Application of Numerical Prognostic Heights to Surface Temperature Forecasts[1]

By WILLIAM H. KLEIN and BILLY M. LEWIS, CURTIS W. CROCKETT, and ISADORE ENGER, United States Weather Bureau, Washington, D.C.

(Manuscript received February 29, 1960)

Abstract

Multiple regression equations for predicting 5-day mean surface temperature in the United States from prior anomalies of 700-mb height and surface temperature are derived by a screening procedure for fall, winter, and spring. The heights required for an objective forecast are approximated with the aid of daily numerical prognoses prepared at the 500-mb level by a barotropic model and at the 700-mb level by a baroclinic model; the required temperatures are estimated by combining observed daily values with short-range prognoses. The resulting objective temperature predictions are verified on independent data at a number of cities, and the sources of error are analyzed. Comparison is made with forecasts prepared by local temperature persistence, continuity of error, and conventional methods. It is concluded that the objective method can produce in a short time temperature predictions which have appreciable skill beyond persistence. Thus a valuable aid and base for further improvement has been made available to the forecaster.

1. Introduction

Considerable research on the relation between temperature and circulation for extended periods was conducted in the Extended Forecast Section of the United States Weather Bureau about ten years ago by MARTIN and LEIGHT (1949) and MARTIN and HAWKINS (1950). During the past few years interest in this problem has been revived because of the availability of high-speed electronic computers (KELLEY and MALONE, 1955). In a previous paper (KLEIN, LEWIS, and ENGER, 1959) a statistical screening procedure was used to obtain multiple regression equations for predicting 5-day mean surface temperature as a linear function of selected 700-mb heights. Application of these equations to heights obtained from barotropic prognoses produced temperature predictions of positive skill during the test winter of 1957—58. The forecasts were considerably improved by including as a predictor the local value of current 5-day mean surface temperature.

The last paper described above was primarily exploratory in nature and deficient in several respects. The only temperature predictor used in the regression equations was the local temperature; i.e., at the station for which the forecast was made. Furthermore, the final equations were derived on desk calculators without use of the screening technique and were therefore limited to 11 cities, 1 season, and 3 predictors. The resulting objective forecasts were tested on only 16 independent cases.

The present paper will describe an objective method which did not suffer from the aforesaid limitations. In addition to persistence at the local station, temperatures at a large

[1] Part of this paper was presented on May 7, 1959 at the joint meeting of the American Meteorological Society and the American Geophysical Union in Washington, D.C.

network of cities in the United States were considered as possible predictors. By use of a high speed electronic computer, the screening technique was applied to the field of surface temperatures as well as 700-mb heights. Regression equations were derived for 3 seasons and 39 cities, and as many as 7 predictors were included in each equation. The forecasts were tested on a large sample of independent data consisting of a total of 226 cases.

We shall first discuss the objective method, including the data used to derive it (section 2), its meteorological interpretation (section 3), and the screening procedure (section 4). A series of experiments with various numerical prognoses as input data will be described in section 5. The results of applying the objective method during a 2-year trial period will be presented in sections 6 and 7, and some comparison will be made with forecasts produced by other methods. Limitations of the present objective method and suggestions for improvement will be discussed in section 8. The principal findings are summarized in section 9.

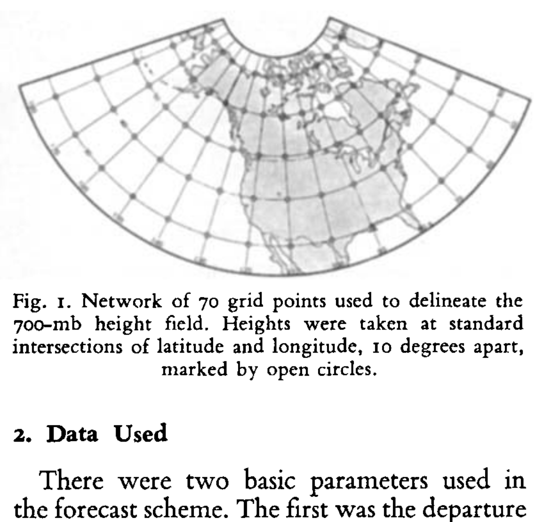

Fig. 1. Network of 70 grid points used to delineate the 700-mb height field. Heights were taken at standard intersections of latitude and longitude, 10 degrees apart, marked by open circles.

2. Data Used

There were two basic parameters used in the forecast scheme. The first was the departure from normal of 5-day mean 700-mb height over North America and adjacent oceanic areas. Heights[2] were taken at a network of 70 grid points, 10 degrees apart, extending from 30° to 70° north and from 50° to 180° west (figure 1).

[2] All references to heights and temperatures throughout this paper should be understood to apply to their anomalies, or deviations from local normal, rather than to their absolute values.

Fig. 2. Network of 39 stations used for surface temperature.

The second parameter was the anomaly of 5-day mean surface temperature at a network of cities covering all parts of the country. For fall and spring the 39 stations shown in figure 2 were used, but for winter only 30 were available (see fig. 2 of KLEIN and others, 1959). In selecting cities, consideration was given to such factors as homogeneity of record, representativeness of exposure, and availability of teletype data.

In order to clarify the time periods involved in the forecasts, a schematic calendar has been prepared, as shown in figure 3. Let us call the day on which the prediction is made day 0. The 5-day mean forecasts routinely prepared in the Extended Forecast Section of the United States Weather Bureau apply to the period beginning 2 days after forecast day (day 2) and ending 6 days later (day 6); i.e., to the 5-days centered 4 days after forecast day and therefore designated T_4. Since it is difficult to predict accurately the 700-mb height field for a period this far in advance, the 5-day mean heights required for the objective forecast were taken from a period 2 days earlier, extending from day 0 to day 4. These heights are designated H_2 because they are centered 2 days after forecast day. This period was chosen because it gave good results in our earlier study (KLEIN and others, 1959), when it was noted that surface temperatures tend to lag behind upper level heights, due in part to the time required for advection to operate.

In operational practice, the daily barotropic forecasts prepared by the Joint Numerical Weather Prediction Unit (JNWP) at Suitland, Maryland (CRESSMAN, 1959) were com-

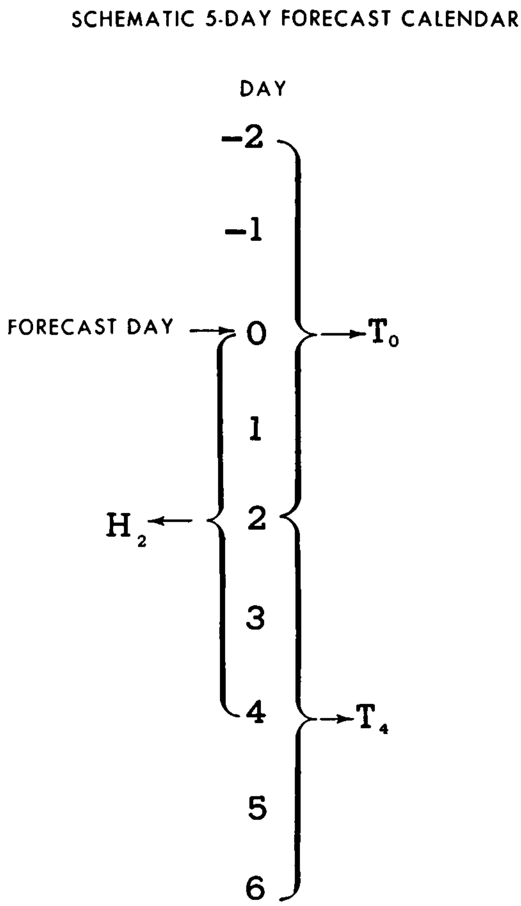

Fig. 3. Schematic calendar with various 5-day mean periods depicted relative to forecast day.

bined into a 5-day mean map by simply averaging the last observed map (day 0) with the forecasts for each of the next 4 days from day 1 to day 4. The resulting objective estimate of H_2 was readily available because it has been prepared on a routine basis in the Extended Forecast Section for the past few years, as explained by NAMIAS and Collaborators (1958).

The temperature predictor used in the objective method was centered on forecast day and therefore designated T_0. It covered the period from 2 days prior to forecast day to 2 days later and thus overlapped T_4 by one day (day 2). Half the data required for T_0 had already been observed, and the other half was estimated from forecasts for the next 2 ½ days routinely prepared at Weather Bureau district forecast centers throughout the United States. Thus the objective forecast scheme utilized 5-day mean surface temperatures centered on forecast day and 700-mb heights centered 2 days later to predict 5-day mean surface temperatures centered 4 days later, or:

$$T_4 = f(T_0, H_2) \qquad (1)$$

In the forecast routine, H_2 and T_0 were estimated from the numerical and district forecasts. Since these estimates were not available for a long period of record, the research was conducted with only *observed* values of both quantities. The fall season was defined as the 14-week period from mid-September to mid-December, and one case was chosen each week over a 10-year period from 1948 to 1957, making a total of 140 cases. A similar procedure was followed for the other seasons, so that 140 winter cases were selected from mid-December 1947 to mid-March 1957, and 140 spring cases from mid-March 1949 to mid-June 1958.[3]

3. Synoptic Implications

The first step in deriving the objective method was to obtain simple linear correlation coefficients between 5-day mean surface temperature (T_4) at each city plotted in Figure 2 and 5-day mean 700-mb height 2 days earlier (H_2) at each of the 70 grid points shown in Fig. 1. The resulting correlation fields are illustrated in figure 4 for New York City during

[3] A more refined treatment would have consisted of standardizing the data, so as to equalize the variability of different months within the same season. Better yet would have been to subdivide the data by months, rather than seasons, but an insufficient number of years was available for this purpose.

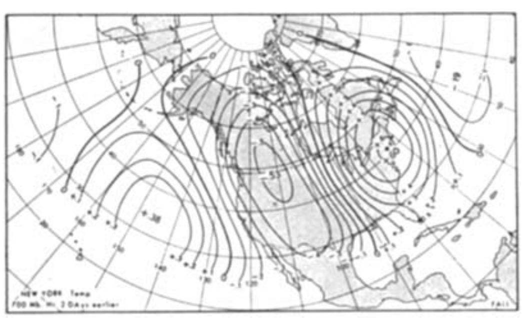

Fig. 4. Simple linear correlation coefficients between 5-day mean surface temperature anomaly at New York City (located by asterisk) and 5-day mean 700-mb height anomaly 2 days earlier for 140 fall cases.

the fall season. The isopleths of equal correlation indicate that fall temperatures at New York during the 10 years studied were most closely related to 700-mb heights in the immediate vicinity. The magnitude of the correlation was .60 and the sign was positive, thereby showing that above normal heights were associated with above normal temperatures, while low heights were followed by cold weather. This is a reflection of the fact that the troposphere usually approaches an equivalent-barotropic state (CHARNEY, 1949), with the contours tending to be in phase with the thickness lines. In addition, surface temperature is generally well correlated with thickness in the lower troposphere (ROWE, 1944).

There is another important center of maximum correlation in figure 4. In the Northern Rocky Mountain States the highest correlation was .53 but the sign was negative, thus indicating that above normal heights in this area were followed by below normal temperatures at New York, while low heights accompanied warm weather. One synoptic interpretation is as follows: Since the Rockies are normally about half a wave length upstream from New York, a ridge of large amplitude in the West usually results in strong northwesterly flow of cold polar continental air into New York; conversely, low heights over the Rockies imply westerly or southerly flow in the East and thus milder air masses.

Figure 5 A summarizes the results obtained by constructing similar correlation fields for all 39 cities during the fall season. The heavy dots show the points of maximum correlation between 700-mb heights and surface temperatures 2 days later at cities located near the tip of the arrows. For example, at Bismarck, North Dakota, the maximum correlation was −.64 with 700-mb heights in the Alaskan peninsula. Similarly, most cities in central and northwestern sections of the United States had points of highest correlation about 1,000 to 2,000 miles to the northwest. On the other hand, most cities in the East and Southwest were like New York, with the maximum correlation positive and located close to the station. All cities were basically similar, however, in exhibiting one center of positive correlation near the station and one center of negative correlation about half a planetary wave length upstream.

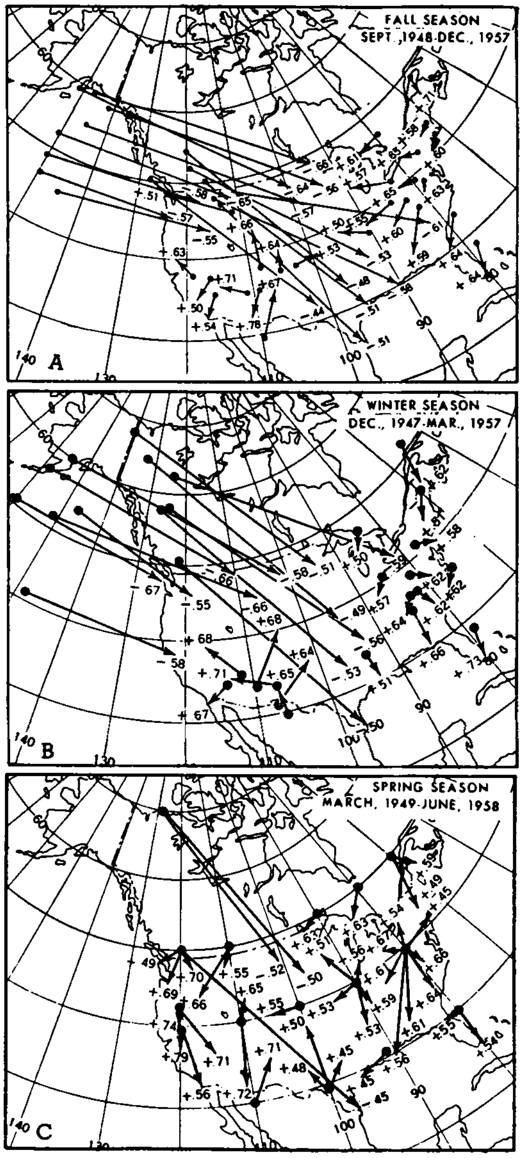

Fig. 5. Maximum simple linear correlation between 5-day mean anomalies of surface temperature and 700-mb height 2 days earlier for A) fall, B) winter, C) spring. The point of highest correlation is shown as a heavy solid circle, the city is located at the tip of each arrow, and the value of the correlation is written next to each city.

The above statements apply equally well to the winter season, for which the corresponding diagram is presented in figure 5 B. The marked similarity between figures 5 A and 5 B suggests that fall and winter tend to be climatologically

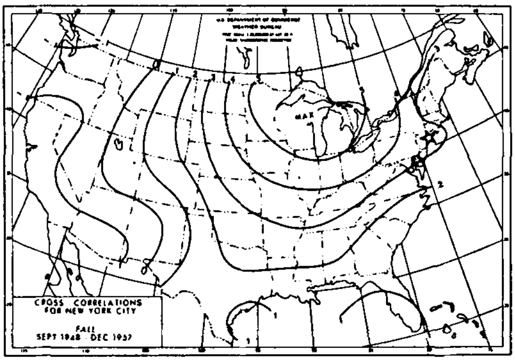

Fig. 6. Simple linear correlation coefficients between 5-day mean surface temperature anomaly at New York City (starred) and 5-day mean surface temperature anomaly at network of 39 cities 4 days earlier for 140 fall cases.

homogeneous with respect to the interrelation between temperature and circulation. In spring, on the other hand, a somewhat different relationship prevails since at all but three cities the correlations in figure 5 C are positive, and the arrows originate in the vicinity of the station. This indicates an increased importance of local height, at the expense of heights upstream, during the spring months. This characteristic is even more pronounced in summer, according to an earlier study by MARTIN and HAWKINS (1950), when local factors like cloudiness and vertical stability have a greater effect on surface temperature than does advection of air masses.

A second step in deriving the objective method was to obtain simple correlations between 5-day mean temperature at each city (T_4) and 5-day mean temperature for a period (overlapping by one day) centered 4 days earlier (T_0) at other cities. For example, the correlation field of figure 6 shows that at New York City during the fall the persistence of local 5-day mean temperatures 4 days apart gave a correlation of only +.33. Higher correlations were obtained by considering the temperature 4 days earlier at points to the west, and the maximum correlation (+.54) was given by a point in the upper Great Lakes. The positive sign shows that low temperatures in the Great Lakes tend to be followed by cold weather in New York 4 days later, and conversely for warm weather. This is an expression of the normal westerly steering flow prevailing at New York and may be associated with the fact that 5-day mean troughs and ridges move slowly eastward on the average (NAMIAS, 1947).

Figure 7 A summarizes the results of a similar analysis at all 39 cities during the fall season by showing the points of maximum correlation between temperatures 4 days apart. For example, at Albany, New York, the results are similar to those at New York City, with the highest correlation being +.57 with temperatures in the Upper Lakes region. The distance to the points of maximum correlation ranged from a few miles for cities in the Great Basin to as much as 800 miles for eastern cities like Jacksonville, Florida, and Buffalo, New York. It is interesting to note that many of the arrows point in a southerly as well as an easterly direction, suggesting that 5-day mean temperature anomalies tend to move southeastward over intervals of 4 days. This result may be due in large part to the strong tendency for cold polar anticyclones to move southward as well as eastward in North America (KLEIN, 1957). Figures 7 B and 7 C show that similar relations apply during the winter and spring seasons.

4. Screening Procedure

Thus far we have considered simple linear correlation coefficients which tell how well we could forecast temperature by using only one predictor. The next task is to improve these results by selecting additional predictors which will contribute significantly and independently to the temperature. This was done by a step-wise method of multiple regression known as the screening procedure which had been programmed for the IBM-704 electronic computer by MILLER (1958). As explained in our previous paper (KLEIN and others, 1959), the machine automatically selected the second most important variable by maximizing the multiple correlation between temperature and the combination of the first predictor plus any other. This process was repeated, and additional variables were selected one by one.

For example, at New York City in fall the most important single variable was the 700-mb height 2 days earlier at the intersection of 40° N and 70° W. The predictor which contributed the most additional information was the surface temperature 4 days earlier at International Falls, Minnesota. Combination of this

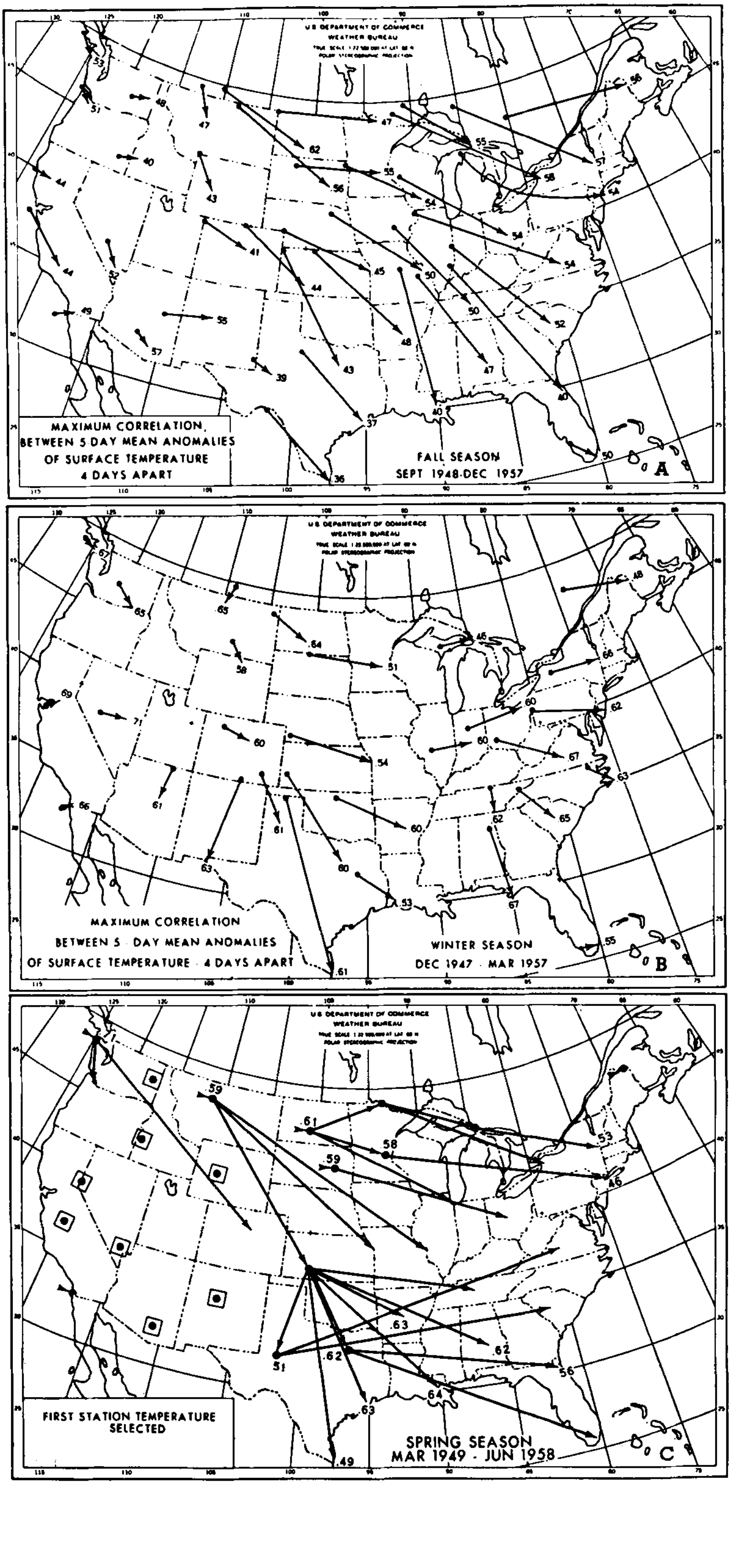

Fig. 7. A) Maximum simple linear correlation between 5-day mean anomalies of surface temperature 4 days apart for 140 fall cases. The point of highest correlation is shown as a heavy solid circle, the city is located at the tip of each arrow, and the value of the correlation is written next to each city. B) Same as A) for 140 winter cases. C) Similar to A) for 140 spring cases, except that the data were obtained from the first station temperature selected by the screening program rather than from the analyzed correlation fields. The long arrows originate at the predictor city and terminate at the predictand city, the short arrows indicate that the city itself was selected, and the squares indicate that no station temperature was selected as a predictor.

variable with the first yielded a multiple correlation of .68. The maximum multiple correlation using the combination of these first two predictors and any other one was .75, produced by adding the 700-mb height 2 days earlier at 50° N, 110° W. Since no additional variable was able to increase the explained variance by more than 2½ %, the screening process was stopped with these three predictors, all of which were located near the centers of maximum on the original correlation fields (figure 4 and 6). The linear multiple regression equation used in preparing the objective forecast during the fall season can then be written as:

$$T_4\,(\text{New York}) = .75 + +.104\,H_2(40\text{ N}, 70\text{ W}) + .263\ T_0\,(\text{Int. Falls}) - .097\,H_2\,(50\text{ N}, 110\text{ W}) \quad (2)$$

where temperatures are in degrees Fahrenheit and heights in tens of feet.

Multiple regression equations of the type illustrated in equation (2) were derived separately for fall, winter, and spring at each of the cities shown in figure 7. Physical interpretation of these equations was similar to the discussion already given for New York. In nearly all cases at least 2 heights were selected, one located in the vicinity of the remote center of correlation to the northwest, and the other near the local center. In addition, at least one temperature was usually selected, generally from the center of maximum correlation near or northwest of the city. At most cities the screening procedure was continued after the third step, with one to three additional heights and temperatures (generally to the west of the station) being selected as significant predictors on the basis of synoptic reasoning as well as statistical tests. The final equations contained from 2 to 7 variables (heights and temperatures in about equal proportion) and explained an average (for all cities and seasons) of 60 percent of the temperature variance of the developmental sample.

5. Tests on Independent Data

Since standard statistical tests of significance are not strictly applicable to this type of work, it is important to test the results on independent data. This was done by programming the multiple regression equations for the IBM-704 and then making temperature forecasts at each of the 39 cities on each Thursday for 12 weeks of the 1958 fall, from October 2 to December 18, for a total of 468 forecasts. The forecasts at each city were verified in terms of the explained variance.

$$EV = 1 - \frac{\sum_{i=1}^{12} (F_i - o_i)^2}{\sum_{i=1}^{12} (N - o_i)^2} \quad (3)$$

where F_i is the forecast, o_i the observed, and N the normal temperature. Values of EV were averaged for all 39 cities; the results are shown in table 1.

The first test was made by using only observed 5-day mean values of 700-mb height 2 days earlier (H_2) and surface temperature 4 days earlier (T_0) as predictors of T_4. Since this was exactly the same type of data used in the original derivation, it provided a test of the stability of the regression equations. The average EV obtained was 54.7 percent (line 1, table 1), which is almost as good as the EV of 60 percent explained on the original sample of 10 years (section 4). This result indicates that the regression equations are stable features which hold up well on new data.

Thus far we have discussed *observed* values available after the fact. How would the equations fare under actual operating conditions; i.e., using only *estimated* values of 5-day mean height and temperature available on forecast day? The required values of H_2 and T_0 were

Table 1. Percent of Variance of 5-day Mean Temperature Explained by Multiple Regression Equations during Fall of 1958 (mean of 12 cases at 39 cities).

1. Observed 700-mb height, observed surface temperature	54.7
2. Barotropic 500-mb height, estimated surface temperature	9.3
3. Observed 700-mb height, estimated surface temperature	49.2
4. Observed 500-mb height, estimated surface temperature	36.4
5. Baroclinic 700-mb height, estimated surface temperature	21.0

obtained with the aid of the barotropic height (0000 GMT) and district temperature forecasts as explained in section 2. The second line of table 1 shows that only 9.3 percent of the temperature variance at the 39 cities was explained in this way, a decrease of 45 percent from the results obtained in the first test with observed heights and temperatures.

In order to investigate the sources of this decrease, two additional tests were made. In the first, estimated values of temperature were substituted for observed values, but observed 700-mb heights were retained. The percent of variance explained was 49.2 (line 3), only 5 percent less than the *EV* obtained in the first test using observed temperatures. This result indicated that the estimated temperatures were good approximations to the observed temperatures for the five days centered on forecast day (T_0). It therefore appears that most of the errors in the temperature forecasts were due to errors in estimating the heights. A similar conclusion can be drawn from the results obtained with a different sample in our earlier study (KLEIN and others, 1959).

One source of height error is the fact that the barotropic forecasts were made at the 500-mb level, while the objective method was derived from heights at 700 mb. In all cases the required conversion was made by simply multiplying the 500-mb height by the ratio of normal 700 to 500-mb height for each month at each point.[4] To evaluate the inaccuracies of this conversion, a second test was made by using 5-day mean heights observed at the 500-mb level and reduced to the 700-mb level, in conjunction with the estimated temperatures. Only 36.4 percent of the variance was explained in this way (line 4), compared to 49.2 percent for observed 700-mb heights. The difference of 13 % can be attributed to improper reduction from 500 to 700 mb.

In an attempt to diminish this error, the regression equations were applied to baroclinic forecasts of 700-mb height rather than to barotropic forecasts of 500-mb height. During 1958 JNWP introduced into its operations a two-level model which predicts the thickness between 850 and 500 mb by taking account of horizontal advection and vertical motion. This prognostic thickness is subtracted from the barotropic 500-mb height to yield a prognostic 850-mb height, and the 700-mb height is then obtained by interpolation. The equation currently used by JNWP is:

$$Z_7 = 153 + .37\ Z_5 + .61\ Z_{8.5} \tag{4}$$

where Z is the height in feet at the 700-mb, 500-mb, and 850-mb levels, respectively.

Since the baroclinic model is run routinely up to only 36 hours in advance, it was not possible to obtain directly a 5-day mean map for the H_2 period at the 700-mb level. However, a good approximation could be obtained from the 700-mb 36-hour prognostic chart routinely prepared by JNWP at 1200 GMT on forecast day since it is valid on day 2 (fig. 3). This is the middle day of the required 5-day period, and therefore its map resembles the 5-day mean more closely than any other daily map (NAMIAS, 1947). Because the variability of daily maps is greater than that of 5-day mean maps, a damping factor of .5 was applied to the prognostic daily height anomalies before they were put into the regression equations. This is equivalent to taking H_2 as the mean of the 36-hour prognosis and the normal height at each point. Use of this H_2 as input data, again in conjunction with the estimated mean temperatures, led to a marked improvement. An average of 21.0 percent of the temperature variance was now explained (line 5), more than double the 9.3 percent explained by the barotropic 500-mb maps.

Part of this improvement was due to the fact that it was operationally feasible to base the 700-mb prognostic chart on observations made 12 hours later than those used as the starting point of the 500-mb mean map. In order to analyze additional causes of the improvement, experiments were performed with daily barotropic 500-mb prognoses prepared from 1200 GMT data on forecast day. Each of these was used as input for the regression equations (in conjunction with the same estimated value of T_0 used previously) to see how much of the temperature variance they could explain. The results are shown in figure 8 for the initial observed map (time 0) and the prognostic charts valid 24, 36, 48, and 72

[4] In symbols:

$$Z_7 = Z_5 \left(\frac{N_7}{N_5}\right)$$

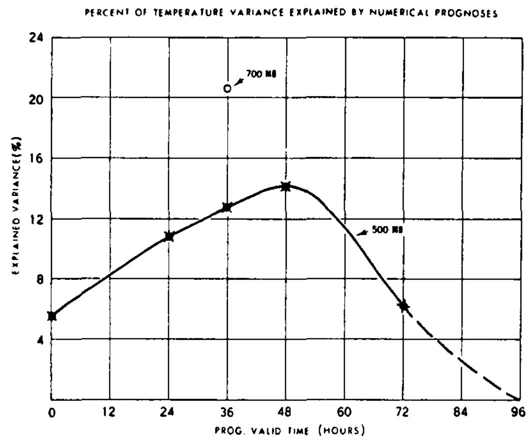

Fig. 8. Percent of total variance of 5-day mean surface temperature anomaly explained by barotropic forecasts of 500-mb height from 0 to 4 days in advance. The curve is extrapolated (dashed) after 72 hours. The open circle represents the variance explained by the 36-hour numerical forecast of 700-mb height. All values are averages of 39 cities during 12 fall cases from October 2 to December 18, 1958.

hours later. The points are easily fitted by a smooth curve which has a maximum explained variance of 14.2 percent at 48 hours, followed by a rapid decline to (extrapolated) zero around 96 hours. Thus, none of the daily barotropic forecasts was able to explain as much of the surface temperature variance as the 21.0 percent explained by the baroclinic 36-hour 700-mb map, which is plotted as an open circle in figure 8.

6. Verification in Terms of Temperature Classes

Guidance forecasts of 5-day mean temperature issued to Weather Bureau District centers by official forecasters of the Extended Forecast Section are expressed in five classes; much below, near, above, and much above normal. The center three have a climatological probability of $^1/_4$ each; the first and last, $^1/_8$ each. In order to express the objective temperature forecasts in terms of these classes, they were first multiplied by the reciprocal of the multiple correlation at each city since:

$$R = \frac{\sigma_f}{\sigma_0} \qquad (5)$$

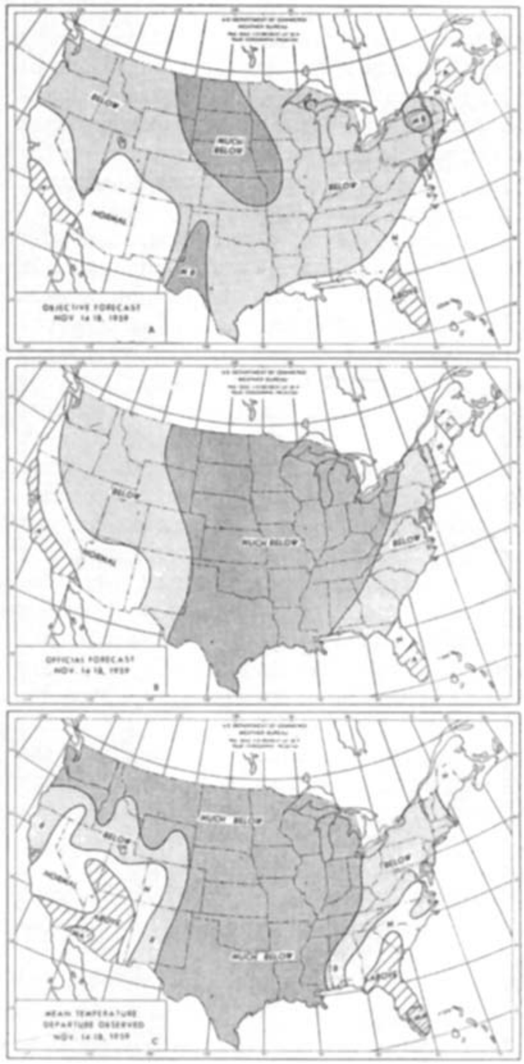

Fig. 9. Forecast and observed temperature anomalies for the 5-day period from November 14 to November 18, 1959. Analysis is in terms of the five temperature classes customarily used in the U.S. Weather Bureau (A) is the objective prediction, (B) is the official forecast, and (C) is the verifying map. The objective forecast had a skill score of 28 with 97 % correct within one class; the official forecast had a skill score of 53 with 96 % correct within one class.

where R is the multiple correlation coefficient, σ_f is the standard deviation of the forecast temperatures, and σ_0 is the standard deviation of the observed temperatures (EZEKIEL, 1941). The regression forecasts were inflated in this way in order to equate their variability with

Table 2. Verification of 5-day mean forecasts in terms of 5 temperature classes at 100 cities in United States during 3 seasons (1958—1959).[1]

Line	Forecast Method	Skill Score Fall	Winter	Spring	All	Percent in correct class Fall	Winter	Spring	All	Percent within 1 class Fall	Winter	Spring	All
1.	Objective-barotropic 500-mb mean	14.4	12.9	7.9	11.6	33.1	32.9	28.6	31.4	75.9	76.0	71.2	74.3
2.	Objective-baroclinic 700-mb daily	14.4	14.9	14.6	14.7	33.3	34.5	34.0	33.9	78.8	79.3	77.8	78.6
3.	Persistence-estimated mean temperature	9.9	9.3	6.8	8.6	29.6	30.4	27.7	28.3	73.1	73.6	67.8	71.4
4.	Continuity of error-graphical	16.4	13.7	6.3	11.9	34.7	33.6	27.3	31.7	77.2	75.9	69.2	73.9
5.	Official-Extended Forecast Section	15.2	15.3	13.9	14.7	33.7	34.8	33.3	33.9	80.3	79.9	77.7	79.2
6.	Official — 1952—1957	15.6	15.1	13.0	14.5	34.3	33.1	32.0	33.1	76.5	76.2	74.7	75.8
7.	Chance-random guess				0.0				21.9				59.4

[1] Fall — 32 Cases from September 30, 1958 to December 11, 1958
Winter — 39 Cases from December 14, 1958 to March 12, 1959
Spring — 39 Cases from March 15, 1959 to June 11, 1959

that observed and thus prevent an excessive number of forecasts in the near normal category.

Objective forecasts were prepared routinely three times a week on the IBM-704 from estimated values of T_0 and H_2; i.e., only information actually available on forecast day was used. Two different estimates of H_2 were tested: the 5-day mean maps prepared from the barotropic 500-mb forecasts at 0000 GMT, and the daily 36-hour forecasts prepared at the 700-mb level from the baroclinic model at 1200 GMT. The inflated forecasts (equation 5) were converted into one of the five temperature classes according to the appropriate class limits at each city (NAMIAS, 1947). After the forecast class for each city was plotted on a map, lines were drawn by interpolation delineating the distribution of the five classes over the entire country. A sample of a good objective forecast is shown in fig. 9A. 110 forecast maps covering three seasons were verified at a grid of 100 points fairly evenly spaced over the United States.

The verification results are summarized in table 2 on a seasonal and over-all basis. The first line shows that when the regression equations were applied to the barotropic 500-mb mean maps, an average (for all three seasons) of 31.4 % of the objective forecasts were in exactly the correct temperature class, for a skill score[5] (BRIER and ALLEN, 1951) of 11.6. Credit is sometimes given for a near hit, such as forecasting much below normal when below verifies. On this basis, the last column shows that 74.3 % of the objective forecasts were either exactly correct or correct within one class for all seasons combined. Line 2 shows that application of the objective method to the baroclinic 700-mb 36-hour forecasts resulted in an improvement to 33.9 % in the correct class, an average skill score of 14.7, and 78.6 % correct within one class. The scores in line 2 were not only higher than those in line 1 but also more consistent, with little variation from one season to the next.

The remainder of the table contains various controls against which to measure the accuracy of the objective forecasts. Line 3 shows the scores obtained by local persistence; i.e., using the estimated mean temperature (T_0) as a forecast of T_4. Although these scores were definitely higher than those expected by chance (line 7), they were consistently lower than the scores given by either of the objective fore-

[5] The skill score is

$$S = \left(\frac{R-E}{T-E}\right) 100$$

where R is the number of correct forecasts, T is the total number of forecasts, and E is the number expected correct by chance. S is 100 for perfect forecasts, and it is positive, zero, or negative for forecasts better, equal, or worse than chance forecasts.

casts. Thus the skill of the latter cannot be ascribed solely to persistence of a short-range forecast.

Line 4 gives the scores obtained by a method based upon simple two-parameter graphs (MARTIN and HAWKINS, 1950) and corrected by applying continuity of error in the manner described by NAMIAS and Collaborators (1958). This method was approximately equal in skill to the objective forecasts based upon the 500-mb barotropic mean (line 1), but was not as good as the objective method applied to the baroclinic 700-mb maps (line 2).

Line 5 contains the verification of the official guidance forecasts issued by experienced forecasters of the Extended Forecast Section. During the 9-month test period, these forecasters were able to make use of objective temperature predictions based on the barotropic 500-mb mean maps, local persistence, and continuity of error; but the objectives made from the baroclinic 700-mb 36-hour charts were not available to them. As a result, the scores in line 5 were decidedly superior to those in lines 1, 3, or 4 but merely equal to those in line 2. It is encouraging to note that both the official and baroclinic objective forecasts were superior to the average of the official forecasts during the 5 years from 1952 to 1957 (line 6), especially when near hits were counted (last column).

7. More Recent Results

Numerous attempts have been made to improve the results described in the preceding two sections by obtaining better forecasts of the 700-mb mean circulation. Although the details of these experiments will not be presented here, they can be summarized by stating that in no case has any significant increase in accuracy been attained. The results clearly indicate that the 36-hour forecast of 700-mb heights made by JNWP at 1200 GMT of forecast day is the best numerical input currently available for the objective temperature method described in this paper.

Beginning on September 15, 1959, objective temperature forecasts have been prepared on an operational basis, using these 700-mb heights, instead of the 500-mb mean barotropic heights, as input to the regression equations. The resulting predictions have been verified in terms o temperature classes (section 6) for 116 cases through June, 1960. The results are summarized in table 3, along with comparative verification for other methods.

Comparisons of tables 2 and 3 (line 2) shows that the objective temperature forecasts during the fall of 1959 were definitely superior to corresponding forecasts made the previous fall. This improvement was accomplished despite a sharp drop in the skill of the forecasts based upon either persistence (line 3) or continuity of error (line 4). The objective predictions continued to exhibit superiority over both persistence and continuity of error during the winter and spring (table 3), despite a marked difference in the character of these seasons. The winter of 1959—60 was unusually persistent in nature, and all temperature fore-

Table 3. Verification of 5-day mean forecasts in terms of 5 temperature classes at 100 cities in United States during 3 seasons (1959—60).[1]

Line	Forecast Method	Skill Score				Percent in correct class				Percent within 1 class			
		Fall	Winter	Spring	All	Fall	Winter	Spring	All	Fall	Winter	Spring	All
1.	Objective-barotropic 500-mb mean	Not computed											
2.	Objective-baroclinic 700-mb daily	19.0	21.3	13.9	18.1	37.0	38.1	32.8	36.0	81.8	83.3	80.8	82.0
3.	Persistence-estimated mean temp.	8.9	19.4	9.7	12.8	29.3	36.8	29.6	32.0	72.4	74.7	70.5	72.5
4.	Continuity of error-graphical	11.3	18.0	10.7	13.4	31.1	35.6	30.3	32.4	73.1	76.7	73.3	74.4
5.	Official-Extended Forecast Section	20.4	27.0	17.2	21.6	38.1	42.6	35.4	38.8	79.6	85.3	79.3	81.4

[1] Fall — 37 Cases from September 15, 1959 to December 8, 1959
Winter — 40 Cases from December 10, 1959 to March 10, 1960
Spring — 39 Cases from March 13, 1960 to June 9, 1960

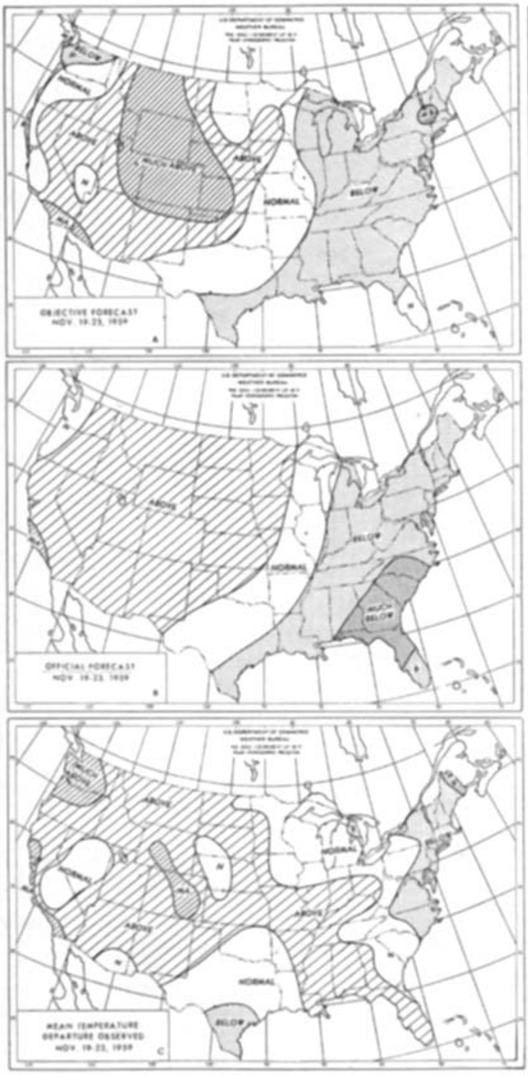

Fig. 10. Objective forecast (A), official forecast (B), and observed (C) temperature anomaly classes for the 5-day period November 19—23, 1959. The skill scores were 20 for the objective and 33 for the official, but both forecasts had 83 percent correct within one class.

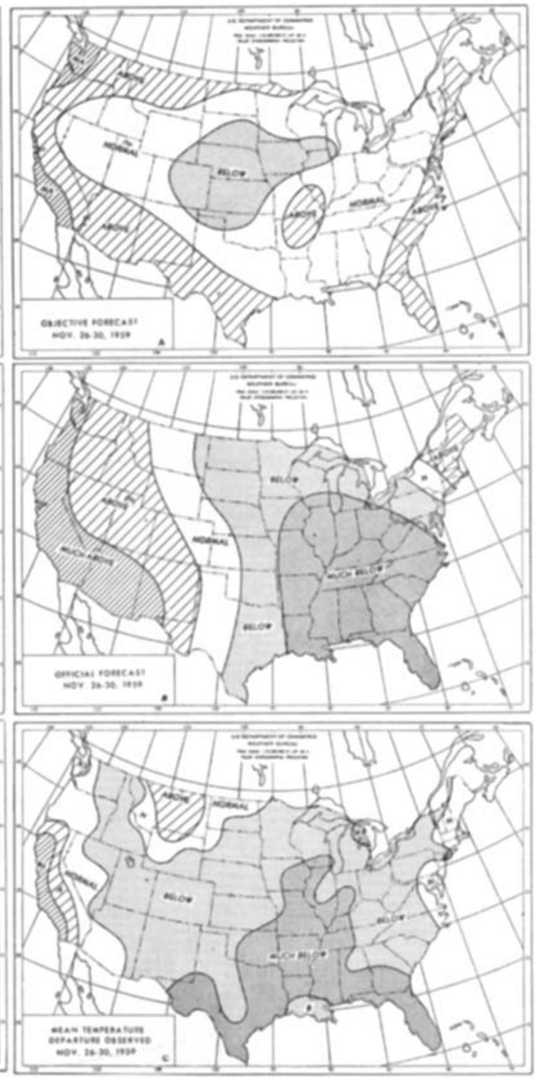

Fig. 11. Objective forecast (A), official forecast (B) ,and observed (C) temperature anomaly classes for the 5-day period, November 26—30, 1959. The skill scores were 4 for the objective and 14 for the official, the percentages correct within one class were 67 for the objective and 75 for the official.

casts experienced a sizeable increase in skill. The spring of 1960 was rather difficult and changeable, and the accuracy of the forecasts diminished sharply. For all three seasons combined, the objective temperature predictions had a skill score of 18.1, a decided improvement over the corresponding score of 14.7 obtained during the preceding year (table 2). Because many factors could have contributed, and because only two years are involved, it is difficult to know to what extent this improvement will continue to hold in the future.

Line 5 contains the verification of the guidance forecasts issued to the districts by official forecasters of the Extended Forecast Section. The average skill score of 21.6 was definitely better than the average skill score of the objective forecasts (18.1), and this superiority was maintained during each season. It was also a marked increase over the official score of the previous year (14.7, table 2), and the highest score on record for any 9-month period since 1941.

It is difficult to ascertain the exact causes of this increased skill over previous years. The official forecasts are based primarily upon prognostic charts of the 5-day mean 700-mb circulation for the H_4 period, plus numerous additional tools, as explained by NAMIAS in several papers (1947, 1958). One of these tools is the objective temperature forecast based on the 36-hour 700-mb baroclinic heights. These were first made available to the official forecasters in the fall of 1959. Although this factor may account for part of the increased skill, it is also possible that improvement in the other tools or differences in the general circulation may have been responsible.

It is interesting to note that the official forecasters were able to improve the objective predictions when verified in terms of skill score and percent exactly correct, but not in terms of percent correct within one class (last column of table 3). The objective forecasts had an over-all score of 82.0 in this category, a slight improvement over the official score of 81.4, and a decided improvement over the objective score of 78.6 during the previous year (table 2).

Two cases in point are illustrated in figures 9 and 10. In the first forecast, prepared Nov. 12, 1959, a large area of below normal on the objective (fig. 9 A) was modified to much below normal on the official forecast (fig. 9 B), in part because of extremely cold air and high sea level pressure initially available in the polar source region over western Canada, and nearly all of this much below normal actually verified (fig. 9 C). A few days later, on Nov. 17, the objective prediction (fig. 10 A) called for a very large change, from much below to much above normal, in the Great Plains, but the official forecast (fig. 10 B) correctly modified this to above normal in view of the initial presence of cold air and snow cover in much of the area. A case when the official forecaster successfully called for a large change in regime despite a poor objective is shown in fig. 11, while a case where the official forecaster incorrectly modified the objective is illustrated in fig. 12.

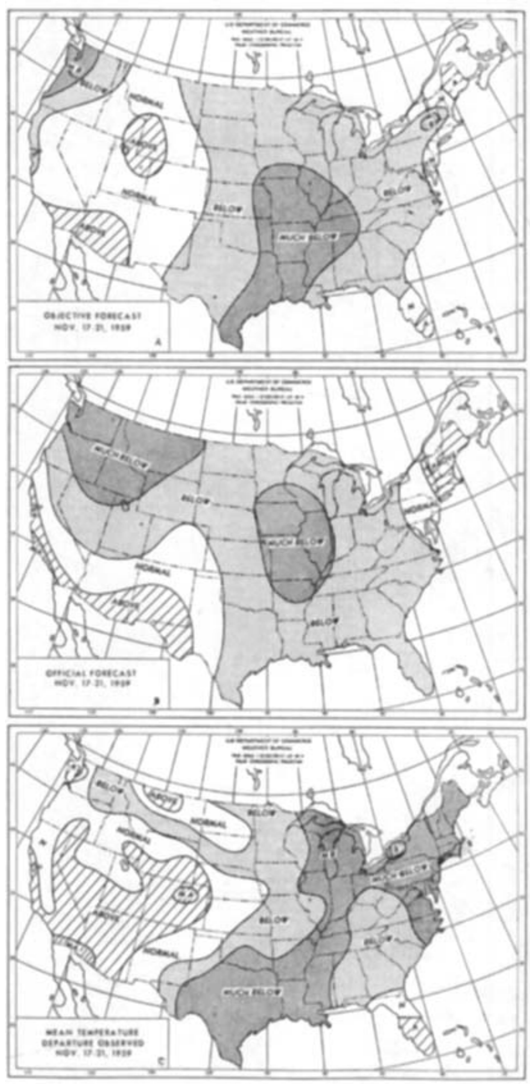

Fig. 12. Objective forecast (A), official forecast (B), and observed (C) temperature anomaly classes for the 5-day period, November 17—21, 1959. The skill scores were 22 for the objective and 3 for the official; the percentages correct within one class were 92 for the objective and 73 for the official.

8. Suggestions for Improvement

There are several sources of error in the objective method as currently applied. In the first place, the 36-hour baroclinic forecasts (plus the normal) are imperfect representations of the desired 5-day mean heights for the H_2 period. Even if these forecasts were completely accurate for the first 36 hours, there would still be errors in estimating H_2 because of changes in the circulation pattern which may occur during the next few days. A smaller error can be attributed to the fact that the current 5-day mean surface temperatures (T_0) are not perfectly estimated because of inaccuracies in the district forecasts for the next 2 ½ days. As a result of these limitations in estimating values of H_2 and T_0, the objective temperature forecasts tend to be conservative; i.e., they are more accurate for the beginning than the end of the 5-day period. In fact, a separate test was made which showed that the objective forecasts explained 40 percent of the temperature variance for the 5-day period centered two days after forecast day (T_2), almost twice as much as the 21 percent they explained for the T_4 period (section 5).

Even if perfect forecasts of H_2 and T_4 were available as input for the objective method, it could explain at most only an average of 60 percent of the temperature variance by the relatively simple technique used here (section 4). A more sophisticated approach utilizing actual air trajectories and stratifying the data might result in greater accuracy. Further improvement will occur when better forecasts of the 700-mb circulation for the H_4 period and each of its daily components become available. Consideration should also be given to the initial thermal distribution over a region larger than the contiguous United States, with particular reference to abnormalities of temperature in air mass source regions. Finally, and perhaps most important, the effects of parameters other than 700-mb height and surface temperature should be included, such as prognostic thickness, sea level pressure, snow cover, ocean temperature, and cloudiness associated with vertical motion.

9. Conclusion

An objective method of forecasting 5-day mean temperatures over the United States has been developed. For this purpose a statistical screening technique was applied to a large mass of historical data to produce a fairly reliable and stable set of multiple regression equations. It has been demonstrated that the best numerical forecast of upper level heights currently available as input for these equations is the 36-hour baroclinic 700-mb prognosis regularly prepared by JNWP from 1200 GMT data on forecast day. By use of these heights, a reasonable level of skill in predicting 5-day mean surface temperatures has been maintained over a 2-year test period.

Since the objective temperature forecasts can be prepared in a few minutes with the aid of the IBM-704, they should serve as a valuable and practical forecast tool. Results during the single year that this objective aid has been available to the official forecaster are encouraging. It is hoped that continued improvement will occur in the future as a result of greater accuracy of the numerical height and district temperature forecasts, a greater familiarity on the part of the forecasters with pitfalls and strong points of the objective method in individual situations, and refinements in the method itself.

Acknowledgments

The authors wish to acknowledge the valuable aid received from the following individuals in various phases of data processing and machine programming: Capt.John Brown, Capt. Russel Harris, and Mr. Frank Lewis of the U.S. Air Weather Service; and Mrs. Mary Abernathy, Mrs. Marion Hess, Miss Myrtle Wagner, Mrs. Elizabeth Scott, Mr. Thomas Carpenter, Mr. Julius Bosen, Mr. Charles White, Mr. Burton Lovan, and Mr. Michael Gaidurgis of the U.S. Weather Bureau.

REFERENCES

BRIER, G. W. and ALLEN, R. A., 1951: Verification of weather forecasts, *Compendium of Meteorology*, T. F. Malone, ed., Boston, Mass. American Meteorological Society, pp. 841—848.

CHARNEY, J., 1949: On the physical basis for numerical prediction of large-scale motions in the atmosphere, *J. Meteor.* **6**, pp. 371—385.

CRESSMAN, G. P., 1958: Barotropic divergence and very long atmospheric waves, *Monthly Weather Review* **86**, pp. 45—52.

EZEKIEL, M., 1941: *Methods of correlation analysis*, New York, N.Y., John Wiley and Sons, Inc., 531 pp.

KELLY, E. A. and MALONE, T. F., 1955: Application of synoptic climatology to five-day weather prediction, talk presented at meeting of Amer. Meteor. Soc. on January 26, 1956, in New York. Abstract in *Bull. Amer. Meteor. Soc.* **36**, p. 545.

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.

KLEIN, W. H., LEWIS, B. M. and ENGER, I., 1959: Objective prediction of five-day mean temperatures during winter, *J. Meteor.* **16**, pp. 672—682.

MARTIN, D. F. and HAWKINS, H. F., JR., 1950: Forecasting the weather: The relationship of temperature and precipitation over the United States to the circulation aloft, *Weatherwise* **3**, pp. 16—19, 40—43, 65—67.

MARTIN, D. F. and LEIGHT, W. E., 1959: Objective temperature estimates from mean circulation patterns, *Mon. Wea. Rev.* **77**, pp. 275—283.

MILLER, R. G., 1958: The screening procedure, Part II of *Studies in Statistical Weather Prediction*, final report, contract no. AF 19(604) 1590 Travelers Weather Research Center, Hartford, Conn., pp. 86—136.

NAMIAS, J., 1947: *Extended forecasting by mean circulation methods*, Washington, D.C., U.S. Weather Bureau, 89 pp.

NAMIAS, J. and Collaborators, 1958: Application of numerical methods to extended forecasting practices in the U.S. Weather Bureau, *Mon. Wea. Rev.* **86**, pp. 467—476.

ROWE, W. M., 1944: Relationship between surface temperature and mean virtual temperature in the lower troposhere, U.S. Weather Bureau *Research Paper* No. **22**, 8 pp.