A Numerical Prediction Experiment Involving Data Paucity and Random Errors

By EDWARD O. JESS, U.S. Air Force

(Manuscript received July 4, 1958, revised May 5, 1959)

Abstract

500-mb prognoses computed with a barotropic model from progressively lessening amounts of data, showed increasing statistical differences. These differences also grew with the passage of time. With increasing paucity of data, the smaller meteorological features disappeared, while larger features were smoothed and suffered a change in speed of motion.

Random errors, which simulated a natural error source, were added to the data. This appeared to intensify and decelerate the features on the prognostic charts.

Introduction

Errors in observational data are still a challenging problem despite great progress being made in eliminating them. Attention has also been focused upon other contributors to forecast inaccuracy. Variations between analyses of the weather central of the Royal Swedish Air Force and the Deutscher Wetterdienst, Frankfurt, Germany, grew into operationally significant differences in 48-hour and 72-hour forecasts (BEST, 1956). Within the same group of forecasters, analyses vary greatly (BERGGREN, 1957) with anticipated forecast differences.

All forecasters agree that sparsity of data leads to analysis error with concomitant forecast error. Isolated examples showing how extreme these can be are familiar to all practicing meteorologists. Each veteran forecaster has his favorite story about a forecast that was (or was not) a "bust" simply because of the absence (or presence) of one weather observation in an area of no data. But are these critical observations as common and serious as these entertaining tales suggest?

A 300-kilometer distance between grid points is considered suitable for numerical weather prediction of the mid-tropospheric flow. Such grid points may be considered as weather reporting stations, each conveniently separated from adjacent "weather stations" by a 300 kilometer distance. But is this the only satisfactory upper air data from meteorological and economic points of view? Would not station separations two, three or four times as great produce essentially the same forecasts or forecasts which are nearly as accurate? If the weather stations were progressively 300, then 600, then 900 and finally 1,200 kilometers apart, how would the shape and movement of the forecast meteorological systems be affected? Using the forecast computed from the more plentiful 300-km spaced data as a standard, one could compare it with the other forecasts and consider the differences as errors. These would be "forecast errors" caused by increasing data paucity.

But many sources of error, not just one, are involved in forecasting. For a proper understanding, it is necessary to know how these interact. This information could contribute toward determining if further reduction of various errors is warranted. The simplest interaction to study would be that between only two error sources such as data paucity and random errors. This paper attempts to shed some light upon the effect on forecasts rendered by increasing the paucity of meteorological data and by allowing a random error source to operate on that data.

Procedure

During January and March 1956 the Swedish Air Force machine-computed 24-, 48- and 72-hour forecasts on a routine basis. Each forecast was calculated from an objective analysis (BERGTHORSSON and DÖÖS, 1955) having a grid interval of 300-km at 50° N latitude.

In this experiment these Swedish Air Force routine numerical predictions were used to verify experimental forecasts. The latter were derived in the following manner. From the objective analysis all data or "observations" were removed except those at the grid points which were 600 kilometers apart. This formed a 600-km grid of data. These remaining data were analyzed in the conventional manner (by hand). The same procedure was repeated to obtain analyses with data spacings of 900 and 1,200 km. Each chart was analyzed by a different analyst who had not become biased by seeing any of the other charts.

Then pressure-height values were carefully interpolated at all intervening grid points in order to make machine computations. The differences between these new interpolated values and the original objective analysis values, which were removed, can be looked upon as the error of analysis introduced by increased paucity of data. Numerical forecasts were made from these artificially derived charts and the results compared with the corresponding verifying forecasts. Numerical weather prediction gives a unique solution for each synoptic situation. One may then attribute any difference between the artificial forecasts thus obtained and the corresponding verifying forecasts to be due only to the differing abundance of "observational data". Intuition suggests that this difference will grow with increasing data sparsity and will be accompanied by a corresponding increase in forecast error.

The routine forecasts accomplished by the Swedish Air Force and the experimental numerical predictions made for this project were computed over a rectangular 32×41 grid of 1,312 points on a polar stereographic projection. See Figure 1. (This experiment dealt with varying distances between data but the forecasts computation grid interval was always 300 kilometers.) The complex and very subjective procedure of conventional forecasting was standardized by utilizing the digital computer BESK and the same barotropic forecasting model (BOLIN, 1955) for every forecast.

The synoptic time for all analyses and forecasts was 1500 Greenwich Meridian Time and all discussion concerns the 500-mb level only. Forecasts computed from data spaced at 600-, 900-, 1,200- and 300-km is referred to as the 600-, 900-, 1,200- and 300-km (Swedish Air Force verifying) forecasts, respectively.

Data Spacing

A 300-km spacing of data exists over the United States while west-central Europe with the British Isles can boast of a 200-km spacing of data. The 900-km spacing corresponds roughly to the distance between the following Atlantic Ocean weather ships:

Weather	Ship	A —	Weather	Ship	C
»	»	A —	»	»	I
»	»	C —	»	»	J
»	»	B —	»	»	C
»	»	J —	»	»	K

A 1,200-km spacing corresponds roughly to:

Weather	Ship	C —	Weather	Ship	D
»	»	D —	»	»	E
»	»	A —	»	»	J
»	»	I —	»	»	C

Other spacings between weather ships are greater than 1,200 kilometers. On the Pacific Ocean distances between weather ships are generally greater than in the Atlantic region.

Adding Random Errors

Paucity of data is not the only source of error in a forecast. There are several others. In order to observe how two such sources of

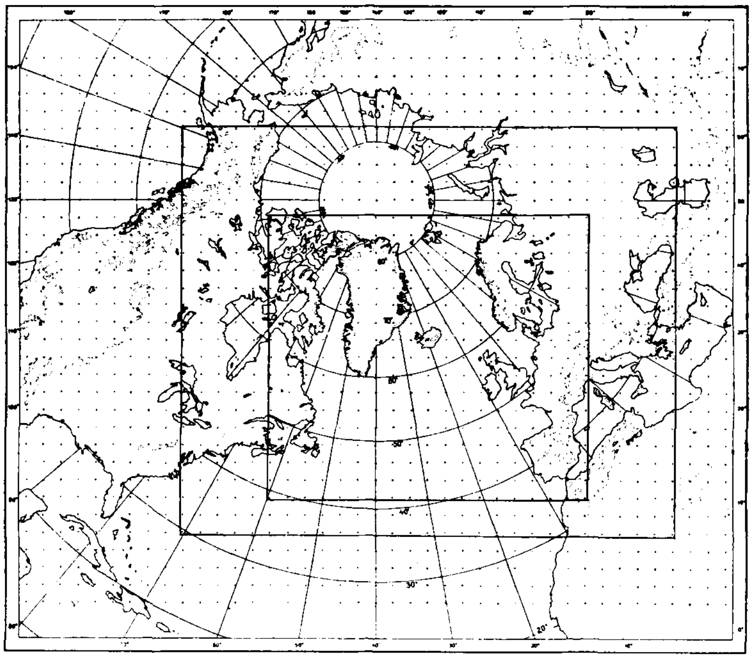

Fig. 1. Location of the grids used in the numerical predictions and in the statistical verifications. The numerical predictions were made over the grid containing all the dots. The verifications were done over the two smaller areas enclosed by the rectangles.

error might operate simultaneously on a series of forecasts, a group of random errors were added in a random fashion to the 1,312 grid points on the original Swedish Air Force objective analysis. The procedure of deleting certain data, making a conventional subjective analysis from the remaining data and interpolating values for the deleted points, was repeated. Each analyst was warned that random errors had been added to the pressure-heights but was asked to use caution in discarding data. The resulting analyses contained errors due to both paucity of data and the added random errors. Numerical forecasts were made from these doubly contaminated data. The random errors were chosen (CONRAD and POLLOCK, 1950) so as to have a normal distribution with a standard deviation of 20 meters.

The Cases

Three cases were compared.

Case I 13 January 1956
Case II 24 January 1956
Case III 28 March 1956

Case II was a relatively high-index synoptic situation whereas the others were low-index. Three cases are insufficient in number to guarantee the following results to be statistically significant but they may substantiate or contradict intuitive speculations. Occasionally it is beneficial to critically scrutinize such intuitive impressions.

Statistical Comparison

The charts were compared statistically for the following:

1. Direction of flow
2. Speed of flow
3. Change of flow

The linear correlation coefficient between corresponding pressure-heights on two charts

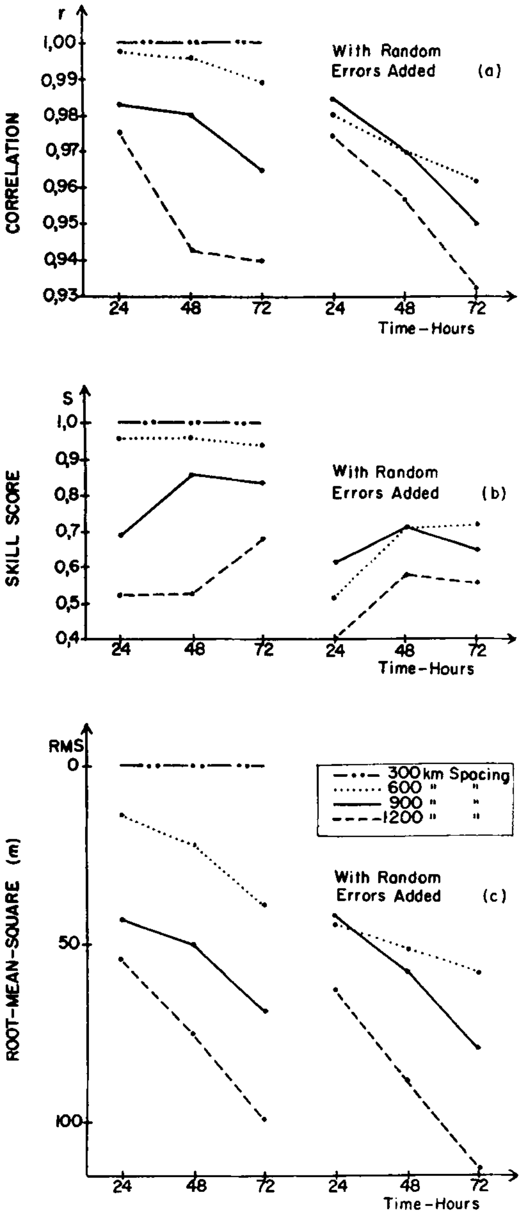

Fig. 2. Direction of flow comparison. (a) Correlation between forecast and verifying pressure-heights. (b) Pettersen's Skill Score for forecast and verifying pressure-heights. (c) Root-mean-square of the differences between forecast and verifying pressure-heights.

is a measure of similarity of the direction of flow. When the correlation coefficient is 1.0, the charts concerned have the same direction (but not necessarily speed) of flow.

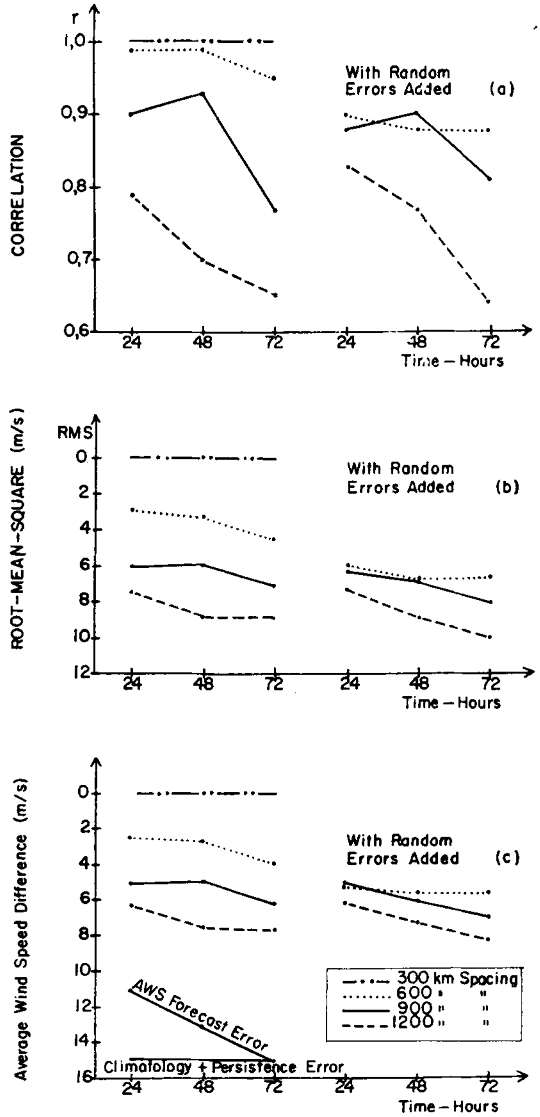

Fig. 3. Speed of flow comparisons. (a) Correlation between forecast and verifying vector wind speeds. (b) Root-mean-square of the differences between forecast and verifying vector wind speeds. (c) Average vector wind speed differences between forecast and verifying charts caused by differing paucities of data. Average vector wind speed errors in USAF Air Weather Service Forecasts, average vector wind speed errors resulting from Climatology and from Persistence Forecasts (ELLSAESSER, 1956).

A skill score (PETTERSEN, 1957) is also used in order to relate forecasts of the high and low index situations. If one lets P_{iv} and P_{fv} denote the correlation coefficients between the initial objective analyses and verifying charts and between the experimental forecast and verifying charts, respectively, then Pettersen's correlation skill score S is defined as:

$$S = \frac{P_{fv} - P_{iv}}{1 - P_{iv}}$$

S will vary between 1.0 and $-\infty$ where 1.0 is a perfect score. Positive values show skill and 0.0 indicates no skill or a forecast of persistence.

The pressure-height differences between corresponding points on two charts also compare circulation or pressure-height configuration. The root-mean-squares of these differences have been included. See Figure 2.

Speed of flow was the second measure of forecast excellence. In this experiment the geostrophic vector flow was computed at each grid point and the scalar magnitudes of the flow (speed) were compared for corresponding grid points. The linear correlation coefficient between geostrophic wind speeds at corresponding points on two charts is presented along with the root-mean-square vector difference and the average vector difference. See Figure 3. In addition an average of the maximum vector wind speed difference on each chart is presented as illustrating maximum forecast wind speed error in an individual forecast. See Table I.

Table 1. Max Wind Speed Difference m/s

Data Spacing	Random Error	Forecasts 24^h	48^h	72^h
600	Without	9	9	14
900	»	18	20	22
1,200	»	22	22	23
600	With	16	20	18
900	»	24	18	20
1,200	»	22	24	26

The third measure of forecast excellence, change of flow, is suggested by two statistical summaries of pressure-height changes with time. These include the linear correlation coefficient between height-changes at corresponding points on the forecast and verifying charts at 24, 48 and 72 hours. Also the root-mean-square of the differences between verifying and forecast height changes is shown under the conventional symbol ε. See Figure 4.

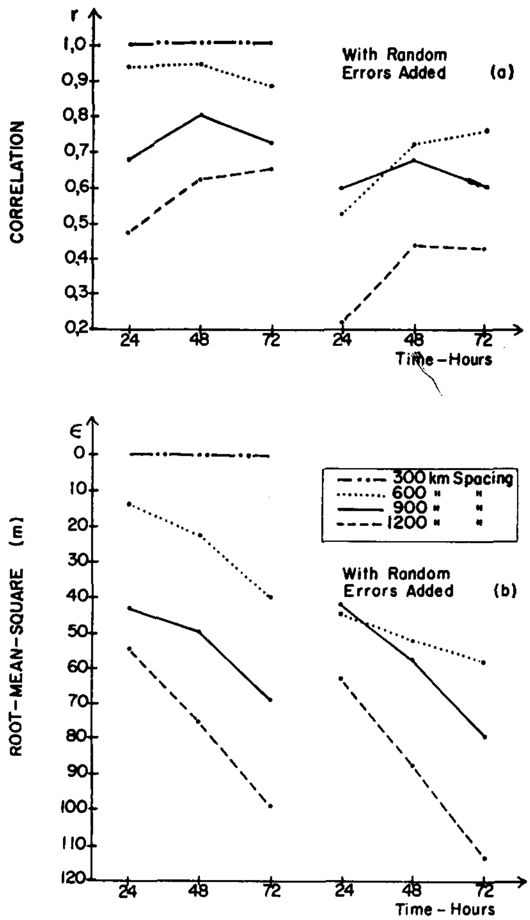

Fig. 4. Height Change Comparison. (a) Correlation between forecast and verifying changes (with time) of pressure-heights. (b) Root-mean-square of the difference between verifying and forecast height changes, ε.

All statistics were computed over the two areas depicted in Figure 1. The smaller has $19 \times 17 = 323$ grid points; the larger area has $29 \times 24 = 696$ grid points. Only the results computed over the smaller area are shown. The results over the larger area were generally similar.

Statistical Results

a. *Direction of flow.* Figure 2 verifies intuitive speculation by showing that the difference between forecast and verifying charts increased with the distance between observations. It also increased with time, with a slight acceleration in growth of the difference between 48 and 72 hours.

When the random-errors-effect is added, the 24-hour forecast computed from 900-km data spacing is initially least different from the verifying 300-km forecasts. However, the 600-km forecasts show greatest similarity to the verifying 300-km forecasts at 48 and 72 hours. In general the 600- and 900-km charts resemble each other more closely than the verifying 300-km or the 1,200-km forecasts.

b. *Speed of Flow.* An idea of the importance of the change in forecast wind speed due to a decrease in number of observations can be gained by a comparison with the average vector speed error of winds forecast in the U.S.A.F. Air Weather Service Forecast Capabilities (ELLSAESSER, 1956). In the 24-, 48- and 72-hour 500-mb AWS forecasts, the errors averaged 11, 13, 15 m/s, respectively. Climatological and persistence forecasts each gave average errors of about 15 m/s. These compare with changes of 3—4, 5—6 and 6—8 m/s caused by using the 600-, 900- and 1,200-km fields of data, respectively, instead of the objective analysis. When both random error and paucity of data are contaminating the analysis simultaneously, the corresponding changes are 5—6, 5—7 and 6—8 m/s. These statistics are depicted in Figure 3 c.

Generally speaking the errors caused in this experiment by varying the data density alone and by varying data density while also including random errors, are from one-quarter to one-half of the magnitude of errors observed in wind speed forecasts. If these errors are additive then it may be claimed that the 600-km grid of data is deficient while the 1,200-km grid of data is unsatisfactory for wind speed forecasts. From 24 to 72 hours the decrease in accuracy of a wind forecast is less than the change caused by varying data spacing by 600 or 900 kilometers. Adding random errors to the 600-km spaced data nearly doubles the amount of change. On the other hand, it makes little difference when the random errors are added to the 1,200-km spaced data.

One explanation is immediately apparent. Plentiful data give good definition of features, even of the short error waves. On the other hand, data paucity induces smoothing which obliterates the very short waves while mildly affecting the longer meteorological waves. A pattern of random errors consists of many short error waves which are not well described in a 1,200-km data field. On the other hand, there is less smoothing in the 600-km data so its error field is better defined and has a conspicuously deleterious effect on the forecast.

Among the curves for sparse data (900-km and 1,200-km) forecasts, a few points show increased "accuracy" after random errors have been added. This may be a statistical oddity or it might be partly due to the errors temporarily producing a more "accurate" forecast. When the scale of the initial error field is much less than the scale of the meteorological perturbations, the forecast error may decrease temporarily (THOMPSON, 1957). Also it might be due to the verifying forecasts being correct only by definition or assumption. BERGGREN (1957) showed that 29 analysts will produce 29 appreciably different analyses from the same set of synoptic data. So the "correct" forecast or analysis must be one that is assumed or defined to be such.

Table I shows the maximum vector wind speed difference which illustrates the magnitude of wind speed error associated with data paucity and random error effects. This statistic does not have such a clear-cut time or data dependence. It is largely a function of only two or four adjacent pressure-height data.

c. *Change of flow.* DÖÖS (1956) and BERGTHORSSON (1955) have observed ε to be about 55, 115 and 175 meters for 24-, 48- and 72-hour forecasts, respectively. These values were numerically computed for the eastern third of the small area in Figure 1 during November 1954 to February 1955. A comparison with values of ε from this study shows that decreasing data abundance to 600-, 900- and 1,200-km spacings gives differences whose RMS-values amount to roughly ¼, ½ and ⅔, respectively, of the Döös et als forecast errors. See Figure 4. Including the random-errors-effect increases the differences only slightly except for the 600-km data spacing in which case, the forecast error is more than doubled.

d. *Statistical Conclusions.* The direction, speed and change of flow were compared in forecasts computed from data with different spacings; i.e., 300, 600, 900, and 1,200 km. Differences between these forecasts increased with the data

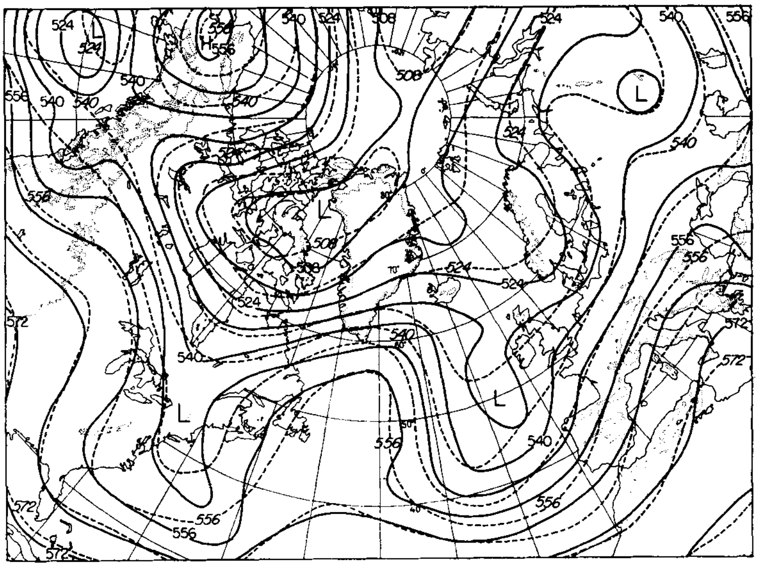

Fig. 5. Analysis of 13 January 1956. Solid lines indicate 300-km data-spaced objective analysis. Dashed lines indicate 1,200-km data-spaced analysis.

paucity progressively from roughly ¼ to about ½ of the ordinary forecast error. So scarcity of data does make a difference in forecasts.

Adding errors to the data had relatively innocuous effects when such data were widely dispersed but when the data were plentiful the effect was serious. This was especially true during the first twenty-four hours of forecasting, for the addition-of-random-errors-effect is most noticeable initially but in time tends to fade away.

Subjective Comparisons

As far as verification is concerned, statistical comparisons are scarcely sufficient, at best. A visual comparison is necessary. Characteristics common to all the forecasts may be seen by examining a few of the charts.

Figure 5 shows the analysis of 13 January 1956 with the 300- and the 1,200-km data spacing. The greater data spacing smoothes most features and essentially loses the trough over Oregon. Forecasts up to 72 hours were made from these, as well as all the other analyses. See Figure 6.

In the three cases studied, the adding of random errors to the analyses resulted in many forecast systems being intensified. In the example shown of 72-hour forecasts made from 1,200-km spaced data (Figure 7), this effect is illustrated by the dashed isopleths over the Iberian peninsula. (Compare with the solid isopleths depicting the corresponding forecast without the random-error-effect). The random errors contributed to forming a cut-off low between two prominent ridges. The ridge north of Alaska also illustrates the same intensifying process.

One may mention several factors encouraging this intensification. Adding random errors introduces short (error) waves into the analysis. These waves express false small-scale vorticity patterns which are advected by the numerical forecasting procedure. Since the barotropic model conserves absolute vorticity, these error features gain cyclonic vorticity while moving southward and lose it while moving northward. So they will tend to intensify the troughs and ridges through which they are being advected.

Fjørtoft (1953) has developed a theory applicable to two-dimensional flow of a homogeneous, non-divergent, inviscid fluid in which there is a net transfer of energy from the short waves to the long waves. When applied to the phenomena mentioned above, this theory describes how most energy transfer consists

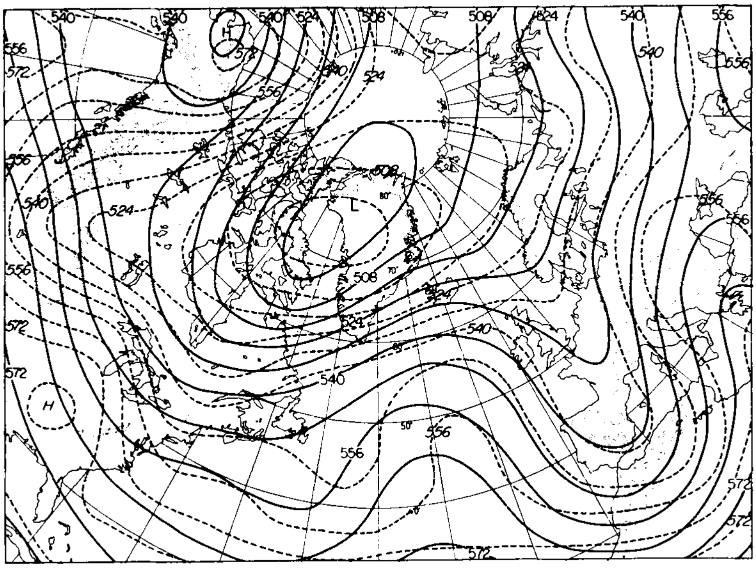

Fig. 6. 72-hour forecasts valid 16 January 1956. Solid lines indicate the forecast computed from the 1,200-km data-spaced analysis. Dashed lines indicate the forecast computed from the 300-km data-spaced objective analysis.

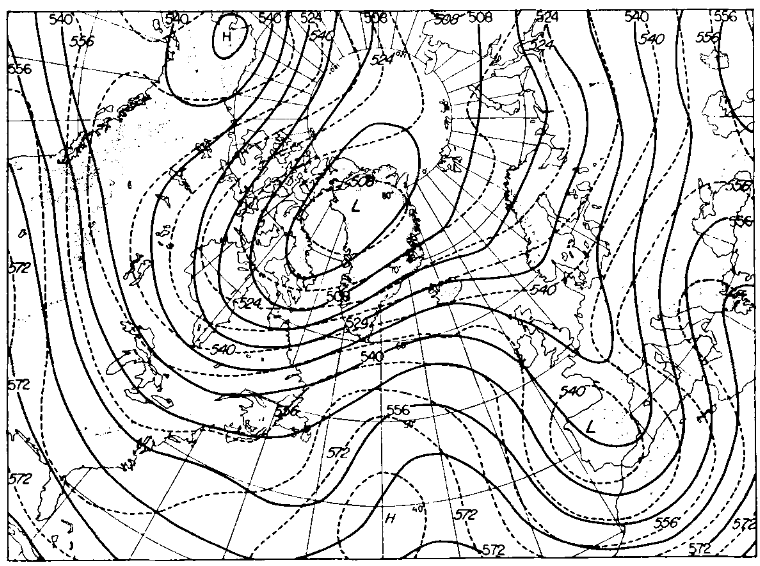

Fig. 7. 72-hour forecasts computed from 1,200-km data-spaced analyses. Dashed lines indicate the forecast computed from an analysis containing the random errors. Solid lines indicate the forecast computed from an analysis without the random errors.

of that passing from the small error waves to the larger (intensifying) meteorological system. In addition, the smaller amount of energy transfer, that from meteorological to error waves, is counteracted by the smoothing function, (BOLIN, 1955) built into the numerical forecasting process. Also with this portion of the energy gradually transferring to shorter and shorter waves, it is eventually lost with those waves which are too short to be defined by the grid. In this way the net transfer of energy is from small perturbations to the longer waves.

Fjørtoft's idea originated in a theoretical development from the vorticity equation. PHILLIPS (1958) considered this phenomenon in a different light by studying the finite difference operator. Non-linear interactions in the operation of a barotropic forecast model produce waves of various lengths. Some of these waves, whose lengths are shorter than the grid interval, may be incorrectly interpreted by the grid as long waves. This mechanism gives a false transfer of energy from short to long waves which also contributes to the intensification of the longer meteorological features.

Over some populated areas data is spaced at about every 300 kilometers; over oceans, at about every 1,200 kilometers. Forecasts from these data-spacings are illustrated by the 72-hour forecasts (Figure 6). Note the trough over western Europe. The smoothing effect of 1,200-km data spacing is illustrated by the Atlantic trough and ridge and the flat flow over Eurasia. Also the Atlantic trough-ridge system was decelerated in its eastward motion by the greater data spacing. Over western North America is a trough which initially lay over the mid-Pacific region where it had a wave length so small that it slipped through the 1,200-km grid.

Conclusions

An increase in spacing between data smoothes large features and eliminates smaller ones. This was demonstrated by conventionally drawn subjective analyses. Increasing data paucity redistributes the flow, decelerating the faster moving pressure systems and accelerating the slower moving features while the mean flow remains constant. It also expands the differences shown between the resulting forecasts. These differences grow with the passage of time as well as with the spacing between data.

The addition of errors to the data intensifies meteorological systems in the forecast. In the experiment this activity counteracted the smoothing tendency of increasing data spacing. The intensification is accompanied by deceleration of motion of the systems affected.

A decrease in spacing between data provides a much better definition of random errors and of small-scale weather features of limited local interest. It results in a lesser improvement in the definition of large scale meteorological features which is of general interest.

So we may conclude that the advantage gained by more abundant data will be lost unless data accuracy is maintained or increased. An increase in data quantity must go hand in hand with an increase in data quality. It is unwise to improve one while neglecting the other.

Acknowledgements

The author is indebted to several people for assistance with the essay from which this paper is abstracted. He wishes to thank Mr. Aksel Wiin-Nielsen for his great interest and constant encouragement. Gratitude is also extended to the staff of the International Institute and Dr. K. R. Johannessen for valuable discussions. Finally the author thanks Mr. A. Bring for advice in coding the computations, Mrs I. Hedberg for preparing the tapes and Mrs. S. Bring for drafting the figures.

Postscript

Some time after submitting this manuscript the author received the results of a similar experiment being performed by Mr. Charles L. Bristor of the U.S. Weather Bureau. Those results of the two experiments which were comparable, were generally compatible.

REFERENCES

BERGGREN, R., 1957: On the Accuracy of 500 mb Analysis with Reference to Numerical Forecasting. *Tellus*, **9**, pp. 323—340.

BERGTHORSSON, P., et. als., 1955: Routine Forecasting with the Barotropic Model, *Tellus*, **7**, pp. 272—274.

BERGTHORSSON, P., Döös, B., 1955: Numerical Weather Map Analyses, *Tellus*, **7**, pp. 329—340.

BEST, W. H., 1956: Differences in Numerical Prognoses Resulting from Differences in Analyses, *Tellus*, **8**, pp. 351—356.

BOLIN, B., 1955: Numerical Forecasting with the Barotropic Model, *Tellus*, **7**, pp. 27—49.

CONRAD, V., and POLLAK, L. W., 1950: Methods in Climatology, Cambridge; Harvard University Press, 459.

DÖÖS, B. R., 1956: Automation of 500 mb Forecasts Through Successive Numerical Map Analyses, *Tellus*, **8**, pp. 76—81.

ELLSAESSER, H. W., 1956: Errors in Upper Level Wind Forecasts, *AWS Technical Report*, No. 105—140/1, Washington, U.S.A.F., p. 17.

FJØRTOFT, R., 1953: On the Changes in the Spectral Distribution of Kinetic Energy for Two-Dimensional, Non-Divergent Flow, *Tellus*, **5**, pp. 225—230.

PETTERSEN, S., 1957: A Note on Verification of Prognostic Charts, *Tellus*, **9**, pp. 313—315.

PHILLIPS, N. A., 1958: An Example of Non-Linear Computational Instability. To be published in Rossby Memorial Volume.

THOMPSON, P. D., 1957: Uncertainty of Initial State as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns, *Tellus*, **9**, pp. 275—295.