United States
Environmental Protection
Agency

Environmental Monitoring
Systems Laboratory
P.O. Box 15027
Las Vegas NV 89114

EPA-600/4-79-069
October 1979

Research and Development

EPA

Optimum Meteorological and Air Pollution Sampling Network Selection in Cities:

Volume II - Evaluation of Wind Field Predictions for St. Louis

RESEARCH REPORTING SERIES

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

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

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

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

EPA-600/4-79-069
October 1979

OPTIMUM METEOROLOGICAL AND AIR POLLUTION SAMPLING NETWORK SELECTION IN CITIES

Volume II: Evaluation of Wind Field Predictions for St. Louis

Fred M. Vukovich and C. Andrew Clayton
Research Triangle Institute
P. O. Box 12194
Research Triangle Park,
North Carolina 27709

Contract No. 68-03-2187

Project Officer

James L. McElroy
Monitoring Systems Research and Development Division
Environmental Monitoring Systems Laboratory
Las Vegas, Nevada 89114

ENVIRONMENTAL MONITORING SYSTEMS LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U. S. ENVIRONMENTAL PROTECTION AGENCY
LAS VEGAS, NEVADA 89114

DISCLAIMER

FOREWORD

Protection of the environment requires effective regulatory actions that are based on sound technical and scientific data. This information must include the quantitative description and linking of pollutant sources, transport mechanisms, interactions, and resulting effects on man and his environment. Because of the complexities involved, assessment of specific pollutants in the environment requires a total systems approach that transcends the media of air, water, and land. The Environmental Monitoring Systems Laboratory-Las Vegas contributes to the formation and enhancement of a sound monitoring data base for exposure assessment through programs designed to:

- develop and optimize systems and strategies for monitoring pollutants and their impact on the environment
- demonstrate new monitoring systems and technologies by applying them to fulfill special monitoring needs of the Agency's operating programs

This report is the second in a series (see EPA-600/4-78-030) on a method for designing meteorological and air quality monitoring networks and the application of the method to the metropolitan St. Louis area. It is concerned with the evaluation of the meteorological (wind field) network selected for St. Louis. Regional or local agencies may find this method useful in planning new or adjusting existing aerometric monitoring networks. The Monitoring Systems Design and Analysis Staff may be contacted for further information on the topic.

George B. Morgan
Director
Environmental Monitoring Systems Laboratory
Las Vegas

PREFACE

This document is the second in a series on the development of a methodology for designing optimum meteorological and air quality monitoring networks and the application of the methodology to the metropolitan St. Louis area. It deals with the evaluation of the meteorological (wind field) network. The first document (EPA-600/4-78-030) considered the theoretical aspects of the methodology and the network(s) established for St. Louis. Subsequent reports will be concerned with verification of the methodology with regard to the air quality.

James L. McElroy
Project Officer
Environmental Monitoring Systems Laboratory
Las Vegas

SUMMARY

This report is the second in a series treating a method for developing optimum meteorological and air pollution networks and the application of the methodology for St. Louis (EPA-600/4-78-030 describes the method and the network for St. Louis). This particular report deals with the evaluation of the wind field determined from the optimum network. For this purpose, wind data obtained through summer (August 1975) and winter (February-March 1976) field programs were reduced and validated. The basic objective of the evaluation was to determine the precision and accuracy of the procedures used for estimating the wind field. The procedures for determining the wind field involved applying stepwise regression to a class of statistical models and data from a 19-station network; the network (the optimum network) and the class of models (linear statistical models involving subsets of a specific set of 13 terms) were determined during the theoretical phase of the study. Evaluation included the selection of a large class of model forms to compare with the 13-term class. For this purpose, a basic set of 23 terms which were dictated by the results of the theoretical phase of the study was chosen. The evaluation also included estimations based on data from all reporting stations--up to a total of 26 stations.

The principal conclusion of this study was that application of stepwise regression to the 13-term model together with wind data from the 19-station optimum network produced predicted wind fields comparable to those obtained by more general procedures (the 23-term model) applied to a larger network (at most 26 stations). This substantiated through observed data the results of the theoretical analysis conducted in the above-mentioned report.

An exhaustive evaluation was not feasible largely due to numerous analytical and data limitations. Less than 50% of the total data collected could be used for the analysis due to unreported and invalid data. The wind data associated with the winter field program was atypical for that period in that the period was characterized by southwesterly winds. According to available statistical information, the winter period in the St. Louis region is normally characterized by northwesterly winds. Furthermore, relative measures had to be utilized in the evaluation since the best model was unknown, and since only a small number of additional (i.e., non-network) wind monitoring stations were available in St. Louis. Also, errors which arose from network deficiencies could not be isolated from errors arising from other sources (e.g., model deficiencies, measurement errors).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

U	west-east wind component in meters per second (mps)
V	south-north wind component in meters per second (mps)
W	wind speed (mps)
(U,V)	wind vector with components U and V
x	west-east geographic coordinate relative to a given origin, in kilometers (km)
y	south-north geographic coordinate relative to a given origin (km)
h or $h(x,y)$	terrain elevation in meters (m) at the point (x,y) relative to a fixed base plane
k	wind component index (k = 1 for U-component, k = 2 for V-component
i	station index
t	time index
j	index that identifies modeling procedures
(x_i,y_i)	geographic coordinates of i^{th} station
(U_m,V_m)	observed wind components at m minutes away from a nominal time point (at 10 or 30 m above ground level)
$(\overline{U}_o,\overline{V}_o)$	20-minute average of wind components, at 10 or 30 m above ground level
$(\overline{U}'_o,\overline{V}'_o)$	20-minute average of wind components at 10 m above ground level, as estimated from observations at 30 m above ground level
(U^*,V^*)	west-east, south-north components of the friction velocity
e	elevation above ground level (m)
T_o	mean roughness length (m)
Z_{kt} or $Z_{kt}(x,y)$	20-minute average for wind component k at 10 m above gound level, at time t and at the point (x,y)

$Z_{kt}(i)$	$Z_{kt}(x_i,y_i)$--i.e., observed value of wind component k at station i at time t
$\underline{x}$	a 23 x 1 vector involving functions of x and y
$\underline{\beta}_{kt}$	a 23 x 1 vector of unknown parameters associated with wind component k at time t
ε_{kt} or $\varepsilon_{kt}(x,y)$	the deviation, at the point (x,y), of wind component k at time t from an assumed model of the form $\underline{x}'\underline{\beta}_{kt}$
$\underline{x}_0$	a 13 x 1 vector consisting of the first 13 elements of $\underline{x}$
$\underline{\beta}_{0kt}$	a 13 x 1 vector of unknown parameters associated with wind component k at time t
ε_{0kt} or $\varepsilon_{0kt}(x,y)$	the deviation, at the point (x,y), of wind component k at time t from an assumed model of the form $\underline{x}_0'\underline{\beta}_{0kt}$
$\underline{Z}_{kt}$	a vector containing the $Z_{kt}(i)$, i=1,2,... .
X_t^*	a matrix for which the i^{th} row consists of the $\underline{x}'$ vector evaluated at (x_i,y_i)
Q	a arbitrary subset of stations
F	a network consisting of all (reporting) stations
R	a subset of stations consisting of all (reporting) stations in the RTI network
$n_t(Q)$	the number of observations (i.e., reporting stations) at time t in the network Q, where Q = R or F
$p_{kt}(j,Q)$	the number of terms in the model for wind component k at time t when modeling procedure j is applied to data in network Q, where Q = R or F
$\hat{\underline{\beta}}_{jkt}^{(Q)}$	a vector of $p_{kt}(j,Q)$ estimated parameters for component k at time t, obtained by applying modeling procedure j to network Q, where Q = R or F
$\underline{x}_{jkt}$	a vector obtained by retaining those elements of $\underline{x}$ which correspond to the $\hat{\underline{\beta}}_{jkt}^{(Q)}$ elements
X_{jkt}	a matrix for which the i^{th} row consists of the $\underline{x}_{jkt}'$ vector elevated at (x_i,y_i)

$\hat{Z}_{kt}(j,Q,i)$	the predicted value of the k^{th} wind component at time t at the point (x_i,y_i) when modeling procedure j is applied to data in network Q, where Q = R or F
$\hat{e}_{kt}(j,F,i)$	the deviation between the observed wind component, $Z_{kt}(i)$, and the predicted component, $\hat{Z}_{kt}(j,F,i)$
$s^2_{kt}(j,Q)$	the residual variance for component k from the model based on procedure j applied to wind data from network Q (Q = R or F) at time t
$R^2_{kt}(j,Q)$	the proportion of the total variation in wind component k at time t (over network Q) accounted for by the model resulting from modeling procedure j when it is applied to network Q (Q = R or F), i.e., an R^2 statistic
$A^2_{kt}(j,Q)$	the adjusted R^2 statistic based on $R^2_{kt}(j,Q)$
C	an arbitrary subset of cases (i.e., t values)
$s^2_{kC}(j,Q)$	the pooled residual variance over C, obtained as a weighted average of the $s^2_{kt}(j,Q)$ values
$R^2_{kC}(j,Q)$	the pooled R^2 statistic over C, i.e., the proportion of the total within-case variation in wind component k accounted for by applying modeling procedure j to data from network Q (Q = R or F)
$A^2_{kC}(j,Q)$	the pooled adjusted R^2 statistic based on $R^2_{kC}(j,Q)$
N_{CQ}	the number of wind observations in the intersection of C and Q
$W_t(i)$	observed wind speed at (x_i,y_i) at time t
$\theta_t(i)$	observed wind direction at (x_i,y_i) at time t
$\hat{W}_t(j,Q,i)$	predicted wind speed at (x_i,y_i) at time t, based on applying modeling procedure j to data from network Q (Q = R or F)
$\hat{\theta}_t(j,Q,i)$	predicted wind direction at (x_i,y_i) at time t, based on applying modeling procedure j to data from network Q (Q = R or F)
s(j)	the square root of $s^2_{1C}(j,R) + s^2_{2C}(j,R)$, where C consists of all cases

$r(j)$	the pooled vector root mean square error associated with procedure j -- pooled over both wind components, all cases, and all stations not in the RTI network
$r^*(j)$	same as $r(j)$ but over all interior stations not in the RTI network
$f_\alpha(j)$	a weighted average of $[r(j)]^2$ and $[s(j)]^2$, where α is the weight attached to the former
$g_\alpha(j)$	a weighted average of $[r^*(j)]^2$ and $[s(j)]^2$, where α is the weight attached to the former

ACKNOWLEDGMENTS

This report was prepared by the Research Triangle Institute (RTI), Research Triangle Park, North Carolina, under contract No. 68-03-2187 for the U.S. Environmental Protection Agency (EPA). The project officer was Dr. James L. McElroy. Many individuals from RTI participated in this project. Mr. J. W. Dunn was responsible for developing the computer algorithm for processing and reducing the wind data. Mr. Bobby Crissman was responsible for the initial data reduction. Mr. Clifford Decker was responsible for management of the field program.

We would also like to acknowledge the cooperation of Mr. Robert Browning of EPA, Research Triangle Park, North Carolina, for providing us with the Regional Air Pollution Study (RAPS) data; and Mr. Ashwin Gajjar, St. Louis County Air Pollution Control Agency, for providing data from the St. Louis City and County air pollution stations.

SECTION 1

INTRODUCTION

This report provides an evaluation of one aspect of an overall methodology for generating estimated pollution concentration surfaces over an urban area. This methodology, if successful, would avoid three of the major problems typically encountered in estimating such surfaces directly from observed air quality data; these problems occur because:

(a) reliable estimation (for a single pollutant) requires a high resolution network of air quality monitoring stations,

(b) "optimal" networks for two different pollutants would generally be different because of different emission sources, and

(c) an "optimal" network (for a single pollutant) remains "optimal" only in the short-term because of changes in the emission sources.

The proposed methodology has the potential of overcoming these problems by utilizing the emissions source inventory as a primary source of data and by establishing a network which is "optimal" for estimating wind fields. The model development phase of the proposed methodology, as well as its implementation in the St. Louis, Missouri area, is described by Vukovich et al. (1978).

The following subsections provide a brief description of the overall concept, and summarize the statistical model form and sampling network which resulted from applying the methodology in St. Louis. The specific objectives of this report are then described, along with a description of the organization of the remainder of the report.

OVERVIEW OF THE PROPOSED METHODOLOGY

The proposed methodology involves six major steps:

(1) Utilize a three-dimensional hydrodynamic model to generate simulated wind fields for the (urban) area under a variety of (initial) meteorological conditions.

(2) Determine a class of statistical model forms relating winds to geographic location and topography which will yield a reasonable approximation to the simulated results for any of the initial conditions.

(3) Using the results of (2), determine an "optimal" set of sites for monitoring winds.

(4) Establish wind and air quality monitoring stations at the indicated sites.

(5) Estimate wind fields by fitting statistical models based on the class of forms determined in (2) to the observed data.

(6) Utilize an objective variational analysis model to estimate pollutant concentrations over the area by combining the emissions source inventory, the observed pollutant concentrations, and the estimated wind fields.

With minor modifications resulting from practical and economic constraints, the first four steps above have been completed for the St. Louis area; the following section describes the class of statistical models and the network established in the St. Louis area.

SUMMARY OF PREVIOUS RESULTS

Consider an arbitrary point in the St. Louis region with coordinates (x,y) relative to a fixed origin, where x denotes distance in kilometers (km) in the east direction and y, in the north direction. Let $h \equiv h(x,y)$ denote the elevation in meters (m) at (x,y) relative to a fixed base plane at river elevation of approximately 100 m. Let $Z_{kt} \equiv Z_{kt}(x,y)$ denote the value in meters per second (mps) of the k^{th} wind component (k=1 for the west-east component, U; k=2 for the south-north component, V) at time t. The model form proposed by the Research Triangle Institute (RTI) in Vukovich et al. (1978), which formed the basis for determining the sampling network, was

$$Z_{kt} = \underline{x}_0' \, \underline{\beta}_{0kt} + \varepsilon_{0kt} \qquad (1)$$

where

$\underline{x}_0' = (1 \; x \; y \; x^2 \; y^2 \; xy \; x^3 \; y^3 \; x^2y \; xy^2 \; x^4 \; y^4 \; h)$,

$\underline{\beta}_{0kt}$ = a 13 x 1 vector of unknown parameters for component k at time t, and

$\varepsilon_{0kt} \equiv \varepsilon_{0kt}(x,y)$ = random deviation in component k at time t at the point (x,y).

The proposed network, which was subsequently established and which is herein referred to as the RTI network, involves 19 stations (see Figure 1). Because of on-going data collection activities in the local area, it was only necessary for RTI to set up three stations for this evaluation. Sixteen existing stations were situated in close proximity to "optimal" locations established during the theoretical phase of the study. Table 1 shows the (x,y) coordinates and elevations (h) of the 19 stations in the RTI network. Four of the 19 stations in the network are St. Louis city/county stations (denoted by the STL prefix in the station names), twelve are Regional Air Pollution Study (RAPS) Stations of the United States Environmental Protection Agency (denoted by the EPA prefix in the station names), and three stations (denoted by the RTI prefix) were temporary stations set up by RTI specifically for this research project. The RTI stations were located on the grounds of Incarnate Word Academy in northwest St. Louis county; on the grounds of Ken-

TABLE 1. GEOGRAPHIC LOCATIONS AND TERRAIN ELEVATIONS FOR STATIONS IN THE RTI NETWORK*

Station Name	x (km)	y (km)	h (m)
STL008	0.	16.	45.
RTI202	- 4.	8.	79.
STL009	- 7.	2.	44.
STL006	-20.	- 3.	46.
RTI205	- 6.	- 6.	37.
STL002	0.	-10.	12.
RTI207	10.	-10.	11.
EPA101	6.	1.	24.
EPA102	5.	7.	5.
EPA104	9.	- 2.	13.
EPA105	4.	- 3.	50.
EPA106	1.	- 1.	36.
EPA108	10.	12.	9.
EPA109	18.	0.	13.
EPA110	9.	- 6.	6.
EPA113	0.	10.	55.
EPA118	5.	-16.	28.
EPA119	- 9.	- 8.	56.
EPA120	-16.	8.	37.

* Locations are defined relative to an origin at the intersection of Lindell Blvd. and King's Highway in St. Louis. Elevations are defined relative to a local river elevation of approximately 100 m.

rick Seminary in southwest St. Louis county; and on the grounds of the East Side Sanitary District's South Pumping Station in East St. Louis, Illinois.

The major emphasis of the second phase of the research project involved the preparation and execution of a summer and winter field program in St. Louis. These field programs were held during a period when EPA was performing an intensive study in St. Louis: August, 1975, and February and March, 1976. During this time, there was a concerted effort to maintain a high level of performance of the RAPS stations.

OBJECTIVES AND SCOPE OF THE CURRENT RESEARCH

The scope of the current effort is limited to an analysis of the wind data which were obtained during the summer and winter field programs. These data consisted of the horizontal wind components as measured at various time intervals and at 27 sites within the St. Louis region. These 27 sites included the 19 stations involved in the RTI network. The objectives of the study are

(1) to develop an easily-automated estimation procedure, based on the model form in equation (1), for generating estimated wind fields, and

(2) to evaluate the performance of this procedure and of the RTI network.

Thus, in terms of the six major steps involved in the methodology, this phase of the research involves a demonstration of step 5, and an evaluation of the overall methodology up through step 5.

Section 2 describes the available data, its limitations, and the editing procedures employed in preparing the data for analysis. The analytical approach is described in Section 3 and the results are summarized in Section 4. Section 5 presents the conclusions, findings, recommendations, and analytical limitations of the study.

OBJECTIVES AND SCOPE OF REMAINING RESEARCH

Assuming validation of the procedures and network for making wind field predictions, the next step in the research project will involve an evaluation of the objective variational analysis model (OVAM) used to derive the estimated air pollution distribution. The OVAM uses the estimated wind field as an input parameter, along with the emissions inventory and the air pollution concentrations as measured at the network stations. Carbon monoxide (CO) will be used in the evaluation.

The evaluation of the OVAM will be made on a case study basis, with each case study covering a 12- to 24-hour period. The selected case studies will be chosen so as to represent a variety of wind conditions (speeds, directions) and of CO concentration distributions over the monitoring stations. The basic evaluation parameters will consist of correlations and root mean squared errors between observed and predicted CO concentrations at stations outside of the RTI network. As a part of this study, it will be determined if it is necessary to monitor CO at each of the 19 network stations.

SECTION 2

SUMMARY OF AVAILABLE WIND DATA

DESCRIPTION OF RAW DATA

In addition to stations in the RTI network, eight additional stations provided data. Coordinates and elevations of these stations are shown in Table 2.

TABLE 2. GEOGRAPHIC LOCATIONS AND TERRAIN ELEVATIONS FOR STATIONS NOT IN THE RTI NETWORK*

Station Name	x (km)	y (km)	h (m)
STL003	2	6	39
STL004	-2	-1	39
STL007	-10	10	81
STL010	-8	-12	62
EPA103	10	3	16
EPA107	2	3	44
EPA111	1	-7	19
EPA112	-4	2	44

* Locations are defined relative to an origin at the intersection of Lindell Blvd. and King's Highway in St. Louis. Elevations are defined relative to a local river elevation of approximately 100 m.

The total set of 27 stations, whose locations are shown in Figure 1, will be referred to as the full network; the above set of eight stations will be referred to as non-network stations (meaning non-RTI-network stations). Stations STL007, STL010, and EPA103 will be referred to as outer-non-network stations, since they are located on the border of the innermost grid (see Figure 1), whereas the remaining five non-network stations will be called inner-non-network stations. These two sets of non-network stations are distinguished because it was shown in the first report in this series (Vukovich, et al., 1978) that, if wind data from

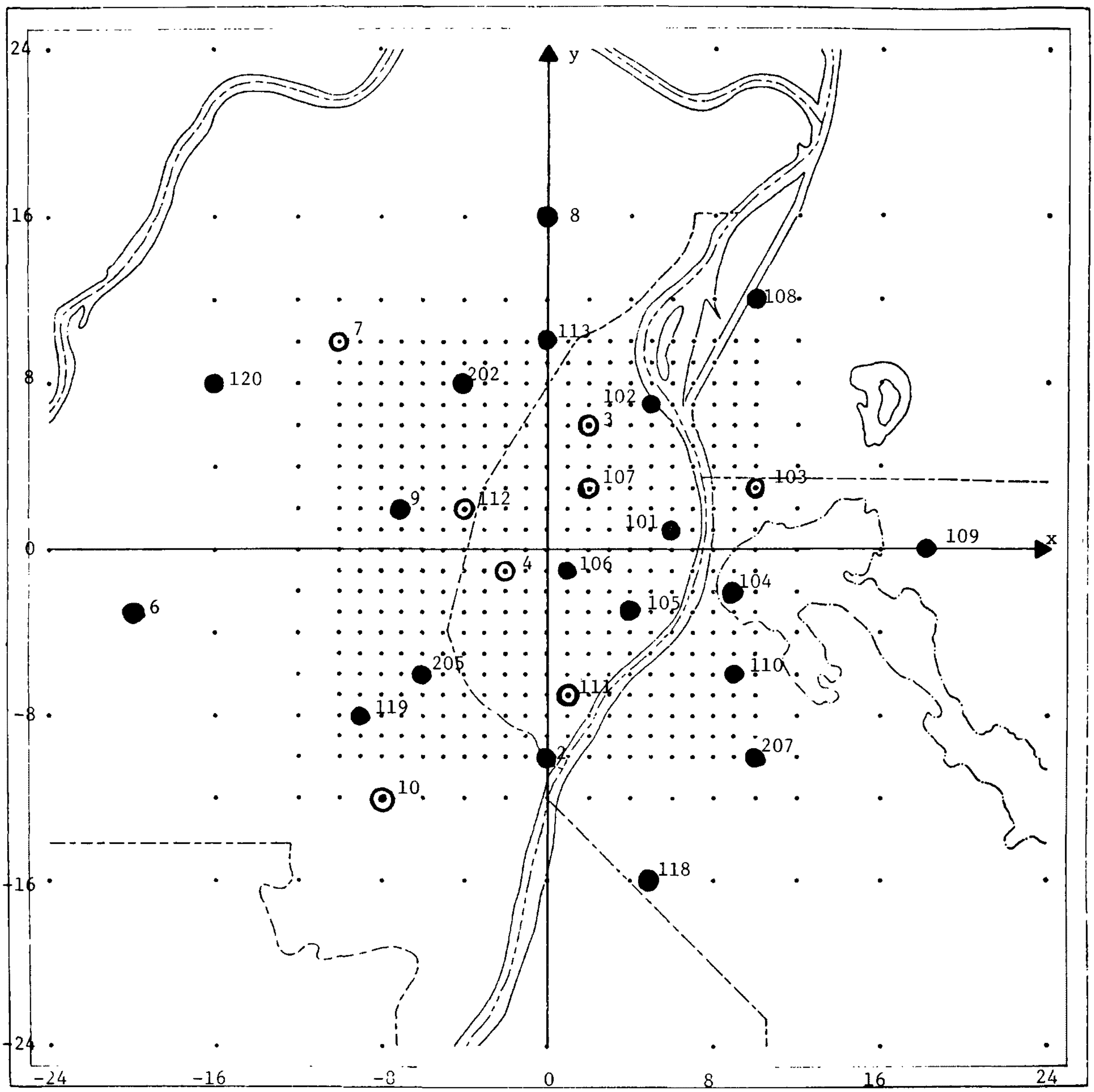


Figure 1. Location of stations in the RTI network (solid dots) and other non-network stations (open dots) used in the evaluation (interior grid spacing = 1 km)

the RTI network were used to produce predictions at the non-network stations, considerably better predictions should be achieved for inner-non-network stations than for outer-non-network stations.

The raw wind data consisted of 1-minute and 5-minute average values from the EPA stations, 3-minute average values from the St. Louis (STL) city/county stations and 5-minute average values from the RTI stations. Five-point averages centered at each half-hour were constructed. The nominal 20- to 25-minute averaging period is consistent with the averaging performed in the hydrodynamic model, which produced the simulated wind fields upon which the RTI network was based. For the EPA and RTI stations, these averages were computed from the 5-minute averages (for the U and V components, respectively) as

$$\overline{U}_o = (U_{-10} + U_{-5} + U_0 + U_5 + U_{10})/5$$

and

$$\overline{V}_o = (V_{-10} + V_{-5} + V_0 + V_5 + V_{10})/5,$$

where the subscripts indicate deviations in minutes between the nominal (hour or half hour) time point and the midpoint of the averaging interval for the raw data. For the city/county stations,

$$\overline{U}_o = (U_{-9} + U_{-6} + U_0 + U_6 + U_9)/5$$

and

$$\overline{V}_o = (V_{-9} + V_{-6} + V_0 + V_6 + V_9)/5.$$

In either case, at least three of the five readings were required to be present in order for an average wind to be used.

Winds at the three RTI stations and at all of the St. Louis city/county stations were measured at 10 m above ground level. This was also the case for three of the RAPS stations: EPA108, EPA110, and EPA118. Measurements at the remaining thirteen RAPS stations, however, were made at 30 m above ground level. The wind data from these stations were therefore inappropriate for evaluating the methodology. To alleviate this problem, a profile equation for the surface boundary layer (Estoque and Bhumralkar, 1969) was used to generate estimated 10-m winds at these thirteen stations using the winds at the 30-m level. The estimated wind components at 10 m at a particular station and time were determined as:

$$\left.\begin{aligned} \overline{U}'_o &= \overline{U}_o/3 + 2.5U^* \, [L_{10}-L_{30}] \\ \overline{V}'_o &= \overline{V}_o/3 + 2.5V^* \, [L_{10}-L_{30}] \end{aligned}\right\} \qquad (2)$$

where

$\overline{U}'_o$ and $\overline{V}'_o$ are, respectively, the west-east and south-north components of the wind velocity at the 10-m level;

$\overline{U}_o$ and $\overline{V}_o$ are, respectively, the west-east and south-north components of the wind velocity at the 30-m level;

$$L_e = \ln\left[\frac{e+T_o}{T_o}\right];$$

T_o is the mean roughness length associated with the particular site;

e is the elevation (m) above ground level; and

U* and V* are, respectively, the west-east and south-north components of the friction velocity.

The U* component was determined as

$$U^* = \left[A + B|\overline{U}_o| + C\overline{U}_o^2\right] \quad \text{sign}(\overline{U}_o) \qquad (3)$$

where the coeffients A, B, and C were based on data relating the mean wind speed to the friction velocity (J.F. Clarke, EPA-RTP, personal communique, 1978). A similar formula was used for the V*-component.

The coefficients in Eq. (3) were determined separately for each season (summer and winter) and for each of three types of stations (urban, suburban, rural). They are based on comparative analyses between measured turbulence parameters and wind speeds that were performed at numerous RAPS stations and were consolidated for the purposes of this study. The analyses were performed by, and the results acquired from, the U.S. Environmental Protection Agency. Values of the coefficients are shown in Table 3 below:

TABLE 3. COEFFICIENTS USED FOR DETERMINING FRICTIONAL VELOCITY COMPONENTS

Season	Region	Coefficient A	Coefficient B	Coefficient C
Summer	Urban	-0.04591	0.18763	-0.01036
	Suburban	-0.05006	0.13023	-0.00212
	Rural	-0.01640	0.05419	0.00102
Winter	Urban	-0.07601	0.16372	-0.00469
	Suburban	-0.04947	0.12742	-0.00275
	Rural	0.02616	0.02902	0.00243

Originally, coefficients were also determined as a function of stability. However, the values obtained were judged to be sufficiently similar so that such additional differentiation was unnecessary.

Table 4 indicates the type of each station and its mean roughness length (T_o), as used in the above conversion formulae. The roughness lengths were determined using the technique developed by Lettau (1969), with parameters developed specifically for St. Louis (Vukovich et al., 1976).

The estimated $\overline{U}'_o$ and $\overline{V}'_o$ values determined for the 13 RAPS stations from equations (2) and (3), along with the observed $\overline{U}_o$ and $\overline{V}_o$ values for the other 14 stations, constituted the basic wind data upon which the evaluations were performed.

TABLE 4. MEAN ROUGHNESS LENGTHS (T_o), BY STATION

Type	Station	T_o(m)
Urban	EPA101	0.72
	EPA104	0.39
	EPA106	1.08
	EPA107	1.32
Suburban	EPA102	0.20
	EPA103	0.20
	EPA105	0.60
	EPA111	0.24
	EPA112	0.48
	EPA113	0.66
	EPA119	0.66
Rural	EPA109	0.20
	EPA120	0.45

DATA EDITING

For those time points (cases) in which only a few of the 19 RTI-network stations provided data, the estimation of wind fields would be quite tenuous; furthermore, evaluation of the performance of the network for providing good predictions would be unrealistic in such cases. Consequently, as a first step in preparing the data for analysis, all cases in which more than one-third of the RTI-network stations failed to furnish wind data were deleted from further consideration. With this requirement imposed on the data set -- namely, that data be available for at least 13 stations in the RTI network -- there were 260 cases available from the summer field program and 654 from the winter field program.

A manual screening of these data was then performed. Inconsistencies in the city/county data relative to the remaining data in the first six summer cases, which were scattered across 11 days (July 29 to August 8, 1975), led to the exclusion of these six cases from the basic summer data set. These inconsistencies were apparently the result of calibration problems. Out of the remaining 254 summertime cases, the manual screening resulted in the deletion of

(a) all wind data from EPA102, which appeared highly inconsistent with data at nearby stations,*

(b) three extremely peculiar wind values at other stations, and

(c) forty-three consecutive observations for EPA120 and eight for RTI205 in which instrument failures were apparently responsible for producing zero values for both wind components.

Among the 254 summer cases, no data were available for two stations: STL010 and EPA111.

A similar editing of the winter data resulted in the exclusion of all wind data from STL010, and partial exclusion of data from six other stations. Counts of these exclusions, which also resulted from instrument failures, are shown below:

Station	Initial No. of Reported Cases	No. of Cases Deleted	No. of Cases Retained
RTI202	454	184	270
RTI205	457	7	450
RTI207	549	4	545
STL008	639	550	89
STL007	639	543	96
STL010	500	500	0
EPA107	616	3	613

The final edited data sets covered an 8-day period in August 1975 (August 9-16) and a 25-day period in the winter of 1976 (February 10 - March 5). Out of the 8 x 48 = 384 potential cases which could have occurred during the 8-day summer field program, only 254 were actually

---

* Major repairs were performed on the wind monitoring equipment at EPA102 between the times of the summer field program and the subsequent winter program.

retained after all editing; only 654 wintertime cases were available, out of a possible 25 x 48 = 1200. Thus, the field programs not only were of short duration (especially the summer program) but also failed to provide "sufficient" data in many cases. This is depicted in the left-hand portion of Table 5, which shows the distribution of available cases, by date and time-of-day. It is clear from this table that a large degree of clustering of cases within time periods occurred. Consequently, increasing the number of cases by constructing three rather than two 20-minute averages per hour would not have enhanced the data base in terms of its coverage of additional wind conditions (see Table 8).

Unfortunately, a substantial amount of missing data occurred even within the 908 cases for which 13 or more stations in the RTI network reported data (because of the above-described editing, or because the data were simply not reported). For instance, the full set of 19 stations in the RTI network furnished coincident data in only 14 of the 908 cases; these cases all occurred during a 2-day period within the winter field program, as shown in the right-hand portion of Table 5. Eighteen or more stations in the network reported in only 105 out of the 908 cases.

The high incidence of missing data was not confined to stations in the RTI network, as evidenced in Table 6. Two of the non-network stations, STL007 and EPA111, had particularly low reporting rates (after editing). Assuming 8 full days for the summer program and 25 full days for the winter program, the reporting rates in terms of individual observations were as shown in Table 7.

TABLE 5. DISTRIBUTION OF CASES, BY DATE AND TIME-OF-DAY AND BY DATE AND NUMBER OF RTI NETWORK STATIONS REPORTING

Date	Time-of-Day Reporting				Total	No. of Reporting Stations in RTI Network							
	0000-0530	0600-1130	1200-1730	1800-2330		12	13	14	15	16	17	18	19
75/08/09	6	6	0	11	23	--	--	--	23	--	--	--	--
10	12	12	11	10	45	--	--	--	45	--	--	--	--
11	12	12	12	12	48	--	11	13	11	13	--	--	--
12	12	12	12	12	48	--	--	--	3	42	3	--	--
13	12	9	6	7	34	--	--	1	8	12	2	11	--
14	5	5	9	9	28	--	--	--	--	1	7	20	--
15	12	11	1	0	24	--	--	1	--	3	3	17	--
16	0	0	0	4	4	--	--	--	4	--	--	--	--
76/02/10	10	11	10	0	31	--	--	--	--	11	20	--	--
11	0	0	2	3	5	--	--	--	--	--	5	--	--
12	7	8	12	7	34	--	--	--	3	17	14	--	--
13	6	0	0	0	6	--	1	5	--	--	--	--	--
14	0	7	12	8	27	--	--	--	1	4	22	--	--
15	10	12	12	1	35	--	5	2	7	10	11	--	--
16	4	12	12	3	31	--	--	--	2	7	22	--	--
17	0	3	7	9	19	--	--	--	1	2	16	--	--
18	12	4	3	11	30	--	1	3	5	8	13	--	--
19	11	1	0	2	14	--	--	3	3	2	6	--	--
20	6	6	4	0	16	--	--	--	--	6	6	4	--
21	3	9	5	0	17	--	1	4	1	4	7	--	--
22	0	1	8	0	9	--	--	--	--	--	--	9	--
23	3	11	9	12	35	--	--	3	4	4	20	4	--
24	12	5	2	10	29	--	5	4	18	2	--	--	--
25	11	11	12	11	45	--	1	3	12	19	10	--	--
26	11	1	1	11	24	--	1	--	--	4	15	4	--
27	12	12	12	12	48	--	--	1	4	4	24	10	5
28	1	1	12	12	26	--	--	--	--	5	12	--	9
29	12	12	10	12	46	--	3	1	7	19	14	2	--
76/03/01	2	4	12	12	30	--	--	1	2	7	20	--	--
02	11	12	12	12	47	2	--	4	2	14	19	6	--
03	12	1	0	0	13	--	1	--	--	1	11	--	--
04	0	0	11	12	23	1	--	1	7	5	9	--	--
05	12	2	0	0	14	--	--	1	2	1	6	4	--
Summer Total	71	67	51	65	254	0	11	15	94	71	15	48	0
Winter Total	168	146	180	160	654	3	19	36	81	56	302	43	14
Overall Total	239	213	231	225	908	3	30	51	175	227	317	91	14

TABLE 6. NUMBER OF CASES FOR WHICH VALID WIND DATA ARE REPORTED, BY STATION

Network	Station	No. of Cases Summer	Winter	Total
RTI:	STL008	254	89	343
	RTI202	223	270	493
	STL009	254	639	893
	STL006	254	639	893
	RTI205	242	450	692
	STL002	253	639	892
	RTI207	250	545	795
	EPA101	116	618	734
	EPA102	0	641	641
	EPA104	240	639	879
	EPA105	243	647	890
	EPA106	254	638	892
	EPA108	252	645	897
	EPA109	253	514	767
	EPA110	248	530	778
	EPA113	245	636	881
	EPA118	164	635	799
	EPA119	70	628	698
	EPA120	203	630	833
Non-RTI:				
Outer:	STL007	254	96	350
	EPA103	253	580	833
Inner:	STL003	254	638	892
	STL004	254	638	892
	EPA107	231	613	844
	EPA111	0	596	596
	EPA112	237	624	861

NOTE: Station STL010, in the outer non-network, is omitted because no "valid" data were reported from this station.

TABLE 7. NUMBER OF AVAILABLE AND POTENTIAL OBSERVATIONS, BY NETWORK AND SEASON +

Subset of Stations	No. of Stations	Summer Field Program No. Obs.	Summer Field Program Potential No. Obs.	Summer Field Program Rate (%)	Winter Field Program No. Obs.	Winter Field Program Potential No. Obs.	Winter Field Program Rate (%)
RTI Network	19	4018	7296	55.1	10672	22800	46.8
Non-Network*	8	1483	3072	48.3	3785	9600	39.4
Outer*	3	507	1152	44.0	676	3600	18.8
Inner	5	976	1920	50.8	3109	6000	51.8
Full Network	27	5501	10368	53.0	14457	32400	44.6

\+ Potential no. observations = no. stations x 48 cases per day x no. days

* For completeness, STL010 is counted as a potential station, although no valid wind data were obtained from this station.

It is clear that the limited time span, the limited number of non-network stations, and the large amount of missing data impose some severe limitations on the model and network evaluations. Fortunately, analysis of historical, seasonal wind roses for the National Weather Service Station at Lambert Air Field in St. Louis showed that the winds which occurred during the 8-day summer field program were typical of the wind conditions which are prevalent in St. Louis in the summer (i.e., predominantly south to southwest winds of low velocity). The distribution of prevailing wind speeds/directions for these 254 cases is shown in the upper portion of Table 8. The prevailing wind speeds and directions for a particular case were based on the average wind vector over the following outlying stations: STL006, STL007, STL008, EPA118, EPA119, EPA120. This definition is maintained throughout this report. On the other hand, the winter data also showed a predominant southerly wind (of

TABLE 8. DISTRIBUTION OF CASES, BY SEASON AND PREVAILING WIND CONDITIONS *

Season	Prevailing Speed (mps)	Prevailing Direction N	NE	E	SE	S	SW	W	NW	Total
Summer	0-1	1	--	--	7	24	14	8	1	55
	1-2	--	--	--	2	60	46	7	6	121
	2-3	--	--	--	9	35	19	7	--	70
	3-4	--	--	--	1	3	1	2	--	7
	4-5	--	--	--	--	--	--	1	--	1
	Total	1	--	--	19	122	80	25	7	254
Winter	0-1	--	--	1	3	3	4	1	--	12
	1-2	1	--	--	4	28	8	7	1	49
	2-3	--	--	--	18	47	19	5	--	89
	3-4	--	--	1	32	114	36	10	--	193
	4-5	--	--	--	11	114	33	18	--	176
	5-6	--	--	--	3	40	28	18	--	89
	6-7	--	--	--	--	11	13	6	--	30
	7-8	--	--	--	--	4	4	4	--	12
	8-9	--	--	--	--	4	--	--	--	4
	Total	1	--	2	71	365	145	69	1	654
Combined	0-1	1	--	1	10	27	18	9	1	67
	1-2	1	--	--	6	88	54	14	7	170
	2-3	--	--	--	27	82	38	12	--	159
	3-4	--	--	1	33	117	37	12	--	200
	4-5	--	--	--	11	114	33	19	--	177
	5-6	--	--	--	3	40	28	18	--	89
	6-7	--	--	--	--	11	13	6	--	30
	7-8	--	--	--	--	4	4	4	--	12
	8-9	--	--	--	--	4	--	--	--	4
	Total	2	--	2	90	487	225	94	8	908

* Prevailing wind speed and direction for a particular case is based on the average wind vector over the following outlying stations: STL006, STL007, STL008, EPA118, EPA119, EPA120.

somewhat higher velocity), which contrasts with the wintertime pattern of northerly winds typical of the St. Louis area. Thus the model and network evaluation is also limited in terms of the types of cases covered.

An overall summary of the observed, edited data is given in Table 9, which shows sample sizes (N) and the means and standard deviations (s.d.) of the wind components (denoted by U and V) and wind speeds (denoted by W) for each station. Some care must be exercised in comparing the means of two or more stations, since the averages are not necessarily taken over the same set of cases due to the presence of missing data.

TABLE 9. SUMMARY STATISTICS BY STATION FOR OBSERVED WIND DATA OVER ALL CASES

STATION	N	U MEAN (mps)	V MEAN (mps)	W MEAN (mps)	U S.D. (mps)	V S.D. (mps)	W S.D. (mps)
SIL008	343	1.220	0.738	1.925	2.081	0.926	1.874
RI1202	493	0.800	1.331	2.176	1.641	1.252	1.389
SIL009	893	1.820	1.781	3.076	1.423	1.953	1.689
SIL006	893	1.642	1.551	3.133	2.001	1.941	1.746
RI1205	692	-0.373	2.168	2.496	1.222	1.562	1.594
SIL002	892	1.205	2.929	3.725	1.802	1.782	1.604
RI1207	795	0.231	3.100	3.718	2.056	2.206	2.220
EPA101	734	0.647	3.189	3.848	1.837	1.659	1.360
EPA102	641	0.317	4.559	5.527	2.869	2.274	1.929
EPA104	879	1.383	3.787	4.667	2.231	2.189	2.058
EPA105	890	0.992	3.227	3.810	1.634	1.778	1.646
EPA106	892	1.246	2.570	3.411	1.720	1.386	1.182
EPA108	897	0.706	4.233	5.125	2.671	2.968	2.844
EPA109	767	0.630	3.046	3.595	1.677	1.881	1.760
EPA110	778	1.145	2.969	3.830	2.246	1.904	2.029
EPA113	881	1.011	3.080	3.918	1.976	1.885	1.615
EPA118	799	0.479	4.498	5.201	2.426	2.588	2.448
EPA119	698	0.727	2.890	3.602	1.916	1.730	1.601
EPA120	833	0.682	2.305	3.007	1.718	1.464	1.352
SIL007	350	0.153	2.014	2.460	1.165	1.287	1.018
SIL010	0	0.0	0.0	0.0	0.0	0.0	0.0
EPA103	833	1.201	4.522	5.483	2.590	2.437	2.115
SIL003	892	1.456	2.228	3.321	1.839	1.632	1.448
SIL004	892	0.550	3.095	3.851	2.230	1.776	1.782
EPA107	844	0.796	2.756	3.314	1.479	1.480	1.274
EPA111	596	0.514	3.582	4.313	2.189	1.905	1.703
EPA112	861	0.726	2.680	3.383	1.793	1.584	1.411

SECTION 3

EVALUATION TECHNIQUES

The development of the RTI network was based on the 13-term model form (Eq. (1)), as described in Vukovich et al. (1978). Based on the simulated data, it appeared reasonable to assume that a model of this order of variability would yield adequate wind field predictions in all, or virtually all, cases which might occur within the St. Louis area--if actual and simulated winds behave similarly. It was also apparent from the simulated data that this order of variability (i.e., the full model) would not necessarily be required in all, or even in most, cases. That is, some simpler model would be sufficient in the majority of cases. Because fitting the full model in a case in which a simpler submodel is appropriate can substantially decrease the precision of a predicted value, the modeling procedure not only must provide estimates of the regression coefficients but also must establish, through some variable selection technique, the form of the model. It should be noted that a proper evaluation of the theoretical phase results requires that wind fields be estimated via some submodel of the model given in Eq. (1). In an actual implementation of the technique, however, other surface-fitting techniques could be utilized for estimating the winds over the region of interest.

There are many possible variable selection procedures; in general, three basic steps are involved:

(1) specification of a class of potential model forms from which the selection is to be made,

(2) determination of a single "good" model form from within this class, and

(3) estimation of the parameters of this model form.

Step (1) was performed as a part of the network selection during the theoretical phase of the study. In the current context, the proposed methodology can be considered successful in terms of selecting a model form for estimating wind fields only if submodels of Eq. (1) can provide "adequate" fits to the wind component data at any point in time. In terms of both model and network selection, the methodology can be considered successful for estimating wind fields if applying this modeling procedure to wind data from stations in the RTI network provides accurate prediction of the winds over the region or, in practice, at particular sites not in the RTI network.

The first requirement for the evaluation is therefore to define the wind field modeling procedure i.e., to define precisely this aspect of the methodology. For this evaluation, the definition must be compatible with the results of the theoretical phase and must therefore utilize the model of Eq. (1) as its basis. The second requirement is a definition of alternative modeling procedures against which this procedure can be compared. The development of these modeling procedures is described in the subsection below.

The next step in the evaluation is to define measures of model "adequacy" for making the comparisons among the alternative procedures. Finally, measures of accuracy for judging the success of the overall methodology (excluding the air quality predictions) are needed. These two steps are described in the last two subsections of this section.

As indicated in the previous section, the 908 cases available for evaluating the methodology do not constitute a probability sample of time intervals; consequently, it was not possible to make valid statistical inferences to the population of wind conditions occurring in St.

Louis during some given period of time (e.g., one year). On the other hand, consideration of a few selected cases would also not appear to be sufficient for evaluating the methodology. Hence, the basic strategy adopted for the evaluation involves generating estimated wind fields for all 908 cases, generating measures which reflect the precision and accuracy of these estimates, and then summarizing these measures -- in terms of descriptive statistics -- over all cases and over various subsets of cases.

SELECTION OF MODELING PROCEDURES

The class of model forms indicated by the analysis of the simulated wind data consists of all possible subsets of the following twelve terms (as defined in Eq. (1)):

$$(x,\ y,\ x^2,\ y^2,\ xy,\ x^3,\ y^3,\ x^2y,\ xy^2,\ x^4,\ y^4,\ h) \qquad (4)$$

This assumes that a constant or intercept term would be required in any selected model. Thus there are $2^{12} = 4{,}096$ possible model forms (i.e., subsets of terms), which range in complexity from a constant, one-term model (corresponding to the selection of no terms from (4)) to a full 13-term model corresponding to the selection of all 12 terms in (4).

Many algorithms can be used for selecting variables; however, most of these procedures provide the user with a list of candidate models. The user must then apply some additional criterion in order to arrive at a single model form. This is regarded as a major advantage of these techniques; in the present context, however, such techniques are not practical unless one can also automate the additional criterion because of the large number of cases. For instance, in this study, the user of such a technique would have to examine 1,816 lists of candidate models

(908 cases x 2 wind components). The burden on the user during an actual implementation of the methodology would also be extreme if such an approach were to be adopted. Hence, one practical constraint on the variable selection procedure to be used is that it be fully automated in the sense that it incorporates its own stopping criterion and therefore yields a single model for each individual case. Even so, such an approach cannot be advocated for general implementation unless, for selected cases, one can carry out (a) an examination of residuals, and (b) a comparison with alternative modeling procedures (such as all-possible regressions). Regardless of what procedure might be implemented, it would also be essential that screening the wind data for erroneous values precede the model fitting.

Three sequential variable selection techniques which meet the above described constraint are the forward selection technique, the backward elimination technique, and the stepwise technique. Draper and Smith (1966) and Barr et al., (1976), for example, provide descriptions of these techniques and their relative merits. The stepwise procedure is generally considered to be superior to either of the other techniques. Also, the backward elimination technique, which successively deletes terms from an assumed larger, "full" model, would encounter estimability problems when the number of model terms exceeded the number of reporting stations. Hence, the stepwise regression approach was selected for use in the evaluations.

This procedure requires the use of two parameters referred to as the "inclusion" and "retention" parameters. As with the forward selection approach, the stepwise procedure begins by finding the best 2-variable model; this assumes an intercept is always included and is counted as one of the variables. Here, variable A is considered better than variable B if its correlation with the dependent variable (i.e., the observed wind component data) is higher, or more generally, if the partial F-statistic associated with variable A is larger than that for variable B. The variable with the largest F-statistic is retained in the model if the significance probability associated with the F-statistic is less than the "retention" parameter. If so, partial F-statistics associated with the remaining independent variables are computed and their significance probabilities are compared with the "inclusion" parameter; the variable with the smallest significance probability is added if this probability is less than the "inclusion" parameter. After such a variable is added, partial F-statistics are computed for all variables currently in the model to determine if any variable should be deleted from the model. A previously included variable is dropped if its associated significance probability exceeds the "retention" parameter. After any such deletions have been made, the F-values for the remaining variables are again determined to see if any meet the inclusion criterion. This process is continued until no variable can meet the inclusion criterion or until deletion of the last included variable occurs.

Two pairs of inclusion and retention parameters were used:

Modeling Procedure	Inclusion Parameter	Retention Parameter
1	0.10	0.10
2	0.20	0.20

These values were chosen, as opposed to smaller values, because of the small effects expected for many of the candidate terms and because of the small sample sizes -- namely, about 15 stations per case for the RTI network. In such situations, the use of smaller parameter values is generally not recommended because the derived models will tend to omit one or more "good" predictors.

Part of the evaluation procedure is thus a determination of which of these procedures is the more appropriate. Obviously, procedure 2 generates larger (i.e., more terms) models than does procedure 1; also, procedure 2 produces larger R^2 statistics (the square of the multiple correlation coefficient) and smaller residual sums of squares than procedure 1 achieves. However, procedure 1 may produce better predictions if procedure 2 tends to select "too many" terms.

Another key question to be addressed in the evaluation involves the choice of the initial class of model forms. For instance, is there another class of model forms which contains models that would provide substantially better approximations to the wind fields in the St. Louis area? Obviously, this aspect of the evaluation can be carried out only to a limited degree since there are an infinite number of possible model forms which could be investigated. The problem of evaluation is compounded by the fact that many cases are involved. In order to provide some evaluation of this potential source of error in the methodology, several other modeling procedures are considered. Whereas procedures 1 and 2 above are consistent with the proposed methodology, these additional procedures, in one way or another, are inconsistent with it. Hence, if performance of one of these additional procedures was judged to be substantially superior to procedures 1 and 2, it would indicate a deficiency in the proposed methodology. On the other hand, "good" performance by procedure 1 or 2 relative to the additional procedures would tend to support this aspect of the methodology but would not, of course, provide absolute proof of it because of the limitations involved in the evaluation.

The four additional modeling procedures used for the evaluation are defined below:

Modeling Procedure	Description
0	Fit the full 13-term model by ordinary least squares.
3	Apply stepwise regression to a larger class of model forms, utilizing the same "inclusion" and "retention" parameters as used for procedure 1.
4	Same as procedure 3, but using the parameters of procedure 2 rather than those of procedure 1.
5	Fit a flat surface (i.e., a one-term model involving the constant term) by ordinary least squares.

As with procedures 1 and 2, the above procedures are applied on a case-by-case basis for each horizontal wind component.

Procedures 0 and 5 represent the extremes of the previously-defined class of model forms used in procedures 1 and 2. These two procedures are not considered likely candidates for modeling winds, but are defined here because summary statistics based on these procedures are used for comparative purposes in the evaluation.

Procedures 3 and 4 differ from procedures 1 and 2, respectively, only in the choice of initial terms from which a model is developed. This initial class of terms for procedures 3 and 4 involves a total of 22 terms; in addition to all 12 terms shown in (4), the following 10 are also included:

$$x^3y \qquad xy^3 \qquad x^2y^2 \qquad xh \qquad yh$$
$$x^5 \qquad y^5 \qquad x^6 \qquad y^6 \qquad h^2$$

The basis for selecting these additional terms was the analysis of the simulated data, as described in Vukovich, et al. (1978). This analysis indicated that such terms, while less important than the 12-term set, were nevertheless useful for explaining some of the variation in some of the simulated cases. The class of models based on the 22-term set contains 2^{22} potential model forms; hence, this class is 1,024 times larger than the class based on 12 terms. Because the 12-term set is a subset of the 22-term set, it is clear that models based on procedure 3 (or 4) will generally explain more variation (i.e., larger R^2 values and smaller residual sums of squares) than procedure 1 (or 2). However, in terms of accuracy of predictions, models based on procedure 1 or 2 could still be superior to models based on procedures 3 or 4.

All six of the modeling procedures described above can be regarded as six different techniques for selecting a subset of 23 terms which consists of an intercept plus 22 specific terms. Let $\underline{x}$ denote the column vector of these 23 terms at an arbitrary location in the St. Louis area; that is,

$$\underline{x}' = (1, x, y, x^2, y^2, xy, x^3, y^3, x^2y, xy^2, x^4, y^4, h, x^3y, xy^3, x^5, y^5, x^6, y^6, x^2y^2, xh, yh, h^2) \quad (5)$$

Let $\underline{\beta}_{kt}$ denote a 23 x 1 vector of unknown coefficients for wind component k at time t. The general model can therefore be expressed as

$$Z_{kt} = \underline{x}' \, \underline{\beta}_{kt} + \varepsilon_{kt}, \quad (6)$$

where $Z_{kt} = Z_{kt}(x,y)$ is the observed value of the k^{th} wind component at (x,y) and time t, and ε_{kt} is a random deviation in component k at time t at the point (x,y). Table 10 summarizes the six modeling procedures with respect to this general model formulation. As indicated in this table, each procedure involves an assumption as to which coefficients are negligible (i.e., which terms in the $\underline{x}$ vector are deleted). Model (6) reduces to model (1), for example, when the last ten terms of

(5) are assumed negligible. Also, procedures 1 through 4 may determine, on the basis of the statistical tests involved in the stepwise algorithm, that other parameters can reasonably be assigned a zero value. In these cases, the selected model form will depend on what data

TABLE 10. SUMMARY OF MODELING PROCEDURES*

	Modeling Procedure					
	0	1	2	3	4	5
Coefficients assumed to be non-zero:	1-13	1	1	1	1	1
Coefficients assumed to be zero:	14-23	14-23	14-23	none	none	2-23
Coefficients which may be zero, as determined by stepwise regression:	none	2-13	2-13	2-23	2-23	none
Stepwise regression parameters -						
Inclusion:	N/A	0.1	0.2	0.1	0.2	N/A
Retention:	N/A	0.1	0.2	0.1	0.2	N/A

* Term numbers appearing as tabular entries assume that terms are ordered as in definition (5)

set is utilized, e.g., the full network or the RTI network. Once the model form has been determined, ordinary least squares is used for estimating the parameters.

CRITERIA FOR EVALUATING MODELING PROCEDURES

Each of the modeling procedures is applied, at a given point in time, to two sets of wind data--the data from stations in the RTI network and the data from all stations (i.e., the full network). Thus, for each case, twelve estimated wind fields are produced (2 networks x 6 procedures), as illustrated below:

Network Used for Model Estimation	Modeling Procedure					
	0	1	2	3	4	5
RTI						
Full						

For evaluating the modeling procedures, data from the full network (F) are utilized to determine the model forms and to estimate parameters. For each case (t) and wind component (k), six models are therefore estimated.

Assume that there are $n_t(F)$ stations in the full network which provide "valid" wind data at time t. Let $p_{kt}(j,F)$ denote the number of terms in the (selected) model when procedure j is applied to this set of data. Let X_t^* denote a matrix consisting of $n_t(F)$ rows (one row corresponding to each reporting station) and 23 columns; the i^{th} row consists of the $\underline{x}'$ vector (5) evaluated at the coordinates of the i^{th} station. Once the form of the model has been established, the least squares estimates are determined as

$$\hat{\underline{\beta}}_{jkt}^{(F)} = (X_{jkt}' X_{jkt})^{-1} X_{jkt}' \underline{Z}_{kt}$$

where

$\hat{\underline{\beta}}_{jkt}^{(F)}$ is the vector of $p_{kt}(j,F)$ estimated coefficients from procedure j, applied to the full network (F),

$\underline{Z}_{kt}$ is the vector of observed data, $Z_{kt}(x_i,y_i)$, $i=1,2,\ldots,n_t(F)$, and

X_{jkt} is a matrix obtained by deleting those columns (terms) of X_t^* that are associated with zero regression coefficients, as indicated by the particular procedure (see Table 10).

At an arbitrary point (x,y) in the region, the six predicted values of the wind component are obtained as

$$\underline{x}'_{jkt}\ \hat{\underline{\beta}}^{(F)}_{jkt} \qquad (7)$$

where $\underline{x}'_{jkt}$ consists of the relevant model terms. Hence, if coordinates of the i^{th} station, (x_i, y_i), are substituted into (7), predicted values for this station are determined. Let $\hat{Z}_{kt}(j,F,i)$ denote the predicted value of the k^{th} wind component for case t and station i, when procedure j is applied to the full network (denoted by F). The observed wind component at station i for case t is denoted by $Z_{kt}(i)$, i.e., $Z_{kt}(i) = Z_{kt}(x_i, y_i)$. For each value of k, t, and i, there are six deviations between observed and predicted values:

$$\hat{e}_{kt}(j,F,i) \equiv Z_{kt}(i) - \hat{Z}_{kt}(j,F,i) \qquad j=0,1,\ldots,5. \qquad (8)$$

These deviations form the basis for evaluating the modeling procedures.

It should be noted that the mean of these deviations is zero when the average is taken over stations in the full network (denoted by $i\varepsilon F$); that is,

$$\sum_{i\varepsilon F} \hat{e}_{kt}(j,F,i) = 0 \text{ for all k, t, and j.} \qquad (9)$$

In this same situation, the residual sums of squares corresponds to the sum of the squared deviations:

$$\sum_{i\varepsilon F} \hat{e}^2_{kt}(j,F,i) = [n_t(F) - p_{kt}(j,F)]s^2_{kt}(j,F) \qquad (10)$$

where

$n_t(F)$ = number of stations in network F providing valid data at time t,

$p_{kt}(j,F)$ = number of terms in the selected model when procedure j is applied to data from network F, for component k at time t, and

$s^2_{kt}(j,F)$ = the residual variance from the model based on procedure j when it is applied to data from network F, for component k at time t.

To simplify notation, let SSE(j) denote the sums of squared deviations appearing in (10)--for an arbitrary k and t. Then, as previously indicated, the following conditions must hold:

$$SSE(0) \leq SSE(2) \leq SSE(1) \leq SSE(5) \tag{11}$$

$$SSE(4) \leq SSE(3) \leq SSE(5). \tag{12}$$

The following conditions also usually, though not necessarily, hold:

$$SSE(3) \leq SSE(1) \tag{13}$$

$$SSE(4) \leq SSE(2). \tag{14}$$

For an individual case and wind component, typical ways for evaluating the fits of various models are

(a) comparison of individual residuals

(b) comparison of frequency distributions of the residuals or of absolute values of residuals -- or equivalently, the proportion of residuals less than some constant

(c) comparison of residual variances

(d) comparison of R^2 statistics

(e) comparison of adjusted R^2 statistics.

The residual variances are defined in (10) and can be rewritten as

$$s^2_{kt}(j,F) = \frac{SSE(j)}{n_t(F)-p_{kt}(j,F)} \tag{15}$$

The R^2 statistics for a particular case are defined as

$$R^2_{kt}(j,F) = \frac{SSE(5)-SSE(j)}{SSE(5)}. \tag{16}$$

The adjusted R^2 statistics for a particular case are given by

$$A^2_{kt}(j,F) = 1-(1-R^2_{kt}(j,F)) \frac{n_t(F)}{n_t(F)-p_{kt}(j,F)} \tag{17}$$

It should be noted that R^2 statistics are highly dependent on model size in situations where the number of parameters is large relative to the number of observations. The same is true, to a lesser degree, for the residual variance criterion. The adjusted R^2 statistics avoid this problem.

The general strategy for comparing modeling procedures <u>over cases</u> involves (a) computing the above-described statistics for each case and summarizing the distributions of such statistics over all cases or over relevant subsets of cases or (b) computing analogous statistics "pooled" over all cases or over relevant subsets of cases. The subsets of primary interest are the following:

- season (i.e., the winter or summer field program),
- prevailing wind speed categories (0-2 mps, 2-4 mps, 4-6 mps, >6 mps),
- prevailing wind direction categories (E & SE, S, SW, other).

The pooled residual variance over an arbitrary subset of cases (say, C) is defined as

$$s^2_{kC}(j,F) = \frac{\sum\limits_{t\varepsilon C} \left[n_t(F)-p_{kt}(j,F)\right] s^2_{kt}(j,F)}{\sum\limits_{t\varepsilon C} \left[n_t(F)-p_{kt}(j,F)\right]} . \tag{18}$$

Note that this is a weighted average of the individual residual variances. The corresponding overall R^2's are obtained as the proportions

$$R^2_{kC}(j,F) = \frac{\sum\limits_{i\varepsilon C} SSE(5) - \sum\limits_{i\varepsilon C} SSE(j)}{\sum\limits_{i\varepsilon C} SSE(5)} . \tag{19}$$

The associated adjusted R^2's are computed as

$$A^2_{kC}(j,F) = 1-(1-R^2_{kC}(j,F)) \frac{\sum\limits_{i\varepsilon C} n_t(F)}{\sum\limits_{i\varepsilon C} \left[n_t(F)-p_{kt}(j,F)\right]} \tag{20}$$

Because the above-described criteria are based on residuals which result from fitting models to the full network data, they cannot provide a thorough evaluation of the modeling procedures. One option, for example, for extending the evaluation would be to apply the modeling procedures to various subsets of the full network obtained by deleting one or several data points and to examine the distributions of residuals occurring at the omitted points. Except for one special case, such a procedure was not employed because of the large number of cases involved. The special case involved selecting the subset of stations to be the RTI network. Then, such a procedure leads to a joint evaluation of the modeling procedures and the RTI network--that is, an overall evaluation of the proposed methodology for estimating wind fields. This is discussed in the following subsection.

CRITERIA FOR EVALUATING THE RTI NETWORK

The remaining six predicted wind fields--those based on data from stations in the RTI network--are utilized for evaluating the performance of the RTI network. This network evaluation can be carried out only in a limited sense. In particular, most of the evaluative measures must be judged in terms of absolute, rather than relative, units, since observed data are available at only a limited number of sites. That is, there is little "feel" for how well some other network of comparable size might have performed. Although comparing the performance of the RTI network with that of the full network is useful in terms of the overall wind prediction capability of the network, it does not provide a separate evaluation of the RTI network. Rather, differences in the evaluative criteria for the two networks generally represent measures which reflect both network and model differences. Such differences do, however, yield a limited comparative evaluation.

The evaluative criteria for this aspect of the evaluation are of two basic types. Both types are based on the deviations between observed and predicted wind component values, where the predicted values are determined by applying the modeling procedures to data from stations <u>in the RTI network</u>.

The first type of criteria, and their formulation and properties, are completely analogous to those described in the previous subsection. These criteria are obtained by using "R" to represent the RTI network and substituting R for F into (7) through (20). Such criteria do not, of course, provide measures of the accuracy of the predictions but simply characterize the predictions over the RTI network itself.

Criteria of the second type do provide such measures of accuracy and, consequently, are regarded as the more important type. Let Q denote some subset of the non-network stations. For instance, Q might represent

(a) the non-network,
(b) the inner non-network,
(c) the outer non-network, or
(d) an individual station within the non-network.

Let C define subsets of cases, as previously described. The accuracy measures are of three basic types:

(1) means of deviations, over C and Q,
(2) means of squared deviations (or the square root thereof), over C and Q, and
(3) frequency distributions of deviations, over C and Q.

The mean deviations are defined, for each modeling procedure (j) and each wind component (k), as

$$\frac{1}{N_{CQ}} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} \hat{e}_{kt}(j,R,i) = \frac{1}{N_{CQ}} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} \left[Z_{kt}(i)-\hat{Z}_{kt}(j,R,i)\right], \quad (21)$$

where $N_{CQ} = \sum_{t\varepsilon C} n_t(Q)$ = the number of observed values occurring at stations in the Q subset and within the set of cases C.

These means represent average biases over the particular subsets of cases and non-network stations. The root mean squared error (RMSE) criteria are determined as the square root of

$$\sum_{t\varepsilon C} \sum_{i\varepsilon Q} \hat{e}^2_{kt}(j,R,i)/N_{CQ} \, . \quad (22)$$

A mean squared error criterion for the Q and C subsets which encompasses errors in both the U- and V-components is obtained by summing (22) over k. This criterion, referred to as the vector mean square error, represents an average of the squared differences between the observed and predicted wind vectors in terms of distances in the (U, V) plane. The vector MSE can be partitioned into two components which represent the mean squared errors in predicting wind speeds and in predicting wind directions:

$$\begin{aligned}
\text{Vector MSE} &= \frac{1}{N_{CQ}} \sum_{k=1}^{2} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} [Z_{kt}(i)-\hat{Z}_{kt}(j,R,i)]^2 \\
&= \frac{1}{N_{CQ}} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} \left\{ \sum_{k=1}^{2} Z^2_{kt}(i) + \sum_{k=1}^{2} \hat{Z}^2_{kt}(j,R,i) \right. \\
&\qquad \left. -2\sum_{k=1}^{2} Z_{kt}(i)\, \hat{Z}_{kt}(j,R,i) \right\} \\
&= \frac{1}{N_{CQ}} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} \left[W_t(i)-\hat{W}_t(j,R,i)\right]^2 \\
&\quad + \frac{2}{N_{CQ}} \sum_{t\varepsilon C} \sum_{i\varepsilon Q} W_t(i)\hat{W}_t(j,R,i) \left\{1-\cos\left[\theta_t(i)-\hat{\theta}_t(j,R,i)\right]\right\} \quad (23)
\end{aligned}$$

where $W_t(i)$ and $\theta_t(i)$ are the observed wind speed and direction, respectively, for case t at station i, and

$\hat{W}_t(j,R,i)$ and $\hat{\theta}_t(j,R,i)$ are the corresponding predicted values based on applying procedure j to data from the RTI network.

The first component in (23) is the MSE associated with predicting wind speeds; the second term is the MSE associated with direction errors.

SECTION 4

EVALUATION RESULTS

Although it would be desirable to have evaluative measures which would isolate the effects of the modeling procedures from those of the network, this is not really feasible--because evaluation of the procedures is conditional on what network is used and because evaluation of the network requires that a given model be employed across a number of alternative networks. Consequently, the evaluation is organized as follows:

(1) Comparing modeling procedures over the full network when data from the full network are used to establish model forms and parameter estimates.

(2) Comparing modeling procedures over the RTI network when data from the RTI network are used to establish the model forms and parameter estimates and comparing results to those of (1) above.

(3) Evaluating the accuracy of predicted results at non-network stations, when estimation has been carried out using data from the RTI network stations.

It should be noted that (1) and (2) above, in contrast to (3), are basically concerned with precision. Also, (1) deals with the criteria discussed in the second subsection of Section 3, whereas (2) and (3) deal with the two types of criteria discussed in the last subsection of that section.

Before discussing the results of these evaluations, it is useful to characterize the modeling procedures involving stepwise regression in terms of the model forms that resulted from applying the algorithms. This is the purpose of the subsection below.

SUMMARY OF SPECIFIC MODELS SELECTED BY ALTERNATIVE APPROACHES

Each of the four stepwise regression procedures (procedures 1-4 of Table 10) was applied to data from the RTI and the full networks. The RLSTEP subroutine of the International Mathematical and Statistical Libraries, Inc. (1975) was utilized to perform the stepwise regressions. Potentially, eight different model forms could result for each specific case and wind component. Table 11 characterizes the models selected by the four procedures in terms of the frequency with which various size models result. This table demonstrates that models containing more than four or five terms are rarely selected. As expected, models from procedure 4 are larger than those from procedure 3; similarly, models from procedure 2 are larger than those from procedure 1. The pattern of these distributions is similar for both the U- and V-components; however, smaller size models are much more frequent for the U-component. When the same modeling procedure is applied to the two different networks, there is a tendency for the full network cases to yield slightly larger models; this, of course, is not surprising because of the increase in statistical power which results from the larger number of stations used in the full-network estimations.

Table 11 also indicates the number of times (out of 908 cases) that flat-surface models are chosen (i.e., the number of 1-term, constant models). This is summarized, in terms of percentages, below:

Percentage of Cases in Which One-Term Models are Selected

Modeling Procedure	RTI Network Data		Full Network Data	
	U	V	U	V
1	36.2	15.2	30.4	10.2
2	15.1	5.5	11.2	4.0
3	23.3	10.1	20.5	7.8
4	8.6	3.3	4.8	2.5

TABLE 11. DISTRIBUTION OF CASES BY MODEL SIZE--FOR FOUR MODELING PROCEDURES APPLIED TO WIND COMPONENT DATA FROM STATIONS IN THE RTI NETWORK AND THE FULL NETWORK

No. of Terms in Selected Model	Using Data From RTI Network, Modeling Procedure 1	2	3	4	Using Data From Full Network, Modeling Procedure 1	2	3	4
U-Component								
1	329	137	257	78	276	102	186	44
2	362	273	328	200	348	231	302	148
3	136	223	190	231	161	233	227	243
4	54	119	81	169	64	144	112	193
5	19	86	24	79	48	110	52	112
6	6	33	14	59	7	45	16	72
7	2	21	9	37	4	23	7	43
8	0	6	2	22	0	15	2	29
9	0	6	2	19	0	1	3	12
10	0	3	1	9	0	4	1	5
11	0	1	0	2	0	0	0	6
12	0	0	0	3	0	0	0	1
V-Component								
1	138	50	92	30	93	36	71	23
2	287	127	234	74	237	84	189	59
3	263	238	277	167	318	187	267	155
4	155	193	172	184	162	233	220	190
5	43	149	82	160	62	173	98	175
6	13	79	33	112	27	122	49	133
7	5	35	12	76	6	48	9	94
8	2	25	5	40	2	19	3	45
9	2	8	1	32	1	5	2	19
10	0	4	0	21	0	1	0	12
11	0	0	0	4	0	0	0	0
12	0	0	0	5	0	0	0	1
13	0	0	0	0	0	0	0	2
14	--	--	0	1	--	--	0	0
15	--	--	0	1	--	--	0	0
16	--	--	0	1	--	--	0	0

These percentages indicate how frequently each procedure yields a model like the procedure 5 model. These cases are important in that no variation is accounted for by such models (i.e., $R^2=0$).

Table 12 provides pairwise comparisons of the network/modeling procedures in terms of their model sizes and model forms. The two methods involved in a comparison (denoted by method A and method B in Table 12) can differ in several ways, as indicated below:

Type of Comparison	Initial Class of Model Forms	Stepwise Regression Parameters	Network Used as Data Base
I	Different	Same	Same
II	Same	Different	Same
III	Different	Different	Same
IV	Same	Same	Different
V	Different	Same	Different
VI	Same	Different	Different
VII	Different	Different	Different

As might be expected, similar size and similar form models occur more frequently when the two methods being compared are more alike -- for example, types I, II, and IV as compared to type VII. Table 13 shows the results of Table 12 relating to the similarity of model forms in terms of percentages.

Out of the 908 cases, the number of times that each of the 23 model terms occurred in a selected model is shown in Table 14. Because many of the potential terms are highly correlated, the inclusion of a particular term in a model is highly dependent on what other terms are involved in the model; also, there are likely to be many models with essentially the same predictive capability. Thus the results of Table 14 merely provide a descriptive summary of the selected models and should be so interpreted.

TABLE 12. PAIRWISE COMPARISONS OF MODELING PROCEDURES IN TERMS OF MODEL SIZES AND MODEL FORMS

Type of Comparison	Methods+ A	Methods+ B	U-Component: Number of Cases With: Method A Model Smaller				V-Component: Number of Cases With: Method A Model Smaller			
			Method A Model Smaller	Method B Model Smaller	Same Model Size	Same Model Form	Method A Model Smaller	Method B Model Smaller	Same Model Size	Same Model Form
I	1R	3R	212	35	661	515	258	67	583	401
	2R	4R	367	91	451	263	426	123	359	167
	1F	3F	262	47	599	422	254	51	603	412
	2F	4F	393	119	396	192	367	123	418	187
II	1R	2R	488	0	420	420	536	0	372	369
	3R	4R	566	2*	340	336	603	2*	303	296
	1F	2F	516	0	392	388	568	0	340	335
	3F	4F	573	0	335	329	585	0	323	308
III	1R	4R	662	14	232	172	678	14	216	122
	3R	2R	410	95	403	258	457	97	354	186
	1F	4F	691	14	203	125	688	19	201	121
	3F	2F	419	93	396	217	476	74	358	172
IV	1R	1F	212	112	578	511	267	156	485	390
	2R	2F	281	162	465	337	331	224	353	225
	3R	3F	265	140	503	405	296	206	406	265
	4R	4F	333	215	360	223	318	314	276	119
V	1R	3F	398	77	433	278	401	126	381	221
	2R	4F	486	128	294	104	486	198	224	67
	1F	3R	248	186	474	322	291	229	388	204
	2F	4R	353	226	329	119	419	256	233	52
VI	1R	2F	566	22	320	262	617	45	246	177
	3R	4F	640	41	227	154	616	70	222	116
	1F	2R	431	36	441	365	490	80	338	250
	3F	4R	483	64	361	256	570	69	269	157
VII	1R	4F	722	18	168	94	708	42	158	76
	3R	2F	506	96	306	166	543	112	253	80
	1F	4R	608	44	256	146	648	62	198	87
	3F	2R	343	170	395	194	429	174	305	137

\+ The notation 1R means modeling procedure 1 applied to the RTI network; similarly, 2F means procedure 2 using data from the full network of stations. See Table 10 for definitions of modeling procedures.

* This number is not zero, because procedure 4R failed to meet round-off tolerances in these cases and the "selected" model from procedure 4R was <u>defined</u> to be the constant, one-term model.

TABLE 13. PERCENTAGE OF 908 CASES IN WHICH NETWORK/MODELING PROCEDURES RESULTED IN THE SAME MODEL FORM

U-Component

Method	2R	3R	4R	1F	2F	3F	4F
1R	46.3	56.7	18.9	56.3	28.9	30.6	10.4
2R		28.4	29.0	40.2	37.1	21.4	11.5
3R			37.0	35.5	18.3	44.6	17.0
4R				16.1	13.1	28.2	24.6
1F					42.7	46.5	13.8
2F						23.9	21.1
3F							36.2

V-Component

Method	2R	3R	4R	1F	2F	3F	4F
1R	40.6	44.2	13.4	43.0	19.5	24.3	8.4
2R		20.5	18.4	27.5	24.8	15.1	7.4
3R			32.6	22.5	8.8	29.2	12.8
4R				9.6	5.7	17.3	13.1
1F					36.9	45.4	13.3
2F						18.9	20.6
3F							33.9

TABLE 14. NUMBER OF CASES FOR WHICH SPECIFIC MODEL TERMS ARE SELECTED, BY WIND COMPONENT, MODELING PROCEDURE AND NETWORK

Term Number*	Models for U-Component Predictions								Models for V-Component Predictions							
	Using Data From RTI Network				Using Data From Full Network				Using Data From RTI Network				Using Data From Full Network			
	Modeling Procedure				Modeling Procedure				Modeling Procedure				Modeling Procedure			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	908	908	908	908	908	908	908	908	908	908	908	908	908	908	908	908
2	97	168	75	132	121	192	105	158	389	506	272	355	460	537	392	463
3	74	135	64	106	88	156	91	144	120	193	96	193	107	189	97	175
4	34	98	28	76	75	171	56	119	39	135	31	79	66	152	41	99
5	140	247	123	195	151	249	114	152	187	290	154	252	177	270	136	193
6	53	129	40	116	77	172	57	125	75	206	54	133	83	222	78	132
7	74	175	27	74	97	193	40	99	79	209	39	151	123	270	71	151
8	62	119	29	67	58	115	22	50	91	169	39	85	98	185	36	83
9	34	112	42	112	48	133	61	132	104	213	88	176	156	276	143	238
10	138	212	117	166	107	188	94	151	242	279	245	287	193	229	179	219
11	42	111	8	28	62	129	12	38	38	96	10	50	49	133	14	42
12	68	136	38	74	99	162	37	62	98	187	46	95	129	226	70	110
13	98	196	71	142	130	221	89	150	113	175	86	135	159	240	124	186
14			48	112			63	129			51	119			44	114
15			38	95			45	98			44	107			34	98
16			78	149			89	169			54	130			71	139
17			38	65			37	60			58	125			65	118
18			38	83			51	100			29	74			45	99
19			35	62			47	96			64	122			67	116
20			125	220			168	244			126	206			137	189
21			54	109			70	126			186	272			98	183
22			41	126			63	156			89	195			119	222
23			53	133			58	136			51	137			50	109

* Terms are assumed to be ordered as in definition (5).

COMPARISON OF ALTERNATIVE MODELING PROCEDURES USING WIND DATA FROM ALL STATIONS

Overall analyses of the data from the summer and winter field programs are shown in Table 15. For these analyses, all data from each season are pooled together. The total sums of squares among the 254 summertime cases and the 654 wintertime cases are each partitioned into a between-case and a within-case component. The various modeling procedures are applied on a case-by-case basis and can therefore have no effect on the between-case component of variation. The within-case component, which corresponds to fitting a constant for each case (i.e., using modeling procedure 5), can then be partitioned into a "pooled regression" and a "pooled residual" component for each of the modeling procedures. Only the latter of these two components is actually shown in Table 15.

The results shown in Table 15 are utilized to compute values of the "pooled" criteria described in the second subsection of Section 3. These results are presented in Table 16 for each of the two seasons. As expected, all of the stepwise procedures result in smaller pooled residual standard deviations and larger adjusted R^2's than either procedure 0 or procedure 5. In terms of these measures of overall precision, modeling procedure 4 is clearly superior for both wind components and both seasons. Modeling procedure 2 yields models which achieve, on average, virtually the same precision as modeling procedure 3. However, it requires an average of about one more term per case than procedure 3 requires. On the average, procedure 1 involves fewer terms than the other stepwise procedures and, in terms of the pooled precision measures, appears the least favorable among the four stepwise procedures.

TABLE 15. SUMMARY OF ANALYSIS OF VARIANCE RESULTS BASED ON ESTIMATIONS FROM THE FULL NETWORK

Source of Variation*	Summer Field Program			
	Degrees of Freedom		Mean Squares (mps^2)	
	U	V	U	V
Total	5500	5500	1.3928	1.9584
Between Cases	253	253	16.7408	21.9731
Within Cases	5247	5247	0.6528	0.9933
Residual (0)	2199	2199	0.6279	0.7480
Residual (1)	4942	4773	0.5574	0.6770
Residual (2)	4628	4450	0.5033	0.5996
Residual (3)	4829	4708	0.5029	0.6373
Residual (4)	4490	4327	0.4495	0.5437

Source of Variation*	Winter Field Program			
	Degrees of Freedom		Mean Squares (mps^2)	
	U	V	U	V
Total	14456	14456	5.2119	4.5891
Between Cases	653	653	94.9021	64.3676
Within Cases	13803	13803	0.9688	1.7610
Residual (0)	5955	5955	0.9316	0.8924
Residual (1)	12995	12477	0.7289	0.8471
Residual (2)	12341	11671	0.6721	0.7447
Residual (3)	12752	12231	0.6865	0.7750
Residual (4)	11866	11245	0.6095	0.6599

* The notation "Residual (j)" means the pooled residual variation from fitting models determined by modeling procedure j. It should be noted that "Within-Cases" is equivalent to "Residual (5)".

TABLE 16. VALUES OF POOLED EVALUATIVE CRITERIA BY SEASON, WIND COMPONENT AND MODELING PROCEDURE BASED ON THE FULL NETWORK ESTIMATIONS

Statistic	Wind Component	Season*	Modeling Procedure 0	1	2	3	4	5
Average No. of Model Terms (intercept included)	U	S	13.0	2.2	3.4	2.6	4.0	1.0
		W	13.0	2.2	3.2	2.6	4.0	1.0
	V	S	13.0	2.9	4.1	3.1	4.6	1.0
		W	13.0	3.0	4.3	3.4	4.9	1.0
Pooled Residual Std. Dev. (mps)	U	S	0.79	0.75	0.71	0.71	0.67	0.81
		W	0.97	0.85	0.82	0.83	0.78	0.98
	V	S	0.86	0.82	0.77	0.80	0.74	1.00
		W	0.94	0.92	0.86	0.88	0.81	1.33
Pooled R^2	U	S	0.60	0.20	0.32	0.29	0.41	0.00
		W	0.59	0.29	0.38	0.35	0.46	0.00
	V	S	0.68	0.38	0.49	0.42	0.55	0.00
		W	0.78	0.57	0.64	0.61	0.69	0.00
Pooled Adjusted R^2	U	S	0.04	0.15	0.23	0.23	0.31	0.00
		W	0.04	0.25	0.31	0.29	0.37	0.00
	V	S	0.25	0.32	0.40	0.36	0.45	0.00
		W	0.49	0.52	0.58	0.56	0.63	0.00

* S = summer field program;
W = winter field program.

The following tendencies should also be noted:

(a) there is greater within-case variation from station-to-station in the V-component,

(b) there is greater total variation in both components in the winter than in the summer,

(c) a smaller percentage of the total variation is accounted for by the models in the summer than in the winter for both the U- and V-components.

Tables 17 and 18 present, respectively, the distributions of the residual standard deviations and the distributions of the adjusted R^2 values. These distributions are based on all 908 cases. Particular note should be made of the similarity of the distributions for procedures 2 and 3 in both of these tables.

Tables 19 and 20 show distributions of the individual residuals resulting from the full-network estimations. Table 19 clearly indicates the smaller summertime variation, as compared to that of the wintertime. Table 20 combines the distributions of Table 19 over seasons. In addition, the distributions of deviations between observed and predicted wind speeds are shown. Large positive deviations in the wind speeds appear more frequently than large negative deviations, indicating (when such errors occur) a tendency toward underprediction of the wind speeds. The majority of the wind speed residuals, however, are less than 1.5 mps, as shown by the percentages below that are derived directly from Table 20:

Modeling Procedure	Percentage of Observations With: $\lvert W-\hat{W}\rvert \leq 1.5$ mps	$\lvert W-\hat{W}\rvert > 1.5$ mps
1	93.62	6.38
2	95.45	4.55
3	94.24	5.76
4	96.20	3.80
5	83.54	16.46

TABLE 17. DISTRIBUTIONS OF RESIDUAL STANDARD DEVIATIONS OVER THE 908 CASES FOR FOUR MODELING PROCEDURES APPLIED TO DATA FROM ALL STATIONS

Residual Std. Dev. (mps)	Percentage Frequency Distributions Modeling Procedure 1	2	3	4	Cumulative Percentages Modeling Procedure 1	2	3	4
				U-Component				
0.0 - 0.5	14.5	18.6	18.0	24.2	14.5	18.6	18.0	24.2
0.5 - 1.0	66.9	66.9	66.1	65.0	81.4	85.5	84.0	89.2
1.0 - 1.5	18.0	14.1	15.4	10.4	99.3	99.6	99.4	99.6
1.5 - 2.0	0.7	0.4	0.6	0.4	100.0	100.0	100.0	100.0
				V-Component				
0.0 - 0.5	5.5	8.0	7.2	13.8	5.5	8.0	7.2	13.8
0.5 - 1.0	68.9	72.5	71.8	72.1	74.4	80.5	79.0	85.9
1.0 - 1.5	24.1	19.1	20.4	14.0	98.6	99.6	99.3	99.9
1.5 - 2.0	1.4	0.4	0.7	0.1	100.0	100.0	100.0	100.0

TABLE 18. DISTRIBUTIONS OF ADJUSTED R^2 STATISTICS OVER THE 908 CASES FOR FOUR MODELING PROCEDURES APPLIED TO DATA FROM ALL STATIONS

Adjusted R^2	Percentage Frequency Distributions				Cumulative Percentages			
	Modeling Procedure				Modeling Procedure			
	1	2	3	4	1	2	3	4
				U-Component				
0.0	30.4	11.2	20.5	4.8	30.4	11.2	20.5	4.8
0.0 - 0.2	30.3	36.2	27.6	26.9	60.7	47.5	48.1	31.7
0.2 - 0.4	23.7	29.0	29.0	32.7	84.4	76.4	77.1	64.4
0.4 - 0.6	10.2	14.9	14.3	20.2	94.6	91.3	91.4	84.6
0.6 - 0.8	5.1	8.0	7.5	12.0	99.7	99.3	98.9	96.6
0.8 - 1.0	0.3	0.7	1.1	3.4	100.0	100.0	100.0	100.0
				V-Component				
0.0	10.2	4.0	7.8	2.5	10.2	4.0	7.8	2.5
0.0 - 0.2	12.6	10.5	11.0	8.6	22.8	14.4	18.8	11.1
0.2 - 0.4	24.1	21.3	21.8	17.6	46.9	35.7	40.6	28.7
0.4 - 0.6	30.7	32.2	30.3	29.0	77.6	67.8	70.9	57.7
0.6 - 0.8	19.7	26.8	24.8	30.5	97.4	94.6	95.7	88.2
0.8 - 1.0	2.6	5.4	4.3	11.8	100.0	100.0	100.0	100.0

TABLE 19. PERCENTAGE FREQUENCY DISTRIBUTIONS OF RESIDUALS BY SEASON, WIND COMPONENT, AND MODELING PROCEDURES OVER ALL STATIONS BASED ON FULL NETWORK ESTIMATIONS

Summer Field Program (5501 Observations)

Wind Comp.	Modeling Procedure	Deviation Between Observed and Predicted Value (midpt. of interval) ≤-5	-4	-3	-2	-1	0	1	2	3	4	≥5
U	1	--	0.04	0.11	1.75	18.21	61.23	16.05	2.16	0.44	0.02	--
	2	--	0.04	0.07	1.42	16.11	65.53	14.72	1.80	0.29	0.02	--
	3	--	0.04	0.09	1.51	16.12	65.12	15.02	1.75	0.35	0.02	--
	4	--	0.04	0.07	1.25	13.65	70.57	12.60	1.53	0.27	0.02	--
	5	--	0.05	0.16	2.65	20.85	54.01	18.91	2.80	0.53	0.04	--
V	1	--	--	0.22	2.16	21.20	54.99	18.00	3.04	0.38	0.02	--
	2	--	--	0.16	1.45	19.14	59.86	16.60	2.51	0.27	--	--
	3	--	--	0.18	1.78	20.45	56.95	17.25	3.02	0.35	0.02	--
	4	--	--	0.15	1.15	17.31	63.95	14.96	2.27	0.22	--	--
	5	--	0.02	0.78	5.33	21.61	45.36	21.12	4.29	1.38	0.09	0.02

Winter Field Program (14457 Observations)

Wind Comp.	Modeling Procedure											
U	1	0.01	0.04	0.32	2.91	20.70	52.32	19.92	3.42	0.28	0.08	0.01
	2	0.01	0.03	0.30	2.46	19.10	56.51	18.52	2.78	0.22	0.08	0.01
	3	0.01	0.04	0.32	2.68	19.33	55.50	18.84	2.94	0.27	0.06	0.01
	4	0.01	0.02	0.24	2.19	16.74	61.38	16.99	2.17	0.19	0.06	0.01
	5	0.02	0.06	0.64	4.57	22.65	45.53	20.43	4.89	0.93	0.27	0.01
V	1	--	0.04	0.26	3.33	22.43	48.77	20.88	3.72	0.47	0.09	--
	2	--	0.01	0.18	2.66	20.43	54.08	19.36	2.89	0.35	0.04	--
	3	--	0.04	0.18	2.87	21.33	51.97	19.89	3.25	0.39	0.08	--
	4	--	0.03	0.10	2.13	18.32	59.29	17.36	2.49	0.26	0.03	--
	5	0.22	0.48	1.61	7.40	24.85	35.07	18.27	8.13	3.09	0.73	0.14

TABLE 20. PERCENTAGE FREQUENCY DISTRIBUTIONS OF RESIDUALS BY WIND COMPONENT AND MODELING PROCEDURE OVER ALL CASES AND ALL STATIONS BASED ON FULL NETWORK ESTIMATIONS

Wind Comp.*	Modeling Procedure	Deviation Between Observed and Predicted Value (midpt. of interval) ≤-5	-4	-3	-2	-1	0	1	2	3	4	≥5
U	1	0.01	0.04	0.26	2.59	20.02	54.78	18.85	3.07	0.33	0.06	0.01
	2	0.01	0.03	0.24	2.17	18.28	58.99	17.47	2.51	0.24	0.06	0.01
	3	0.01	0.04	0.26	2.36	18.44	58.15	17.78	2.61	0.29	0.05	0.01
	4	0.01	0.03	0.20	1.93	15.89	63.91	15.78	1.99	0.22	0.05	0.01
	5	0.02	0.06	0.51	4.04	22.15	47.87	20.01	4.31	0.82	0.21	0.01
V	1	--	0.03	0.25	3.01	22.09	50.49	20.09	3.53	0.45	0.07	--
	2	--	0.01	0.18	2.32	20.07	55.68	18.60	2.79	0.33	0.03	--
	3	--	0.03	0.18	2.57	21.09	53.34	19.16	3.19	0.38	0.07	--
	4	--	0.02	0.11	1.86	18.04	60.57	16.70	2.43	0.25	0.02	--
	5	0.16	0.36	1.38	6.83	23.96	37.90	19.06	7.07	2.62	0.56	0.11
W	1	--	0.01	0.09	1.86	18.45	52.87	22.30	3.92	0.41	0.09	0.01
	2	--	--	0.04	1.25	16.69	58.25	20.51	2.96	0.27	0.03	0.01
	3	--	0.01	0.08	1.60	17.75	55.15	21.34	3.61	0.38	0.07	0.01
	4	--	--	0.02	0.94	15.28	62.34	18.58	2.61	0.22	0.02	--
	5	0.01	0.06	0.49	4.43	22.94	39.78	20.82	7.70	2.98	0.68	0.11

* W denotes wind speed.

COMPARISON OF ALTERNATIVE MODELING PROCEDURES USING WIND DATA FROM STATIONS IN THE RTI NETWORK

The modeling procedures applied to data from the RTI network stations yield similar results to those described in the previous subsection. Tables 21 through 24, which are analogous to Tables 15 through 18, respectively, provide a summary of the major results. It should again be emphasized that these results, like those of the preceding subsection, relate to the precision of the modeling procedures rather than to their accuracy.

Comparison of these results to those of the previous subsection indicates that in general the criteria values based on the RTI network estimations are slightly less consistent than those for the full network. More specifically, the results can be summarized as follows:

(a) In the winter, the within-case variation over the RTI network stations is somewhat larger than the within-case variation over all stations -- for both wind components; in the summer, the within-case variation for the V-component is smaller for the RTI network than for the full network.

(b) For the V-component, the residual variances over the RTI network are usually smaller than the corresponding quantities for the full network; for the U-component, they are about the same in the summer and larger in the winter than the corresponding full-network residual variances.

(c) Among the stepwise regression procedures, the pooled adjusted R^2 statistics from the RTI network estimations are quite comparable to those of the full network; the distributions of the adjusted R^2 statistics show that more large <u>and</u> more small adjusted R^2 values occur for the RTI-network estimations than for the full-network estimations.

TABLE 21. SUMMARY OF ANALYSIS OF VARIANCE RESULTS BASED ON ESTIMATIONS FROM THE RTI NETWORK

Source of Variation*	Summer Field Program			
	Degrees of Freedom		Mean Squares (mps^2)	
	U	V	U	V
Total	4017	4017	1.2919	1.7273
Between Cases	253	253	10.5446	13.9551
Within Cases	3764	3764	0.6700	0.9044
Residual (0)	716	716	0.5483	0.6147
Residual (1)	3515	3421	0.5419	0.6324
Residual (2)	3200	3096	0.4909	0.5380
Residual (3)	3434	3294	0.5041	0.5529
Residual (4)	3131	2747	0.4500	0.4175

Source of Variation*	Winter Field Program			
	Degrees of Freedom		Mean Squares (mps^2)	
	U	V	U	V
Total	10671	10671	5.2568	4.8337
Between Cases	653	653	69.4364	49.3675
Within Cases	10018	10018	1.0734	1.9309
Residual (0)	2170	2170	1.2154	0.8338
Residual (1)	9353	8786	0.8168	0.8065
Residual (2)	8744	8028	0.7411	0.6861
Residual (3)	9138	8576	0.7611	0.7199
Residual (4)	8209	7557	0.6635	0.5898

* The notation "Residual (j)" means the pooled residual variation from fitting models determined by modeling procedure j. It should be noted that "Within Cases" is equivalent to "Residual (5)".

TABLE 22. VALUES OF POOLED EVALUATIVE CRITERIA BY SEASON, WIND COMPONENT, AND MODELING PROCEDURE BASED ON RTI NETWORK ESTIMATIONS

Statistic	Wind Component	Season*	Modeling Procedure 0	1	2	3	4	5
Average No. of Model Terms (intercept included)	U	S	13.0	2.0	3.2	2.3	3.5	1.0
		W	13.0	2.0	2.9	2.3	3.8	1.0
	V	S	13.0	2.4	3.6	2.9	5.0	1.0
		W	13.0	2.9	4.0	3.2	4.8	1.0
Pooled Residual Std. Dev. (mps)	U	S	0.74	0.74	0.70	0.71	0.67	0.82
		W	1.10	0.90	0.86	0.87	0.81	1.04
	V	S	0.78	0.80	0.73	0.74	0.65	0.95
		W	0.91	0.90	0.83	0.85	0.77	1.39
Pooled R^2	U	S	0.84	0.24	0.38	0.32	0.44	0.00
		W	0.75	0.29	0.40	0.35	0.49	0.00
	V	S	0.87	0.37	0.51	0.47	0.66	0.00
		W	0.91	0.63	0.72	0.68	0.77	0.00
Pooled Adjusted R^2	U	S	0.18	0.19	0.27	0.25	0.33	0.00
		W	0.00	0.24	0.31	0.29	0.38	0.00
	V	S	0.32	0.30	0.41	0.39	0.54	0.00
		W	0.57	0.58	0.64	0.63	0.69	0.00

* S = summer field program;
W = winter field program.

TABLE 23. DISTRIBUTIONS OF RESIDUAL STANDARD DEVIATIONS OVER THE 908 CASES FOR FOUR MODELING PROCEDURES APPLIED TO DATA FROM STATIONS IN THE RTI NETWORK

Residual Std. Dev.	Percentage Frequency Distributions: Modeling Procedure 1	2	3	4	Cumulative Percentages: Modeling Procedure 1	2	3	4
				U-Component				
0.0 - 0.5	15.3	19.7	18.9	28.1	15.3	19.7	18.9	28.1
0.5 - 1.0	61.3	61.5	60.7	56.6	76.7	81.2	79.6	84.7
1.0 - 1.5	21.4	17.7	19.2	14.4	98.0	98.9	98.8	99.1
1.5 - 2.0	2.0	1.1	1.2	0.9	100.0	100.0	100.0	100.0
				V-Component				
0.0 - 0.5	10.1	15.5	15.4	31.6	10.1	15.5	15.4	31.6
0.5 - 1.0	67.9	68.8	66.4	57.6	78.0	84.4	81.8	89.2
1.0 - 1.5	20.5	15.1	17.4	10.2	98.5	99.4	99.2	99.4
1.5 - 2.0	1.5	0.6	0.8	0.6	100.0	100.0	100.0	100.0

TABLE 24. DISTRIBUTIONS OF ADJUSTED R^2 STATISTICS OVER THE 908 CASES FOR FOUR MODELING PROCEDURES APPLIED TO DATA FROM STATIONS IN THE RTI NETWORK

Adjusted R^2	Percentage Frequency Distributions				Cumulative Percentages			
	Modeling Procedure				Modeling Procedure			
	1	2	3	4	1	2	3	4
				U-Component				
0.0	36.2	15.1	28.3	8.8	36.2	15.1	28.3	8.6
0.0 - 0.2	23.8	31.4	21.5	24.2	60.0	46.5	49.8	32.8
0.2 - 0.4	21.0	24.2	24.3	26.2	81.1	70.7	74.1	59.0
0.4 - 0.6	12.4	17.4	15.3	19.8	93.5	88.1	89.4	78.9
0.6 - 0.8	5.4	8.7	7.8	12.7	98.9	96.8	97.2	91.5
0.8 - 1.0	1.1	3.2	2.8	8.5	100.0	100.0	100.0	100.0
				V-Component				
0.0	15.2	5.5	10.1	3.3	15.2	5.5	10.1	3.3
0.0 - 0.2	9.9	11.5	9.1	7.3	25.1	17.0	19.3	10.6
0.2 - 0.4	19.1	16.2	15.6	10.2	44.2	33.1	34.9	20.8
0.4 - 0.6	24.2	23.7	24.0	21.6	68.4	56.8	58.9	42.4
0.6 - 0.8	24.7	29.3	27.3	27.6	93.1	86.1	86.2	70.0
0.8 - 1.0	6.9	13.9	13.8	30.0	100.0	100.0	100.0	100.0

ACCURACY OF PREDICTED WIND FIELDS

Evaluation of the accuracy of the modeling procedures depends upon the deviations between observed and predicted values at the non-network stations, when the estimation is based on data from the RTI network. The means of these deviations by wind component are shown in Table 25 for each of the seven non-network stations; for comparative purposes, the corresponding mean deviations resulting from the full network estimations are given in Table 26. It is apparent from these results that the largest discrepancies between the mean deviations of the two tables occur for the outer-non-network stations STL007 and EPA103. Except for these two stations, the corresponding mean deviations of Tables 25 and 26 usually differ by less that 0.1 mps.

Pooled root mean square errors (RMSE's) at each of the non-network stations are presented in Table 27. These are shown for each wind component and season; the pooled vector RMSE's, denoted by (U,V), provide a convenient method of summarizing the errors over the two components, as described at the end of Section 3. In order to evaluate the magnitude of the errors occurring at the non-network stations, the pooled vector RMSE's shown in the last five rows of Table 27 are plotted in Figure 2 along with the corresponding RMSE's for the RTI network stations. Root mean square errors based on the full-network (F) estimations are also shown. These appear to the left of each vertical line, whereas those based on the RTI network data are shown on the right. This plot clearly demonstrates the trend of decreasing RMSE's for the full-network estimations when going from procedure 1 to procedure 4 and the similar trend <u>for RTI network stations</u> based on the RTI network estimations. The greatest improvement in precision in going from proce-

TABLE 25. MEANS OF DEVIATIONS BETWEEN OBSERVED AND PREDICTED VALUES AT NON-NETWORK STATIONS, BY WIND COMPONENT AND MODELING PROCEDURE--BASED ON ESTIMATIONS FROM RTI NETWORK DATA

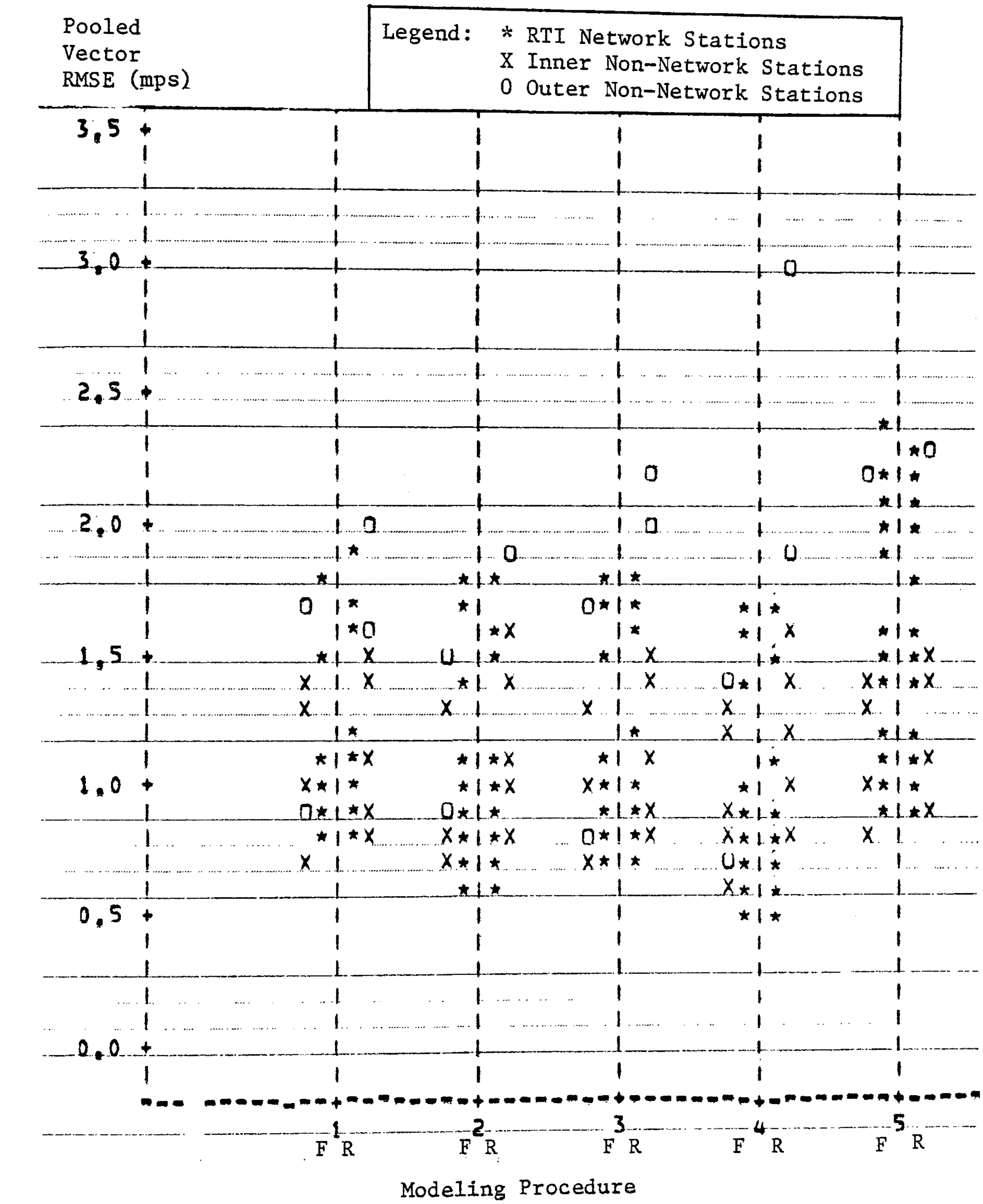
Wind Comp.	Modeling Procedure	Station						
		STL003	STL004	STL007	EPA103	EPA107	EPA111	EPA112
U (mps)	1	0.543	-0.395	-0.700	0.328	-0.044	-0.519	-0.202
	2	0.526	-0.440	-0.761	0.310	-0.083	-0.517	-0.239
	3	0.520	-0.418	-0.560	0.314	-0.067	-0.549	-0.238
	4	0.483	-0.466	-0.612	0.306	-0.113	-0.585	-0.308
	5	0.566	-0.338	-0.428	0.332	0.000	-0.542	-0.138
V (mps)	1	-1.663	0.452	0.414	1.387	-0.146	0.210	0.175
	2	-0.693	0.531	0.558	1.232	-0.150	0.248	0.272
	3	-0.678	0.429	0.833	1.380	-0.182	0.225	0.193
	4	-0.733	0.506	1.273	1.228	-0.210	0.279	0.307
	5	-0.702	0.167	-0.215	1.709	-0.263	0.168	-0.197

TABLE 26. MEANS OF DEVIATIONS BETWEEN OBSERVED AND PREDICTED VALUES AT NON-NETWORK STATIONS, BY WIND COMPONENT AND MODELING PROCEDURE--BASED ON ESTIMATIONS FROM FULL NETWORK DATA

Wind Comp.	Modeling Procedure	Station						
		STL003	STL004	STL007	EPA103	EPA107	EPA111	EPA112
U (mps)	1	0.545	-0.367	-0.416	0.292	-0.030	-0.468	-0.171
	2	0.510	-0.387	-0.359	0.258	-0.078	-0.430	-0.186
	3	0.492	-0.384	-0.229	0.261	-0.069	-0.479	-0.200
	4	0.452	-0.381	-0.173	0.193	-0.102	-0.466	-0.221
	5	0.571	-0.333	-0.433	0.331	0.008	-0.519	-0.133
V (mps)	1	-0.716	0.392	0.273	1.126	-0.216	0.144	0.143
	2	-0.720	0.461	0.285	0.951	-0.205	0.164	0.233
	3	-0.683	0.387	0.328	1.131	-0.207	0.157	0.165
	4	-0.664	0.451	0.291	0.913	-0.190	0.209	0.255
	5	-0.731	0.138	-0.301	1.667	-0.290	0.180	-0.227

TABLE 27. ROOT MEAN SQUARE ERRORS (MPS) FOR EACH NON-NETWORK STATION BASED ON ESTIMATIONS FROM THE RTI NETWORK, BY WIND COMPONENT, SEASON, AND MODELING PROCEDURE

Wind Comp.	Season	Modeling Procedure	Station STL003	STL004	STL007	EPA103	EPA107	EPA111	EPA112	All Non-Network Stations	Inner Non-Network Stations
U	Summer	1	0.960	1.109	1.017	0.870	0.426	--	0.516	0.863	0.817
		2	0.927	1.090	1.438	0.848	0.443	--	0.581	0.955	0.813
		3	0.942	1.113	1.034	0.869	0.426	--	0.542	0.867	0.817
		4	0.925	1.104	1.311	0.876	0.443	--	0.607	0.933	0.822
		5	1.015	1.073	0.678	0.901	0.435	--	0.469	0.810	0.816
U	Winter	1	0.849	0.973	1.359	1.014	0.494	0.783	0.663	0.833	0.772
		2	0.850	1.041	1.382	0.981	0.493	0.781	0.696	0.846	0.795
		3	0.842	0.991	1.790	1.030	0.487	0.817	0.672	0.864	0.782
		4	0.841	1.069	2.001	1.021	0.510	0.867	0.767	0.911	0.833
		5	1.859	0.903	1.461	1.112	0.571	0.794	0.641	0.851	0.765
V	Summer	1	0.774	1.118	1.120	1.937	0.501	--	0.509	1.119	0.777
		2	0.817	1.075	1.363	1.755	0.521	--	0.541	1.116	0.781
		3	0.837	1.051	1.946	1.871	0.510	--	0.493	1.279	0.768
		4	0.990	1.027	2.847	1.678	0.566	--	0.534	1.521	0.821
		5	0.807	1.099	0.788	2.231	0.444	--	0.453	1.156	0.762
V	Winter	1	1.220	1.183	1.088	1.689	0.659	0.771	0.704	1.095	0.944
		2	1.202	1.259	1.147	1.582	0.641	0.786	0.780	1.092	0.971
		3	1.193	1.164	1.265	1.695	0.639	0.795	0.681	1.092	0.930
		4	1.191	1.238	1.592	1.609	0.635	0.823	0.777	1.109	0.968
		5	1.245	0.989	1.589	1.971	0.729	0.770	0.725	1.163	0.918
(U,V)	Summer	1	1.233	1.575	1.513	2.124	0.657	--	0.725	1.413	1.127
		2	1.236	1.531	1.981	1.949	0.684	--	0.794	1.469	1.128
		3	1.260	1.531	2.204	2.063	0.665	--	0.733	1.545	1.122
		4	1.355	1.508	3.135	1.893	0.719	--	0.808	1.784	1.162
		5	1.297	1.536	1.040	2.406	0.621	--	0.652	1.412	1.116
(U,V)	Winter	1	1.486	1.532	1.741	1.970	0.824	1.099	0.967	1.376	1.220
		2	1.472	1.633	1.796	1.861	0.809	1.108	1.045	1.381	1.255
		3	1.460	1.529	2.192	1.984	0.804	1.140	0.956	1.392	1.215
		4	1.458	1.636	2.558	1.906	0.815	1.196	1.092	1.436	1.277
		5	1.512	1.339	2.158	2.263	0.926	1.106	0.967	1.441	1.195
(U,V)	Combined	1	1.419	1.544	1.579	2.018	0.782	1.099	0.907	1.387	1.198
		2	1.409	1.605	1.932	1.888	0.777	1.108	0.982	1.406	1.226
		3	1.406	1.529	2.201	2.008	0.768	1.140	0.900	1.437	1.194
		4	1.429	1.601	2.988	1.902	0.790	1.196	1.022	1.542	1.250
		5	1.454	1.398	1.436	2.308	0.853	1.106	0.892	1.433	1.177



NOTE: Some observations are not shown since computer would not overprint.

(F = Full network
R = RTI network)

Figure 2. Pooled vector RMSE's for individual stations by modeling procedure

dure 1 to procedure 4 appears to occur for stations in the outlying areas of the St. Louis region.

Also apparent from Figure 2 is the increase in the RMSE's for the non-network stations based on the RTI network estimations over the corresponding values for the full network estimations. These increases tend to be most pronounced for procedures 3 and 4 and are quite dramatic for the outer-non-network station, STL007. Thus, among the non-network stations, there is a general trend of increasing RMSE's in going from procedure 1 to procedure 4, as contrasted to the reverse trend for the RTI network stations. As shown in Table 27, at least one of the first three procedures yields a smaller RMSE than procedure 4 at each of the non-network stations. This suggests that the apparently higher precision of model 4 relative to the other procedures is obtained by overfitting (i.e., including too many terms) in some cases; this results in a loss in accuracy relative to the first three procedures. Figure 2 also shows that, although the flat-surface models (procedure 5) provide predictions at the non-network stations which are nearly comparable in accuracy to those of procedures 1, 2, and 3, the precision of procedure 5, as measured by the RMSE's at the RTI network stations, is substantially poorer than that of the stepwise regression procedures.

Among the four stepwise regression procedures, procedure 1 would appear to yield the most accurate results across all seven non-network stations; procedure 3 appears more accurate across the five inner-non-network stations. These general conclusions are supported by the results of Tables 28 and 29, which show various statistics that summarize the distributions of the RMSE's over all cases. Table 28 provides these

TABLE 28. CHARACTERIZATION OF THE DISTRIBUTIONS OVER THE 908 CASES OF RMSE'S ACROSS ALL NON-NETWORK STATIONS BASED ON ESTIMATIONS FROM RTI NETWORK DATA

Wind Comp.	Modeling Procedure	Pooled RMSE (mps)	Mean RMSE (mps)	Std. Dev. of RMSE (mps)	Maximum RMSE (mps)	Percentage of Cases with RMSE (mps):				
						≤0.5	≤1.0	≤1.5	≤2.0	≤2.5
U	1	0.842	0.778	0.313	2.193	19.2	79.0	97.6	99.7	100.0
	2	0.878	0.794	0.364	5.038	19.1	76.9	97.0	99.3	99.8
	3	0.865	0.787	0.349	3.488	19.2	77.6	96.1	99.3	99.8
	4	0.917	0.822	0.394	3.588	17.2	74.8	94.2	98.5	99.2
	5	0.840	0.772	0.324	2.202	18.7	79.6	96.6	99.7	100.0
V	1	1.102	1.039	0.365	2.478	5.7	47.5	89.6	98.8	100.0
	2	1.098	1.035	0.365	2.898	4.8	49.7	90.0	98.5	99.8
	3	1.147	1.066	0.419	5.494	4.6	46.0	87.8	97.6	99.4
	4	1.239	1.116	0.531	6.441	4.7	44.6	85.2	96.1	98.6
	5	1.161	1.095	0.374	2.357	4.8	41.3	86.7	97.9	100.0
W	1	1.086	1.023	0.361	2.245	5.7	50.4	90.1	98.8	100.0
	2	1.061	0.995	0.361	3.434	5.3	54.3	91.0	98.9	99.8
	3	1.097	1.027	0.374	2.763	5.5	51.0	89.0	98.3	99.8
	4	1.113	1.030	0.406	3.737	5.7	50.9	88.9	97.6	99.2
	5	1.188	1.118	0.385	2.453	4.5	39.6	84.9	98.1	100.0
(U,V)	1	1.387	1.328	0.389	2.844	0.2	20.0	70.5	94.5	99.3
	2	1.406	1.336	0.427	5.117	0.2	20.5	70.0	93.2	98.8
	3	1.437	1.359	0.455	5.525	0.2	18.9	69.6	91.7	98.5
	4	1.542	1.428	0.565	6.467	0.2	18.2	65.0	89.0	96.1
	5	1.433	1.373	0.394	3.114	0.3	15.3	65.2	92.8	99.7

TABLE 29. CHARACTERIZATION OF THE DISTRIBUTIONS OVER THE 908 CASES OF RMSE'S ACROSS STATIONS IN THE INNER-NON-NETWORK BASED ON ESTIMATIONS FROM RTI NETWORK DATA

Wind Comp.	Modeling Procedure	Pooled RMSE (mps)	Mean RMSE (mps)	Std. Dev. of RMSE (mps)	Maximum RMSE (mps)	Percentage of Cases with RMSE (mps): ≤0.5	≤1.0	≤1.5	≤2.0	≤2.5
U	1	0.783	0.717	0.319	2.372	26.0	83.4	97.8	99.7	100.0
	2	0.800	0.728	0.331	2.372	25.0	82.2	97.2	99.7	100.0
	3	0.791	0.722	0.326	2.372	25.1	82.6	97.2	99.7	100.0
	4	0.830	0.751	0.350	2.372	23.1	79.5	95.9	99.3	100.0
	5	0.778	0.714	0.318	2.275	26.1	84.1	97.9	99.6	100.0
V	1	0.907	0.819	0.377	2.645	19.7	72.7	94.3	99.2	99.8
	2	0.929	0.842	0.376	2.645	17.5	70.5	94.2	98.9	99.9
	3	0.894	0.811	0.362	2.306	19.2	74.7	94.7	99.3	100.0
	4	0.935	0.851	0.376	2.254	17.0	69.4	93.6	99.2	100.0
	5	0.883	0.809	0.344	2.418	18.3	75.2	96.0	99.6	100.0
W	1	0.830	0.759	0.342	2.411	23.9	77.9	96.7	99.6	100.0
	2	0.835	0.765	0.333	2.445	22.5	78.9	96.8	99.7	100.0
	3	0.816	0.746	0.332	2.410	24.2	79.2	97.1	99.7	100.0
	4	0.845	0.775	0.334	2.397	20.6	76.9	96.7	99.8	100.0
	5	0.836	0.768	0.341	2.359	22.1	78.0	97.1	99.3	100.0
(U,V)	1	1.198	1.127	0.397	2.773	2.4	42.6	83.2	97.6	99.4
	2	1.226	1.153	0.402	2.850	2.2	40.4	82.6	96.6	99.2
	3	1.194	1.124	0.393	2.587	2.3	43.1	83.0	97.1	99.8
	4	1.250	1.176	0.411	2.570	2.1	38.8	79.1	95.9	99.8
	5	1.177	1.113	0.380	2.594	2.6	41.7	85.1	97.4	99.7

measures for all of the non-network whereas Table 29 provides comparable results for the inner-non-network only.

COMBINED EVALUATIVE MEASURES

The overall merit of a procedure must be judged by some combined measure of its estimation error (precision) and its prediction error (accuracy). An overall measure of the precision of a procedure is the square root of the sum of the pooled residual variances for the two wind components, based on estimations from the RTI network data; these values are shown below and are denoted by s(j), where j indicates the particular procedure:

Modeling Procedure (j)	Pooled Residual Variances (mps^2) U	Pooled Residual Variances (mps^2) V	Total	s(j): Square Root of Total (mps)
1R	0.7417	0.7577	1.4994	1.224
2R	0.6741	0.6449	1.3190	1.148
3R	0.6909	0.6736	1.3645	1.168
4R	0.6045	0.5439	1.1484	1.072
5R	0.9632	1.6508	2.6140	1.617

A compatible measure that reflects the accuracy of a procedure is the pooled vector RMSE over stations not in the RTI network. Values of these quantities were shown for the entire non-network in Table 28 and for the inner non-network in Table 29.

Let α, where $0 \le \alpha \le 1$, be used as a weighting factor to reflect the importance of accuracy relative to precision; define

$$f_\alpha(j) = \alpha[r(j)]^2 + (1-\alpha)\ [s(j)]^2 \qquad (24)$$

and

$$g_\alpha(j) = \alpha[r^*(j)]^2 + (1-\alpha)\ [s(j)]^2, \qquad (25)$$

where r(j) and r*(j) represent, respectively, the pooled vector RMSE's over all non-network stations and over inner-non-network stations. Note that $\alpha=0$ corresponds to assuming that precision of a particular procedure is of paramount importance and that accuracy can be completely ignored. Choosing $\alpha=1$, on the other hand, would completely ignore how well the particular procedure actually fit the data which were used to produce the estimates (i.e., the RTI network data). Regarding estimation and prediction errors to be of equal importance (i.e., $\alpha=0.5$) would result in the selection of procedures 2 or 4 as the "best" procedure, depending upon whether (24) or (25) is used as the criterion. As indicated by Figure 3, however, for this choice of α, there is little difference among the four stepwise procedures. Figure 4, which shows values of the $g_\alpha(j)$ versus α, indicates little preference among the stepwise procedures when $\alpha \gtrsim 0.75$. Although the choice of a particular α value is arbitrary, values in the range 0.5 to 0.8 would appear to be most reasonable; this corresponds to assuming that prediction errors are at least as important as estimation errors and may be up to 4 times more important. It should be noted that values of $f_\alpha(2)$ and $f_\alpha(3)$ are close for all values of α. The same holds true for $g_\alpha(2)$ and $g_\alpha(3)$. For $\alpha > 0.5$, values of $f_\alpha(2)$ and $g_\alpha(2)$ are also close to values of $f_\alpha(1)$ and $g_\alpha(1)$, respectively.

Figures 3 and 4 suggest that procedure 4, because of its tendency to produce inaccurate results, is the least preferable of the stepwise procedures. Among the first three procedures, there is no clear preference: larger values of α tend to support procedure 1 whereas smaller values tend to support procedures 2 or 3. Over the range 0.5 to 0.8, procedure 2 might be selected because when it is not "best", its f_α and g_α values are never "much larger" than the corresponding values for the procedure with the smallest f_α and g_α values. The same can be said for

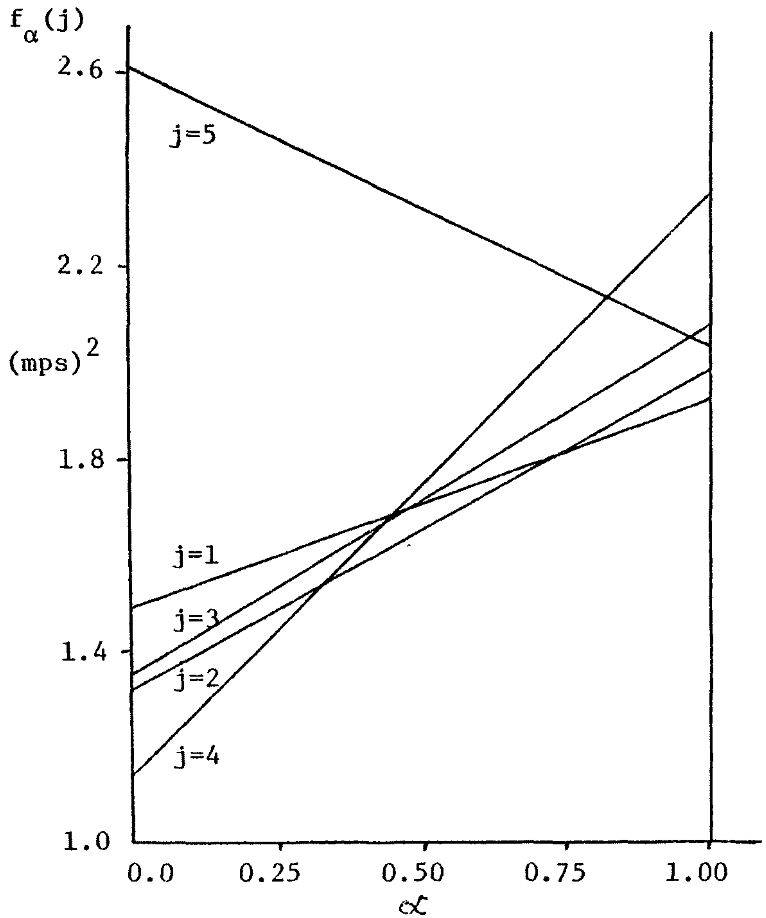


Figure 3. Plot of $f_{\alpha}(j)$ versus α, for five modeling procedures (j)

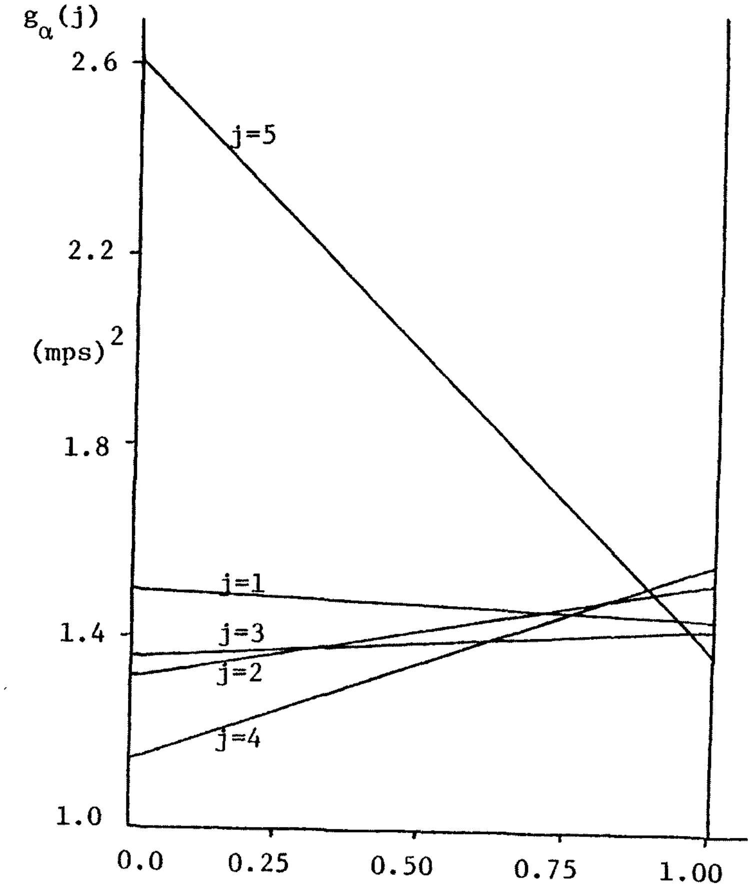


Figure 4. Plot of $g_{\alpha}(j)$ versus α, for five modeling procedures (j)

procedure 1, however, for α values greater than 0.6 or 0.7. Also, the consistency of procedure 1 over the non-network stations would tend to support its use.

Tables 30 and 31 show frequency distributions of the deviations between observed and predicted values for procedures 1 and 2 (and, for comparative purposes, for procedure 5). Table 30 shows these distributions for each non-network station and for the non-network as a whole. Table 31 shows the distributions over the RTI network, the full network, and the outer and inner non-networks. In both tables, all available observations are considered. Again, no strong preference for procedure 1 over procedure 2, or vice versa, is discernable.

SELECTED CASES AND CONDITIONS

Results of the previous subsections have presented various evaluative measures which, for the most part, have represented averages over a large number of cases. Such summaries, while quite essential for reducing the volume of data, can also be misleading in some situations. For instance, the importance of a difference of 0.1 mps in the average RMSE's of two procedures averaged over all cases may be difficult to judge. Such a difference could be caused by a few extreme cases or could be the result of small differences in a large number of cases. Although the various frequency distributions shown in previous sections provide some insight regarding the relative performance of the modeling procedures on a more case-specific basis, an overall "feel" for how the procedures might perform in a specific situation may be lacking.

Consequently, this subsection provides some additional detail and illustrative examples that should prove useful. Two major types of results are presented:

TABLE 30. PERCENTAGE FREQUENCY DISTRIBUTIONS OF DEVIATIONS BETWEEN OBSERVED AND PREDICTED WIND COMPONENTS AT NON-RTI NETWORK STATIONS -- ESTIMATIONS BASED ON DATA FROM RTI NETWORK STATIONS

Modeling Procedure	Station	Deviation Between Observed and Predicted U-Components (midpoint of interval, in mps) N	<-4	-4	-3	-2	-1	0	1	2	3	4	>4
1	STL003	892	--	--	--	0.56	6.28	37.67	48.32	7.06	0.11	--	--
	STL004	892	--	0.22	1.23	9.53	33.07	41.03	12.33	2.24	0.34	--	--
	STL007	350	--	0.29	2.86	12.57	41.43	36.86	5.14	0.86	--	--	--
	EPA103	833	--	--	0.12	1.56	13.21	48.14	26.65	7.92	2.16	0.24	--
	EPA107	844	--	--	--	0.24	13.74	74.05	11.73	0.24	--	--	--
	EPA111	596	--	0.17	0.17	3.69	45.97	46.31	3.69	--	--	--	--
	EPA112	861	--	--	0.12	1.05	28.11	61.32	8.71	0.58	--	0.12	--
	Total	5268	--	0.08	0.46	3.41	23.50	50.51	18.55	3.02	0.42	0.06	--
2	STL003	892	--	--	--	0.67	6.17	38.79	46.86	7.40	0.11	--	--
	STL004	892	--	0.45	1.46	10.99	32.40	40.47	11.77	2.13	0.34	--	--
	STL007	350	0.86	0.57	3.14	12.86	39.14	35.43	6.86	1.14	--	--	--
	EPA103	833	--	--	0.12	1.56	14.05	48.02	26.65	8.40	1.08	0.12	--
	EPA107	844	--	--	--	0.47	15.40	74.17	9.95	--	--	--	--
	EPA111	596	--	0.17	0.17	4.19	45.64	46.48	3.36	--	--	--	--
	EPA112	861	--	--	0.12	1.28	32.40	56.68	8.71	0.70	--	0.12	--
	Total	5268	0.06	0.13	0.51	3.83	24.28	49.77	18.00	3.13	0.25	0.04	--
5	STL003	892	--	--	--	0.22	6.61	37.00	47.09	8.86	0.22	--	--
	STL004	892	--	0.22	0.90	7.40	32.29	43.83	12.89	1.91	0.56	--	--
	STL007	350	--	0.57	2.57	4.86	35.14	45.71	10.00	1.14	--	--	--
	EPA103	833	--	--	0.48	2.16	12.48	47.42	26.17	7.92	3.24	0.12	--
	EPA107	844	--	--	--	1.30	12.91	70.97	14.34	0.47	--	--	--
	EPA111	596	--	0.34	--	4.19	46.14	46.81	2.52	--	--	--	--
	EPA112	861	--	--	0.12	0.58	23.69	65.16	9.87	0.46	--	0.12	--
	Total	5268	--	0.11	0.42	2.73	22.06	51.54	19.15	3.03	0.65	0.04	--

TABLE 30 (continued)

Modeling Procedure	Station	Deviation Between Observed and Predicted V-Components (midpoint of interval, in mps) N	<-4	-4	-3	-2	-1	0	1	2	3	4	>4
1	STL003	892	--	0.22	1.91	13.90	43.72	31.61	7.29	1.35	--	--	--
	STL004	892	--	--	--	3.36	15.70	33.52	30.94	12.33	3.81	0.34	--
	STL007	350	--	--	0.29	3.71	13.71	33.71	35.71	10.86	1.43	0.57	--
	EPA103	833	--	--	0.12	0.48	4.44	15.13	29.89	36.25	11.52	2.16	--
	EPA107	844	--	--	0.12	1.90	24.17	62.32	11.26	0.24	--	--	--
	EPA111	596	--	--	--	0.34	17.11	46.98	32.05	3.02	0.50	--	--
	EPA112	861	--	--	--	0.12	13.47	59.23	24.51	2.56	0.12	--	--
	Total	5268	--	0.04	0.38	3.61	19.68	40.64	23.01	9.57	2.64	0.44	--
2	STL003	892	--	0.11	1.35	14.69	45.63	30.04	7.17	1.01	--	--	--
	STL004	892	--	--	0.11	3.59	12.89	32.51	32.74	13.12	4.37	0.67	--
	STL007	350	--	--	0.86	3.14	11.71	32.57	32.86	15.71	1.71	0.86	0.57
	EPA103	833	--	--	0.12	0.72	5.64	16.09	34.93	32.41	8.40	1.68	--
	EPA107	844	--	--	--	1.54	25.71	61.37	11.14	0.12	0.12	--	--
	EPA111	596	--	--	--	0.50	14.77	48.49	31.54	4.19	0.50	--	--
	EPA112	861	--	--	--	0.23	11.96	52.73	31.36	3.48	0.23	--	--
	Total	5268	--	0.02	0.32	3.76	19.32	39.24	24.94	9.62	2.30	0.44	0.04
5	STL003	892	--	0.56	1.46	15.25	45.07	29.15	7.06	1.46	--	--	--
	STL004	892	--	--	--	5.38	20.40	37.67	27.13	7.74	1.68	--	--
	STL007	350	0.29	0.29	2.29	9.14	22.57	40.57	22.57	2.29	--	--	--
	EPA103	833	--	--	0.24	0.84	3.12	11.16	22.45	37.45	21.13	3.48	0.12
	EPA107	844	--	--	0.12	2.73	31.28	56.28	9.60	--	--	--	--
	EPA111	596	--	--	--	0.34	19.46	47.48	28.86	3.52	0.34	--	--
	EPA112	861	--	--	--	2.21	29.50	56.68	11.03	0.58	--	--	--
	Total	5268	0.02	0.11	0.46	5.07	25.11	39.43	17.44	8.12	3.66	0.55	0.02

TABLE 30 (continued)

Modeling Procedure	Station	Deviation Between Observed and Predicted Wind Speeds (midpoint of interval, in mps) N	<-4	-4	-3	-2	-1	0	1	2	3	4	>4
1	STL003	892	--	0.22	1.23	5.94	30.83	42.94	16.70	2.02	0.11	--	--
	STL004	892	--	--	--	1.23	11.21	36.32	35.99	13.57	1.46	0.22	--
	STL007	350	--	0.29	1.14	5.43	12.86	47.14	27.14	5.71	0.29	--	--
	EPA103	833	--	--	--	--	1.32	8.64	31.33	40.46	15.49	2.52	0.24
	EPA107	844	--	--	0.12	3.08	25.36	58.41	12.91	0.12	--	--	--
	EPA111	596	--	--	--	0.50	17.95	53.02	26.68	1.68	0.17	--	--
	EPA112	861	--	--	--	0.23	11.61	60.63	25.32	1.74	0.23	0.12	0.12
	Total	5268	--	0.06	0.30	2.16	16.17	43.19	24.91	9.91	2.79	0.46	0.06
2	STL003	892	--	0.11	1.01	6.39	32.74	42.38	15.47	1.79	0.11	--	--
	STL004	892	--	--	--	0.78	9.08	34.75	40.70	12.67	1.79	0.22	--
	STL007	350	0.57	--	2.57	3.43	12.57	42.86	30.29	6.86	0.86	--	--
	EPA103	833	--	--	--	0.12	2.40	11.04	35.65	38.54	10.32	1.68	0.24
	EPA107	844	--	--	0.12	2.13	27.73	59.36	10.55	0.12	--	--	--
	EPA111	596	--	--	--	0.84	14.77	55.03	26.34	2.85	0.17	--	--
	EPA112	861	--	--	--	0.23	11.03	54.47	31.82	1.97	0.23	0.12	0.12
	Total	5268	0.04	0.02	0.36	1.94	16.21	42.29	27.03	9.66	2.07	0.32	0.06
5	STL003	892	--	0.45	0.90	6.61	29.82	42.60	16.82	2.58	0.22	--	--
	STL004	892	--	--	--	2.80	17.38	40.25	30.04	7.96	1.46	0.11	--
	STL007	350	0.57	1.14	3.43	9.71	13.14	49.43	21.43	1.14	--	--	--
	EPA103	833	--	--	--	0.12	0.96	6.12	21.01	40.94	25.81	4.68	0.36
	EPA107	844	--	--	0.12	4.98	31.87	50.47	12.44	0.12	--	--	--
	EPA111	596	--	--	--	1.01	19.13	54.70	23.49	1.68	--	--	--
	EPA112	861	--	--	--	2.09	29.62	52.96	14.05	1.05	0.12	0.12	--
	Total	5268	0.04	0.15	0.40	3.51	21.13	41.21	19.63	8.71	4.38	0.78	0.06

TABLE 31. PERCENTAGE FREQUENCY DISTRIBUTION OF DEVIATIONS BETWEEN OBSERVED AND PREDICTED VALUES BASED ON ESTIMATIONS FROM RTI NETWORK DATA

Variable	Modeling Procedure	Subset of Stations	No. Obs.	Deviation Between Observed and Predicted Values (midpoint of interval, in mps) <-4	-4	-3	-2	-1	0	1	2	3	4	>4
U-Comp.	1	Outer-non	1183	--	0.08	0.93	4.82	21.56	44.80	20.29	5.83	1.52	0.17	--
		Inner-non	4085	--	0.07	0.32	3.01	24.06	52.17	18.04	2.20	0.10	0.02	--
		RTI	14690	0.01	0.03	0.41	3.08	19.61	54.27	18.97	3.18	0.37	0.06	0.01
		Full	19958	0.01	0.04	0.42	3.17	20.64	53.28	18.86	3.14	0.38	0.06	0.01
	2	Outer-non	1183	0.25	0.17	1.01	4.90	21.47	44.29	20.79	6.26	0.76	0.08	--
		Inner-non	4085	--	0.12	0.37	3.53	25.09	51.36	17.18	1.81	0.10	0.02	--
		RTI	14690	0.01	0.02	0.34	2.58	17.41	59.24	17.71	2.40	0.22	0.06	0.01
		Full	19958	0.03	0.05	0.39	2.91	19.22	56.74	17.78	2.60	0.23	0.06	0.01
	5	Outer-non	1183	--	0.17	1.10	2.96	19.19	46.91	21.39	5.92	2.28	0.08	--
		Inner-non	4085	--	0.10	0.22	2.67	22.89	52.88	18.51	2.55	0.17	0.02	--
		RTI	14690	0.02	0.05	0.62	4.73	22.10	46.39	20.37	4.59	0.87	0.25	0.01
		Full	19958	0.02	0.07	0.57	4.20	22.09	47.75	20.05	4.25	0.81	0.19	0.01
V-Comp.	1	Outer-non	1183	--	--	0.17	1.44	7.19	20.63	31.61	28.74	8.54	1.69	--
		Inner-non	4085	--	0.05	0.44	4.24	23.30	46.44	20.51	4.01	0.93	0.07	--
		RTI	14690	--	0.01	0.21	2.97	21.25	51.42	21.05	2.80	0.25	0.03	--
		Full	19958	--	0.02	0.26	3.14	20.84	48.57	21.57	4.59	0.88	0.14	--
	2	Outer-non	1183	--	--	0.34	1.44	7.44	20.96	34.32	27.47	6.42	1.44	0.17
		Inner-non	4085	--	0.02	0.32	4.43	22.77	44.53	22.23	4.46	1.10	0.15	--
		RTI	14690	--	--	0.12	1.91	18.49	59.24	18.05	1.99	0.19	0.02	--
		Full	19958	--	0.01	0.17	2.40	18.71	53.96	19.87	4.01	0.75	0.13	0.01
	5	Outer-non	1183	0.08	0.08	0.85	3.30	8.88	19.86	22.49	27.05	14.88	2.45	0.08
		Inner-non	4085	--	0.12	0.34	5.58	29.82	45.09	15.99	2.64	0.42	--	--
		RTI	14690	0.21	0.45	1.71	7.39	23.08	35.22	22.04	6.89	2.33	0.57	0.12
		Full	19958	0.16	0.36	1.38	6.77	23.61	36.33	20.82	7.22	2.69	0.57	0.09

1. Summaries over various subsets of cases, and
2. Detailed results for several individual cases.

The results shown are limited to modeling procedures 1 and 2, since the combined measures of the previous subsection indicate that these two procedures are certainly competitive with the two procedures that utilize the larger class of model terms.

The prevailing wind speeds and directions are utilized to group the 908 cases into subsets over which the various evaluation measures are computed. Four prevailing wind speed categories and four prevailing wind direction categories are used; these are shown in Table 32 below, along with their relevant sample sizes (number of cases and observations).

TABLE 32. SAMPLE SIZES, BY PREVAILING WIND SPEED AND DIRECTION CATEGORIES

Prevailing Wind Condition		No. of Cases	Number of Observations		
			RTI Network	Non-Network	Inner-Non-Network
Speed: (mps)	<2	237	3813	1377	965
	2-4	359	5733	2057	1629
	4-6	266	4377	1572	1274
	>6	46	767	262	217
Direction:					
	E,SE	92	1565	560	436
	S	487	7787	2829	2208
	SW	225	3599	1268	973
	Other	104	1739	611	468
Total		908	14690	5268	4085

Table 33 presents values of three evaluation measures that characterize the magnitude of the <u>estimation</u> errors. These are shown for both procedures 1 and 2 applied to, and evaluated over, the stations in the RTI network. Parts A and B of this table indicate that the estimation errors tend to be larger for the higher wind speed cases. This is true,

TABLE 33. SUMMARY OF ESTIMATION ERRORS BY PREVAILING WIND SPEED AND DIRECTION CATEGORIES

A. Pooled Residual Standard Deviations (mps)

Wind Comp.	Modeling Procedure	Prevailing Wind Speed (mps)				Prevailing Wind Direction			
		<2	2-4	4-6	>6	E,SE	S	SW	Other
U	1R	0.612	0.806	1.022	1.221	0.868	0.862	0.884	0.800
	2R	0.580	0.766	0.973	1.170	0.813	0.839	0.828	0.710
V	1R	0.707	0.832	0.978	1.203	0.784	0.854	0.893	0.969
	2R	0.663	0.764	0.898	1.115	0.697	0.782	0.827	0.923
(U,V)	1R	0.935	1.158	1.415	1.714	1.170	1.213	1.256	1.256
	2R	0.881	1.082	1.324	1.617	1.071	1.147	1.171	1.165

B. Percentage of Cases With Residual Standard Deviations Less Than 1.0 mps

Wind Comp.	Modeling Procedure	Prevailing Wind Speed (mps)				Prevailing Wind Direction			
		<2	2-4	4-6	>6	E,SE	S	SW	Other
U	1R	95.8	84.7	56.0	34.8	75.0	76.8	72.0	87.5
	2R	97.0	88.9	62.4	47.8	84.8	79.5	77.8	93.3
V	1R	92.4	84.4	63.9	34.8	90.2	79.5	75.1	66.3
	2R	94.5	90.5	74.4	41.3	93.5	87.3	81.8	68.3

C. Pooled Adjusted R^2 Statistics

Wind Comp.	Modeling Procedure	Prevailing Wind Speed (mps)				Prevailing Wind Direction			
		<2	2-4	4-6	>6	E,SE	S	SW	Other
U	1R	0.253	0.217	0.240	0.218	0.253	0.147	0.247	0.458
	2R	0.331	0.291	0.311	0.282	0.345	0.191	0.339	0.573
V	1R	0.397	0.472	0.600	0.640	0.237	0.532	0.625	0.492
	2R	0.470	0.555	0.663	0.690	0.396	0.607	0.678	0.539

even though a larger percentage of the variation is typically accounted for in the high-speed cases, as is demonstrated in Part C of Table 33. On an absolute scale, smaller estimation errors occur for the east/southeast category than for the other wind direction categories; this appears to be a reflection of the trend noted above for wind speed, since the average wind speed over cases in this wind direction category is less than that for the other direction categories. As might be expected, only a small percentage of the variation is accounted for in the less-dominant wind component--for instance, in the U-component when a prevailing southerly wind occurs. In terms of estimation errors, the improvement of procedure 2 over procedure 1 appears to be quite consistent across all eight (4 speed and 4 direction) categories and both wind components. Differences in the pooled adjusted R^2 statistics for procedures 1 and 2, for example, range from about 0.05 to about 0.12.

Table 34 provides two basic measures of the prediction errors--namely, pooled RMSE's over all non-network stations (Part A), and over all inner-non-network stations (Part B)--categorized by prevailing wind speed and by prevailing wind direction. The pooled RMSE's for procedure 1 are usually slightly smaller than the corresponding RMSE's for procedure 2. As with the estimation errors, there is a definite pattern of larger RMSE's for the higher wind speed cases as compared to the lower speed cases; this trend appears to be more pronounced for the inner-non-network stations (Part A). Figure 5 illustrates this trend and permits a visual comparison of the relative magnitudes of the estimation and prediction errors to be made for the various wind speed categories.

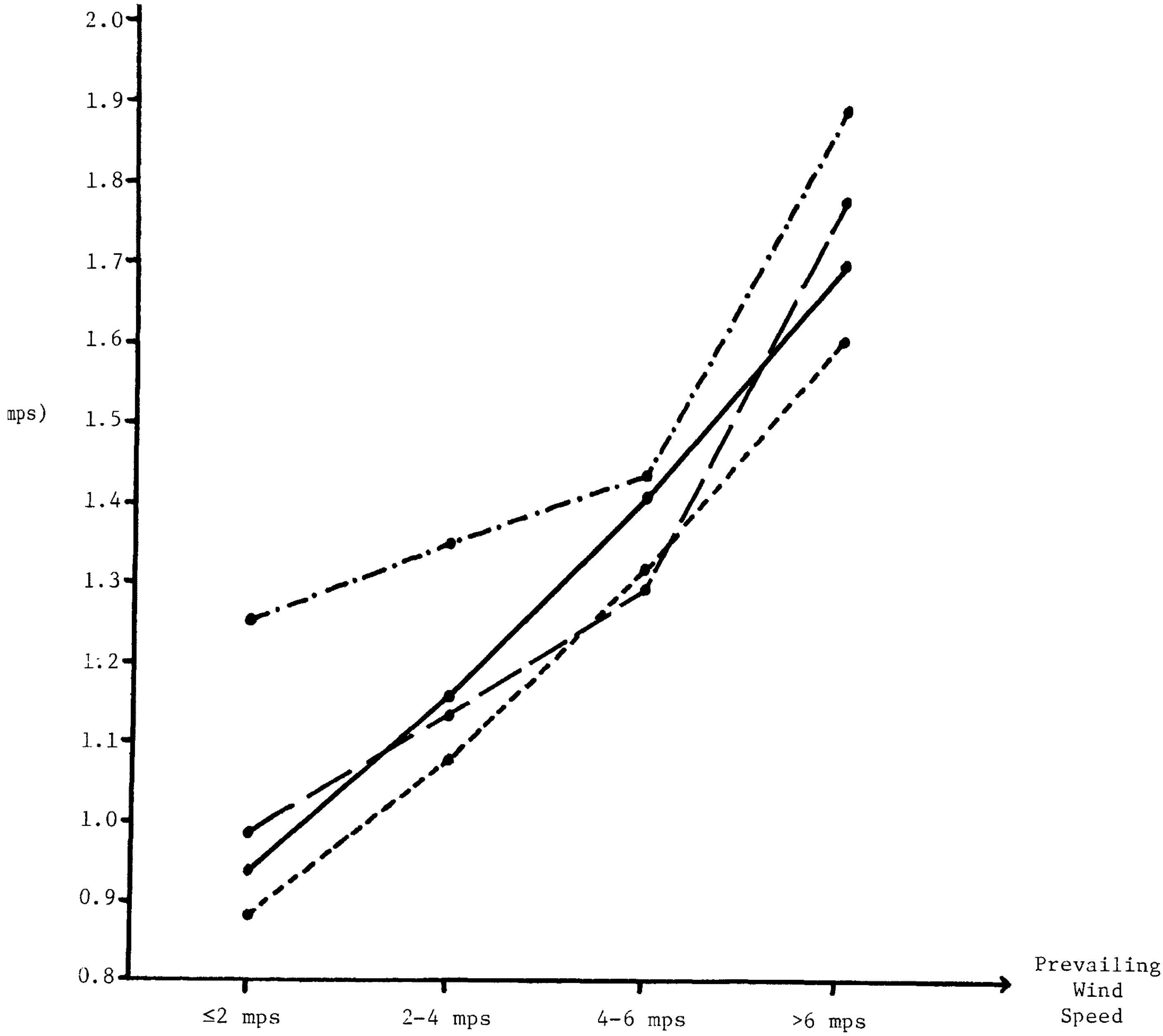
TABLE 34. SUMMARY OF PREDICTION ERRORS BY PREVAILING WIND SPEED AND DIRECTION CATEGORIES

A. Pooled Root Mean Square Errors Over All Non-Network Stations (mps)

Wind Comp.	Modeling Procedure	Prevailing Wind Speed (mps)				Prevailing Wind Direction			
		<2	2-4	4-6	>6	E,SE	S	SW	Other
U	1R	0.716	0.834	0.927	0.968	0.958	0.758	0.876	1.009
	2R	0.744	0.900	0.941	0.957	1.047	0.820	0.874	0.970
V	1R	1.026	1.062	1.106	1.643	0.923	1.098	1.192	1.077
	2R	1.009	1.057	1.119	1.622	0.948	1.078	1.198	1.106
W	1R	1.032	1.062	1.081	1.499	0.985	1.082	1.131	1.096
	2R	0.985	1.042	1.072	1.452	0.991	1.054	1.099	1.074
(U,V)	1R	1.251	1.351	1.443	1.907	1.330	1.334	1.479	1.476
	2R	1.253	1.388	1.462	1.884	1.412	1.355	1.482	1.472

B. Pooled Root Mean Square Errors Over All Inner-Non-Network Stations (mps)

Wind Comp.	Modeling Procedure	Prevailing Wind Speed (mps)				Prevailing Wind Direction			
		<2	2-4	4-6	>6	E,SE	S	SW	Other
U	1R	0.682	0.776	0.851	0.841	0.933	0.743	0.803	0.770
	2R	0.671	0.800	0.875	0.857	1.018	0.754	0.805	0.766
V	1R	0.710	0.821	0.983	1.585	0.747	0.823	1.043	1.099
	2R	0.724	0.833	1.028	1.584	0.795	0.841	1.059	1.129
W	1R	0.741	0.792	0.841	1.283	0.792	0.786	0.922	0.864
	2R	0.716	0.798	0.871	1.265	0.813	0.790	0.918	0.874
(U,V)	1R	0.985	1.130	1.300	1.795	1.195	1.109	1.316	1.343
	2R	0.987	1.156	1.350	1.801	1.292	1.130	1.330	1.364



———— Pooled Vector Residual Std. Dev., Procedure 1R

- - - - - Pooled Vector Residual Std. Dev., Procedure 2R

— — — — Pooled Vector RMSE (Inner-Non-Network)*

-·-·-·-·- Pooled Vector RMSE (Non-Network)*

* Shown for procedure 1R; a very similar curve occurs for procedure 2R.

Figure 5. Pooled measures of estimation and prediction errors versus prevailing wind speed

Percentage errors in predicting wind speeds at non-network stations are summarized in Table 35 by prevailing wind speed categories. These percentage errors are shown for procedures 1 and 2 applied to data from the RTI network and, for comparative purposes, for procedure 4 applied to the full network data. The percentage errors tend to decrease with increasing wind speed for all three of these procedures. Over the inner-non-network, the percentage errors for procedures 1R and 2R are about 20% greater than the corresponding percentage errors for procedure 4F. Percentage errors for procedures 1R and 2R, over all non-network stations, are roughly 30% larger than the percentage errors for procedure 4F.

In order to further demonstrate the performance of the modeling procedures in specific cases, three particular cases were selected for detailed examination:

			Prevailing Winds	
Case	Date	Time	Speed	Direction
I	8/12/75	1800	2.56 mps	158°
II	2/20/76	1600	4.83 mps	136°
III	2/21/76	1700	5.59 mps	274°

It should be emphasized that these cases were picked arbitrarily. They do not necessarily reflect "typical" cases from among the 908 cases; Case I, for example, was purposely chosen as a worst case situation for modeling procedure 2, in that the vector RMSE over all non-network stations for this case was much larger than for any other case.

The prediction models determined by procedures 1 and 2 for these cases are given in Table 36. These particular models illustrate the typical pattern of larger, more complex models for modeling procedure 2, as compared to procedure 1. It should also be noted that in two of the

TABLE 35. PERCENTAGE ERRORS IN WIND SPEED PREDICTIONS AT NON-NETWORK STATIONS, BY PREVAILING WIND SPEED CATEGORIES

Subset of Stations	Prevailing Wind Speed Category	Mean Wind Speed (mps) ($\bar{W}$)	100% x (RMSE(W)/$\bar{W}$) For Modeling Procedure: 1R	2R	4F
Inner-non Network Stations	≤2 mps	1.882	39.4	38.0	32.7
	2-4 mps	3.344	23.7	23.9	19.5
	4-6 mps	4.732	17.8	18.4	15.0
	>6 mps	6.383	20.1	19.8	16.4
	Overall	3.593	23.1	23.2	19.1
All Non-Network Stations	≤2 mps	2.129	48.5	46.3	35.2
	2-4 mps	3.622	29.3	28.8	21.6
	4-6 mps	5.036	21.5	21.3	16.5
	>6 mps	6.896	21.7	21.1	17.0
	Overall	3.816	28.5	27.8	21.3

TABLE 36. PREDICTION MODELS FOR THREE SPECIFIC CASES

Case	Modeling Procedure	Prediction Model Based on Data From RTI Network
I	1	$\hat{U} = -\ 0.30969 - 0.00127536xy^2$
		$\hat{V} = 3.23072 - 0.00027154y^3 + 0.00154632x^2y$
	2	$\hat{U} = 0.95191 - 0.170683x + 0.0095480x^2 - 0.0454584xy + 0.00445697x^2y - 0.00176493xy^2 - 0.038375h$
		$\hat{V} = 3.26253 + 0.00111345x^2y + 0.00103673xy^2 - 0.000027405y^4$
II	1	$\hat{U} = -\ 3.81314 - 0.085727x$
		$\hat{V} = 4.76958 - 0.0118349y^2 - 0.00026648x^3 + 0.00232374xy^2$
	2	$\hat{U} = -3.60660 - 0.195394x + 0.00045859x^3$
		$\hat{V} = 4.76958 - 0.0118349y^2 - 0.00026648x^3 + 0.00232374xy^2$
III	1	$\hat{U} = 5.82123 + 0.00113637xy^2$
		$\hat{V} = 0.09533 - 0.000014250x^4$
	2	$\hat{U} = 5.84047 + 0.162089x + 0.0043193xy - 0.00056576x^3 - 0.0174165h$
		$\hat{V} = 0.09533 - 0.000014250x^4$

cases (Cases II and III) the same model was selected for the V-component by both procedures.

Table 37 summarizes the fit of the models over the RTI network stations (i.e., over the set of stations actually used for determining the model form and parameter estimates). Modeling procedure 2 generally accounts for more of the variation in winds among these stations (i.e., larger R^2 values). That is, the predicted surfaces from procedure 2 will typically have more hills, valleys, ridges, etc. than those from procedure 1, and therefore, if these are "real" (e.g., as demonstrated by comparing predicted values with observed data from the non-network stations), it would be the preferred procedure. On the other hand, because procedure 2 yields more complex polynomials, it is more likely to produce spurious hills, valleys, ridges, etc. in the wind field over those areas not in the vicinity of one or more RTI network stations--for instance, in the outlying areas of the region. This is well illustrated by Case I, in which the RMSE's over the inner-non-network are quite comparable for the two procedures (see Table 38, Part B), whereas the RMSE for procedure 2 over the entire non-network is extremely large relative to the corresponding RMSE from procedure 1 (see Table 38, Part A). The large deviation in observed and predicted winds at station STL007 accounts for this discrepancy; it should be noted that wind data were not available for any RTI network station near to the STL007 site.

The observed data for Case I, as shown in Figure 6(A), indicate that the wind flow is generally out of the south-southeast with wind speeds across the city ranging from about 1 to 6 mps and averaging about 3.3 mps. The flow pattern suggests the influence of a heat island

TABLE 37. ANALYSIS OF VARIANCE RESULTS FOR THREE SPECIFIC CASES

Case	No. Stations in RTI Network	Modeling Procedure	Wind Comp.	No. of Terms in Model	Sums of Squares $(mps)^2$ Total	Regression	Residual	Residual Variance*	R^2	F Value+
I	16	1	U	2	26.5783	6.8707	19.7087	1.4078	0.258	4.881
			V	3	18.2802	8.5297	9.7504	0.7500	0.467	5.686
		2	U	7	26.5793	21.5190	5.0603	0.5623	0.810	6.379
			V	4	18.2802	11.2206	7.0596	0.5883	0.614	6.358
II	18	1	U	2	47.8158	11.6919	36.1238	2.2577	0.245	5.179
			V	4	33.4045	24.5170	8.8875	0.6348	0.734	12.873
		2	U	3	47.8158	17.3583	30.4575	2.0305	0.363	4.274
			V	4	33.4045	24.5170	8.8875	0.6348	0.734	12.873
III	17	1	U	2	20.2199	7.6288	12.5911	0.8394	0.377	9.088
			V	2	19.8347	6.6933	13.1415	0.8761	0.337	7.640
		2	U	5	20.2199	15.9330	4.2869	0.3572	0.788	11.150
			V	2	19.8347	6.6933	13.1415	0.8761	0.337	7.640

* The residual variance is calculated by dividing the residual sum of squares by the number of residual degrees of freedom. This degrees of freedom is the number of stations minus the number of model terms.

\+ The F-value is calculated as the ratio of the regression mean square to the residual variance. The degrees of freedom for the regression mean square is one less than the number of model terms.

TABLE 38. ROOT MEAN SQUARE ERRORS (MPS) FOR THREE SPECIFIC CASES

A. RMSE's Over all Non-Network Stations

Wind Comp.	Modeling Procedure	Case I	Case II	Case III
U	1	1.245	1.485	1.160
	2	5.038	1.663	0.779
V	1	1.418	1.384	0.759
	2	0.896	1.384	0.759
W	1	1.247	1.487	1.140
	2	3.434	1.511	0.806
(U, V)	1	1.886	2.030	1.387
	2	5.117	2.164	1.088

B. RMSE's Over All Inner-Non-Network Stations

Wind Comp.	Modeling Procedure	Case I	Case II	Case III
U	1	1.840	1.648	0.801
	2	1.891	1.910	0.606
V	1	0.574	1.085	0.816
	2	0.524	1.085	0.816
W	1	1.068	1.210	0.785
	2	0.048	1.321	0.655
(U, V)	1	1.928	1.973	1.143
	2	1.963	2.196	1.016

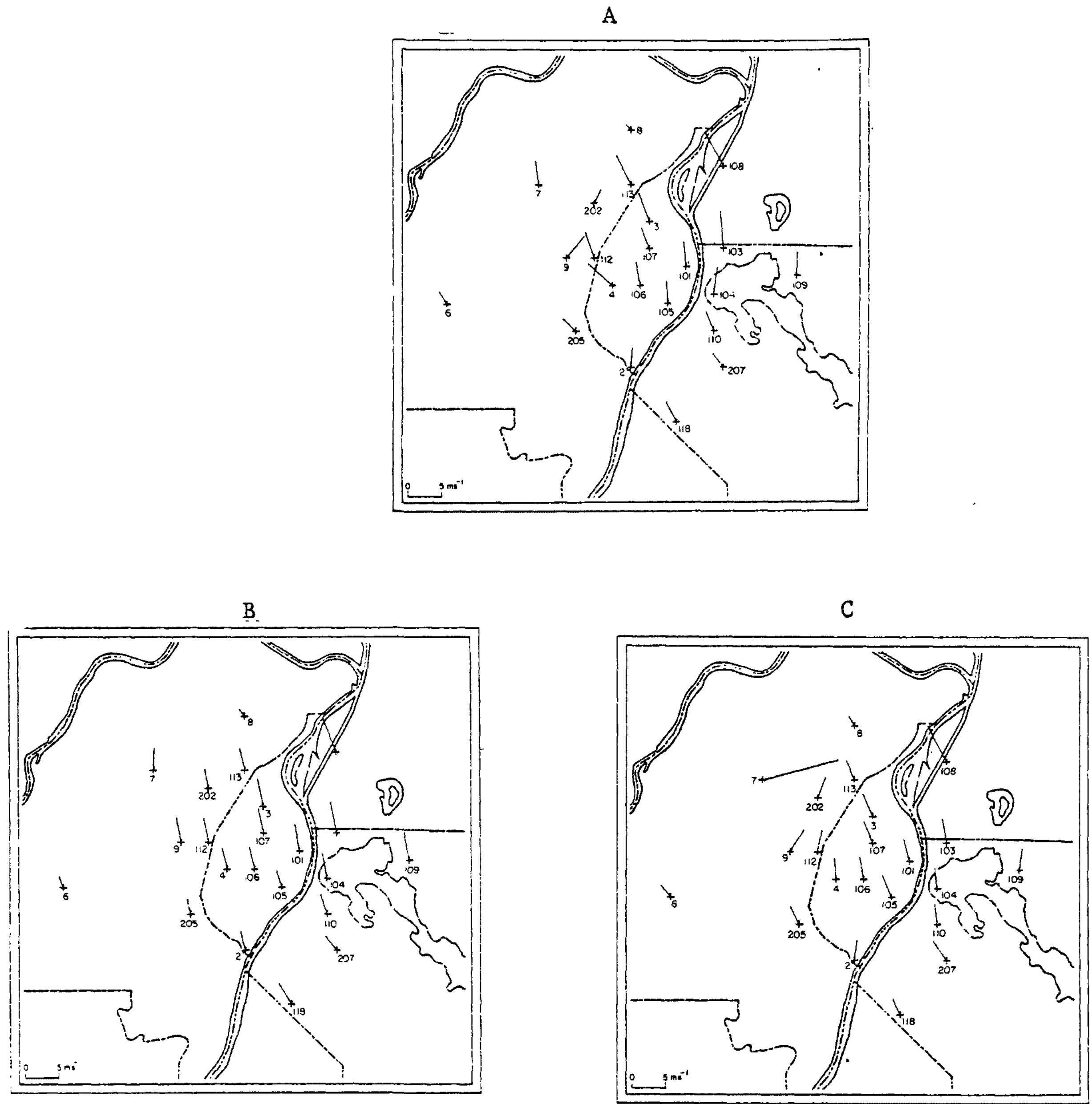


Figure 6. Observed and predicted winds for case I: (A) observed data; (B) predicted winds using procedure 1; (C) predicted winds using procedure 2

circulation having strong convergence in the northwestern part of the city. Such flow patterns associated with heat island circulation have been observed before in the city of St. Louis (Vukovich, Dunn et al., 1979).

The wind predictions for Case I were based on only 16 of the 19 stations in the RTI network. The three non-reporting stations were RAPS stations 102, 119, and 120. Station 119 is located at the outer boundary of the southwestern portion of the network and station 120, at the outer boundary of the northwestern portion (see Figure 1). Large errors in the predicted wind field might be expected in these regions due to the absence of wind data from these areas of the network. That is, the predictions in these areas would essentially represent extrapolation of the polynomial models outside the range of the data; it is well known that such extrapolation is highly error prone.

The predicted wind fields for Case I determined by modeling procedures 1 and 2 are shown in Figures 6(B) and 6(C), respectively. Figure 6(B) shows a general flow pattern from the south-southeast with wind speeds ranging from about 1.4 to 5.1 mps. Although some convergence in the flow downstream of the city is evident, it is not as intense as that appearing in the observed flow field. The predicted wind field from procedure 2 (Figure 6(C)) also shows the general south-southeasterly flow pattern; the predicted wind speeds range from about 1.6 to 11.7 mps. A southwesterly wind with a speed of 11.7 mps is predicted for STL007. This station is in the northwestern zone of the region and thus represents an area in which extrapolation occurs when no data are available from EPA120. If STL007 is excluded, the predicted wind speeds range from 1.6 to 6.2 mps across the other stations; the predicted field also shows the strong convergence downstream of the city that was apparent in the observed data. Except for the problem of extrapolation caused by the missing data, it would therefore appear that modeling procedure 2 performed better than procedure 1 in this case.

The wind data for Case II (Figure 7(A)) shows a general southeasterly flow with wind speeds ranging from 1.8 to 9.4 mps. The average wind speed was 6.0 mps. There is some indication of convergence immediately downstream of the city which may be associated with the heat island circulation; this convergence is not as significant as that found in Case I. There is also an apparent speed convergence over the city, probably due to the increased friction in that region.

The predicted wind fields were based on 18 of the 19 stations in the RTI network. The missing station was RTI202, which is located in the interior of the network domain (see Figure 1). Both of the predicted wind fields for this case, shown in Figures 7(B) and 7(C), appear to pick up the indicated speed convergence over the central portion of the city. Procedure 2 appears to indicate the convergence downstream of the city somewhat better than procedure 1.

The observed data for Case III (Figure 8(A)) indicate flow from the west with wind speeds ranging from 4 to 8 mps. The wind distribution shows no significant distortion of the flow pattern due to the presence of the city except for a slight decrease in wind speed over the central portion of the city again due to the increased friction in that region.

The predicted wind fields for Case III were obtained by utilizing 17 of the RTI network stations. Data were missing from RTI stations 202 and 205, which are located in the interior of the network. The flow pattern obtained from modeling procedure 1 (Figure 8(B)) is very similar

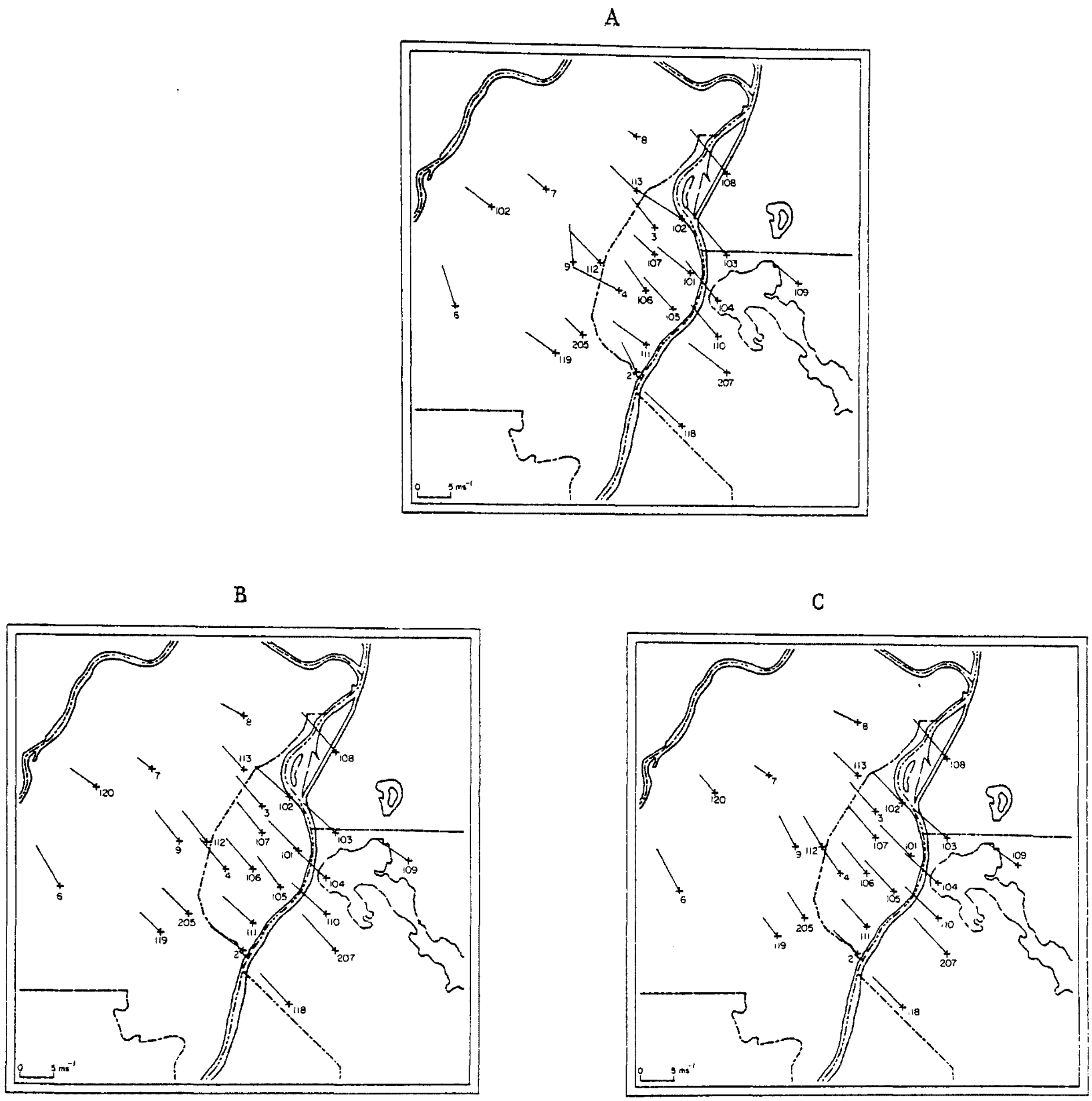


Figure 7. Observed and predicted winds for case III: (A) observed data; (B) predicted winds using procedure 1; (C) predicted winds using procedure 2

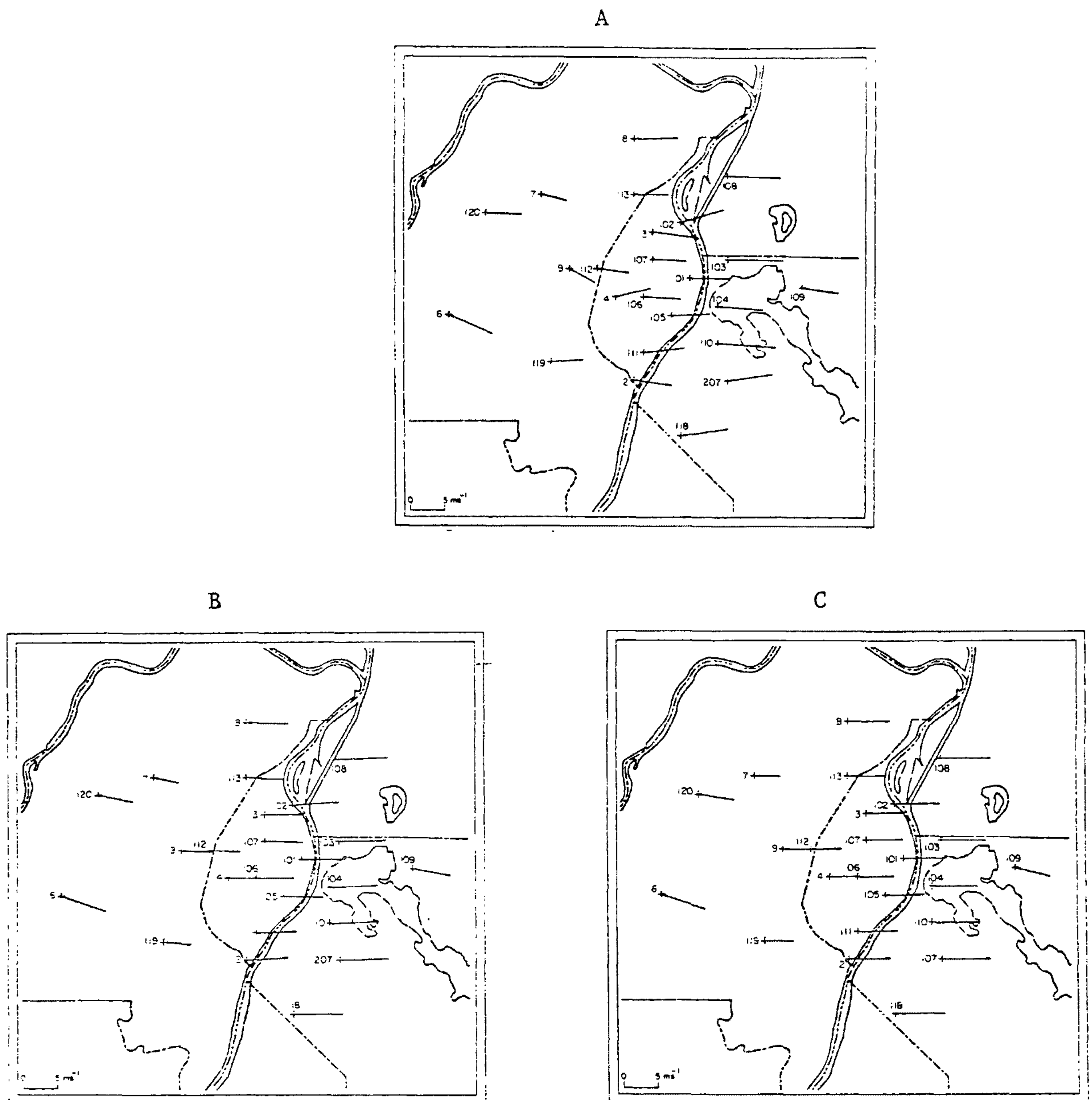


Figure 8. Observed and predicted winds for case III: (A) observed data; (B) predicted winds using procedure 1; (C) predicted winds using procedure 2

to the observed data, although the lower wind speeds in the central city (relative to the surrounding regions) are not as obvious as those of Figure 8(A). The flow field based on procedure 2 (Figure 8(C)) is also quite similar to the observed flow field; in this case, the lower wind speeds over the urban region are somewhat more evident.

Based on these three cases, it appears that modeling procedure 2 may produce predicted wind fields with general characteristics more similar to the observed wind field than the procedure 1 predictions. The results also indicate that missing data may lead to substantially poorer predictions in some areas within the region, particularly when the missing data occur at the boundaries of the network. In such cases, it will be necessary to redefine the network domain so as to avoid the effects of extrapolations.

SECTION 5

DISCUSSION OF RESULTS

CONCLUSIONS AND FINDINGS

The primary conclusion of this study is that a polynomial model derived by stepwise regression on 13 model terms and applied to the 19-station RTI network could produce predicted wind fields for St. Louis comparable to those produced by similar procedures applied to a larger class of model terms and a larger network (i.e., 23 terms and 26 stations). The 13-term model and the 19-station network were selected in the theoretical phase of this research program (Vukovich et al., 1978) based on the argument that the addition of terms in the model and/or stations in the network would not markedly improve the analysis of the wind field. This hypothesis has now been substantiated using observed data. The conclusion of this study is based on the following findings:

In terms of estimation errors (precision), the results of applying four stepwise regression procedures to wind data from the RTI network (the "optimum" network) and from the full network indicate that comparable results are obtained for the RTI network and the full network, although the estimations for the full network yield somewhat more consistent adjusted R^2 values across the various cases.

The four stepwise regression procedures are clearly superior to both procedure 0 (fitting the full 13-term model) and procedure 5 (fitting a flat surface). This indicates that stepwise regression techniques offer a practical method for automating the model form determination over a large number of cases; prior screening of the data for outliers, however, may hamper implementation of any automated, quick-response method for model estimation.

Among the four stepwise regression procedures, the procedure permitting the most complex model forms (i.e., procedure 4) yields the smallest estimation errors.

Procedure 2, which utilizes a class of model forms consistent with the overall methodology, yields residual variances that are comparable to those of procedure 3 which utilizes a larger class of model forms.

Procedure 1, which differs from procedure 2 only in that it uses a stepwise regression parameter of 0.1 instead of 0.2, appears the least favorable of the four stepwise regression procedures in terms of estimation errors.

Pooled residual standard deviations for the individual wind components obtained from procedure 1 are about 0.08 mps larger than those for procedure 4 and about 0.04 mps larger than those for procedures 2 and 3.

In terms of predictions at the non-RTI-network stations (accuracy), procedure 4 is clearly less accurate than the other procedures (which tended to produce simpler models than those of procedure 4). The mean square errors over the entire set of seven non-network stations and over all cases are somewhat better for procedure 1 than for procedure 2 or 3; over the subset of five non-network stations in the interior portion of the St. Louis region, however, procedure 3 appears more favorable.

Over interior non-network stations, percentage errors for predicting wind speeds by procedures 1 and 2 averaged 23%, when data from the RTI network are utilized. This compares favorably with a corresponding error of 19% for procedure 4 applied to the full network of stations.

A subjective weighting to reflect the relative importance of estimation errors and prediction errors was utilized to judge the overall performance of the various estimation procedures. If the prediction errors are condisdered the more important of the two types, either of the two procedures consistent with the overall methodology (i.e., procedures 1 and 2) or procedure 3 may be considered "best" depending upon the particular criterion chosen (e.g., the particular weight chosen and the particular set of stations and/or cases considered). It is clear, therefore, that little improvement is achieved by expanding the class of models from the 13-term set up to the 23-term set of candidate terms.

Magnitudes of the average and pooled root mean square errors for procedures 1 and 2 -- across all non-network stations and all cases -- are roughly 0.1 to 0.2 mps larger than the corresponding pooled standard deviations over the RTI network.

Individual case studies, which serve to illustrate the analysis for several wind directions and wind speeds, indicated that procedure 2 performed better than procedure 1, and that the technique yielded estimates of the wind field that closely compared to the observed data.

Over the range of wind directions encountered, there was little change in the estimation error due to wind direction. Unfortunately only three major wind directions occurred with regularity, (i.e., flow from the southeast, south, and southwest). The variation of the errors with wind speed was also not substantial although larger absolute errors and smaller percentage errors tended to occur for cases with high wind speeds.

It was not possible to determine whether the 19-station RTI network was "the" optimum network for the city of St. Louis because there were not sufficient data or a sufficient number of auxiliary stations to test the 19-station network against all other possible networks. Furthermore, the theoretical phase showed that the network chosen for St. Louis is likely to be near-optimal only for the procedures used; if another procedure was used, it is likely that a different network would have been selected. The reliability of the wind field analysis will also depend on the results of the prediction of the air pollution analysis model, since the wind field is an input parameter to that model. However, the results of this study have, in our opinion, demonstrated that the methodology can be used to determine the locations of a reasonable number of stations from which wind data can yield reasonable wind field estimates over the domain of the network.

ANALYTICAL LIMITATIONS

The data available for demonstrating the wind field estimation procedures and for evaluating the sampling network have several limitations. These data had, due to economic constraints, a limited time span (i.e., a total of 33 days). Even within this period, there were large amounts unreported, invalid, and unusable data. For example, only 908 cases out of potential number of 1584 cases were usable; in terms of individual observations, less than 50% of the potential number were available. In addition, for 13 of the 16 RAPS stations, the winds at 10 m above ground had to be estimated from winds observed at 30 m above the ground level. The wind fields associated with the winter field program were atypical for that time of the year according to statistical analysis of wind data obtained from the National Climatic Center for the synoptic weather station (Lambert Field); the winter regime is generally characterized by northwesterly winds but, during the winter data collection period, southwesterly winds predominated.

The model/network evaluations were also limited by several practical constraints. First, only seven stations not in the RTI network furnished data for which comparisons between observed and predicted winds could be made. Secondly, comparisons involved in the evaluation typically had to be made in terms of absolute measures such as root mean squared errors rather than relative measures, since the "best" model was unknown and since only a relatively small number of potential network designs could be judged (i.e., those which were subsets of the full network). Furthermore, errors which arose from deficiencies in the network could not be isolated from those than were effected by other sources (e.g., measurement errors, model deificiencies, etc.).

The findings outlined above provide, in our opinion, an accurate assessment of the major results of this study; they are obviously made within the context of the limitations described above.

REMARKS

The evaluation of the RTI network was based on data obtained from the U.S. Environmental Protection Agency's Regional Air Pollution Study, the St. Louis City/County Air Pollution Network, and three stations set up by the Research Triangle Institute. Overall, there were 26 stations utilized, including the 19-station "optimum" network. Though the economic burden to obtain these data was significant, the data were not sufficient to make a complete evaluation of the network.

In the application of this technique for other cities, an evaluation of the network will certainly be necessary. It is unfeasible for future evaluations to face the same economic burden as the present evaluation. Nevertheless, after establishing the optimum network, a period should be set aside in which data are collected at the network stations and at locations not in the network. Non-network data can be collected by a mobile van during periods when the wind is in quasi-steady state. Case studies should be examined in which wind speeds and wind directions differ from case to case. The results of the case study analyses will yield estimates of the reliability of the network. This technique should also be applied in evaluation of the air pollution distribution obtained from the objective variational analysis model.

The 13-term class of model forms was used in the evaluation of the wind field in order to test and validate the methodology developed in the theoretical phase of this research project. Now that this aspect is complete and the results are positive, other surface fitting procedures for estimating the wind field should be investigated. For example, one procedure which would avoid the extrapolation problems of the polynomial models is gravitational-weighted (inverse of distance squared) interpolation. This approach uses only those data points close to the grid point for which a wind prediction is being made; generally, this allows extrapolation into locations a small distance outside the domain of the network without large error.

The wind analysis is an input parameter to the objective variational analysis model (OVAM) to be used to derive the air pollution distribution. The evaluation of that model will take place in the next phase and will utilize wind field predictions at selected grid points. Figure 9 provides on illustration (Case II, Procedure 2 in the last subsection of Section 4) of the predicted wind field as it would be used in the OVAM. The predicted winds at each grid point in the 20-km x 20-km area would be utilized as inputs. A grid spacing of 2 km is utilized in this figure.

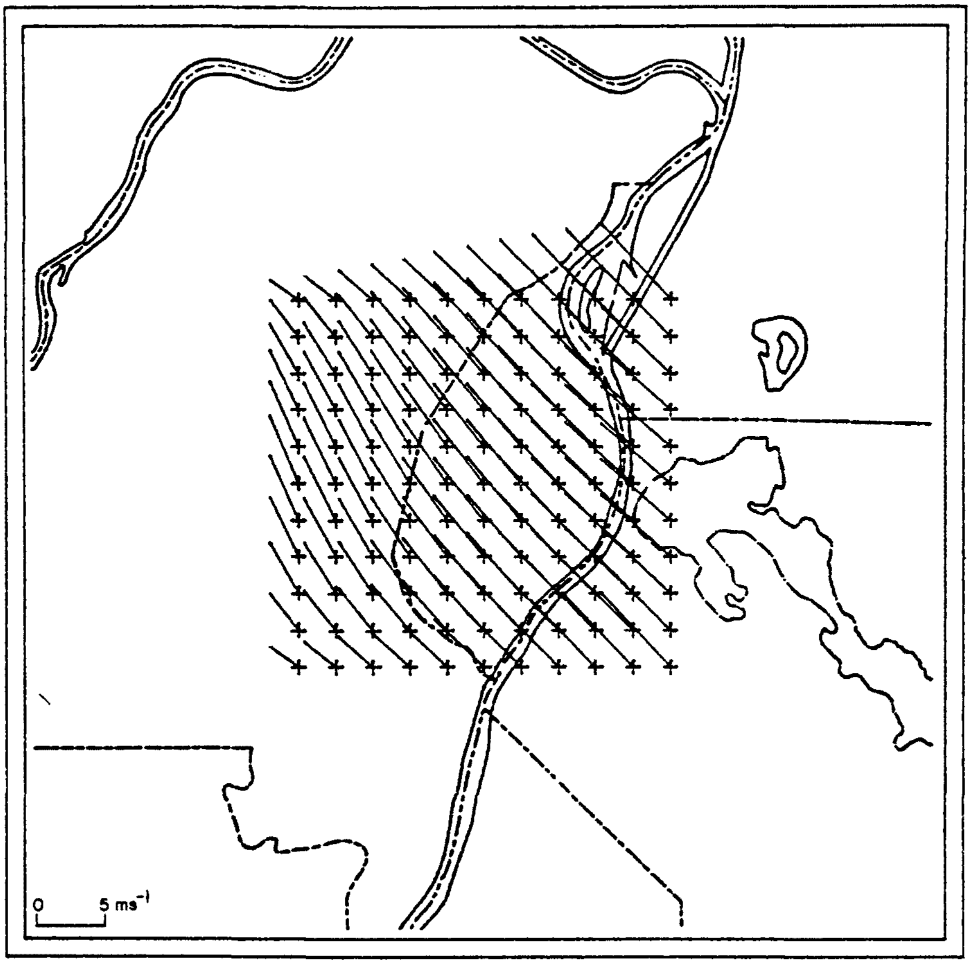


Figure 9. Distribution of predicted winds on a 2-km by 2-km grid for case II using procedure 2

REFERENCES

Barr, A.J., J.H. Goodnight, J.P. Sall, J.T. Helwig, 1976: A User's Guide to SAS - 76, SAS Institute, Sparks Press, Raleigh, N. C.

Draper, N.R. and H. Smith, 1966: Applied Regression Analysis, John Wiley and Sons, New York.

Estoque, M.A. and C.M. Bhumralkar, 1969: "Flow Over a Localized Heat Source", Monthly Weather Review, 97, 850-859.

International Mathematical and Statistical Libraries, 1975: IMSL Library Reference Manual.

Lettau, H., 1969: "Note on Aerodynamic Roughness-Parameter Estimation on the Basis of Roughness-Element Description", J. Appl. Meteor., 8, 822-832.

Vukovich, F.M., J.W. Dunn, and B.W. Crissman, 1976: "A Theoretical Study of the St. Louis Heat Island: The Wind and Temperature Distribution", J. Appl. Meteor., 15, 417-440.

Vukovich, F.M., W.D. Bach, Jr., and C.A. Clayton, 1978: "Optimum Meteorological and Air Pollution Sampling Network Selection in Cities, Volume I: Theory and Design for St. Louis", Environmental Monitoring Series, EPA-600/4-78-030.

Vukovich, F.M., J.W. Dunn, and W.J. King, 1979: Observations and Simulations of Diurnal Variation of the Urban Heat Island Circulation and Associated Ozone Variations: A Case Study. Submitted to J. Appl. Meteor.

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/4-79-069	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE OPTIMUM METEOROLOGICAL AND AIR POLLUTION NETWORK SELECTION IN CITIES: Volume II - Evaluation of Wind Field Predictions for St. Louis		5. REPORT DATE October 1979 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Fred M. Vukovich and C. Andrew Clayton		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Triangle Institute P.O. Box 12094 Research Triangle Park, North Carolina 27709		10. PROGRAM ELEMENT NO. 1HE775 11. CONTRACT/GRANT NO. 63-03-2187
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency--Las Vegas, NV Office of Research and Development Environmental Monitoring and Support Laboratory Las Vegas, NV 89114		13. TYPE OF REPORT AND PERIOD COVERED period ending Feb. 1979 14. SPONSORING AGENCY CODE EPA/600/07

15. SUPPLEMENTARY NOTES

This report is the second in a series on this topic (see EPA-600/4-78-030). For further information contact J.L. McElroy, Project Officer (702)736-2969, X241, Las Vegas

16. ABSTRACT

This report is the second in a series on the development of a method for designing optimum meteorological and air pollution sampling networks and its application for St. Louis, Missouri (see EPA-600/4-78-030). It involves the evaluation of the wind field network and utilizes wind data collected during special summer and winter field programs.

The evaluation considers the precision and accuracy of the procedure used for estimating the wind field. The basic procedure for determining the wind field involves applying stepwise regression to a class of linear statistical models involving subsets of 13 specific terms and data from a 19-station network; determined during the theoretical phase of the study. The evaluation includes the selection of a larger class of model forms and a basic set of 23 terms to compare with the 13-term class and includes estimations based on data from all reporting stations--up to a total of 26 stations.

The results demonstrate that application of 13-term modeling procedures to wind data from the 19-station network can produce predicted wind fields comparable to those produced by similar but more general procedures applied to a larger (26-station) network and that the method can objectively provide a reasonable estimate of the wind field over the domain of the network. An exhaustive evaluation was not feasible due largely to numerous analytical and data limitations.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
mathematical models wind field air pollution meteorology	sampling network St. Louis, Missouri	43F 55C 68A 72E

18. DISTRIBUTION STATEMENT RELEASE TO THE PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 114
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

U.S. GOVERNMENT PRINTING OFFICE 683-091/2209