1964-04

The Evaluation of Several Cyclone Prediction Techniques

Borsting, J.R.; Sheehan, F.F.

Journal of Applied Meteorology, Volume 3, pp. 132-135.
http://hdl.handle.net/10945/45786

The Evaluation of Several Cyclone Prediction Techniques

J. R. BORSTING
U. S. Naval Postgraduate School, Monterey, Calif.
AND F. F. SHEEHAN
San Francisco State College, San Francisco, Calif.

(Manuscript received 21 October 1963)

ABSTRACT

A comparison is made of several schemes for making predictions of the location and intensity of cyclones. One scheme (due to K. W. Veigas and F. P. Ostby) is based on a linear model developed by a stepwise regression technique; another (due to J. J. George and P. M. Wolff) is based on an empirical study of linear and non-linear relations between selected meteorological variables. A third scheme, making use of climatological averages, was developed to use as a basis for determination of skill. The prediction methods are compared on the basis of operationally meaningful criteria. The climatological procedure gave consistent, reasonably accurate results over the entire Northern Hemisphere. The procedure of George and Wolff was superior to climatology throughout the hemisphere. No significant difference in accuracy was detected between the Wolff-George and the Veigas-Ostby procedures in the region for which Veigas-Ostby was developed (United States East Coast).

1. Introduction

The U. S. Naval Fleet Numerical Weather Facility at Monterey, Calif., has for some time been making 12-, 24-, 36-, and 48-hr predictions of cyclone location and pressure changes. Several different schemes using different statistical techniques such as graphical correlation or stepwise regression have been used. Each scheme that has been employed is objective in the sense that it is programmed for computer operation. In this paper, the schemes are compared, using several different criteria.

2. Basic techniques

The three basic prediction techniques considered are described in this section. The Wolff-George scheme (WG) as described in George (1960) is a mathematical model based on an empirical investigation of relations between various meteorological variables. The prediction is accomplished in two steps: First the categorical forecast of deepening or non-deepening of a storm is made; then the amount of change is predicted. This method is basically a graphical correlation technique and is not restricted by any assumption concerning the functional form of the relations.

The second technique considered (designated VO) is one developed by Veigas and Ostby (1963). It consists of a set of three mathematical models for predicting latitude, longitude and pressure. The linear equations that define the VO models were developed for use on the United States East Coast. The small number (5–10) of meteorological variables used in these models were selected from a collection of several hundred by the well known step-wise regression method. The unique features of this application of step-wise regression were the use of a moving coordinate system and Miller's cut-off procedure (Miller, 1962).

A third technique based on climatology was introduced as a basis for the assessment of skill. The climatological forecast procedure was based on a study of ocean cyclones.[1] In the AROWA report, charts illustrating the average 24-hr behavior of cyclones over a major portion of the Northern Hemisphere are presented for each month. The charts are based on ten-years data from 1929 through 1938. The climatological procedure is as follows:

1) Locate initial position of the storm;
2) Find initial position on corresponding monthly chart;
3) Predict 24-hr location by reading average eastward movement in degrees longitude and average direction of movement from the chart.

With respect to pressure, the AROWA charts are partitioned into two categories: deepening storms and filling storms. Thus, to use the charts an initial decision about category would have to be made. No intuitively appealing method for making this decision from avail-

[1] AROWA Technical Report No. 13, Climatology of ocean cyclones. Dec., 1952.

able data was found. Therefore, no climatological prediction of pressure was made.

3. Criteria

To evaluate these techniques, some criteria must be selected. The criterion to be used for any particular purpose (e.g., comparison of several techniques or determination of bias in a prediction scheme) must be selected on a subjective basis. Clearly, different criteria may lead to different conclusions.

The criteria on which the evaluations reported here are based are described below. For comparing competing techniques the criterion adopted was "*per cent accurate.*" "Per cent accurate" refers to the per cent of forecasts that fall within some agreed upon distance from actual occurrence. This criterion has several characteristics to recommend it:

a) Differences in size for large errors are neglected (all forecasts in error by more than some preassigned amount are labeled "inaccurate").
b) The thresholds of accuracy may be selected to be operationally meaningful.
c) Well known statistical tests of significance may be applied in making comparisons.

The particular thresholds of accuracy used in this work were chosen after consultation with meteorologists responsible for operational forecasts and examination of some typical distributions of forecast error. The choices were: within 2 deg and within 1 deg for latitude and longitude forecasts, within 4 mb and within 2 mb for pressure forecasts. The statistical test used to compare forecast techniques on the basis of the "per cent within" criterion was a conservative t-test ("conservative" in the sense of assumption of maximum possible variance).

The criterion used for assessment of bias in forecasts made by any scheme was based on a 95 per cent confidence interval for the mean of the distribution of forecast error. Any such interval that does not include zero is taken as evidence of bias in the forecast scheme with which it is associated.

For certain special comparisons it was convenient and more efficient to compare techniques storm-by-storm (paired comparisons). Here the criterion used was the simple one of recording the "winner," i.e., the technique which achieved the smaller size error for a particular prediction. The statistical techniques that were used in conjunction with this choice of criterion are the standard sign test and a truncated sequential procedure due to Bross (1952).

In several instances it was considered worthwhile to calculate the traditional measure *root-mean-square error*. This measure is indeed sensitive to differences of accuracy between schemes; however, it displays this sensitivity in a marked manner on occasions where forecast errors are large (i.e., where neither of the techniques being compared yields an operationally useful forecast). Further, in dealing with biased forecast procedures, standard statistical tests are not applicable. Thus one cannot readily tell whether an observed difference in root-mean-square error is significant or attributable to chance.

Another intuitively appealing criterion with attributes similar to those of root-mean-square error is *mean absolute error*. This estimate of the average size of the error committed in making a forecast was calculated for a number of situations, and the values are recorded where they seem to be informative.

The following example illustrates the sensitivity of evaluations to the choice of criterion. Veigas-Ostby (1963) describes two different sets of equations: Model A being a selection from about 200 meteorological variables, and Model B being a selection from about 400 meteorological variables. The accuracy of the two models was compared in the report on the basis of root-mean-square error (calculated for an independent sample of 106 cases). The two models were adjudged to be equally accurate. However, when the "per cent accurate" criteria described above were applied to their data, some significant differences appeared. The results are displayed in Table 1. The observed values of the criteria were found to be significant (at the 5 per cent level) with respect to latitude forecasts and pressure forecasts.

Table 1. Per cent accurate criteria applied to VO model Types (independent sample, 106 cases).

	Model A	Model B
Latitude forecasts		
% within 2°	55	66
% within 1°	29	43
Longitude forecasts		
% within 2°	58	65
% within 1°	33	29
Pressure forecasts		
% within 4 mb	57	35
% within 2 mb	30	21

4. Evaluation of basic techniques

The WG technique was compared with the climatological forecast scheme for the Northern Hemisphere, except Asia. The precise region from which the sample was taken was Latitude 20 to 70N, Longitude 20 to 120E, via 180 deg. Sample size was in excess of 300 (April–May, 1962) for each technique. The results are summarized in Table 2.

To note the slight effect of deleting storms in Asia from the sample and to show the accuracy of 12-hr forecasts, the behavior of WG in the entire Northern

TABLE 2. Climatology vs. WG—24-hr forecasts.

Forecast	Technique	% within 2°	% within 1°	Mean magnitude of error (deg)	Root mean square error (deg)
Latitude	Climatology	36.0	23.6	4.2	5.3
	WG	50.1*	32.3*	3.0	3.9
Longitude	Climatology	24.2	17.1	6.0	7.5
	WG	34.2*	21.2	5.1	6.5

* Significantly better than climatology (5 per cent level).

TABLE 3. Evaluation of WG, Northern Hemisphere (sample size: 488).

Forecasts		% within 2°	% within 1°	95% Confidence interval for mean of errors
Latitude	12 hr	73	51	(−0.10, 0.42)
Latitude	24 hr	52	34	(0.11, 0.83) bias
Longitude	12 hr	56	39	(−1.34, −0.56) bias
Longitude	24 hr	35	23	(−2.11, −0.98) bias
		% within 4 mb	% within 2 mb	
Pressure	12 hr	67	41	(−1.70, 0.18)
Pressure	24 hr	39	22	(−1.95, −0.35) bias

TABLE 4. Analysis of distribution of errors in forecasts. U. S. East Coast, sample size 55. April–May 1962.

Forecast	Technique	% within 2°	% within 1°	Range of errors (deg.)	Mean magnitude of errors (deg.)	95% Confidence interval for mean of errors	rms error
a) 12 hr lat.	VO	80	55	13	1.65	(−0.24, 1.00)	2.32
12 hr lat.	WG	76	62	13	1.53	(−0.82, 0.38)	1.54
24 hr lat.	VO	56	38	13	2.49	(−0.19, 1.53)	3.24
24 hr lat.	WG	71	38	14	2.27	(−0.38, 1.22)	2.98
b) 12 hr long.	VO	71	47	24	2.31	(−1.27, 0.59)	3.48
12 hr long.	WG	62	42	17	2.44	(−2.11, −0.37) bias	3.47
24 hr long.	VO	38	20	37	4.87	(−2.95, 0.55)	6.52
24 hr long.	WG	38	24	33	4.85	(−4.60, −1.48) bias	6.55
		% within 4 mb	% within 2 mb	(mb)	(mb)		
c) 12 hr pres.	VO	73	47	24	3.58	(−1.97, 0.55)	4.73
12 hr pres.	WG	56	35	21	4.42	(−3.06, −0.32) bias	5.36
24 hr pres.	VO	44	25	36	6.47	(−3.61, 0.71)	8.18
24 hr pres.	WG	38	18	41	7.82	(−5.72, −0.68) bias	9.87

Hemisphere for the same time period (April-May 1962) is displayed in Table 3.

For storms in the sample that were located on the United States East Coast, the comparison considered was WG vs. VO. The precise region was latitude 30 to 50 N, longitude 50 to 100W. The sample consisted of 55 storms, April–May 1962. The criteria calculated are summarized in Table 4. None of the differences observed for the per cent accurate criteria were large enough to be declared significant.

VO and WG were also compared storm-by-storm (paired comparisons). The technique yielding the smaller error for a particular forecast was labeled "winner." The well-known sign test was applied to the results, and no significant difference between VO and WG was noted.

A third comparison was conducted in terms of the ability of the techniques to predict whether a particular storm would deepen or fill. Whenever the two techniques differed in this decision, the correct one was labeled "winner." The truncated sequential procedure due to Bross (1952) was applied to the results; again, no significant difference was noted. Thus one is led to conclude that the two techniques exhibit comparable accuracy in the region for which VO was built.

Curiosity led to the application of the VO forecast equations to regions other than the United States East Coast. When evaluated over the entire Northern Hemisphere, it was found (as one would expect) that VO did not yield a higher "per cent accurate" value than WG in any forecast category. However, it is interesting to note that the difference was often small, and VO always scored a higher per cent accurate than climatology in the Northern Hemisphere, excluding Asia.

5. Extensions and modifications

Several attempts have been made to improve forecast accuracy. One approach made use of knowledge of position and intensity of storms six hours after the time of collection of initial data. Use of this information with a simple extrapolation process to modify the basic techniques did not yield consistent improvement in forecast accuracy. Occasionally the modification yielded a decrease in accuracy.

Attempts to improve WG by introducing additional meteorological variables (in a linear manner) have not been successful; nor have linear combinations of the outputs of the VO and WG schemes yielded significant improvement in forecast accuracy.

Acknowledgments. We are indebted to Captain Paul Wolff, USN, of the U. S. Naval Fleet Numerical Weather Facility and also to the Computer Center of the U. S. Naval Postgraduate School, Monterey, Calif., for providing us with the data, freedom and facilities that made this evaluation possible.

REFERENCES

Bross, I., 1952: Sequential medical trials. *Biometrics*, 8, 188–205.

George, J. J., 1960: *Weather forecasting for aeronautics*. New York, Academic Press, 673 pp.

Miller, R. G., 1962: Statistical prediction by discriminant analysis. *Meteor. Monogr.*, 4, 25, Boston, Mass., Amer. Meteor. Soc., 53 pp.

Veigas, K. W., and F. P. Ostby, 1963: Application of a moving-coordinate prediction model to East Coast cyclones. *J. appl. Meteor.*, 2, 24–37.