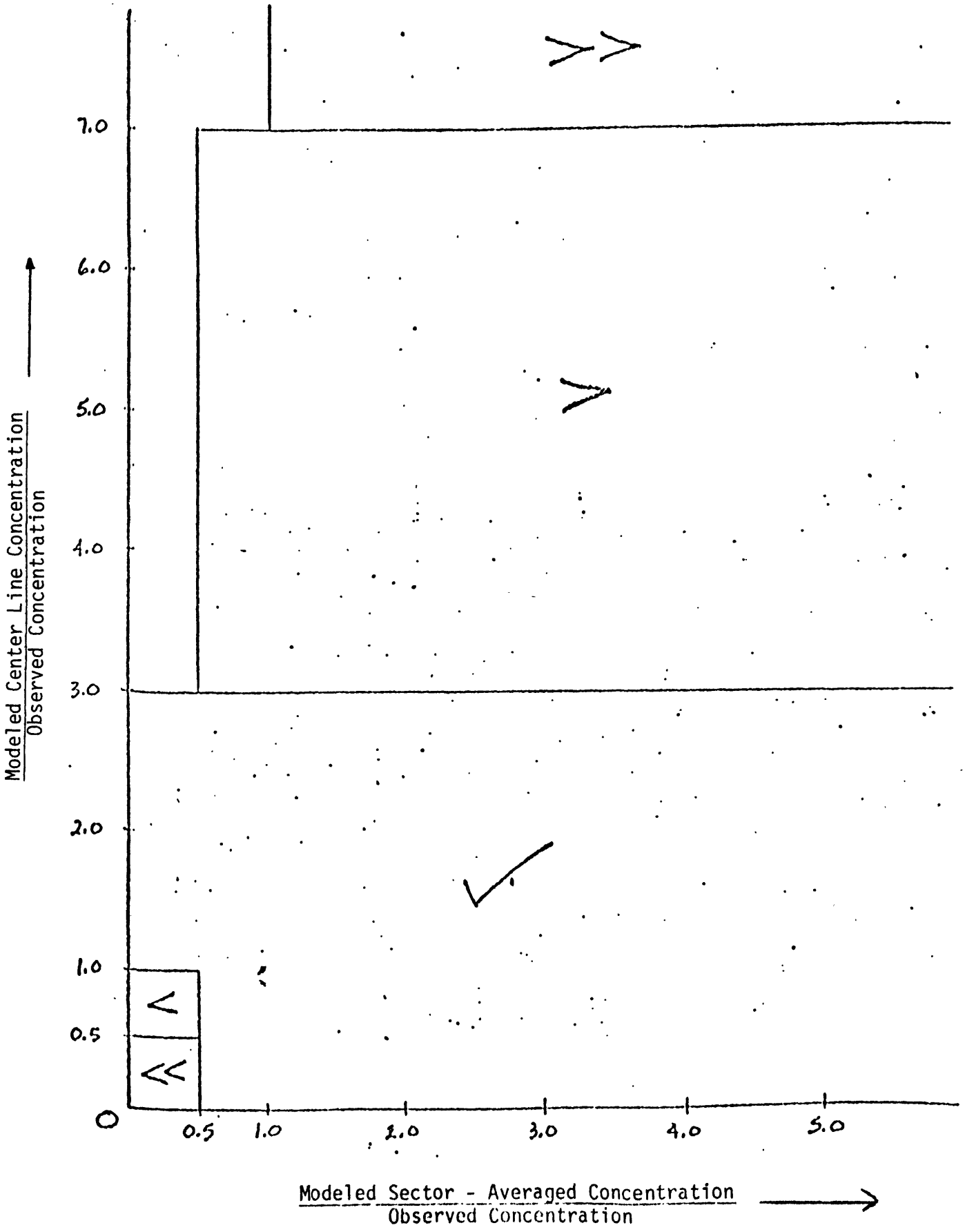

7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.5
0
0.5
1.0
2.0
3.0
4.0
5.0
Modeled Center Line Concentration / Observed Concentration
Modeled Sector - Averaged Concentration / Observed Concentration

Figure 2.1

Table 3.1

Characteristics of Model Modifications

Model Modification Number	Effects of Plume Buoyancy	Constraint on z_o in Stable Hours	Terrain Correction Factors	"Punch-Through" Limit	Additional Background Included	Change of σ's for Rough Terrain
1	None	$\leq$ 150m.	.8,.8,.7,.5,.25	50 m.	0	None
2	$\sigma^2 = \sigma_c^2 + \Delta H^2/10$ (See Pasquill, 1974)	Same	Same	Same	Same	Same
3	ASME curves fit to $(\Delta H/\sqrt{10}, x_c)$ (See text)	No limit	Same	Same	10 ppb.	Same
4	Same	Same	.8,.8,.8,.7,.5	Same	Same	ASME σ's made one class less stable
5	Same	Same	.8,.8,.7,.6,.5	Same	Same	Same
6	Same	Same	.8,.7,.6,.5,.4	Same	Same	Same
7	Same	Same	Same, but modified to be function of approach height	200 m.	Same	None
8	Same	Same	Same	Same	Same	ASME σ's made one class less stable
9	Same	Same	Same	Same	Same	ASME σ's made approximately one class less stable

All modifications used the same power law wind profiles (EPA 4, 1970) reported to have been derived from measurements at the Brookhaven Tower. These profile exponents are 0.10, 0.15, 0.20, 0.25, and 0.30 for Pasquill-turner stability classes A through E, and were associated with the operating model's stability classes 1 through 5. The original implementation of the operating model used wind profiles derived from measurements taken on a 30-m tower in a desert environment, which are probably ill soited to the Chestnut Ridge area. The differences in modeled concentrations resulting from the two sets of profiles were, however, not of great significance. The wind speeds used in the modeling were those measured at the 91.1 m level of the meteorological tower. Any hour in which the average wind wpeed exceeded 25 mph was classified as neutral stability. The wind speed at stack height, which was the actual value used in the model calculations, was estimated by the formula:

$$u_s = u_r \left\{ \frac{H_s}{150} \right\}^p$$

in which u_r is the wind speed at 91.1 m (m/s), H_s is the height of the stack (m), and p is the wind profile exponent appropriate to the stability class of the case.

The assumption was made that the wind speed at tower top, being 91.1 m above the top of Chestnut Ridge, corresponded to the speed that would be measured at a height of 150 meters above a typical plant site. This assumption is difficult to test with the data available. The only wind data measured near the plant sites were taken by a mobile van during the measurements of downwash concentrations at Seward in the Demonstration Period. These observations do not provide data for verification of the extrapolation of the tower-top speed to the top of the stacks, but only for verification of the extrapolation down into the surface layer in the Conemaugh Valley. On one of these two occasions, the wind speeds measured by the van at approximately 15 feet above the ground were estimated within about 10% by the formula. On the other occasion, an instrument or recorder malfunction limited the motion of the recorder pen to less than 18 mph, and the wind speed trace was therefore not averageable.

The program codes used for the model improvement runs reads in one card of data for each case hour, disregards the case if the tower-top wind speed is missing, establishes the stability class, changes the wind speed to meters per second, assigns a default mixing depth if it is missing (1600, 1400, 1200, 1000, and 3000 meters for stabilities 1 through 5, respectively), and calculates both the sector-averaged (3-hour) concentration and the centerline (1-hour) concentration on the assumption that the power plant plume is right over the monitor. The mixing depth was set equal to 3000 m for all stable hours no matter what the "measured" value. The modeled background concentration, which is the sector-averaged value for both calculated impacts, is separately listed though added to both calculated values. Ratios of the sector-averaged and centerline concentrations are also listed.

An objective scheme was needed to evaluate each model modification. The scheme finally used involves assigning each case-hour for a given modification to one of the five categories of success or failure by means of the ratio of calculated to observed concentrations. The five categories are gross underprediction (<<), moderate underprediction (<), success (✓), moderate overprediction (>), and gross overprediction (>>). Because the operating model is used to predict concentrations averaged over three hours and these are compared to one-hour averaged measurements, and becuase it is never certain that the measured concentration is the maximum that occurred at the downwind arc on which the monitor resides, some latitude must be allowed in classifying a modeled case-hour as a success. The criteria used are illustrated in Figure 2-1. The results of each modification run were categorized by stability (unstable, neutral, or stable) and wind speed at tower top ($u \leq 10$ mph, $11 \leq u \leq 21$ mph, $u > 21$ mph) in order to simplify the analysis of the model's faults and to provide guidance for the changes to be incorporated in the next modification.

Modification 1:

The first modification was basically the operating model as implemented with three exceptions. First, the value of σ_2 for stable hours was not constrained to its value at 15 km as specified in the ASME Guide but was restricted not to exceed 150 m, the value given by the ASME curve at approximately 61 kilometers. This was done to reduce the large background concentrations modeled in stable hours that were often much in excess of the values measured. Second, the mixing depth for stable hours was set to 3000 meters in this first modification, as it was in all those which followed. The third distinction from the original model is the use of the CDM wind profiles, a feature common to all subsequent modifications.

Modification 1 underpredicted grossly at Laurel Ridge and Liggett in

light wind, neutral vases and in moderate wind, stable cases. The twenty cases as West Fairfield represent too small a sample to give much weight to the results, which were scattered for this modification as for most others; however, there was a tendency to underpredict in unstable conditions and generally to overpredict somewhat in neutral and stable conditions.

Model Modification 2 added to modification 1 an enhanced plume growth resulting from the entrainment of air caused by the ascent of the plumes under the influence of their buoyancy. This modification, following Pasquill's suggestion literally, added one-tenth of the square of the plume rise to the square of the values given by the ASME σ_y and σ_z curves to give the modified variances; that is,

$$\sigma_\alpha^2 = \frac{\Delta H^2}{10} + \sigma_{\alpha_o}^2 \quad ,$$

in which H is the plume rise given by the appropriate Briggs formula; σ_{α_o} is the value of the plume dimension given by the appropriate ASME curve, σ = y or z; and σ_α is the modified plume dimension, σ = y or z. This modification substantially improved the problem of underprediction at Laurel Ridge and Liggett in neutral conditions, did not affect the stable cases at Liggett, but radically changed the stable cases at Laurel Ridge from underprediction to generally gross overprediction. It made moderate improvements in the unstable and neutral cases at West Fairfield and led to uniformly gross overprediction in the stable cases at that station.

Model Modification 3 removed any constraint on the growth of σ_z in stable conditions because large excesses of modeled background continued; added 10 ppb back into the background in general because the modeled background in neutral and unstable conditions was approximately this much below that observed at upwind monitors, and chnaged the modeling approach to buoyant plume growth. The scheme in Modification 2 simply adds all of the growth resulting from entrainment during the rising phase to the growth of the plume by atmospheric dispersion, which is represented by the usual curves for the dispersion coefficients. These curves for σ_y and σ_z presumably reflect the fact that the rate at which a plume grows depends on the size of the plume as well as the stability. The simple approach of modification 2 to consideration of the enhanced growth of buoyant plumes may not represent this fact most appropriately. Consequently, modification 3 fit the σ curves to bhe "observed" plume standard deviations at the point of final rise, where Pasquill suggests σ_y and σ_z are both equal to $\Delta H/\sqrt{10}$. The appropriate ASME σ_y and σ_z curves were fit at this point by the method of "virtual distances." For Laurel Ridge Modification 3 moved ten of the moderate underpredictions in neutral conditions into the success category, made slight improvements in the three unstable cases, and led to further gross overprediction in the stable cases. The improvement at Liggett in neutral cases was also marked, and gave moderate underpredictions in stable cases. For West Fairfield, this model was excellent in the unstable cases, none of which remained underpredicted; showed a slight improvement in the neutral cases, and did nothing for the gross overprediction in the stable cases.

Model Modification 4 augmented the size of the plumes by modeling them with the ASME σ curves associated with the stability one class less stable than the class derived from the ΔT at the tower. This approach has been reportedly successful in other studies of plume behavior in rough terrain. At the same time, this modification kept the plumes somewhat more removed from elevated locations by using larger terrain correction factors, namely 0.8, 0.8, 0.8, 0.7, and 0.5 for stability classes 1 through 5, respectively. At Laurel Ridge and Liggett, these changes improved the model's performance for neutral cases with light and moderate wind speeds, but they led to underprediction of the high wind speed neutral cases. The three unstable cases at Laurel Ridge and Liggett were somewhat better simulated, the gross overprediction at the elevated monitor and the gross underprediction at Liggett being significantly improved. In fact, this was the first modification that provided any impact from Homer City at Liggett in stable hours. At West Fairfield, this model was successful in all five neutral cases and in all but one of the eight unstable cases; furthermore it moderated the overprediction of the seven stable cases.

Model Modification 5 differed from modification 4 only in that the terrain correction factors for stability classes 3 and 4, were reduced by 0.1. Modification 5 did not make any important changes in model performance, although the few realignments of success categories were all for the better.

Model Modification 6 was again identical to modification 4 except for the terrain correction factors, which were now set to 0.8, 0.7, 0.6, 0.5, and 0.4 for stabilities 1 through 5, respectively. Minor improvements in model performance occurred in neutral conditions at Laurel Ridge and Liggett, while the prediction of stable cases deteriorated somewhat. At West Fairfield, this modification, like the last, had perfect success according to the grading system in use for the unstable and neutral cases, but it still overpredicted the stable cases.

Model Modifications 7, 8, and 9 incorporated a variable lifting of plumes (and mixing heights) over terrain features that depends on the ratio of the height at which the plume approaches the terrain feature to the height of the feature itself. A trial functional relationship was derived on the basis of the constraints that a plume at great elevation ($H_p/z \gg 1$) should not respond to changes in terrain, and that a plume approaching at the height of the terrain feature ($H_p/z = 1$) should be lifted an amount equal to the product of the terrain correction factor and the height of the obstacle. The relationship derived is:

$$\zeta/z = \frac{T_f}{1 + T_f(H_p/z - 1)} ,$$

in which

ζ = the amount plume is lifted above its approach height as it passes over the crest of the obstacle;

z = the height of the obstacle above the stack-base height; $(t_r - t_s)$

t_f = the terrain correction factor, and

H_p = the effective height of the plume in the absence of terrain effects, i.e., $H_p = H_s + \Delta H$, where H_s is the height of the stack and ΔH is the plume rise.

Table 3.3a West Fairfield (1)

		UNSTABLE					NEUTRAL					STABLE				
MODEL	WS	<<	<	✓	>	>>	<<	<	✓	>	>>	<<	<	✓	>	>>
	LOW	3	1	3	1		1		1	1		1			3	
	MED								1	1					3	
	HIGH															
	TOTAL	3	1	5	1		1		2	2		1			6	
(2)	LOW	1	3	3	1			1	1	1						4
	MED								1	1						3
	HIGH															
	TOTAL	1	3	3	1			1	2	2						7
(3)	LOW			7	1				2	1						4
	MED								1	1						3
	HIGH															
	TOTAL			7	1				3	2						7
(4)	LOW	1		7					3						1	3
	MED								2						2	1
	HIGH															
	TOTAL	1		7					5						3	4
(5)	LOW			5					3						1	3
	MED								2						2	1
	HIGH															
	TOTAL			8					5						3	4
(6)	LOW			8					3						1	3
	MED								2						1	2
	HIGH															
	TOTAL			8					5						2	5
(7)	LOW			6	1	1			1	1	1				2	2
	MED								1	1					3	
	HIGH															
	TOTAL			6	1	1			2	2	1				5	2
(8)	LOW			6					1	1	1				1	3
	MED								2						1	2
	HIGH															
	TOTAL			6					3	1	1				2	5
(9)	LOW	1		6	1				1	1	1					4
	MED									2						3
	HIGH															
	TOTAL															

Table 3.3b Laurel Ridge (3)

DELTA WS	UNSTABLE <<	UNSTABLE <	UNSTABLE ✓	UNSTABLE >	UNSTABLE >>	NEUTRAL <<	NEUTRAL <	NEUTRAL ✓	NEUTRAL >	NEUTRAL >>	STABLE <<	STABLE <	STABLE ✓	STABLE >	STABLE >>
(1) LOW						19	2	1					5		
MED		3				2	9	13	2		3	1	1		
HIGH								9	3		1				
TOTAL		3				21	11	23	5		5	1	9		
(2) LOW						6	12	4						1	8
MED		3					8	16	2				1	2	2
HIGH								9	3					1	
TOTAL		3				6	20	29	5				1	4	10
(3) LOW						6	7	9							9
MED		2	1				3	21	2				1	2	2
HIGH								9	3						1
TOTAL		2	1			6	10	39	5				1	2	12
(4) LOW						4	2	16						4	5
MED		3				1	6	19					2	2	1
HIGH						2	6	4						1	
TOTAL		3				7	14	39					2	7	6
(5) LOW						3	3	16						4	5
MED	3					1	6	19					2	2	1
HIGH						2	5	5						1	
TOTAL	3					6	14	40					2	7	6
(6) LOW						2	4	16						3	6
MED	3					1	4	21					1	3	1
HIGH						2	5	5						1	
TOTAL	3					5	13	42					1	7	7
(7) LOW						1	2	14	3	2	6	3			
MED			3				4	17	5		5				
HIGH								6	6		1				
TOTAL			3			1	6	37	14	2	12	3			
(8) LOW							2	19		1					
MED	3					1	5	20					3	6	3
HIGH						2	5	5						2	
TOTAL	3					3	12	44		1			3	9	3
(9) LOW							2	16	3	1					
MED		1	2				1	23	2				4	5	
HIGH							1	11					3	2	
TOTAL													1		

Table 3.3c Liggett (F)

		UNSTABLE						NEUTRAL						STABLE					
MODEL	WS	<<	<	✓	>	>>		<<	<	✓	>	>>		<<	<	✓	>	>>	
(1)	LOW							26	9	3				2	5				
	MED							3	31	24				12		1			
	HIGH								3	25				4					
	TOTAL							29	42	52				20	5	1			
(2)	LOW							15	19	3				2	5				
	MED							1	24	33				14		1			
	HIGH								2	26				4					
	TOTAL							16	45	62				20	5	1			
(3)	LOW							6	26	5				2		5			
	MED							1	16	41				14		1			
	HIGH								1	27				4					
	TOTAL							7	43	73				20		6			
(4)	LOW							2	9	26						5.1	2.1		
	MED							4	22	32				3		7	5		
	HIGH							13	8	7				1		2			
	TOTAL							19	39	65				4		15	7		
(5)	LOW							2	10	25						4	3		
	MED							4	21	33				3		7	5		
	HIGH							14	6	5				1		3			
	TOTAL							20	37	66				4		14	5		
(6)	LOW							2	8	27						3	2		
	MED							4	21	33				3		6	6		
	HIGH							13	7	8				1		3			
	TOTAL							19	36	68				4		12	10		
(7)	LOW								18	16	3			2	1	4			
	MED							1	5	51	1			14	1				
	HIGH								1	26	1			4					
	TOTAL							1	24	93	5			20	2	—			
(8)	LOW							1	4	32						2	5		
	MED							4	21	32	1			3		6	6		
	HIGH							13	7	8				1		3			
	TOTAL							18	32	72	1			4		11	11		
(9)	LOW								7	28	2					5			
	MED							1	9	48				2	3	3			
	HIGH								11	17				1	2	1			
	TOTAL							1	27	93									

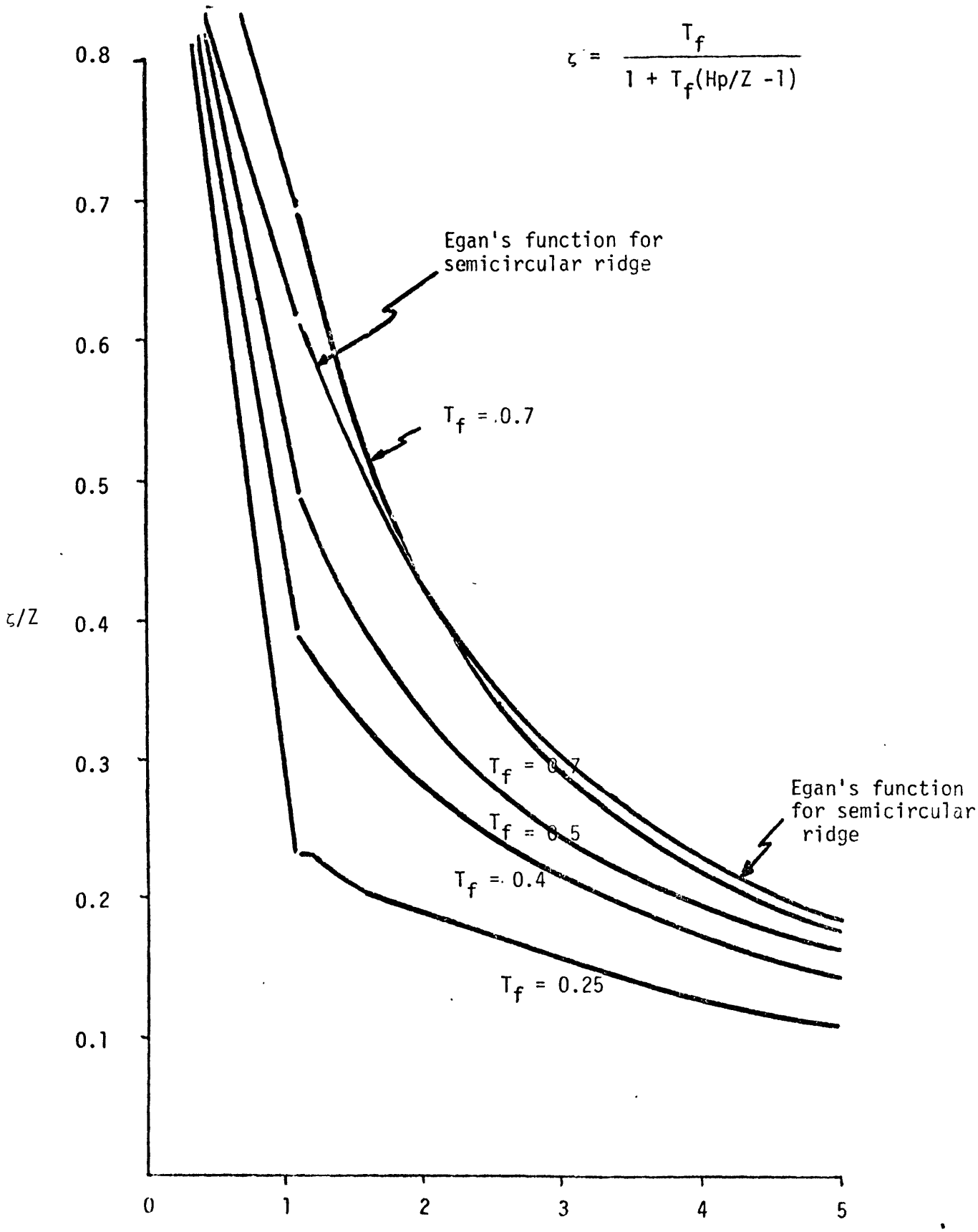

$\zeta = \frac{T_f}{1 + T_f(Hp/Z - 1)}$
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
ζ/Z
Egan's function for semicircular ridge
T_f = 0.7
T_f = 0.5
T_f = 0.4
T_f = 0.25
Egan's function for semicircular ridge
0
1
2
3
4
5
Hp/Z

Figure 3.1

For $H_p < z$, the model code retained the condition $H = T_f H_p$. $1 - (1-T_f)H_p$ This function is plotted as ζ/z vs H_p/z for three values of T_f in Figure 3.1 along with Egan's (Egan, 1, 1975) equivalent function β defined for potential flow solution is quite similar to the relationship used in the model when the terrain correction factor is 0.7 until the approach height H_p is less than a factor of two times the obstacle height. At that point the present model keepts the plume centerline somewhat more removed from the crest of the obstacle. With the terrain correction factor equal to 0.5, the present model allows the plume to approach more closely to the crest, but not as closely as Egan's solution for the flow over a hemisphere (hill). Consequently, for neutral conditions, if the terrain correction factor is set to 0.5, the model is a compromise between the "ridge" and "hill" potential flow solutions. It should be noted, however, that the model retains the wind speed estimated at stack height for the ventilation factor in the Gaussian plume formula. The wind speed in the plume will, in fact, increase in the passage over the obstacle as a result of the squeezing together of the streamlines, but the demands of continuity are such that this effect will be countered in large part by the apparent decrease in the vertical dimension of the plume that results from the same squeezing of the flow. Consequently, there should be little effect on concentrations. As a first approximation to the flow over terrain features with some provision for incorporating the theoretically anticipated effects of thermal stratification, this variable terrain correction factor was implemented in modifications 7, 8 and 9. All three modifications used the same set of stability-dependent terrain correction factors, T_f, namely, 0.8, 0.7, 0.6, 0.5, and 0.4 for stabilities 1 through 5, respectively. Another feature common to modifications 7, 8, and 9 was the change of the "punch-through" to 200 meters to try to overcome the persistent problem of underprediction in cases with restricting mixing depths.

In Model Modification 7, the dispersion coefficients were returned to the standard ASME curves originally associated with each stability class in the first three modifications, that is, no increased dispersion was allowed to account for the rough terrain. This modification resulted in a remarkable improvement in model accuracy at Liggett in neutral conditions, although it did somewhat less well in the neutral cases at Laurel Ridge. The effect of the variable terrain correction factor can be examined for neutral cases by comparison with the results of modification 3. There is a clear improvement at Liggett and more mixed results at Laurel Ridge. The five neutral cases at West Fairfield are overpredicted somewhat more by modification 7 than by modification 3. In the stable cases, there is serious underprediction at Liggett and Laurel Ridge and overprediction at West Fairfield. The three unstable hours at Laurel Ridge were successfully modeled, and two of the eight unstable hours at West Fairfield were overpredicted, one grossly.

Model Modification 8 once more incorporated enhanced diffusion rates as in modifications 4 through 6. In all stability categories, these results may be compared to modification 6 for assessment of the influence of the variable terrain correction. Moderate improvement was achieved over both modifications 6 and 7 for the neutral and stable cases at Laurel Ridge, although all three unstable cases at that site were now grossly underestimated with the enhanced dispersion rates, as they had been with modification 6. At Liggett, the neutral cases tended to be underpredicted more often than with modification 7, and the slight improvement over modification 6 occurred in the light wind speed neutral cases when the plumes were both large and high.

The results in stable cases at this monitor were much superior to those for the previous modification, which lacked the augmented dispersion, but they were not significantly different from modification 6, which had the augmented dispersion parameters but not the variable terrain correction. At West Fairfield, the unstable and stable cases were modeled with success identical to that of modification 6, but the neutral cases were only slightly improved over modification 7 and were worse than the perfect score of modification 6.

Model Modification 9 was a compromise between modifications 7 and 8 in that the dispersion coefficients associated with each stability class were set to lie approximately halfway between the standard ASME curves and those for the class next less stable. This modification achieved success with 50 of the 60 neutral cases at Laurel Ridge with four moderate underpredictions, five moderate overpredictions, and one gross overprediction in a very light wind-speed case. Two of the three unstable cases and eight of 15 stable cases at Laurel Ridge were successfully modeled, with no gross errors. At Liggett, modification 9 modeled 93 of 123 neutral cases successfully with a tendency to underestimate the high wind-speed cases. Only modification 7 was as successful in neutral conditions at the monitor. Fourteen of the 19 stable cases were also modeled properly.

It is notable that all eight of the gross underpredictions at Liggett with modification 9 occurred for case-hours in which neither Homer City unit was operating. The "background" concentrations measured in these eight hours were all at least 90 ppb, with the maximum of 130 ppb being measured during a stable hour with an eight mph wind speed at the top of the tower. The concentrations measured at nearby stations during these hours did not approach the values measured at Liggett, a fact which suggests that a source exists somewhat upwind (west) of Liggett that is not accounted for in the model.

Furthermore, three of the 27 cases of moderate underprediction among the 123 neutral Liggett cases were not supported by ΔT measurements, and the neutral stability category was assigned because of the forecasters' recorded affirmation that the Pasquill stability classification was neutral. One of these cases was, however, one in which the wind at the tower top was 42 mph, so it would have been assigned to the neutral class regardless of the ΔT measurement.

4. Skill Scheme Comparisons

In order to allow determination of which modifications showed the best "skill" as operating models in the SCS application, two scoring schemes were devised and used. They assign scores to the modeled result of each case-hour according to which of the previously defined categories of success or failure that result fell into. "Skill Scheme I" simply assigns a 1 to each "success" (✓), a 0 to each moderate over-or underprediction, and a -1 to each gross over- or underprediction. "Skill Scheme II" differs only in that moderate overpredictions are assigned an 0.25 on the premise that a little conservation is tolerable and perhaps even a virtue.

The scores achieved by the various modifications are presented at random in Tables 4.1a, b, and c by monitoring station, stability classification, and skill scheme. The ranking of the modifications by the two scoring schemes is nearly identical, the only differences being the position of modifications 7 and 8 at West Fairfield.

Table 4.1(a)

West Fairfield (1)

SKILL SCHEME 1

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
2	2	2	-7	-3
9	5	0	-7	-2
1	0	1	-1	0
3	7	3	-7	3
7	5	1	-2	4
8	8	2	-5	5
4	6	5	-4	7
6	8	5	-5	8
5	8	5	-4	9

SKILL SCHEME 2

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
2	2.25	2.5	-7.0	-2.25
9	6.25	0.75	-7.0	0
1	0.25	1.5	0.5	2.25
3	7.25	3.5	-7.0	3.75
8	8.0	2.25	-4.5	5.75
7	5.25	1.5	-0.75	6.0
4	6.0	5.0	-3.75	7.25
6	8.0	5.0	-4.5	8.5
5	8.0	5.0	-3.75	9.25

Table 4.1(b)

Laurel Ridge (3)

SKILL SCHEME 1

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
1	0	2	4	6
2	0	23	-9	14
3	1	33	-11	23
4	-3	32	-4	25
7	3	34	-12	25
5	-3	34	-4	27
6	-3	37	-6	28
8	-3	40	0	37
9	2	49	8	59

SKILL SCHEME 2

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
1	0	4.5	4	8.50
2	0	25.5	-8	17.50
3	1	35.5	-10.5	26.00
4	-3	32.0	- 2.25	26.75
7	3	37.5	-12.0	28.50
5	-3	34.0	- 2.25	28.75
6	-3	37.0	- 4.25	29.75
8	-3	40.0	2.25	39.25
9	2	50.25	9.75	62.00

Table 4.1(c)

Liggett (F)

SKILL SCHEME 1

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
1		23	-14	9
2		46	-14	32
3		66	-14	52
4		46	10	56
5		46	10	56
6		49	8	57
8		54	7	61
7		92	-15	77
9		92	7	99

SKILL SCHEME 2

MODEL	UNSTABLE	NEUTRAL	STABLE	Σ
1		23	-14	9
2		46	-14	32
3		66	-14	52
4		46	12	58
5		46	12	58
6		49	10.5	59.5
8		54.25	9.75	64
7		93.25	-15	78.25
9		92.5	7	99.5

At both Laurel Ridge and Liggett, where a significant number of case-hours were modeled, modification 9 achieved the highest skill rating by a wide margin, according to either scoring scheme. Furthermore, there is an obvious trend of improvement in skill through the course of modifications. This is distinctly not the case at the West Fairfield monitor, however, where the modifications with the plume dimensions shifted one stability class did best in unstable cases and those with the standard ASME σ curves did best in stable cases, and modification 5 scored best overall.

5. Discussion

The results listed in Tables 3.1 and 4.1 suggest that further manipulation of the model might yield even better performance. However, it is never certain that forcing a model to fit observations taken at a few stations at the ground ensures that the model actually represents the physical realities of the behavior of the plume. No data were taken in the present program that can be used directly to validate intermediate results of the model calculations, such as the plume rise, the actual flow over the terrain features, and the σ_y and σ_z values. A comparison of the results of modification 9 with observations of the Homer City plumes taken by aircraft traverses during the LAPPES Program (EPA,5, 1968-1971) indicates that, in stable conditions, at least, the model generally underestimates plume rise and σ_y and overestimates σ_z. Consequently, one would expect that modification 9 would probably overpredict the SO_2 concentrations at Liggett in stable conditions, yet it did not do so to any large extent in any of the 26 stable case-hours at Liggett. It is difficult to tell from the data at hand why this should be true.

Of particular importance to the success of an SCS is the accuracy of the operating model in predicting the very highest observed concentrations. Consequently, the few case-hours exceeding 200 ppb were separately evaluated. No such hours occurred at West Fairfield, three occurred at Laurel Ridge, and 12 at Liggett. The results of each model modification are shown in Tables 5.1a and b. At the elevated monitor on Laurel Ridge, modification 9 excelled; it is the only modification that achieved success with the lone stable case, and it was successful in the moderate wind-speed neutral case and overpredicted the low wind-speed neutral case moderately. At Liggett, modification 9 was also the most skillful, having success with the two stable case-hours and half of the 10 neutral case-hours. Only modification 7 performed better for the neutral cases, achieving success with one more moderate wind-speed case than modification 9, but grossly underpredicting both stable case-hours.

TABLE 5.1a Laurel Ridge (3) 7200 ppb

		UNSTABLE						NEUTRAL						STABLE					
MODEL	WS	<<	<	✓	>	>>		<<	<	✓	>	>>		<<	<	✓	>	>>	
①	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
②	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
③	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
④	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
⑤	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
⑥	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
⑦	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
⑧	LOW																		
	MED																		
	HIGH																		
	TOTAL																		
⑨	LOW																		
	MED																		
	HIGH																		
	TOTAL																		

Table 5.1b Liggett (F) 7200 ppb

		UNSTABLE						NEUTRAL						STABLE					
PASQUILL	WS	<<	<	✓	>	>>		<<	<	✓	>	>>		<<	<	✓	>	>>	
C	LOW							2						2					
	MED							1	7										
	HIGH																		
	TOTAL							3	7					2					
D	LOW								2					2					
	MED								7	1									
	HIGH																		
	TOTAL								9	1				2					
E	LOW								2					2					
	MED								7	1									
	HIGH																		
	TOTAL								9	1				2					
F	LOW								2								2		
	MED								7	1									
	HIGH																		
	TOTAL								9	1							2		
G	LOW								2								2		
	MED							1	7										
	HIGH																		
	TOTAL							1	9								2		
	LOW								2								2		
	MED							1	7										
	HIGH																		
	TOTAL							1	9								3		
	LOW								1/3	1/5				2					
	MED								1	1				2					
	HIGH																		
	TOTAL								2/7								2		
	LOW							1											
	MED								4										
	HIGH																		
	TOTAL								1/2	1/4							2		