On «Results of forecasting with the barotropic model»

Dear Sir,

In a recent paper describing results of forecasting by numerical methods (STAFF MEMBERS, UNIVERSITY OF STOCKHOLM, 1954), it was stated that the correlation between observed and computed changes is not a satisfactory measure of the goodness of a forecast. Previously the same group (Staff Members, University of Stockholm, 1952) had raised the objection that a correlation coefficient measures the similarity in behaviour of two variables without any reference to systematic errors such as a difference in scale of the variations. However there is a more fundamental objection which in no way depends on the properties of a correlation coefficient as such, but applies to any comparison of observed and computed *changes*, whether quantitative or qualitative.

The basis of this objection is illustrated by a hypothetical theory which yields "computed" changes in every case equal and opposite to the corresponding initial departure of the forecast element from its normal value. Any examination of such "computed" changes and those observed would on the whole show considerable agreement between them – it can be shown theoretically that without persistence one would expect a correlation of about 0.7 between observed and "computed" changes, while somewhat smaller values would appear if persistence were important. The "forecasts" resulting from this "theory" are of course identical with the ordinary climatological forecasts, which can be made without any theoretical background. They therefore do not represent any practical progress in spite of the good agreement between observed and "computed" changes.

Now it may be argued that the main purpose of many experimental forecasts is to obtain confirmation or otherwise of a particular theory, and that agreement between observed and computed changes is sufficient from this point of view. However, for the forecast described above, it can be shown that good agreement will also result if the same "computed" changes are compared with the changes observed during any arbitrary forecast interval. In this case there is good agreement without any apparent theoretical justification, so that the good agreement previously found for the proper forecast interval cannot be regarded as confirming the "theory".

Reflection shows that the truth lies between these two positions of extreme optimism and extreme pessimism in the interpretation of the results. Thus if a real physical theory existed which could predict on every occasion a return to normal conditions (which is the case for the hypothetical theory considered), the very important problem of explaining these normal or mean conditions (in effect the problem of the general circulation) would have been solved. However such a theory obviously plays no part in explaining *variations* from mean conditions, which might be considered as the main task of any forecast theory.

The hypothetical case therefore shows good agreement between observed and computed *changes* associated with a "theory" which is of no practical value and which is satisfactory only in a very restricted sense. While the example given is admittedly rather artificial, similar reasoning applied to any other forecast methods will show that that part of the agreement between observation and theory which reflects a "return to normal" effect in the underlying theory will be similarly without practical value. It may be valuable as a contribution to the general circulation problem, but not to the genuine forecast problem of explaining variations from normal conditions.

It follows from this that the relationship between observed and computed *changes* (whether expressed as a correlation or in any other way) cannot be interpreted directly in terms of goodness of a forecast. Even in comparing forecasts by different methods or under different conditions, the comparison usually made (i.e. by the magnitude of the correlation) would have meaning only if the "return to normal" effect contributed equally to all types of forecast. However the possibility cannot be ruled out that an apparently better method of forecasting contains in reality only a stronger tendency for "return to normal", so that the improvement is largely illusory.

This pitfall can be avoided by using a partial correlation coefficient in which effects of variation in the initial departure from normal have been eliminated. The partial correlation, given by $r_p = \frac{r - r_1 r_2}{\{(1 - r_1^2)(1 - r_2^2)\}^{\frac{1}{2}}}$, where r is the correlation between observed and computed changes and r_1 and r_2 are correlations between each of these quantities and the initial departure from normal, depends only on that part of the relationship between observed and computed changes which is independent of the "return to normal" effect. It is therefore a much more satisfactory measure of the goodness of a forecast than is the simple correlation r between observed and computed changes.

The importance of the suggested change inverification procedure when applied to numerical forecasts can be judged only by calculating r_p from the original data, a course unfortunately not available to the present writer. However r_p may differ considerably from the simple correlation r, so that values of the latter in published verifications give little information regarding the true contributions of the various theoretical models towards solution of the forecast problem.

REFERENCES

STAFF MEMBERS, *The Institute of Meteorology, University of Stockholm*, 1952: Preliminary report on the prognostic value of barotropic models in the forecasting of 500 mb height changes. *Tellus*, **4**, 21—30.

STAFF MEMBERS, *The Institute of Meteorology, University of Stockholm*, 1954: Results of forecasting with the barotropic model on an electronic computer (BESK). *Tellus* **6**, 139—149.

January 27, 1955

ALISON M. GRANT
University of Melbourne, Australia

Miss Grant's comments on the verification of forecasts, conventional or numerical, are interesting. The following additional remarks should be given in connection to the numerical forecasts made at the Institute of Meteorology, University of Stockholm.

The partial correlation coefficients have been computed for the 24, 48 and 72 hour forecasts from October 2, 1954, 0300 GCT (cf. BOLIN, 1955). They were 0.89 (0.92), 0.84 (0.92) and 0.79 (0.87), where the figures within the parentheses are the corresponding ordinary correlation coefficients. It is seen that the partial correlation coefficients are consistently lower as would be expected. The coefficients r_1 and r_2 varied between 0.55 and 0.70 in this case. These values of r_1 and r_2 give $r_p = 0$, if $r \approx 0.4$.

For a detailed comparison between various forecasts one probably should also include the partial correlation coefficient. Under the assumption, however, that the persistence of atmospheric flow pattern does not vary too much (which of course is somewhat doubtful), it merely means an expansion of the scale from $r = 0.4$—0.5 meaning a worthless forecast to $r = 1$ meaning a perfect forecast to a scale $r_p = 0$ = failure to $r_p = 1$ = perfect.

It should finally be stressed that any correlation coefficient of this kind gives a very incomplete description of the success or failure of a forecast. This is most clearly seen from the experiment of subjective verification reported in a note in this issue of Tellus (pp. 272—274). Quantities as error, standard deviations, etc. may be of some additional help, but a forecast map is mainly used for evaluating changes of the *flow*, and one should therefore aim at a method which gives some information about the success in this respect.

REFERENCE

BOLIN, B., 1955: Numerical forecasting with the barotropic model. *Tellus*, **7**, 27—49.

BERT BOLIN
Institute of Meteorology
University of Stockholm