A Note on Verification of Prognostic Charts

By SVERRE PETTERSSEN, University of Chicago[1]

(Manuscript received January 2, 1957)

1. Correlations of pressure changes

In recent forecasting experiments it has been customary to verify the prognostic charts by computing the correlation coefficients between the observed and predicted pressure changes. The justification has been that correlations between predicted and observed pressures would not correctly reflect skill, since in many cases high correlations would result from persistence.

While correlations of observed and predicted changes certainly remove similarities resulting from persistence they also tend to remove the measure of similarities resulting from travel of pressure systems. To illustrate, it suffices to consider a sinusoidal pressure wave with length L, amplitude A, and phase velocity C which travels along an axis x. The equation for the wave profile may then be written

$$Y_0 = A \sin \frac{2\pi}{L}(x - Ct) + N_0 \qquad (1)$$

where N_0 symbolizes some kind of noise superimposed upon the primary wave. The noise may be real or due to errors in observations and the analyses. In any case, it will be assumed that the maximum value of N_0 is very much smaller than A. In general, the forecasting problem will involve changes in A, L and C, but for the purpose of illustration it will be assumed that the wave is permanent, in the sense that its speed and shape are maintained. Next, let it be assumed that the movement of the primary wave was predicted perfectly, so that at the end of the forecast period the prediction was

$$Y_p = A \sin \frac{2\pi}{L}(x - Ct) + N_p$$

where N_p signifies some kind of noise ($A > N_p$) which was not predicted.

Now if the forecast period were such that $Ct = \frac{1}{2} L$, the observed and predicted changes would be

$$\Delta Y_0 = -2 A \sin \frac{2\pi}{L} x + \Delta N_0,$$

$$\Delta Y_p = -2 A \sin \frac{2\pi}{L} x + \Delta N_p$$

respectively.

Since the extreme values of N were supposed to be very much smaller than A, it is evident that a correlation close to unity would be found. On the other hand, if the forecast period were such that $Ct = L$, the changes would be

$$\Delta Y_0 = \Delta N_0 \quad \Delta Y_p = \Delta N_p$$

and the correlation would be determined exclusively by the noise.

In practice this would mean that correlation of pressure (or height) changes would

[1]) The research reported in this note has been made possible through support and sponsorship extended by the Geophysics Research Directorate, Air Force Cambridge Research Center, under Contract No. AF 19(604)—1293.

give high credit to forecasts in cases where the final state tended to be out of phase with the initial state, and it would discredit forecasts in which in-phase relationships were involved.

It can readily be shown that verification by pressure (or height) changes tend to give undue credit to poor forecasts in cases where the distance traveled by the system is small in comparison with its dimensions. To illustrate, we return to the sinusoidal wave. Leaving the noise out of consideration, it is readily found that the correlation coefficient $R(\Delta)$ between the observed and predicted changes is expressed by

$$R(\Delta) = \pm \cos \pi \frac{D_0 - D_p}{L} \quad (D_0 D_p \gtrless 0) \quad (2)$$

Here D_0 and D_p signify the observed and the predicted displacements, and plus or minus are chosen according as D_0 and D_p have the same or opposite sign.

Since $D_0 - D_p$ is the error in the forecast displacement, it will be seen that $R(\Delta)$ remains positive as long as the error remains less than one-half of the wavelength.

In connection with the forecasting of upper waves it would seem reasonable to consider a forecast to be useful if the error in the predicted displacement were less than the displacement itself. As an example, let us consider an average upper wave of length 60° latitude and a 24-hour displacement of 10°. With the above definition of usefulness one would find that only correlation coefficients in excess of 0.87 would indicate success. On the other hand, if the wavelength were 30° and the displacement 15°, any positive value of the correlation coefficient would signify success. From the simple cases discussed above it would appear that the interpretation of correlation of changes, in terms of usefulness, is rather complex.

2. Correlation of pressures

Since the main purpose of prognostic charts is to provide a basis for the issuance of specific forecasts, and since the circulation patterns, rather than the change patterns, at the end of the forecast period are important, it appears that the usefulness of prognostic charts can more readily be assessed from correlations (or similar measures) of the agreement between the predicted and the observed pressure distributions. Since the correlations reflect only proportionately, the root-mean-square error, or some similar measure, would supply additional information. Whatever system of verification is used, the zero-level (determined by the degree of persistence) must be considered.

As regards real persistence it must be admitted that considerable skill on the part of the forecaster is required to predict with success that the changes throughout an appreciable period will be insignificant. If, in such a case, an individual prognostic chart verifies no better than straight persistence, a positive measure of skill would seem in order. It appears, therefore, that there can be but little justification for rating *individual* forecasts on a scale in which the persistence pertaining to the *case itself* marks the zero level. Any such scheme, as well as verification by pressure changes, will emphasize the forecasting of out-of-phase relationships and penalize forecasts of in-phase relationships, whether these latter are due to real persistence or accidental similarity due to travel of systems.

A measure of the skill in which similarity between the initial and final chart is accounted for can be justified for a *representative sample* of forecasts, but the purpose of such a score would be to measure the yield of the effort as compared with no effort at all. To illustrate, let it be assumed that no effort is made to predict the future state and that, instead, the actual chart is presented as a prognostic chart for the end of, say, a 24-hour period. Over a representative period a certain correlation would be found. Whether this correlation resulted from real persistence or accidental similarities is immaterial since, obviously, the procedure involved no skill.

Let $R(IV)$ and $R(PV)$ denote the correlation coefficients between the initial and verifying charts and between the prognostic and verifying charts, respectively, both pertaining to a representative sample of forecasts. Obviously, the forecasts would reflect some skill if $R(PV) - R(IV) > 0$. The maximum score obtainable (for perfect forecasts) would be $1 - R(IV)$. A correlation skill score S, similar to that derived from contingency considerations, could then be defined as

$$S = \frac{R(PV) - R(IV)}{1 - R(IV)} \quad (3)$$

The score would be zero if current information were presented as forecasts. A positive score for a representative sample of forecasts would indicate some degree of skill, and different forecasting techniques, applied to the same sample of cases, can be compared by the scores obtained.

In a recent forecast experiment (PETTERSSEN, ESTOQUE and HUGHES, 1956) it was found that $R(IV)$ for 24-hour sea-level forecasts for the United States in winter was about 0.6. For such low values a relatively short period suffices for determining whether or not a forecasting technique is capable of yielding useful results. In the same experiment it was found that $R(IV)$ at the 500 mb level was about 0.9. For such high values of $R(IV)$ a lengthy trial would be required to obtain significant results. Obviously, as $R(IV)$ approaches unity, prediction ceases to be a problem.

3. A scheme of verification

In the aforementioned forecast experiments the following verification measures were used

$R(PV)$—the correlation coefficient between prognosticated and verifying pressures (or heights).

$R(IV)$—the correlation coefficient between the initial and the verifying pressures (or heights).

S —the correlation skill score as defined by Eq. (3).

$R(\Delta)$ —the correlation coefficient between the predicted and observed changes.

E —the root-mean-square error of the predicted changes.

C —the root-mean-square of the observed changes.

N —the number of occasions when $R(PV)$ exceeded $R(IV)$.

M —the number of occasions when E was less than C.

While no single verification measure will describe the efficacy of a forecasting technique, it was found that the measures S, E, C, N and M, used in combination, are adequate for comparing different methods. Examples of such comparison will be given elsewhere (PETTERSSEN, 1957). Here, it may be noted that not only the skill score but also the value of E, in comparison to that of C, has been found useful. In comparing different methods not only the question of average performance but also that of stability has to be considered, and the measures M and N have been found useful for assessing the stability.

4. Conclusion

Forecasting research has now progressed to a stage where it has become necessary to compare different prediction models, techniques and procedures. While the verification scheme outlined here may not be ideal, it is hoped that it will provoke discussion, and result in the establishment of uniform methods. A yardstick for comparison of prognostic techniques is much in need.

REFERENCES

PETTERSSEN, S., ESTOQUE, M. A., and HUGHES, L. A., 1956: An experiment in prognostication. Scientific Report No. 12, Contract AF 19(604)—1293. Department of Meteorology, University of Chicago.

PETTERSSEN, S., 1957: A review and summary of recent forecasting research. Final Report, Contract AF 19(604)—1293. Department of Meteorology, University of Chicago.