U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 79-9

VERIFICATION OF GREAT LAKES MOS WIND FORECASTS

Lawrence D. Burroughs

March 1979

VERIFICATION OF GREAT LAKES MOS WIND FORECASTS

Lawrence D. Burroughs

1. INTRODUCTION

Generalized wind forecasting equations have been applied to 12 sectors of the Great Lakes (see Fig. 1). Forecasts from these Model Output Statistics equations were verified with and without the use of inflation on 5 months of independent data. Scores computed were mean absolute error (MAE), root mean square error (RMSE), bias, and skill score based on chance. Contingency tables for eight wind direction and six wind speed categories were also computed. We found that inflating the forecasts improved the skill score and bias but made the MAE and RMSE worse. Further, we found the MAE of the wind direction forecasts in western Lake Superior to be abnormally high. To determine the reasons, studies were done which showed that a climatological preference for NE and SW winds exists in that sector and that this preference diminishes from April through December. The preference and its variability appear to be related to topography and seasonal differences in the positions of the principle cyclone and anticyclone tracks with relation to Lake Superior. New predictors which may help explain some of the variance due to these factors are described, and recommendations for changing the equations are made.

2. DISCUSSION

A. Verification

In deriving the generalized set of equations for U, V, and S, only the wind with the highest speed available was used for any sector for any given synoptic time. To verify the equations, independent data from the 1977 summer season (April-August) were used. September data were not used because the equations were derived using LFM-I predictors, and the LFM-II became operational in September 1977.

1) Verification Measures

Verification was done using four measures: skill score against chance, MAE, RMSE, and bias. The skill score is given by

$$S = (R-E)/(T-E), \qquad (1)$$

where

$$E = (\Sigma R_{it} C_{it})/T \qquad (2)$$

is the number expected to be correct based on chance, R is the number of correct forecasts, T is the total number of forecasts, R_{it} is the total of the i-th row of the forecast-observed contingency table and C_{it} is the total of the i-th column of the same table (Panofsky and Brier, 1963). Wind directions have been divided into eight equal categories of 45° centered on the compass points used. These categories are:

1-N, 2-NE, 3-E, 4-SE, 5=S, 6-SW, 7-W, 8-NW.

Wind speeds have been divided into six categories:

1) $C_1 \leq 3$ kt	4) $17 < C_4 \leq 33$ kt
2) $3 < C_2 \leq 10$ kt	5) $33 < C_5 \leq 47$ kt
3) $10 < C_3 \leq 17$ kt	6) $C_6 > 47$ kt

The larger the skill score the better. A skill score of 1 is perfect. $S \leq 0$ implies no skill against chance. The bias is defined as the ratio of the total number of forecasts to the number of observations of the event. The nearer the bias is to 1 the better.

2) Inflation Transformation

According to Glahn (1978), inflation was proposed by Isadore Enger and first applied by Klein et al. (1959). Its purpose is to bring the bias of continuous variables closer to 1 and to increase the probability of any regression equation hitting rare events. The inflation transformation for wind speed is given by

$$\hat{S}_w = [(\hat{S}-\bar{S})/R] + \bar{S} \quad (3)$$

where $\hat{S}$ is the original objective estimate of the speed, $\bar{S}$ is the mean value for the speed predictand from the developmental sample, R is the multiple correlation of the speed predictand with the predictors in the forecasting equation, and $\hat{S}_w$ is the inflated forecast of the wind speed (NWS, 1978).

3) Verification Results

We verified the equations with and without inflation. Currently we do not use inflation. We wanted to see if our skill would improve. Tables 1 through 5 give the results of these verifications. Table 1 shows the sample size for each lake sector shown in Fig. 1 and the total of all sectors, the skill score of the uninflated forecasts, and the skill score of the inflated forecasts. Table 2 shows a comparison of the MAE's for the uninflated and inflated forecasts for each sector. Table 3 shows a comparison of the RMSE's for the uninflated and inflated forecasts, and Table 4 shows a comparison of the biases of the uninflated and inflated forecasts. The verification was done for the 18-h and 30-h projections for both 0000 GMT and 1200 GMT cycles. Only the verification for the 0000 GMT cycle 18-h projection is shown. Discounting sectors 1, 2, and 3 which have little or no data, Table 1 shows that inflation improves the skill in most sectors and overall. Also from Table 4 we see the bias is improved, but from Tables 2 and 3 we see the MAE and RMSE enlarge. Glahn and Allen (1966) point this out as an inevitable consequence of using inflation. But if we accept that improved skill and a bias closer to one are desirable, then we must also accept a RMSE and MAE that are larger--here about 14 percent greater, on the average--when using inflation.

Table 6 is taken from Table 5 and shows the number of correct forecasts for both uninflated and inflated wind speed forecasts for each category. Table 7 shows the same set of data normalized by the total number of correct forecasts for each row and multiplied by 1000. Table 8 shows the difference between normalized uninflated and inflated figures from Table 7. The categorical mean for these data is category 3, and we see from Table 8 that 225 hits are taken from the mean category and distributed as hits in other categories and that 63 of those occur in the small craft advisory (category 4) or gale warning (category 5) categories. Essentially this means there are 39 percent fewer hits at the categorical mean. This, we believe, is an acceptable trade-off since it is of more importance to forecast the rarer small craft advisory, gale, and storm warning events, than winds at the categorical mean.

A close examination of Table 2 revealed that the MAE for sector 12, the western sector of Lake Superior, is high compared to the mean of all the sectors. In fact, the difference is 67° compared with 50°. This seems abnormal and resulted in an examination of the climatology and topography of Lake Superior to determine the reason for this difference.

B. Other Studies Resulting from Verification

1) Topographic and Subsynoptic

The orientation of the western sector of Lake Superior is generally NE-SW. There are steep hills and cliffs along the sides ranging in height from about 240 m to 400 m above the surface of the lake. At the western-most end, near Duluth, Minn., and Ashland, Wisc., are valleys which tend to funnel SW winds. Lake breezes tend to veer in the direction of the lake axis because of the steep cliffs and the difference in roughness between lake and land; while land breezes tend to back in the direction of the lake axis due primarily to the difference in roughness between land and lake. In addition to these topographical and subsynoptic effects, other factors such as the tracks of synoptic systems in relation to Lake Superior, may help to explain the large MAE in the western sector of Lake Superior.

2) Climatologically Preferred Wind Directions

To determine if there is a climatological preference for NE and SW winds in western Lake Superior, the frequencies of wind direction along four axes (NE-SW, N-S, E-W, NW-SE) were normalized by the frequencies along the same axes for Lake Superior as a whole. The frequencies were determined from a climatological summary of surface observations over Lake Superior by the National Climatic Center (1975) hereafter referred to as the summary. The data in this summary are divided into four sections, west (W), west central (WC), east central (EC), and east (E); whereas our (TDL) division of Lake Superior is west (w), central (c), and east (e). To conform the summary data to our sector demarcation, weighted mean frequencies were created by taking W + WC/2 = w, WC/2 + EC/2 = c, and EC/2 + E = e.

The weighting for a particular TDL sector for a particular time (t) is given by

$$F_{(\ )t} = (n_{at}\ f_{at} + n_{bt}\ f_{bt})/(n_{at} + n_{bt}), \tag{4}$$

where $F_{(\ )t}$ is the weighted mean frequency for a particular wind direction for a particular TDL lake sector, $n_{(\ )t}$ is the number of observations in summary sector ( ) or portion thereof, and $f_{(\ )t}$ is the mean frequency for a particular wind direction for a particular summary sector or portion thereof. Likewise the weighted mean frequency for a particular wind direction for a particular time for Lake Superior (LS) as a whole is given by

$$F_{LSt} = \frac{n_{Wt}\ f_{Wt} + n_{WCt}\ f_{WCt} + n_{ECt}\ f_{ECt} + n_{Et}\ f_{Et}}{n_{Wt} + n_{WCt} + n_{ECt} + n_{Et}} \tag{5}$$

The weighted mean frequencies of the wind directions along a particular axis were then added together, i.e., the frequency of NE winds and SW winds for a particular time and sector added give the frequency of winds along the NE-SW axis.

The normalized frequency for a particular axis for a particular sector and time is given by

$$N_{(\ )t} = F_{(\ )t}/F_{LSt}, \tag{6}$$

where $N_{(\ )t}$ is the normalized frequency along a particular axis for a particular TDL sector and time. If we assume there is one wind direction climatology for Lake Superior as a whole, then when there is no climatological preference for a particular axis for a particular time, $N_{(\ )t} = 1$. If $N_{(\ )t} > 1$, then there is a preference for winds along that axis. If $N_{(\ )t} < 1$, then there is a preference for winds along other axes. I have no idea what a significant preference might be, and at this point there is no readily available means of testing for significance, but if we arbitrarily choose a figure say $\pm$ 0.10, then we will consider $N > 1.10$ or $N < 0.90$ significant.

Table 9 shows the normalized weighted mean annual frequencies along each axis at each synoptic time for each TDL sector. In western Lake Superior there are significant preferences for winds along the NE-SW axis and away from the N-S and NW-SE axes. No significant preferences are evident in central Lake Superior; while in eastern Lake Superior preferences for winds away from the NE-SW axis and along the N-S and NW-SE axes are evident. The verification data in Tables 2 and 3 indicated that the generalized equation handles the preferences in eastern Lake Superior fairly well in that the MAE and RMSE are close to the averages for the 11 sectors excluding western Lake Superior.

Table 10 shows the normalized weighted mean frequencies for winds along the NE-SW axis over western Lake Superior. Monthly means for April through December are shown along with seasonal means for summer (April-September) and winter (October-December) and the annual mean for April through December. No data are available for January through March. Since the winter months account for only 27 percent of the total number of observations, the annual averages show little influence from the winter data. Inspection of the monthly frequencies shows their values decrease from April through December. Thus, the climatological preference for winds along the NE-SW axis over western Lake Superior varies from being highly significant in April to having no significance in December.

Table 11 gives the normalized monthly weighted mean frequencies of winds along all four axes for western Lake Superior. These frequencies are for observations at all times taken together. Table 12 gives the corresponding monthly normalized weighted mean wind speeds along all axes for western Lake Superior. These speeds are normalized with respect to the annual weighted mean speed for each axis for Lake Superior as a whole, rather than with respect to the monthly weighted mean speed as was done for the normalized weighted mean frequencies. The weighted mean speed, $S\mu$, for a particular axis for a particular summary lake sector is given by

$$S\mu = (f_{(a)}\, s_{(a)} + f_{(b)}\, s_{(b)})/f_{(a)} + f_{(b)}), \tag{7}$$

where $f_{(\ )}$ is the frequency of a particular wind direction and $s_{(\ )}$ is the mean speed when the wind is from a particular direction. For example, if the frequency of north winds in the western summary sector (W) in April is 0.110, and the mean speed for north winds in April is 13.2 kt, and if the frequency of south winds for April is 0.073, and the mean speed for south winds in April is 11.8 kt, then the weighted mean speed, S_W, for winds along the N-S axis is 12.6 kt, and the combined frequency is 0.183. The weighted mean speed $S_{(\ )}$, for a particular TDL sector of Lake Superior is given by

$$S_{(\ )} = (S\mu_{(a)}\, n_{(a)} + S\mu_{(b)}\, n_{(b)})/(n_{(a)} + n_{(b)}), \tag{8}$$

where $S\mu_{(\ )}$ is defined by eq (7) and $n_{(\ )}$ is the number of observations in summary sector ( ). The annual weighted mean speed, S_A, is given by

$$S_A = \sum_{i=1}^{9} (S_{\mu Ei}\, n_{Ei} + S_{\mu ECi}\, n_{ECi} + S_{\mu Wi}\, n_{WCi} + S_{\mu Wi}\, n_{Wi})/(n_{Ei} + n_{ECi} + n_{WCi} + n_{Wi}), \tag{9}$$

where i is the month, S_μ is defined by eq (7) and $n_{(\ )i}$ is the number of observations for a particular month for a particular summary sector ( ). The normalized mean speed, N_s, for a particular month is given by

$$N_s = S_{(\ )}/S_A. \tag{10}$$

3) Cyclone and Anticyclone Tracks--Cyclogenesis and Anticyclogenesis

Klein (1957) shows the principle tracks of cyclones and anticyclones in the Northern Hemisphere. In addition he shows the frequencies of cyclogenesis and anticyclogenesis for the area. Since wind and frequency data are not available for January through March, I will not make any comments about those months.

In April a principle cyclone track ends over east Lake Superior while principle anticyclone tracks are well to the north and south. Very little cyclogenesis or anticyclogenesis takes place in the vicinity. The position of the principle cyclone track would tend to enhance the frequency of winds along the longitudinal axis of the western Lake Superior (see Table 11).

In May and June the principle cyclone tracks are located far to the north and to the east southeast of the lake, while the principle anticyclone tracks move toward the lake from the north and the south. The wind speeds decrease (see Table 12) and no significant cyclogenesis or anticyclogenesis occurs. The position of the principle anticyclone and cyclone tracks allows for some variability in the wind direction, but the frequency of winds along the longitudinal axis of western Lake Superior remains enhanced though less so than in April.

In July and August no primary cyclone tracks exist near the lake while two principle anticyclone tracks converge in time into one track over Lake Superior in August. A frequency maxima of anticyclogenesis exists over the lake in August as well. The wind speeds continue to diminish through August. These factors tend to increase the variability of the wind direction somewhat, but the weakened wind field tends to enhance the land-lake breeze effect and the frictional difference effect. These mesoscale effects counterbalance the synoptic effect somewhat to allow a continued preference for winds along the NE-SW axis of western Lake Superior, but decreased in amplitude.

In September there are no primary cyclone tracks; while there is an anticyclone track just to the south of the lake. No significant cyclogenesis or anticyclogenesis takes place, but the wind speeds do begin increasing.

In October and November the principle anticyclone track moves well south of the lake, while a principle cyclone track to the north and one to the southeast of the lake converges in time. No significant cyclogenesis or anticyclogenesis occurs, and wind speeds continue to increase over the mean for the year. These synoptic factors taken together point to the more even distribution of wind frequencies along all axes over western Lake Superior shown in Table 11.

In December the principle cyclone tracks mentioned above end just to the north and southeast of the lake and a new track begins just to the east. The principle anticyclone tracks lie well to the south and somewhat to the north of the lake. An area of cyclogenesis occurs over the lake, and wind speeds continue to increase. These factors point to the very even distribution of wind frequencies along all axes shown in Table 11.

Examination of the synoptic patterns and subsynoptic patterns over and near Lake Superior, as they vary from month to month, has given some answers to the question raised by the annual variation in the distribution of wind frequencies along all axes over western Lake Superior.

3. CONCLUDING REMARKS

By close scrutiny of the verification data for the Great Lakes wind equations and climatological data for western Lake Superior, two immediate conclusions can be drawn: (1) the use of inflation tends to give higher skill scores on independent data, and (2) two sets of equations should be used--one for western Lake Superior and one for the other sectors combined.

We have already created and tested such equations and have found the 11 sector generalized equation with inflated speed to give better skill scores than its 12 sector counterpart. However, the single station equation for western Lake Superior still has an abnormally high MAE in the wind direction forecast. To counteract this, we have created two sets of predictors. One set creates a wind direction at any level from the U and V components, subtracts the angle of the lake axis from the created wind direction, and computes the absolute value of the cosine of the resultant angle. If the wind direction is along the axis, the predictor will be one and vary to zero when the wind direction is perpendicular to the lake axis. This, we believe, will help to account for the climatological preference for winds along the NE-SW axis in western Lake Superior. Predictors in the other set are interactive in that they multiply the cosine of the day of the year by the predictor described above. This, we believe, will help to explain the seasonal variation in the climatological preference. Neither predictor set has been tested as yet, and no new implementation will take place until they are.

ACKNOWLEDGEMENTS

I'm deeply grateful to Mr. N. A. Pore and Dr. H. R. Glahn for their helpful comments, to Ms. Mercedes N. Bakon and Ms. Mary B. Battle for typing the manuscript, and to Mr. W. C. Herrmann for drafting the figure.

REFERENCES

Glahn, H. R. and R. A. Allen, 1966: A note concerning the "inflation" of regression forecasts. J. Appl. Meteor., Vol. 5, pp. 124-126.

______, 1978: Computer worded public weather forecasts. NOAA Technical Memorandum, NWS-TDL-67, NOAA, National Weather Service, U.S. Department of Commerce, 24 pp.

Klein, W. H., 1957: Principal tracks and mean frequencies of cyclones and anticyclones in the northern hemisphere. U.S. Weather Bureau Research Paper, No. 40, 60 pp.

______, B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. J. Meteor., 16, pp. 672-682.

National Climatic Center, 1975: Summary of Synoptic Meteorological Observations for Great Lakes Areas, Vol. 4, Lake Superior. NOAA, U.S. Department of Commerce, Asheville, N.C., 244 pp.

National Weather Service, 1978: The use of model output statistics for predicting surface wind. Tech. Proc. Bull., No. 229, NOAA, National Weather Service, U.S. Department of Commerce, 12 pp.

Panofsky, H. A. and G. W. Brier, 1963: Some Applications of Statistics to Meteorology. The Pennsylvania State University, University Park, Pa., 224 pp.

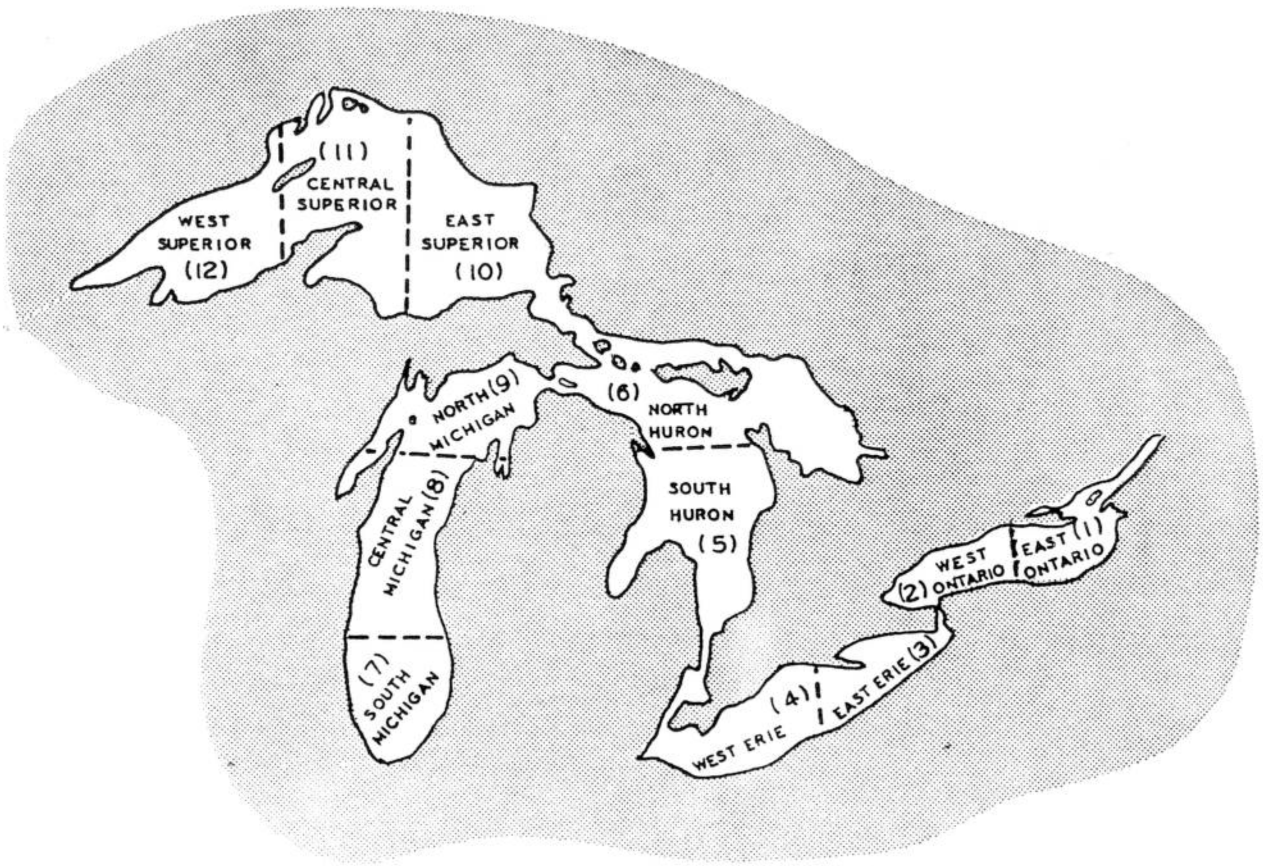

Figure 1. Location of the 12 Great Lakes Sectors for which wind forecasts are made.

Table 1. Skill scores and sample sizes for 18-h 0000 GMT cycle wind speed forecasts.

Lake Sector	Skill Scores		Sample Size
	Uninflated Forecasts	Inflated Forecasts	
1	0	0	0
2	0	0	1
3	.296	.048	10
4	.128	.192	36
5	.243	.270	108
6	.244	.229	66
7	.319	.134	25
8	.078	.219	64
9	.159	.191	50
10	.258	.258	123
11	.107	.243	98
12	.163	.193	104
Total	.191	.230	685

Table 2. Mean absolute errors for 18-h 0000 GMT cycle forecasts.

Lake Sector	Speed Forecasts (kt)		Direction Forecasts (°)
	Uninflated	Inflated	
1	0.0	0.0	0
2	2.0	1.0	30.0
3	3.9	5.2	31.1
4	4.2	4.1	44.1
5	4.4	4.5	47.0
6	4.1	5.3	54.2
7	4.4	4.7	37.6
8	4.5	5.0	47.2
9	4.4	5.3	50.8
10	4.4	5.2	44.7
11	4.4	4.8	48.5
12	4.7	5.4	66.7
Total	4.4	5.0	50.0

Table 3. Root mean square errors for 18-h 0000 GMT cycle wind forecasts.

Lake Sector	Speed Forecasts (kt)		Direction Forecasts (°)
	Uninflated	Inflated	
1	0.0	0.0	0
2	2.0	1.0	30.0
3	5.8	6.0	39.7
4	4.9	5.3	59.0
5	5.5	5.6	66.8
6	5.9	6.9	73.2
7	4.8	5.6	53.3
8	5.7	6.6	63.5
9	5.6	6.6	66.7
10	5.7	6.6	61.2
11	5.3	6.0	61.4
12	5.9	7.3	83.0
Total	5.6	6.4	67.0

Table 4. Bias by category for 18-h 0000 GMT cycle wind speed for all sectors combined.

Speed Category	Uninflated	Inflated
C_1	0.00	2.42
C_2	0.50	1.11
C_3	1.85	1.08
C_4	0.59	0.75
C_5	0.00	0.28
C_6	--	--

Table 5. Contingency tables for 18-h 0000 GMT cycle wind speed forecasts for all sectors combined.

a. Uninflated Forecasts

Observed	C_1	C_2	C_3	C_4	C_5	C_6	Total
C_1	0	6	6	0	0	0	12
C_2	0	57	140	8	0	0	205
C_3	0	27	187	34	0	0	248
C_4	0	12	124	77	0	0	213
C_5	0	0	1	6	0	0	7
C_6	0	0	0	0	0	0	0
Total	0	102	458	125	0	0	685

b. Inflated Forecasts

Observed	C_1	C_2	C_3	C_4	C_5	C_6	Total
C_1	1	8	3	0	0	0	12
C_2	20	109	62	13	1	0	205
C_3	7	81	115	45	0	0	248
C_4	1	79	86	97	0	0	213
C_5	0	0	1	5	1	0	7
C_6	0	0	0	0	0	0	0
Total	29	227	267	160	2	0	685

Table 6. Number of correct forecasts by category for 18-h 0000 GMT cycle wind speed forecasts for all sectors combined (summarized from Table 5).

Speed Category	Uninflated	Inflated
C_1	0	1
C_2	57	109
C_3	187	115
C_4	77	97
C_5	0	1
C_6	0	0
Total	321	323

Table 7. Number of correct forecasts by category for 18-h 0000 GMT cycle wind speed forecasts for all sectors combined found in Table 6 normalized by the total and multiplied by 1000.

Speed Category	Uninflated	Inflated
C_1	0	3
C_2	178	337
C_3	582	357
C_4	240	300
C_5	0	3
C_6	0	0
Total	1000	1000

Table 8. Difference between the uninflated and inflated normalized correct wind speed forecasts by category for 18-h 0000 GMT cycle for all sectors combined (from Table 7).

Speed Category	Difference
C_1	+3
C_2	+159
C_3	-225*
C_4	+60
C_5	+3
C_6	0
Total	0

*Category of the mean.

Table 9. Weighted mean annual frequency of winds along the axes and at the time indicated for each sector normalized by the weighted mean annual frequency for Lake Superior.

Axis	0000 GMT	0600 GMT	1200 GMT	1800 GMT
		Western Superior		
NE-SW	1.33	1.15	1.17	1.35
N-S	.83	.85	.87	.85
E-W	1.00	1.07	1.04	1.05
NW-SE	.84	.93	.92	.75
		Central Superior		
NE-SW	.93	.95	.97	.96
N-S	1.04	1.08	1.06	1.08
E-W	1.07	1.00	1.00	.99
NW-SE	.96	.97	.97	.97
		Eastern Superior		
NE-SW	.75	.90	.86	.69
N-S	1.12	1.08	1.07	1.08
E-W	.94	.93	.95	.96
NW-SE	1.19	1.09	1.12	1.27

Table 10. Weighted mean frequency of winds along the NE-SW axis of western Lake Superior normalized by the weighted mean frequency for Lake Superior.

Period	Hour (GMT)			
	0000	0600	1200	1800
APR	1.57	1.35	1.28	1.43
MAY	1.52	1.24	1.27	1.48
JUN	1.45	1.26	1.28	1.51
JUL	1.51	1.20	1.22	1.42
AUG	1.39	1.13	1.20	1.38
SEP	1.35	1.16	1.06	1.39
OCT	1.35	1.08	1.18	1.32
NOV	1.07	.96	.92	1.05
DEC	1.00	.95	1.05	1.05
APR-SEP (Summer Avg)	1.42	1.20	1.22	1.44
OCT-DEC (Winter Avg)	1.19	1.01	1.07	1.17
APR-DEC (Annual Avg)	1.33	1.15	1.17	1.35

Table 11. Normalized weighted mean frequencies of winds along the axes indicated for western Lake Superior for all synoptic hours combined.

Period	Axis			
	N-S	NE-SW	E-W	NW-SE
APR	.83	1.39	.97	.62
MAY	.92	1.39	1.00	.68
JUN	.85	1.37	.94	.67
JUL	.88	1.36	1.02	.74
AUG	.84	1.27	1.03	.74
SEP	.84	1.25	1.09	.82
OCT	.79	1.21	1.14	.86
NOV	.96	1.10	.92	1.08
DEC	.96	1.01	1.02	1.01

Table 12. Normalized weighted mean wind speeds along the axes indicated for western Lake Superior for all synoptic hours combined.

Period	Axis			
	N-S	NE-SW	E-W	NW-SE
APR	.90	1.12	1.12	.92
MAY	.88	1.01	.93	.86
JUN	.79	.96	.84	.76
JUL	.78	.87	.84	.74
AUG	.75	.90	.83	.73
SEP	.94	.94	1.00	.95
OCT	1.08	1.10	1.10	1.10
NOV	1.15	1.19	1.17	1.22
DEC	1.26	1.25	1.37	1.33