A Statistical Approach to the Short-Period Prediction of Surface Winds[1]

JOHN A. RUSSO, JR., ISADORE ENGER AND EDNA L. SORENSON

The Travelers Research Center, Inc., Hartford, Conn.

(Manuscript received 26 August 1963, in revised form 13 November 1963)

ABSTRACT

The screening-multiple-regression technique is applied to predicting surface *u*- and *v*-wind components at Idlewild International Airport for periods of 2, 3, 5 and 7 hr. The predictors are variables from 11 synoptic stations, easily obtained or derivable from conventional service A teletype data. Additional predictors are used to account for diurnal and seasonal variations. In all, 141 predictors are screened and one prediction equation is obtained for each predictand. Each equation is applicable to any hour of the day and any day of the year.

The regression equations derived from a dependent sample selected randomly from 7 years of data proved significantly better at the 1-per cent level than both persistence and climatology for the 3-, 5- and 7-hr forecasts and at the 5 per cent level for the 2-hr forecasts when tested on 1387 independent cases. The screening-regression root-mean-square errors on this independent set ranged from 3.36 kt to 4.48 kt for the *u*-wind forecasts and from 3.69 kt to 5.57 kt for the *v*-wind forecasts.

Operational 3-, 5- and 7-hr surface-wind forecasts extracted from terminal forecasts made at Idlewild are compared both quantitatively and categorically with corresponding regression forecasts made on a new set of independent data. The screening-regression forecast errors are approximately ⅓ smaller than the subjective errors, and the improvements for all the predictands are statistically significant beyond the 1 per cent level. The categorical comparison concerning only categories of <10 kt and ≥ 10 kt (dictated by the format of the subjective data) resulted in Heidke skill scores of 0.399 for screening regression and 0.249 for the subjective forecasts when applied to 7-hr prediction of the surface-wind speed at Idlewild.

1. Introduction

As part of the weather-data-processing development program sponsored by the Federal Aviation Agency (Contract FAA/BRD-363), The Travelers Research Center has undertaken an extensive program to develop, test, and evaluate certain terminal-weather-forecasting techniques to determine their suitability in a highly-automated aviation weather system.

In this paper, the screening-multiple-regression technique is applied to 2-hr through 7-hr spot-time prediction of the *u*- and *v*-components of the surface wind at Idlewild International Airport (IDL). The technique was chosen because it is readily adaptable to machine computation and is known to be well suited to predicting continuous variables. Idlewild was chosen as the test station because of the large amount of data available there and at surrounding stations and because of its high volume of air traffic.

2. Basic definitions

A *predictand* is defined as a specific element at a specific station for a specific forecast length. For example, 2-hr east-west component at Idlewild is one predictand, 3-hr north-south component at Idlewild is another predictand, etc. There are eight predictands, consisting of the two elements *u* and *v* for each of four forecast lengths. The forecast lengths are 2, 3, 5 and 7 hr.

In general, a predictor variable may be extracted from any information available at the initial hour. In this study, all meteorological *predictors* are "spot-time" variables, defined in the same way as a predictand, and all are variables observed at or from the surface at the initial hour. There are two reasons for this constraint on the type of forecast information to be used. First, it was our aim to develop a method that could produce forecasts operationally every hour, and the only data available are the airways observations. Second, to determine the usefulness of "exotic" predictors (e.g., time or space gradients, advection, translation), it is necessary to know how good the simple predictors are by themselves. Exotic predictors were not included in this study but will be considered in future work.

3. The screening-regression technique

In many instances of multiple regression, most of the linear relationship that resides in a large set of possible

[1] This research was supported by the Federal Aviation Agency under Contract FAA/BRD-363.

predictors can be found in small subsets of these predictors. The screening-regression technique selects such a subset and then produces an ordinary multiple-regression equation between the predictand and the subset of selected predictors. The screening procedure has been described by Miller (1962).

4. Data description

The basic data consisted of a dependent (developmental) sample, covering the six years from May 1951 through April 1957 and an independent sample for the 1-yr period from May 1957 through April 1958. Standard hourly airways observations as punched from WBAN-10A and -10B forms were obtained from Asheville for each of the 11 airways stations shown in Fig. 1. Each station is within a reasonable distance from the predictand station with regard to advection and the forecast intervals. A random selection of 8208 hr was made from the first 6 yr, and 1387 hr from the last year. The predictors are listed in Table 1. Most of the variables are self-explanatory; however, the items *coded values*, *time of day*, and *day of year* require further clarification.

Because the screening-multiple-regression technique requires that all variables be in numerical form, values were assigned to unlimited ceilings and visibilities and to sky conditions such as broken and overcast. In addition, ceiling and visibility were coded in a manner compatible with the standard airways reporting system, to place more importance on low ceilings and visibilities. The coded values assigned to ceiling, visibility, sky conditions, and cloud amounts are given in Table 2.

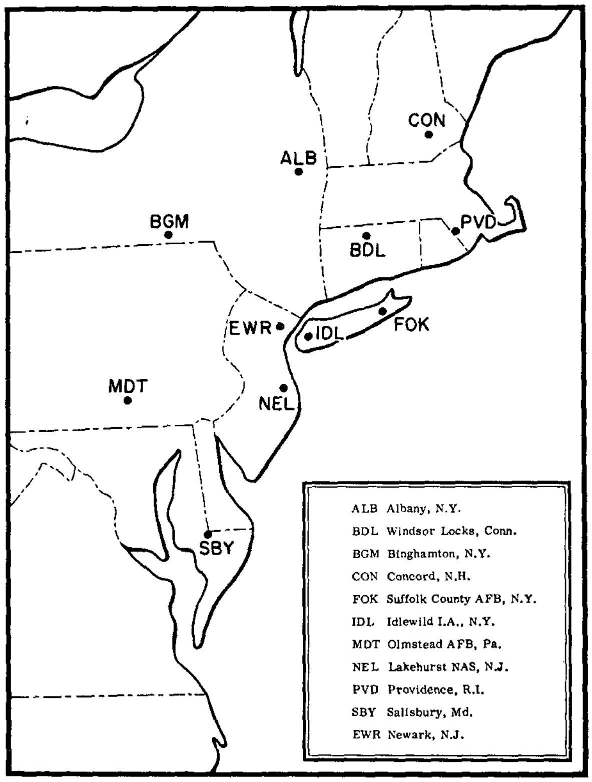

Fig. 1. Idlewild station network.

To compensate for diurnal and seasonal variations, time of day (TOD) and day of year (DOY) variables were used as predictors. This was accomplished by sub-

Table 1. Predictor variables.

Station	Meteorological element*												
	CIG	VIS	UWC	VWC	DBT	DPT	RLH	SLP	STP	SCL	SCU	TCA	OCA
IDL	x	x	x	x	x	x	x	x	x	x	x	x	x
FOK	x	x	x	x	x	x	x	x	x	x	x	x	—
BGM	x	x	x	x	x	x	x	x	x	x	x	x	—
EWR	x	x	x	x	x	x	x	x	x	x	x	x	—
NEL	x	x	x	x	x	x	x	x	x	x	x	x	—
ALB	x	x	x	x	x	x	x	x	x	x	x	x	—
PVD	x	x	x	x	x	x	x	x	x	x	x	x	—
BDL	x	x	x	x	x	x	x	x	x	x	x	x	—
CON	x	x	x	x	x	x	x	x	x	x	x	x	—
SBY	x	x	x	x	x	x	x	x	x	x	x	x	—
MDT	x	x	x	x	x	x	x	x	x	x	x	x	—

TOD $\sin x \cos x \sin 2x \cos 2x$

DOY $\sin z \cos z \sin 2z \cos 2z$

* CIG = ceiling height, coded value.
VIS = visibility, coded value.
UWC = east-west wind component, knots, W is positive.
VWC = north-south wind component, knots, S is positive.
DBT = dry-bulb temperature, F.
DPT = dew-point temperature, F.
RLH = relative humidity, per cent.
SLP = sea-level pressure, mb.
STP = station pressure, in.
SCL = sky condition of lowest cloud layer, coded value.
SCU = sky condition of second cloud layer, coded value.
TCA = total cloud amount, coded value.
OCA = opaque cloud amount, coded value.
TOD = time of day, see text description.
DOY = day of year, see text description.

TABLE 2. Coded values assigned to ceiling, visibility sky conditions, and cloud amounts.

Coded value	Ceiling* (CIG-ft)	Visibility* (VIS-mi)	Sky condition (SCL, SCU)	Cloud amount (TCA, OCA-tenths)
1	0	0	○	<1
2	100	$\frac{1}{16}$	−⦶ or −x	1
3	200	$\frac{1}{8}$	⦶ or +⦶	2
4	300	$\frac{1}{4}$	−⦷	3
5	400	$\frac{1}{2}$	⦷ or +⦷	4
6	500	$\frac{3}{4}$	−⊕	5
7	600	1	⊕ or +⊕	6
8	700	$1\frac{1}{2}$	x	7
9	800	2	—	8
10	900	3	—	9
11	1000	4	—	>9
12	1500	5	—	—
13	2000	6	—	—
14	2500	7	—	—
15	3000	8	—	—
16	5000	9	—	—
17	10,000	10	—	—
18	20,000	11	—	—
19	Unlimited	15	—	—

* The value entered is the lower limit; the upper limit is less than the lower limit of the next class.

TABLE 3. First ten predictors selected for IDL u- and v-wind-component forecasts.

Order of selection	Forecast length (hours) 2	3	5	7
	a) u-wind			
1st	IDL UWC	IDL UWC	IDL UWC	IDL UWC
2nd	EWR UWC	NEL UWC	BGM UWC	BGM UWC
3rd	MDT UWC	BGM UWC	CON STP	CON STP
4th	NEL UWC	SBY DBT	SBY SLP	SBY DBT
5th	SBY DBT	CON SLP	IDL SLP	SBY SLP
6th	BGM UWC	MDT STP	BGM DBT	IDL SLP
7th	PVD SLP	BDL SLP	MDT STP	MDT STP
8th	MDT STP	EWR UWC	ALB STP	ALB STP
9th	EWR RLH	EWR RLH	DOY sin x	BGM DBT
10th	BDL SLP	SBY SLP	BGM SLP	IDL RLH
	b) v-wind			
1st	IDL VWC	IDL VWC	IDL VWC	IDL VWC
2nd	EWR VWC	BGM VWC	BGM VWC	BGM VWC
3rd	NEL VWC	SBY RLH	TOD cos x	IDL OCA
4th	TOD cos x	NEL VWC	IDL OCA	TOD sin x
5th	ALB VWC	EWR VWC	TOD sin x	TOD cos x
6th	NEL RLH	TOD sin x	NEL VWC	DOY cos x
7th	TOD sin x	ALB DBT	DOY cos x	FOK UWC
8th	BGM DBT	IDL DPT	FOK UWC	TOD sin $2x$
9th	SBY DPT	TOD cos x	ALB VWC	ALB VWC
10th	BGM VWC	ALB VWC	TOD sin $2x$	MDT VWC

stituting four predictors for each of the two variables. The new predictors are: sin$x$, cos$x$, sin$2x$, cos$2x$ and sin$z$, cos$z$, sin$2z$, cos$2z$, where

$$x=\frac{h+1}{24}\times 360° \text{ for TOD,} \tag{1}$$

and

$$z=\frac{d}{365}\times 360° \text{ for DOY.} \tag{2}$$

h is the hour of the day in Eastern Standard Time; midnight is 00, and 11:00 p.m. is 23. d is the day of the year; 1 January is 1, and 31 December is 365. The use of these predictors is equivalent to fitting two harmonics to explain both the diurnal and annual cycles of the predictands.

5. Control techniques

The usefulness of a forecast technique may be partially determined by its ability to surpass the forecast skill obtainable by persistence or climatology. Two control techniques depicting the skill achievable by persistence and climatology were developed for comparison with the screening-regression technique. One control technique is persistence-regression:

$$\hat{y}_{i+f}=a_0+a_1y_i, \tag{3}$$

where f is the forecast interval, $\hat{y}_{i+f}$ is the predictand value, a_0 and a_1 are regression coefficients, and y_i is the observation of the predictand variable at the forecast time. The other control technique is persistence-plus-climatology:

$$\hat{y}'_{i+f}=a_0+a_1y_i+a_2\sin x+a_3\cos x+a_4\sin 2x+a_5\cos 2x +a_6\sin z+a_7\cos z+a_8\sin 2z+a_9\cos 2z, \tag{4}$$

where x is the TOD variable and z is the DOY variable and once again the a's are regression coefficients.

6. Dependent-data results

The screening-regression technique was applied to the dependent sample of data, and a set of predictors was selected for each of the eight predictands. The first 10 predictors selected are listed in Table 3. The best single predictor for both u and v for all four forecast lengths is the value of the element itself at forecast time. The second predictor is the predictand element at another station—in general, a close station for a short forecast length and a more distant station for a long forecast period. The most important subsequent predictors are wind and pressure information. The most frequently selected predictor stations are Idlewild and Binghamton (BGM). Binghamton is thought to be important because it is the first station affected by fronts moving in from Canada. In contrast with the u-wind-component predictors selected, the TOD variable is quite important in predicting the v-wind component. This reflects the importance of the sea breeze and its effect at Idlewild.

Three regression equations were obtained for each predictand, one by screening regression and two by the control techniques. Because of the inclusion of the sine and cosine representations of the TOD and DOY variables as predictors, all three equations are applicable to

TABLE 4. Dependent-data results and comparisons with control techniques.

Predictand		Screening regression			Persistence regression*		Persistence+climatology†	
Element	Forecast length (hr)	Number of predictors selected	rms error	Per cent reduction	rms error	Per cent reduction	rms error	Per cent reduction
u-wind	2	23	3.39	79.75	3.89	73.29	3.88	73.44
u-wind	3	21	3.75	75.55	4.40	66.29	4.38	66.58
u-wind	5	20	4.27	67.68	5.05	54.64	5.03	55.02
u-wind	7	26	4.70	61.34	5.68	43.42	5.65	44.03
v-wind	2	28	3.98	78.22	4.59	71.04	4.50	72.15
v-wind	3	32	4.47	72.42	5.24	62.16	5.10	64.11
v-wind	5	32	5.18	62.91	6.23	46.30	5.99	50.34
v-wind	7	19	5.99	51.03	6.98	33.47	6.68	38.98

* 1 predictor.
† 9 predictors.

any hour of the day and any day of the year. Comparisons among the three techniques on the dependent data are shown in Table 4. Included are the number of predictors selected by the screening technique, and the root-mean-square (rms) errors and reduction in variance for each of the eight predictands and each of the three techniques.

The ability of regression to forecast the u-wind component better than the v-wind component is reflected by the smaller rms error and the larger reduction in variance for the former at all four forecast lengths. A detailed investigation of the individual forecasts indicates that this is primarily caused by the sea-breeze effect at Idlewild. During late spring or summer, and primarily in the afternoon (when the East Coast is characterized by a weak pressure gradient), the wind shifts rapidly to a southerly (from-the-sea) component. The statistical technique at times is not capable of discerning this shift, resulting in large v-component errors.

As expected, the forecast accuracy decreases with increasing length for each of the techniques.

7. Verification on independent data

The regression equations for the screening and control techniques from the developmental sample were applied to the independent data sample of 1387 cases selected randomly from the period May 1957–April 1958. The results, including the rms errors for the three techniques and the average absolute errors for the screening technique, are given in Table 5. The average absolute error, a more meaningful statistic to the meteorologist, is presented to convey the magnitude of the errors involved in surface u- and v-wind component forecasts—information surprisingly rare in the meteorological literature.

The paired comparison t-test (Fisher, 1938) was applied to the square roots of the absolute values of the forecast errors. The screening-technique results for each of the predictands were found to be significantly better than those of either control technique at the 5 per cent

TABLE 5. Independent-data results and comparisons with control techniques.

Predictand		Screening regression		Persistence regression*	Persistence +climatology†
Element	Forecast length (hr)	Average absolute error	rms error	rms error	rms error
u-wind	2	2.56	3.36	3.91	3.90
u-wind	3	2.81	3.64	4.21	4.19
u-wind	5	3.24	4.22	5.06	5.05
u-wind	7	3.50	4.48	5.49	5.46
v-wind	2	2.78	3.69	4.20	4.13
v-wind	3	3.16	4.16	5.04	4.92
v-wind	5	3.81	4.94	6.01	5.82
v-wind	7	4.37	5.57	6.56	6.30

* 1 predictor.
† 9 predictors.

level for the 2-hr forecasts and at the 1 per cent level for the 3-, 5- and 7-hr forecasts. Absolute values were used to eliminate the signs of the errors and square roots were taken because they tend to be more normally distributed than the absolute errors themselves. The similarity between the independent- and dependent-data results is proof that the equations are quite stable and should remain so in any future applications.

Although the screening results have been shown to be better than those of either control technique, it remains to be shown that the forecast accuracies are competitive with those produced operationally. This may be accomplished by comparing the statistical forecasts with forecasts made under operational conditions.

8. Comparisons with subjective forecasts

Terminal forecasts (FT1's) prepared under routine operational conditions at Idlewild between 1 October 1960 and 30 September 1961 were collected and processed. The FT1 format ordinarily permits direct extraction of data, but simple interpolation is sometimes required to obtain forecasts for the appropriate forecast lengths. Two types of verification procedures were employed because an FT1 includes a wind forecast only when the wind is expected to equal or exceed 10 kt.

In one procedure, all FT1's lacking wind forecasts were eliminated. This left a sample consisting only of high-wind forecasts but not all high-wind occurrences. These data were converted to 3-, 5- and 7-hr u- and v-wind-component forecasts. The forecast times were 0500, 1100, 1700 and 2300 EST.

The screening-regression equations developed on the dependent sample were applied on this *new* independent sample. Predictions were made *only* for the identical cases for which subjective forecasts were available. The comparisons between the screening-regression forecasts and the subjective forecasts are presented in Table 6.

TABLE 6. Comparisons between the screening technique and subjective 3-, 5-, 7-hr forecasts of u- and v-wind components at IDL.

Predictand		Number of forecasts	rms error (kt)	
Element	Forecast length (hr)		Screening	Subjective
	3	283	4.02	6.34
u-wind	5	314	4.30	6.74
	7	297	5.08	7.74
	3	283	4.92	6.02
v-wind	5	314	5.56	6.37
	7	297	5.37	7.17

The rms errors listed show that the screening-regression forecasts are considerably better than the subjective forecasts. The screening errors are about one-third smaller than the subjective errors. The improvements for all six predictands are statistically significant beyond the 1 per cent level. Of interest, and as shown in Table 6, is the ability of the subjective forecaster to predict the v-wind component better than the u-wind component. This no doubt reflects the capability of the subjective forecaster to discern such things as the sea breeze and suggests a combination of man and machine during certain meteorological situations to attain an optimum forecasting procedure.

Again, because of the apparent rarity of error magnitudes of surface-wind forecasts in the meteorological literature, it was thought desirable to determine and present the magnitudes of the errors involved in predicting wind speeds alone. For the 297 cases (7-hr prediction) in which the subjective forecaster predicted wind speeds of 10 kt or more, his average absolute error was 5.11 kt, as compared with the regression technique's 3.94 kt.

The other evaluation procedure is concerned only with wind speed and was made on a categorical basis. The categories are <10 kt and ≥ 10 kt, dictated by the format of the FT1's and the fact that 10 kt is the critical speed relevant to the crosswind effect on light aircraft. The absence of a wind forecast in the FT1's is, in effect, a forecast of less than 10 kt. To compare the subjective and statistical forecasts, it was necessary to convert the statistical forecasts to categorical forecasts. This was done by a procedure analogous to that shown by Klein, Lewis and Enger (1959). Regression equations tend to "hedge" by not forecasting the extremes as often as they are observed. One method of correcting this tendency is to "inflate" the objective forecasts so that the variability of observed and predicted values is approximately the same. This is done by transforming the u- and v-forecasts:

$$u' = (u-\bar{u})/R_u + \bar{u} \tag{5}$$

and

$$v' = (v-\bar{v})/R_v + \bar{v}, \tag{6}$$

where R is the multiple-correlation coefficient between a predictand and its predictors and $\bar{u}$ and $\bar{v}$ are mean values. All correlations and means are taken from the dependent sample of data. A numerical wind-speed forecast is made by

$$s' = (u'^2 + v'^2)^{\frac{1}{2}} \tag{7}$$

and the s' values are then categorized into two classes, <10 kt and ≥ 10 kt.

Although it is easier to categorize the regression forecasts directly, i.e., use u and v in (7), this is not desirable. We prefer to use (7) as given, for then the number of cases forecast to be in a specific category will tend to agree with the number actually observed. On the other hand, regression minimizes the rms error; this places the emphasis on the size of the forecast error, which is immaterial provided that the correct category is forecast.

Contingency tables comparing the statistical and subjective wind-speed forecasts for the 7-hr forecast period are presented in Table 7. These results are representative of those obtained for the 3- and 5-hr forecasts.

The percentage of hits and the Heidke skill scores (Brier and Allen, 1951) shown in Table 7 indicate that the statistical forecasts are considerably better than the subjective forecasts.

9. Concluding remarks

It is concluded that 2- through 7-hr forecasts of the surface wind at Idlewild by screening multiple regression are reasonably good. The statistical forecasts compare favorably with subjective forecasts produced under operational conditions. However, a large sample of subjective "spot-time" surface-wind forecasts is needed to obtain a more valid evaluation of the statistical procedure. It would also be useful to test the procedure at stations in other climates.

It is felt that improvement in the forecast accuracy is possible by pursuing two lines of investigation—statistical and meteorological. Concentration on methods that

TABLE 7. Contingency tables for the 7-hr prediction of IDL wind speed in two fixed categories (<10 kt, ≥ 10 kt).

Forecast	Observed <10 kt	Observed ≥ 10 kt	Total forecast
	(a) Subjective		
<10 kt	319	165	484
≥ 10 kt	263	389	652
Total observed	582	554	1136
Hits = 708 Percentage of hits = 62 Heidke skill score = 0.249			
	(b) Screening regression		
<10 kt	423	182	605
≥ 10 kt	159	372	531
Total observed	582	554	1136
Hits = 795 Percentage of hits = 70 Heidke skill score = 0.399			

consider nonlinear relationships between the predictand and predictors will be the statistical approach to improving the forecast accuracies obtained in this study. Methods to be included in future studies are curvilinear regression, canonical correlation, and multiple-discriminant analysis.

A preliminary study associating the screening-regression forecast errors with synoptic situations has shown much promise in typing the meteorological situations most prone to errors. Expansion of this study, with the objective of introducing these large error-producing situations (e.g., fronts or the sea breeze) as derived predictors, is contemplated for future studies.

The present study used only simple variables as predictors. It is felt that the results can be improved by applying meteorological reasoning and experience to introduce "richer" predictors, such as time changes, gradients, translation vectors, and advection terms.

Acknowledgments. The authors wish to thank Dr. Robert G. Miller and Dr. Joseph G. Bryan for their helpful advice and guidance. Thanks are also due Miss Ann L. Bussemey and Mr. Robert L. Metcalf for their valuable assistance throughout this study, and to Mrs. Edward A. Miller for typing the manuscript.

REFERENCES

Brier, G. W., and R. A. Allen, 1951: Verification of weather forecasts. *Compendium of meteorology*, Boston, Amer. Meteor. Soc., 841–848.

Fisher, R. A., 1938: *Statistical methods for research workers.* 7th edition, Edinburgh, Oliver and Boyd, 356 pp. (see Chap. V.)

Klein, W. H., B. M. Lewis and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.*, **16**, 672–682.

Miller, R. G., 1962: Statistical prediction by discriminant analysis. *Meteor. Monogr.*, **4**, 25, Boston, Amer. Meteor. Soc., pp. 45–46.