Long-Range Weather Forecasting Using an Analog Approach

ZOLTAN TOTH*

Central Institute for Weather Forecasting, Budapest, Hungary

(Manuscript received 18 September 1987, in final form 18 November 1988)

ABSTRACT

An analog selection method relying on the coincidence of main features (large-scale ridge lines) in the Northern Hemisphere is presented and used for making 30-day weather forecasts for Hungary. Numerous analog model trials were tested, with the aid of the advance selection of the "best circulation analogs" of the Atlantic-European forecast region, for every target month of the 27-yr calibration period and the 5.5 yr test period. The best predictor types are a one pentad (i.e., 5-day) predictor period with spatial smoothing (allowing slight longitudinal shifts between pressure patterns), and a 2 pentad predictor period with time averaging (with a weighting factor of 0.4 on data from outside the forecast region in both cases). A subset of each group of analogs with similar circulations during the forecast period was identified. Using the subset leads to further significant increases in skill.

Monthly weather forecasts for temperature (5-day subperiods) and precipitation quantity (10-day subperiods) in any of three climatologically equal probable categories were given. Different statistics, which were slightly but significantly better than chance expectation and persistence, were employed to assess the skill of the forecasts. By means of the previously chosen best circulation analogs, the potential monthly analog predictability based on our dataset and methods were also determined. Accordingly, the operable forecasting method realizes 30%–60% of potential predictability. Using lengthened data series for selecting analogs, the improvement in both analog predictability and actual forecasting skills was investigated. Extrapolating the experimental data for the future by comparing it with a logistic curve, an estimate was obtained of increased forecast skill from the present 38%–39% to 42% within 15 yr.

1. Introduction

Analog forecasting consists of searching for analogs to a present or preceding situation and then predicting weather for the forthcoming period based on similar cases in the past. It has been a useful tool for synopticians for a long time, although the improvement of numerical weather prediction has resulted in a shift in its application to longer time scales. The main advantage of its use is that it yields a real solution to a difficult problem (Namias 1951, 1968; Bergen and Harnack 1982). It automatically contains the microclimatological influences at every location of the forecast. In addition, analog prediction can be prepared objectively and quickly with the aid of high speed computers.

According to many authors, the biggest drawback to the analog approach is the limited extent of available historical data. Upper air data and sea surface temperature (SST) records, which are usually used for measuring similarity, are available only for the past 30 or 40 yr. That is why one cannot always find "good" analogs.

Lorenz (1969a) investigated the doubling time of the initial errors of the analogs, while Gutzler and Shukla (1984) examined the rms difference between the most closely analogous maps defined over the middle latitudes of the Northern Hemisphere for up to 10 days. Their results were disappointing in short- and medium-range forecasting, as even the persistence forecasts were superior to the analog forecasts. (Lorenz, and Gutzler and Shukla, used 5 and 15 yr of data, respectively.)

However, in the case of long-range forecasting, we can assume that there is no need for such an exact point-by-point coincidence between the analogs during both the predictor and the forecast period. It is not surprising that there are a great number of papers presenting objective approaches to the use of analogs in long-range forecasting, and some of them show real skill over persistence and random forecasts for different locations and periods of the year. Nicholls (1980) and Bergen and Harnack (1982) present a wide review of them.

Following Gutzler and Shukla's (1984) suggestion, the data basis of this study represents only the largest spatial scales of hemispheric data. It consists of the time averaged location of well-developed ridge lines

* Present affiliation: Department of Meteorology, University of Maryland, College Park, Maryland.

Corresponding author address: Dr. Zoltan Toth, Dept. of Meteorology, University of Maryland, College Park, MD 20742.

observed in middle latitudes in the middle troposphere. One of our objectives was to determine the potential monthly analog predictability of central Europe temperature and precipitation. This was achieved for three categories for 5-day (hereafter referred to as pentad) and 10-day (hereafter referred to as dekad) subperiods, respectively. Selecting the 30-day best circulation analogs for every "target" month included in our dataset, the coincidence of their weather categories has been assessed. Even if there is potential for analog long-range forecasting, some questions still remain. For instance, how long should the predictor period be to obtain the best analogs (30-day) found in the predictability experiment? Also, with what weighting factors should the predictor data from outside the Atlantic–European forecast region be considered? Finally, is it useful to employ an additional selection criterion based on the circulation similarity of the chosen analogs during the forecast period? The answers to these questions are not only useful in the construction of a reliable forecasting method but may modestly contribute to the description of large-scale and long-term atmospheric phenomena. Another question may be posed when examining an experimental forecasting scheme based on our 27-yr archive: namely, what is the relationship between the improvement in forecast skill and the length of the data series that can be estimated by using a 5.5 yr independent test period?

2. Data description

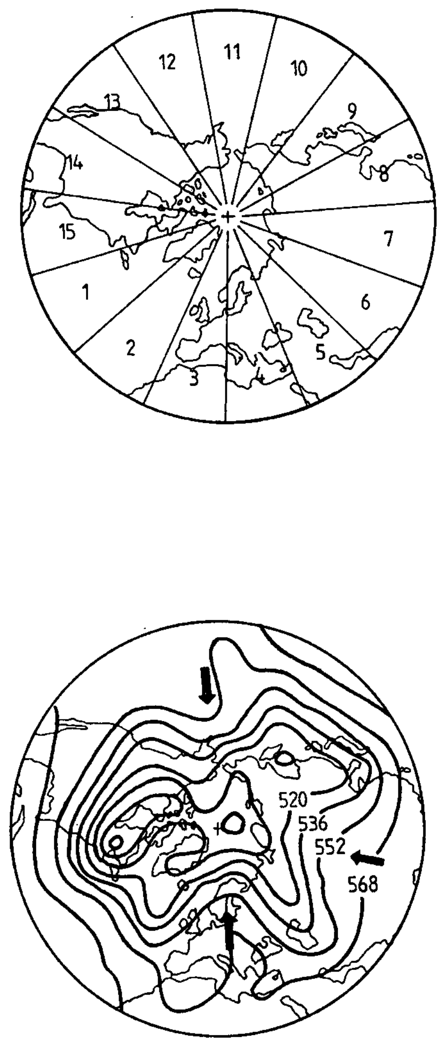
a. Circulation data

In this study the analog search is restricted to the Northern Hemisphere extratropical flow pattern. The 500 hPa level has been chosen to represent the state of the atmosphere. This type of predictor field is often used for selecting analogs. Ratcliff (1974) found that it had more predictive value than analogs of most other types. Since the aim of this study is monthly weather forecasting, a further reduction of data was made. We have attempted to use data representing only the largest spatial scales of hemispheric circulation.

As Sawyer (1970) indicates, high pressure systems are major contributors to fluctuations in atmospheric circulation on time scales of 15 to 60 days. Blocking phenomena have been investigated and catalogued recently by, among others, Treidl et al. (1981), and Knox and Hay (1984). Our dataset is based on the ridge line catalogue of Titkos (1981). He examined the daily Northern Hemisphere 500 hPa heights in order to locate well-developed ridges that had existed for at least 3 days, and divided them into 15 sectors by longitude (see Fig. 1). Since our aim is long-range forecasting, it seemed desirable to filter out short-term fluctuations. We used the time-averaged data of pentads because their duration is comparable with that of natural synoptic periods (see Barry and Perry 1973). Thus, we added up the 15-dimensional binary vectors of Titkos's (1981) catalogue every 5 days. Symbol r_{ij} represents (after the word ridge) these new vectors of our dataset, where $i = 1, \cdots, 15$ denotes the sector (see Fig. 1), and $j = 1, \cdots, 2372$ denotes the pentad in chronological order, beginning with 1–5 January 1953, and ending with 25–29 June 1985.

a

b



c

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
At	At	Eu	Eu	Eu	As	As	As	Pa	Pa	Pa	Am	Am	Am	At
0	0	0	1	0	0	1	0	0	0	1	0	0	0	0

FIG. 1. An example of data reduction made by Titkos (1981): (a) the 15 sectors of the Northern Hemisphere, and (b) 500 hPa height field for 0000 UTC 18 January 1982. Ridges considered are indicated by arrows. (c) The 15-dimensional binary vector yielded from (b) by (a).

The components of the r_{ij} range from 0 to 5, according to the number of days when a ridge was found in the respective sectors (see Fig. 2). (In the event of a leap year, pentad number 12, from 25 February to 1 March, contains 5 days, as data from 29 February are neglected.) The dataset was divided into two parts: a 27-yr calibration period (or archive, $j = 1, \cdots, 1971$) and a 5.5-yr independent test period ($j = 1972, \cdots, 2372$). (Circulation analogs were tested only on data for the 5-yr period, $j = 1972, \cdots, 2336$.)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
At	At	Eu	Eu	Eu	As	As	As	Pa	Pa	Pa	Am	Am	Am	At
0	0	0	5	0	0	3	0	0	0	1	3	0	0	1

FIG. 2. An element of our dataset (including the day illustrated in Fig. 1) representing the state of the atmosphere from 16 to 20 January 1982.

b. Weather elements

As the designed forecasting system was intended for making monthly weather predictions for Hungary, it was convenient to consider the weather elements observed in Budapest. Based on earlier attempts (e.g., Schuurmans 1973), it was decided to forecast the mean temperature of 5-day periods (pentads) within the forecast month in terms of three climatologically equal probable categories. The class limits for every pentad of the year (73 pentads) have been determined empirically by regarding the temperature anomaly distribution of not only the pentad in question (27 anomalies of the archive) but also the symmetrically surrounding 12 pentads' anomalies. Then all observed pentad departures in the dataset were assigned to one of the three equally probable categories; either below, near, or above normal, based on these limits. During the weather forecasting experiment other category limits were employed as well (see section 3b). Those classes are defined so as to occur 30%, 40% and 30% of the time, respectively, over the whole calibration period.

Preliminary experiments with the model show that successful monthly precipitation forecasts can be made for only 10-day subperiods. The 36 dekad class limits of the year (excluding the 73rd pentad's data) were achieved by a method similar to that used for determining temperature limits, and which considered the surrounding 6 dekads precipitation data of the calibration period as well. The corresponding categories for observed 10-day precipitation totals are light, moderate and heavy.

3. Procedure

a. Analog selection methodology

For matching the circulation states of the atmosphere, as represented in our dataset as r_{ij} and r_{ik}, a Euclidean distance function has been used:

$$\Delta_{jk} = (\sum_i w_i(r_{ij} - r_{ik})^2)^{1/2}, \qquad (1)$$

where w_i denotes a weighting factor which can be changed for different experiments. When comparing two (e.g., 6-pentad) series of atmospheric states, the following formula has been used:

$$\Delta_{jk,6} = \Delta_{jk} + \Delta_{j+1,k+1} + \cdots + \Delta_{j+5,k+5}. \qquad (2)$$

There is another way of comparing data from a period longer than 1 pentad. Using the example of six pentads again, let

$$\Delta^a_{jk,6} = (\sum_i w_i((r_{ij} + r_{i,j+1} + \cdots + r_{i,j+5}) - (r_{ik} + r_{i,k+1} + \cdots + r_{i,k+5}))^2)^{1/2}, \qquad (3)$$

where a represents the time averaging method.

In their discussion, Gutzler and Shukla (1984) suggest another way to compare atmospheric states. It is obvious that two circulation patterns which are identical except for a 20° shift in phase would not be considered as a good analog pair in our computations, although their subsequent dynamical evolution might be fairly similar. It is possible to avoid this problem by letting

$$\Delta^s_{jk} = (\sum_i w_i((r_{ij} + r_{i+1,j}) - (r_{ik} + r_{i+1,k}))^2)^{1/2}, \qquad (4)$$

where s denotes spatial smoothing. It can be seen that through this formula, slight phase differences between the large-scale atmospheric phenomena are not considered. Longer series can be compared by altering (4) as (2) or (3).

An analog selection process involves scanning forward and backward in time, comparing a particular predictor period (or, in the predictability experiment, target month) to all other similar periods within our archive. The comparison was made by use of one of the distance functions with a fixed weighting factor. For example, if February forecasts are desired, then data *preceding* the target month are compared with data surrounding the opening pentad of all other Februarys in the data series. (Target month refers to 30-day forecast periods of 1–30 January, 31 January–1 March, . . . , and 27 November–26 December, for which a forecast is desired.)

More precisely we have computed Δ_{jk} for a fixed predictor period (or target month) starting with pentad "j", for analog periods starting with pentad "k", for all

$$|j - k| = p + 73q, \quad \text{where} \quad -6 \leqslant p \leqslant 6 \quad \text{and} \quad 1 \leqslant q \leqslant 26.$$

Shortcomings connected with the shortness of data series are assumed to be of similar magnitude to those deriving from six pentad (30-day) shifts of periodical or quasi-periodical external factors. In other analog trials, Schuurmans (1973), for example, used a narrower window (15 days), while Gutzler and Shukla (1984) used a longer one (as much as 45 days) during their wintertime investigations. Since our aim is weather forecasting, we do not exclude analogs occurring in the same year (and hence not entirely independent of each other) except those within 10 days ($|k_1 - k_2| \leqslant 3$), which represent, evidently, the same cases.

This analog selection process is repeated for data *preceding* every target month in our archive (12 months $\times$ 27 yr − January 1953 without predictor data = 323 target months), as well as for data preceding the 66 target months in the 5.5 yr testing period. Numerous analog models have been executed employing various combinations of predictor period, distance function, and weighting factors. In all cases the 15 analogs having the lowest value of Δ_{jk} have been singled out in every predictor period j. [A similar procedure was also used in the predictability experiment to obtain the 15 best analogs but using the circulation data of the target months; see sections 3c(1) and 4b(1).]

As Craddock et al. (1962) described in their paper, a commonly used approach in analog prediction was the examination of subsequent development of analog cases. Bergen and Harnack (1982), and Livezey and Barnston (1988) also pointed out the benefit of using more than one analog. In this study an attempt has been made, with the aid of a simple hierarchical cluster analysis (see appendix A), to objectively compare the analogs during the forecast month. For this, the circulation data of the 15 top-ranking analogs and the distance function (2) have been used. The clustering process was carried out until the first cluster consisting of six analogs emerged. The six analogs chosen from the 15 original ones are the final analogs used for the purpose of weather forecasting. (Note that these analogs do not necessarily have the lowest values of Δ_{jk}.)

b. *Weather forecasting procedure*

In this study temperature and precipitation forecasts for Budapest are made in three climatologically equal probable categories (described in section 2b) for 5-day and 10-day subperiods, respectively, within the target months. The recorded temperature and precipitation total classes of the six final analogs (and also those of the six best analogs in the predictability experiment, see section 4b) are considered, pentad by pentad and dekad by dekad, respectively. The forecast "extreme category", e.g., below normal (or light), for a pentad (or dekad) is assigned only if it is present in more than one other analog year than the other extreme class. In all other cases the mean category (near normal or moderate) is forecast. For example, a class distribution of (3, 1, 2) for above, near, and below, respectively, would *not* lead to an above forecast but a distribution at (4, 0, 2) would.

This signifies that the mean category is selected, not only when it is the major category, but when no great differences between the analog categories can be found. As indicated earlier, in the temperature forecasting technique, analogs occurring only in either the upper or lower 30% of the temperature range (instead of 33%) were regarded as above or below normal, respectively. (Nevertheless the forecast is given in the climatologically equal probable categories.) These resulted [as we can see in section 4b(2) and 4b(3)] in an undesirably high rate of forecasts falling in the "normal" category. But, at the same time, the reliability of extreme forecasts has increased while no considerable change in the mean category was found. The aforementioned method of weather forecasting, based on the majority category, may be considered as an additional (third) analog selection process.

c. *Forecast verification*

We endeavored to judge our circulation and weather forecasts separately. Unlike Barnett and Preisendorfer (1978), Bergen and Harnack (1982), and Livezey and Barnston (1988), who compared different analog models through forecasts of surface weather elements obtained from them, we intend to pick out the best analog selection models by means of direct assessment of circulation forecasts. We did this through the optimization of the analog selection method, in the hope of avoiding the uncertainty involved in the relationship between large-scale circulation patterns and local weather elements.

1) CHECKING ANALOG QUALITY

First we have selected circulation analogs to the data of every whole target month. In this analysis (and not forecasting) step, only the Atlantic–European region was considered in (2) with $w_1, w_2, \cdots, w_5 = 1$ and $w_6, w_7, \cdots, w_{15} = 0$, because the closest relationship between circulation analogs and their corresponding weather elements in Europe was found to be in these five sectors. [Note that the longitudinal borders of this forecast region coincide with the Eastern–Atlantic teleconnection pattern of Wallace and Gutzler (1981)]. The 15 circulation analogs for every target month having the lowest Δ_{jk} value are the best analogs that can be found in our dataset. After this, we tested various analog forecasting models yielding 15 good analogs to the different predictor fields in every case. Then we checked the lists of the best analogs and the corresponding good analogs to see how many pairs they have in common. The coincidence factor (for all the 323 target months) served as a measure of quality for different analog models. Later, the same checking procedure was used for the six analogs selected from the 15 good ones by cluster analysis.

In both cases the binomial distribution was used to define the significance level at which the method in question described by the total number of coincidences through the calibration and test period is better than blind forecasts gained by randomly selecting analogs (see appendix B). Differences between various analog trials were examined by an appropriate chi-square test (see appendix B).

2) EVALUATING WEATHER FORECASTS

As described in sections 2b and 3b, both 5-day temperature and 10-day precipitation forecasts for Budapest, within a month, are given in three climatologically equal probable categories. The skill score presented here denotes the percentage of correct category forecasts. We computed this for the entire series of the calibration and test periods, and also for particular years and months. The results were compared with random chance using binomial distribution (see appendix B). Using an appropriate chi-square test, we checked whether the models significantly outperform persistence forecasts (see appendix B). Gordon's (1982) scheme for evaluating the skill of categorical forecasts was also computed using a linear penalty factor. In this scheme, predicting the other extreme category is twice as inaccurate as a one-class error (one category wrong forecast).

4. Results

a. Optimizing the analog selection

1) THE LENGTH OF PREDICTOR PERIOD

Although there have been attempts with shorter (mainly 0.5 month or 10-day) periods, most authors have used monthly data for selecting analogs to forecast 1 month ahead (e.g., Bergen and Harnack 1982; Radinovic 1975; Ratcliff 1974; Schuurmans 1973). The optimum length of the predictor period (to the best of the author's knowledge) has been systematically investigated for monthly statistical–dynamic forecasts (Roads and Barnett 1984) and persistence forecasts (Harnack et al. 1986), but not for analog methods. There are also other questions to be answered, including if time averaging or spatial smoothing needs to be employed.

Figure 3 shows our results obtained for the calibration period. The figure shows that the time averaging process [by Eq. (3)] increases analog quality, at least for two-pentad predictor periods. Note that for the investigations described in this section, all weighting factors were fixed to 1. It is also obvious that spatial smoothing between neighboring sectors' ridges is of considerable value for one pentad predictor periods. The difference between the number of best analogs yielded by Eq. (1) (216, as much as the estimation of the chance expectation, see section 4a(2)) and Eq. 4 (252) has a significance level of 10% (according to a chi-square test). Although other differences are not

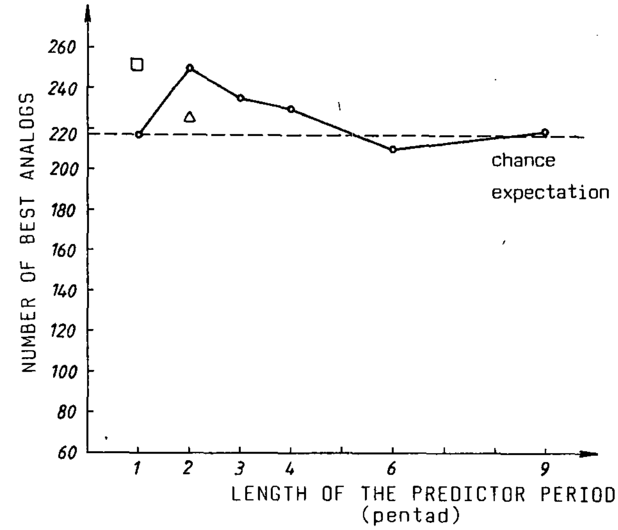


FIG. 3. Number of best analogs (see text) gained by different-length predictor periods during the calibration period (February 1953–December 1979). The square denotes spatial smoothing, circles denote time averaging, the triangle stands for comparison pentad by pentad.

significant, analog selection by Eq. (4) and by Eq. (3) for 2 pentads is superior to other trials (only partly presented in Fig. 3). Note that a further increase in the length of predictor period decreases yearly averaged analog quality, reaching chance expectation at 30 days. Nevertheless, there are some months (from June to November) when a more than two-pentad predictor period proved to be the best.

As can be seen in Table 1, only 3 months did not have optimal predictor periods similar to those found by Harnack et al. (1986) in persistence forecasts. Because the results obtained by Harnack et al. were brought to the author's attention after almost all of the development connected with this work was completed, the yearly standardized procedures were preferred in order to avoid monthly variation in the choice of the length of predictor period (which would be vulnerable to sampling fluctuations). Considering Barnett and Preisendorfer's (1978) and Livezey and Barnston's (1988) suggestions as well, the monthly (or seasonally) adjusted model will be used and tested simultaneously during operational practice.

2) SPATIAL WEIGHTING

For assessment of the role of predictor data from outside our forecast region (sectors 1–5), weighting factors of the other 6–15 sectors were varied. Results for one pentad predictor type with spatial smoothing [Eq. (4)] are shown in Fig. 4. Although there are only slight and insignificant differences, with the exception of the decrease of analog quality for weighting factors greater than 1 in the surrounding sectors, we can conclude that it may be worth considering the data from

TABLE 1. Annual variability of analog quality (number of best analogs) gained by different long time averaged predictor periods (with spatial smoothing for pentad one), for the period of February 1953–December 1979. Values significantly higher (at the 5% level according to test based on the binomial distribution) than the estimate of chance expectation Eq. (18) are set in bold type. Similarity to the results of Harnack et al. (1986) for persistence forecasts is indicated in the right side.

	Pentad						
Month	1	2	3	4	6	9	Type
Jan.	**23**	21	16	17	17	21	similar
Feb.	**31**	22	21	17	15	**29**	no
Mar.	**24**	22	21	**23**	19	15	exact
Apr.	19	**23**	19	20	13	**23**	similar
May	21	**23**	20	16	17	14	similar
June	14	18	14	14	**28**	17	exact
July	16	12	19	19	14	16	no
Aug.	20	21	**27**	17	19	16	no
Sept.	**23**	20	17	**30**	22	17	similar
Oct.	**25**	**25**	**24**	19	14	19	exact
Nov.	18	18	**26**	16	20	17	similar
Dec.	18	**26**	11	22	12	15	similar

outside the forecast region with smaller weighting factors ($w_6, \cdots, w_{15} = 0.4$). This result is consistent both with experiments using the other best predictor type (two pentads with time averaging) and with data of the test period.

3) FURTHER SELECTION

As was indicated in the Introduction and section 3a, the 15 good analogs gained by the nearest neighbor decision rule (NNDR) were subjected to a second selection method based on the analogs' similarity during the forecast month. When we choose the 6 most similar circulation analogs from the 15 good ones gained by NNDR in the case of one pentad predictor with spatial smoothing, letting $w_6, \cdots, w_{15} = 0.4$, we obtain 128 best analogs for the 27-yr calibration period [see section 3c(1)]. When we rely only on the top six ranking analogs gained simply by NNDR, there are only 105 best ones. The difference between these two outcomes (128 and 105) is significant at the 5% level according to the chi-square test. The difference between 128 and the chance expectation of 86 has a significance at the 0.01% level, while the difference in the corresponding values of the 5 yr test period (24 and 16) has a significance at the 1% level.[1] In the case of the other best selection method the results were similar (127 and 24 best analogs in the calibration and test period, respectively). Thus, we have two different ways of choosing analogs (one pentad with spatial smoothing and two pentads with time averaging, in both cases $w_6, \cdots, w_{15} = 0.4$), after which the twice-selected six analogs are ready to be used for weather forecasting purposes.

[1] The chance expectation here was estimated without regard to the specifications detailed in section 3a as numerical experiments showed no significant discrepancies. Accordingly, we estimated chance expectation in the following way: 15 previously chosen best analogs to be hit × 6 cases × 323 target months/26 yr for selection/13 pentad shifts = 86 hit best analogs.

b. Weather forecasting

1) POTENTIAL MONTHLY ANALOG PREDICTABILITY

Before describing the evaluation of the actual weather forecasting method presented in section 3b, and based on the results of 4a, it is worth ascertaining the maximum forecasting skill attainable using our dataset and analog methods. This was achieved by a somewhat different method than that of Barnett and Preisendorfer (1978). They examined the best analogs found directly to their predictand (United States seasonal temperature) fields, while in this work the best circulation analogs were used for this purpose.

First, we executed a new analog selection only for the first pentad of every target month and employing Eq. (1) with $w_1, \cdots, w_5 = 1$, $w_6, \cdots, w_{15} = 0$. The first six analogs were used to determine the relationship between the circulation analogs and their correspond-

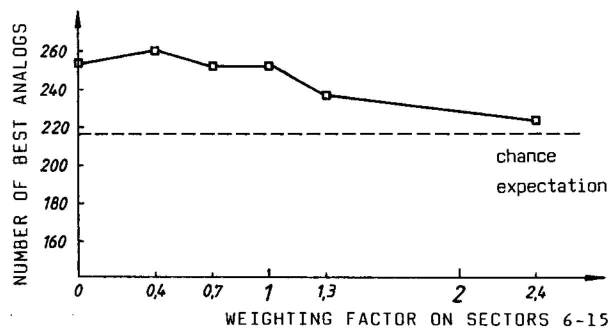


FIG. 4. Number of best analogs gained using one pentad predictor period with spatial smoothing and different weighting factors outside the forecast region (sectors 6–15), during the calibration period (February 1953–December 1979).

TABLE 2. Verification skills (percent correct) of potential predictability during the design period (February 1953–December 1979) using analogs obtained to different long periods.

	Analog selection	
Forecast	The first pentad/decade	The whole month
extreme pentad temperature	49, 2	44, 9
dekad precipitation	41, 4	38, 9
totally	44, 2	41, 8

ing 5-day temperatures. The same hindcasting procedure was also followed, in the case of 10-day precipitation totals, for the first two pentads of every target month in the calibration period [using Eq. (2) altered for two pentads]. Following this, the weather forecasting (or here, rather, hindcasting) procedure (described in section 3b) was executed again using the best circulation analogs found previously for the complete (30-day) target months [see section 3c(1)]. Looking at the hindcasting skills in Table 2 we can estimate the degradation of analog predictability deriving from the extension of the forecasting period from 1 pentad or dekad up to 1 month. Note that in the case of temperature forecasts the near normal forecast category has no skill above chance expectation and therefore is not included in Table 2. [An independent predictability study by the author (Toth 1988) yields a similar result in the near normal range.] The values may of course be strongly related to the dataset type and methods, and also to the length of the archive. [In section 4b(4) we will make an attempt to assess the latter factor.] However, the figures in the second column of Table 2 can be regarded as an upper limit for our weather forecasting scheme or an estimation of analog monthly potential predictability on the basis of our dataset and methods.

2) TEMPERATURE FORECASTS

In this chapter we discuss two good ways of selecting analogs; using a one-pentad predictor with spatial smoothing [see Eq. (4)] and using a two-pentad predictor period with time averaging [Eq. (3) altered for 2 pentads], letting $w_6, \cdots, w_{15} = 0.4$ in both cases. For temperature forecasting, the latter method proved to be more successful so only those results are presented here. Table 3 shows the contingency table of actual versus forecast categories in the calibration and test period (based on 1938 and 396 pentad temperature forecasts, respectively). It is apparent that near normal forecasts show no skill at all. This is to be expected because they have no potential predictability either (see section 3b(1)) and forecasters at the U.S. Climate Analysis Center and in Great Britain obtained similar results (see, e.g., Gilman 1986 and Folland and Varah 1987). It can also be seen from Table 3 that extreme category (above and below normal) forecasts comprise only 46% of all cases, whereas they are observed 67% of the time. This is a result of our endeavor to obtain more reliable extreme forecasts [see section 3(b)] and may be avoided only in a more sophisticated model. In spite of these facts, temperature forecasts are (using Gordon's evaluation scheme) better than random forecasts, with a significance level of 2% and 5% for the calibration and test period, respectively.

From this point, let us consider only the extreme category forecasts. The often used skill score (e.g., Brier and Allen 1951) is 0.044 and 0.076 for the calibration and the test period, respectively. (This score has a value of unity when all forecasts are correct and has a value of zero when the number correct is equal to chance expectation.) We averaged all pentads' skill as no evident decrease could be detected in monthly forecasts with increasing forecast lead time (see Fig. 5). Comparing the model outcome to persistence forecasts (when the temperature category of the pentad preceding the target month is forecast, see Table 4) we can see that the model's skill is superior for the second to sixth pentads (i.e., 6th to 30th days). The difference has a significance level of 5% according to tests based on binomial distribution both in the calibration and test periods. The global skill of 36.3% in the calibration period means that we realize about 30% of 44.9% potential predictability (see the previous section) which is compatible with the results obtained from different terms by Shabbar and Knox (1986).

TABLE 3. Probability of occurrence of pentad temperature forecasts during the design period (February 1953–December 1979) and test period (January 1980–June 1985). Respective number of cases are in parentheses.

	Design period			Test period		
	Cold	Near normal	Warm	Cold	Near normal	Warm
Cold	36, 7	33, 6	27, 4	37, 9	30, 8	28, 0
	(173)	(353)	(115)	(39)	(65)	(23)
Near normal	31, 8	33, 7	36, 8	36, 9	34, 6	32, 9
	(150)	(355)	(154)	(38)	(73)	(27)
Warm	31, 4	32, 4	35, 8	25, 2	34, 6	39, 0
	(148)	(340)	(150)	(26)	(73)	(32)

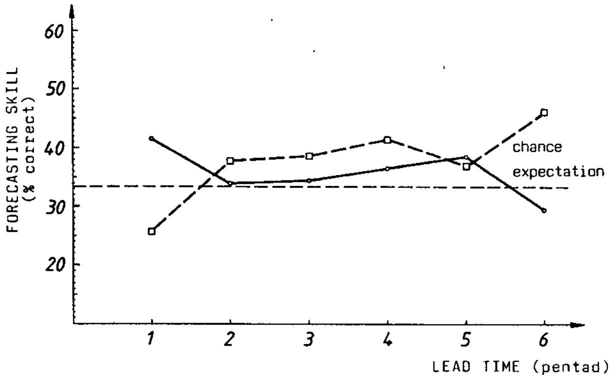


FIG. 5. Forecasting skill of extreme 5-day temperature category predictions with increasing lead time during the calibration period [February 1953–December 1979 (solid line)] and the test period (January 1980–June 1985, dashed line).

TABLE 4. Verification skill (% correct) of extreme pentad temperature forecasts with different lead time during the design period (February 1953–December 1979) and test period (January 1980–June 1985) compared to chance expectation and persistence forecasts from the design period.

	Model forecasts			
Pentad	Design period	Test period	Chance expectation	Persistence forecasts
First pentad	45, 1	30, 3	33, 3	52, 8
Second–sixth pentad	34, 7	40, 1	33, 3	33, 4
Totally	36, 3	38, 4	33, 3	36, 7

Taking the monthly averaged skills in Fig. 6, one could attribute the skill in the test period mainly to January and June results, and therefore possibly to persistence. But the fact is that there were only 4 successful extreme pentad temperature forecasts out of 25 for Januaries and Junes included in the test period when both persistence and analog methods forecast the same category. (We expect slightly more than 4 by chance.) Furthermore, we should keep in mind that in the relatively short test period greater sampling fluctuations can be expected. Consequently, with reference to both lines in Fig. 6 we can find only two notable features: February and May forecasts are no better than chance expectation. The same holds true of all of Bergen and Harnack's (1982) analog monthly temperature forecasting models for May. Thus special care and possibly the use of other methods seem to be needed for May and perhaps for February forecasts during operational forecasting. As in the case of other models cited by Nicholls (1980), our yearly averaged skills fluctuate slightly above chance expectation with only 10 yr out of 32.5 below the level of random forecasts.

3) PRECIPITATION FORECASTS

In the case of forecasting 10-day precipitation totals 1 month in advance, the other predictor type, i.e., one pentad with spatial smoothing [Eq. (4)] letting $w_6, \cdots, w_{15} = 0.4$, turned out to be more efficient than in the case of temperature. Table 5 shows a contingency table similar to Table 3. The model's 969 and 198 predictions are better than chance with a significance level of 2% and 5% in the calibration and test periods, respectively, according to Gordon's (1982) measure. The skill scores (Brier and Allen 1951) are 0.054 and 0.091, respectively. Note that contrary to the results of temperature predictions, forecasts of moderate precipitation are no less accurate than those of extreme precipitation. (This difference between the forecasts of the two weather elements is quite understandable as their governing processes are dissimilar and consequently their distributions are not the same.) The three dekads of the predicted months can be treated together since, similar to 5-day temperature forecasts, no decrease in verification skill is noticeable as lead time increases (see Fig. 7). This is also true for persistence forecasts; therefore, Table 6 consists of one line only. Precipitation forecasts are better than persistence forecasts with a significance level of 2% and 5% in the calibration and test periods, respectively. The global skill of 36.9% in the calibration period is about 60% of the 38.9% potential predictability found in section 4b(1).

As monthly averaged skills have no new peculiarities, in Fig. 8 we show only the yearly averaged skills. Fluctuation is similar to that of temperature forecasts (not shown), although the actual peaks and troughs do not generally occur in the same year. There are only 2 yr, 1970 and 1982, when both types of forecasts have a skill level below chance expectation. It seems encouraging in both cases that forecasting skill does not become less efficient over the independent 5.5 yr test period.

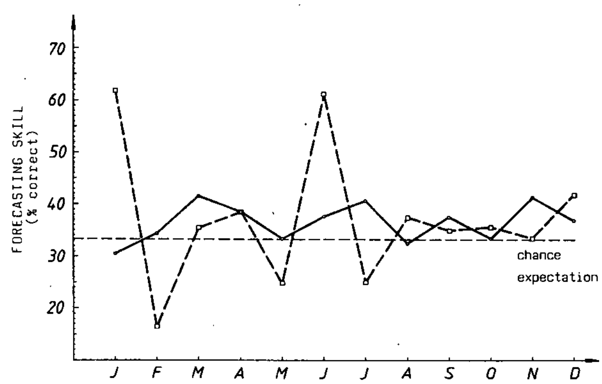


FIG. 6. Monthly averaged forecasting skill of extreme temperature category predictions (for 5-day subperiods within forecast months) during the calibration period (February 1953–December 1979, solid line) and the test period (January 1980–June 1985, dashed line).

TABLE 5. Probability of occurrence (percent) of decade precipitation total forecasts during the design period (February 1953–December 1979) and test period. Respective number of cases are in parentheses.

Precipitation type	Design period			Test period		
	Light	Moderate	Heavy	Light	Moderate	Heavy
Light	35, 5 (94)	31, 8 (156)	31, 0 (68)	44, 9 (22)	37, 5 (36)	35, 8 (19)
Moderate	33, 2 (88)	37, 7 (185)	31, 4 (66)	30, 6 (15)	35, 4 (34)	22, 6 (12)
Heavy	31, 3 (83)	30, 5 (150)	37, 6 (79)	24, 5 (12)	27, 1 (26)	41, 5 (22)

4) IMPROVEMENT OF ANALOG FORECASTS DUE TO EXTENSION OF DATA SERIES

As was indicated earlier we assessed the effect of the brevity of our dataset on weather forecasts, by gradually lengthening the data series for analog selection. By experiments made on the 5.5 yr test period, we determined the potential monthly predictability for the test period (as in section 4a for the calibration period) and the skill of actual forecasting schemes (as in the previous two sections), using, at first, only the last 9 yr (1971–1979) of data for the purpose of analog selection. This database was further supplemented with former data incrementally so that it consisted of 14, 18, 23 and finally all 27 yr of our archive. Provided that no major alteration in the climatic system occurs, an increase in predictability and forecasting skill can be expected due to the obviously better chance of discovering good analogs among the growing number of cases. This increase cannot be expected to monotonically continue. In fact, Livezey and Barnston (1988), adding only 1–3 additional years to their 32-yr case library, found 10%–15% increases and decreases in skill scores due to statistical variations (see their Figs. 8, 10, 11, 12 and 13). As our test period is relatively short and because of the major reduction made on the data series for analog selection, we evaluate both temperature and precipitation forecasts together in the hope of moderating sampling fluctuations.

The varying values of potential predictability and actual forecasting skill are shown in Fig. 9. For interpolation purposes it is sufficient to fit linear curves to the computed values (the correlation coefficients are 0.970 and 0.768, and have a significance level of 1% and 10%, respectively). However, for extrapolation to longer data series it would be appropriate to fit logistic curves, which seem to be more suited to the phenomena investigated (see appendix C). It is evident that the starting level of these curves should be set at chance expectation (33%), but determination of their upper limit is more open to question; therefore, we have determined two boundary levels for it. On the one hand, given that the climate does not change, it is reasonable to assume that the 49% and 44% potential predictability for one pentad/dekad (see Table 1) would increase to some extent with longer data series, and that potential (and, hence, actual with very long data series) predictability for a whole month could reach this higher value, estimated as at least 50%. On the other hand it is also clear from theoretical predictability studies (e.g., Lorenz 1969b, 1969c), that skill will never attain the 100% level. Considering also the obvious deficiencies of the model, the maximum upper limit of 80% was estimated subjectively. This 80% was also compared to the final choice of the upper limit, 65%, that was primarily based on Madden and Shea's (1978) results in order to approximate a symmetrical range of lower- and upper-boundary levels.

Besides the two final logistic extrapolating curves (with the upper limit of 65%) for analog predictability and forecasting skill (the correlation coefficients are 0.974 and 0.792, and have a significance level of 1% and 10%, respectively) the confidence belts at the probability level of 10%, allowing for the variation of the upper limit between 50% and 80% skill (see ap-

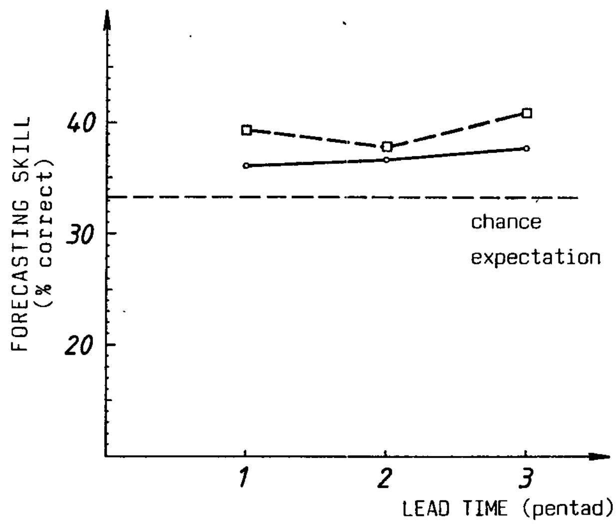


FIG. 7. Forecasting skill of 10-day precipitation-total predictions with increasing lead time during the calibration period (February 1953–December 1979, solid line) and the test period (January 1980–June 1985, dashed line).

TABLE 6. Verification skill (percent correct) of dekad precipitation total forecasts during the design period (February 1953–79) and test period (January 1980–June 1985) compared to chance expectation and persistence forecasts from the design period.

Model forecasts			
Design period	Test period	Chance expectation	Persistence forecasts
36, 9	39, 4	33, 3	35, 7

pendix C) are also presented in Fig. 9. Despite the fact that the real upper confidence limit for actual forecasts may not be higher than that for potential predictability, the computed upper limit is presented in Fig. 9. Naturally the curves are not drawn for the first years. The forecasting scheme designed for 20–30 yr of data is not effective on such a short data series and presumably in this period, logistic curves are not at all adequate.

According to preliminary expectations and the computed correlation coefficients, there is much more uncertainty in the estimation of actual forecasting skills than in potential predictability. As long as the logistic approach is valid and no great climatic change occurs, a theoretical increase of 1%–4% skill in potential predictability per decade (i.e., 10 yr) can be expected at the probability level of 10%. In spite of the greater uncertainty involved, a similar average theoretical increase of 2% skill in actual forecasts per decade can be estimated for future years.

5. Conclusions and discussion

A relatively simple analog methodology has been presented for forecasting of 5-day mean temperature and 10-day precipitation totals for Hungary for the next month in three climatologically equal probable categories. The analog selection is based on large-scale features of the Northern Hemisphere extratropical flow pattern, namely on the location of ridge lines. Numerous circulation analog trials were attempted employing various predictor periods and weighting factors outside the Atlantic–European forecast region and Euclidean distance functions. For measuring analog quality, the best circulation analogs over the forecast region were first singled out for all target months of our calibration (27 yr) and test (5.5 yr) periods. Analogs chosen from the two best predictor types were subjected to further examination in order to select those which have the most similar circulation patterns during the forecast month. Temperature and precipitation forecasts obtained through the six final analogs were statistically compared with chance expectation and persistence. The main conclusions of the study and discussion are:

- There is no need to use more than one or two pentads (i.e., 5–10 days) of atmospheric predictor data in yearly standardized procedures (Fig. 3).
- Gutzler and Shukla's (1984) contention that it is of considerable value to allow slight longitudinal shifts between pressure patterns for short (one pentad) predictor periods in analog selection procedures, proved to be correct.

These results led us to suppose that the initial state of the atmosphere in the context of monthly forecasting may be an averaged map of some days, the duration of which is comparable with that of natural synoptic periods investigated by some authors (see Barry and Perry 1973), and can be compared to different measures of 500 hPa height persistence (Dole and Gordon 1983; Gutzler and Mo 1983; Horel 1985) as well. These quasi-stationary states were approximated in two ways: with the spatial smoothing of one pentad predictor data, (in case of 10-day precipitation total forecasts), and with the time averaging of two pentads' data, (in the case of 5-day temperature forecasts). The difference between the methods of analog selection for the two types of prediction may be connected to higher persistence of temperature forming effects, on the one hand, and less distinct localization in the case of predictable precipitation-producing processes. However, it can be assumed that these states of about 5–10 days do not

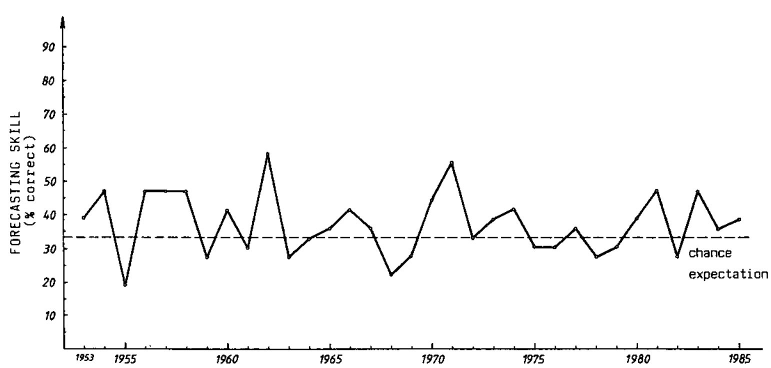


FIG. 8. Annual forecasting skill of monthly precipitation-total predictions for 10-day subperiods (the value for 1985 refers only to data from January–June).

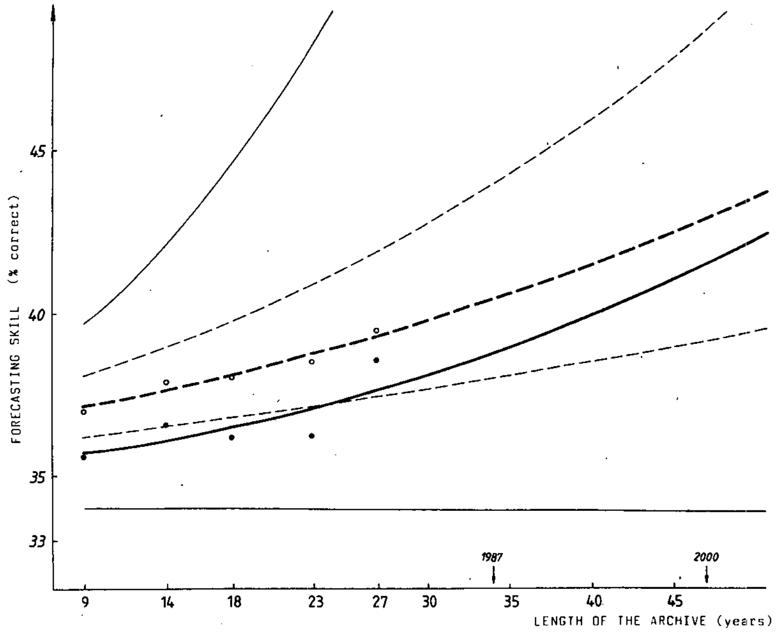


FIG. 9. Monthly potential predictability (circles and dashed line) and actual forecasting skill (dots and solid line) of weather predictions (temperature and precipitation-total categories for 5-day and 10-day subperiods, respectively) during the test period (January 1980–June 1985) using an increasing length of archive to choose analogs from, and the logistic extrapolating curves with upper limit of 65% (thick lines). Confidence belts at the probability level of 10% (allowing the change of upper limit between 50% and 80%) are also shown (thin lines). (Chance expectation is 33.3%).

change simultaneously over the whole hemisphere. On the contrary, they could have a complex space–time structure.

The failure of long (e.g., 1 month) predictor periods to yield good analogs during most of the year can be partly interpreted by the results of Stefanick (1981). In an investigation of daily and monthly variations for temperature and geopotential height at 850 hPa, he concluded that most intermonthly variations are largely statistical residues of averaging daily weather fluctuations which have a time scale of only a few days and are not due to long-period fluctuations. Also, if we require similarity for a longer period in the past, the probability that the resemblance would be maintained in the future may decrease, as long-lived analogs occur rather rarely (Ruosteenoja 1988). However, at least in some regions and during some periods of the year, persistence may also dominate; therefore the annual variability of the optimal length of the predictor period (Table 1) appears to require further investigation.

• There is no great difference between utilizing the whole hemisphere or only the Atlantic–European forecast region in the analog selection process. However a weighting factor of 0.4 on data from outside the forecast region seemed to be most successful (Fig. 4).

As our analog approach relies solely on atmospheric data, external effects are only implicitly regarded through interactions of the atmosphere with the underlying surface. It is obvious that the whole state of the initial atmosphere affects the circulation processes over the forecast region during the predicted month. But the effects of the underlying surface implied in atmospheric data from within the forecast region might have considerably more influence on future atmospheric changes over the region in question than similar effects implied in circulation data from outside the forecast area. The more benefit there is from increased circulation data the less advantage is gained from less defined external information. These two opposing factors give a possible explanation for the very slight variation of analog quality with weighting factors between 0 and 1 outside the forecast region in Fig. 4, and for the results of other analog trials [e.g., Bergen and Harnack's (1982) similar conclusion based on different 700 hPa height predictor domains, see their Fig. 3].

This assumption may be at least partly confirmed by van den Dool et al. (1986) results on the persistence of monthly mean air temperature (MMAT) over the United States. According to them, the persistence of 700 hPa heights accounts to a lesser extent for MMAT anomaly persistence—less than SST and land surface anomalies determined by preceding circulation. These results, in conjunction with the previous point and also with Livezey and Barnston's (1988) findings suggest that in order to develop better methods, we need atmospheric and external predictor data with different time scales and possibly with different domains determined separately.

• It was found that after choosing analogs with the aid of predictor periods and the nearest neighbor decision rule, a further selection among these analogs, by comparing their "future" circulation changes to each other, was of value.

Among others, this idea was outlined by Craddock et al. (1962), and employed subjectively by, e.g., Davis (1978) mainly for weather elements. The underlying philosophy is similar to that of ensemble forecasting (proposed by Leith 1974) which has recently been widely examined and used in dynamical prediction techniques. The benefit of this method can be related to the aforementioned incompleteness of atmospheric data and lack of external data on the one hand and to the shortness of our data series on the other. First, there may be cases with relatively precise degrees of matching which are not good analogs because their less investigated (or uninvestigated) but, nevertheless, important features are dissimilar. Second, in a large number of cases there are no really good analogs. Excluding the deviant cases, there is a chance that the remaining major features represent the most likely future development.

• The presented analog selection method provides further evidence that relatively simple empirical methods based on simple datasets can yield monthly forecasts slightly but significantly better than chance expectation and persistence. The forecasting method realizes 30%–64% of potential predictability during the calibration and test period (36%–39% actual forecasting skill out of 39%–45% skill of potential predictability, cf., Table 2 to Tables 4 and 6). The reduction in skill (2%–8%) is similar to that resulting from the alteration of potential analog prediction from one pentad/dekad to the whole month (2%–5%, Table 2).

The fact that forecasting skills did not decrease in the test period may be due (beyond simple statistical variation) to the checking procedure employed through the optimization of the circulation analog selection [section 3c(1)], which was performed independently of weather forecasting. It is also remarkable that predictor data for the method (large-scale location of ridge lines) can be reliably forecast by numerical models up to about 5 days.

• The investigation of the improvement of analog potential predictability and actual forecasting skills due to the lengthening of data series in time, implies that with 30–40 yr of upper air data available, we are now in a period of increasing forecasting skills. An average increase of 2% skill of both potential predictability and forecasting skills can be expected per decade and hence a forecasting skill about 42%, instead of the present estimated 39%, can be foreseen by the early twenty-first century (Fig. 9).

It follows that even if great improvement occurs in theoretical approaches to long-range forecasting in the forthcoming decades, empirical methods could still presumably compete with them on the basis of extended data series, especially if analog selection techniques can also be improved.

Acknowledgments. This study would not have been possible without the assistance of Gizella Duska of the Computation Center of the Hungarian Meteorological Service, Budapest, and József Boncz, who wrote most of the programs. The computations were carried out using the facilities of Eötvös Loránd University Computation Center, Budapest. The author is especially indebted to Robert E. Livezey and to two anonymous reviewers for their comments and constructive criticism of the draft versions of this paper. Acknowledgment is also due to Tamás Szentimrey of the Central Meteorological Institute, Budapest; to László Hont who assisted the writer with statistical testing; and to Ilona Barabás, Gabriella Karner, and Charlene Mann, respectively, for drafting the figures and typing the manuscript.

APPENDIX A

Central Agglomerative Hierarchical Clustering Method

In the analog selection process, the similarity of the analogs during the forecast period was used as a further criterion. After choosing the 15 top-ranking analogs based on the predictor period (see section 3a), a simple hierarchical clustering process (detailed, e.g., in Anderberg 1973) was performed. In the clustering method, six-pentad (i.e., 30-day) circulation data of the 15 top-ranking analogs were considered in the forecast region of sectors 1–5. Equation (2) served as a measure for comparing the analogs, where [after Eq. (1)]

$$\Delta_{jk} = (\sum_{i=1}^{5} (r_{ij} - r_{ik})^2)^{1/2}.$$

With agglomerative hierarchical clustering methods there are as many clusters as the initial entities in the beginning (this number in our case is 15). The first step of the employed iteration is to compute the similarity matrix of the six-pentad circulation periods (clusters). The most similar pair of clusters are then merged. The circulation data of this new cluster will be the arithmetic mean of those of the two constituent clusters, weighted by the number of their constituent entities (i.e., the center of the clusters). We carried out this process until a cluster consisting of at least six original entities (analogs) emerged. If this cluster contained more than six entities, then that (those) situated most far from the center was (were) neglected.

APPENDIX B

Statistical Tests Used for Evaluating Circulation and Weather Forecasts

1. Testing a hypothesis on the probability of occurrence of a given event

The analogs gained by different selection methods were checked by comparing them to the best analogs chosen previously [see section 3c(1)]. The number of their coincidence (k, or relative frequency of event A), was tested against the estimation of chance expectation (p_0). The success of 5-day temperature and 10-day precipitation total forecasts given in three categories (number of event A), were also tested against chance expectation [p_0, see section 3c(2)].

The null hypothesis is that the probability of occurrence of A ($P(A)$) is equal to p_0,

$$H_0 : P(A) = p_0.$$

If we have n trials with the outcome of k number of event A, then k (the test statistic) has a binomial distribution. After determining the first order error (E), the critical value of the variable (k_{crit}) for

$$P(k|n, p_0) < E,$$

and the zone of acceptance can be computed.

2. Testing a hypothesis on the homogenity of different series of two trials

The number of best analogs and number of successful weather forecasts were tested not only against chance expectations but also against the outcome of other methods [see section 3c(1) and 3c(2)]. The null-hypothesis is that the results of two different methods come from the same probability distribution. If we have two samples with $n_1 = v_s + v_n$ and $n_2 = w_s + w_n$ cases and consider only the successful (v_s and w_s) and non-successful (v_n and w_n) realizations, it can be proved that the variable

$$n_1 n_2 \sum_{i=1}^{2} \frac{1}{v_i + w_i} \left(\frac{v_i}{n_i} - \frac{w_i}{n_2} \right)^2$$

has a chi-square distribution with one degree of freedom, providing $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$. Hence a chi-square test can be executed.

APPENDIX C

Confidence Belt for Logistic Regression Curve

Logistic curves were fitted to the measured 5 values of potential predictability and actual forecasting skill (shown in Fig. 9) through a linearization, after setting the lower (33.3%) and upper (65%) limits of the curve. The extrapolating logistic curve is denoted by

$$Y = Y(t) = \frac{A}{1 + ae^{bt}},$$

where the converted lower limit is 0, t is the independent variable (the length of the archive in years), A the converted upper limit of the curve (potential predictability minus chance expectation) and a and b the two other parameters. In accordance with the phenomenon in question it can be assumed that the residual difference between the logistic estimate $Y(t)$ and the time series $Y_s(t)$ measured at the points of $t_i (i = 1, \cdots, 5)$ decreases with time, hence

$$Y_s(t) = \frac{A}{1 + ae^{bt}e^{E(t)}},$$

where $E(t)$ is a normal white noise process with an expected value of 0. This logistic function can be linearized as the following:

$$Y_s(t) + Y_s(t)ae^{bt}e^{E(t)} = A,$$

$$ae^{bt}e^{E(t)} = \frac{A - Y_s(t)}{Y_s(t)}.$$

When we fix an A, the linear form is

$$\ln a + bt + E(t) = \ln \frac{A - Y_s(t)}{Y_s(t)}.$$

Accordingly,

$$a' + bt + E(t) = Y'_s(t).$$

Let $\hat{b}$ and $\hat{a}'$ be the estimation of parameter b and a', respectively. It can be proven that

$$\frac{(\hat{b} - b) \sum_{i=1}^{5} (t_i - \bar{t}^2)^{1/2}}{\left[\frac{1}{3} \sum_{i=1}^{5} [Y'_s(t_i) - (\hat{a}' + \hat{b}t_i)]^2 \right]^{1/2}} = \frac{(\hat{b} - b)S_1}{S_2}$$

has a Student distribution with three degrees of freedom, where

$$\bar{t} = \frac{1}{5} \sum_{i=1}^{5} t_i,$$

and

$$Y'_s(t_i) = \ln \frac{A - Y_s(t_i)}{Y_s(t_i)}$$

are the linearized values of the five original measurements ($Y_s(t_i)$). Hence,

$$P\left(\hat{b} - x_p \frac{S_2}{S_1} < b < \hat{b} + x_p \frac{S_2}{S_1} \right) = 1 - p,$$

where P is the symbol for probability, p is the confidence level, and x_p is the corresponding value of the Student distribution with three degrees of freedom. As

$$\frac{(\hat{a}' - a')}{S_2\left(\frac{1}{5} + \bar{t}^2 \frac{1}{S_1^2}\right)^{1/2}}$$

also has a Student distribution, we can ascertain that

$$P\left[\hat{a}' - x_p S_2\left(\frac{1}{5} + \bar{t}^2 \frac{1}{S_1^2}\right)^{1/2} < a' < \hat{a}' + x_p S_2\left(\frac{1}{5} + \bar{t}^2 \frac{1}{S_1^2}\right)^{1/2}\right] = 1 - p.$$

The confidence intervals for b and $a = e^{a'}$ were computed at the confidence level of 10%, both for potential predictability and for actual forecasting skill. All intervals were determined for A = 50–33 and 80–33 as preliminarly estimated boundary values for the upper limit of logistic curves. Thus the covering curves of the eight confidence limits presented in Fig. 9 represent uncertainties involved in all three parameters of the logistic regression curves.

REFERENCES

Anderberg, M. R., 1973: *Cluster Analysis for Applications.* Academic Press, 359 pp.

Barnett, T. P., and R. W. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. *J. Atmos. Sci.,* **35,** 1771–1787.

Barry, R. G., and A. H. Perry, 1973: *Synoptic Climatology: Methods and Applications.* Methuen and Co. Ltd., 555 pp.

Bergen, R. E., and R. P. Harnack, 1982: Long-range temperature prediction using a simple analog approach. *Mon. Wea. Rev.,* **110,** 1082–1099.

Brier, G. W., and R. A. Allen, 1951: Verification of weather forecasts. *Compendium of Meteorology,* T. Malone, Ed., Amer. Meteor. Soc., 841–849.

Craddock, J. M., H. Flohn and J. Namias, 1962: The present status of long-range forecasting in the world. WMO-No. 126. TP. 56, Geneva, 23 pp.

Davis, N. E., 1978: Prospects for winter 1978/79. *Weather,* **33,** 480.

Dole, R. M., and N. D. Gordon, 1983: Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation: geographical distribution and regional persistence characteristics. *Mon. Wea, Rev.,* **111,** 1567–1586.

Folland, C. K., and L. D. Varah, 1989: Assessment of probability forecasts of temperature and rainfall for the United Kingdom. *Proc. Second WMO workshop on the dianosis, and prediction of monthly and seasonal variations over the globe,* Toulouse, World Meteor. Org., (in press).

Gilman, D. L., 1986: Expressing uncertainty in long range forecasts. *Namias Symposium,* J. O. Roads, Ed., Ref. Series 86-17, Scripps Institute of Oceanography, 174–187.

Gordon, N. D., 1982: Evaluating the skill of categorical forecasts. *Mon. Wea. Rev.,* **110,** 657–661.

Gutzler, D. S., and K. Mo, 1983: Autocorrelation of Northern Hemisphere geopotential heights. *Mon. Wea. Rev.,* **111,** 155–164.

——, and J. Shukla, 1984: Analogs in the wintertime 500 mb height field. *J. Atmos. Sci.,* **41,** 177–189.

Harnack, R., Ch. Kluepfel and R. Livezey, 1986: Prediction of monthly 700 mb heights using simulated medium-range numerical forecasts. *Mon. Wea. Rev.,* **114,** 1466–1480.

Horel, J. D., 1985: Persistence of the 500 mb height field during Northern Hemisphere winter. *Mon. Wea. Rev.,* **113,** 2030–2042.

Knox, J. L., and J. E. Hay, 1984: Blocking signatures in the Northern Hemisphere: rationale and identification. *Atmos.-Ocean,* **22,** 36–47.

Leith, C. E., 1974: Theoretical skill of Monte Carlo Forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Livezey, R., and A. G. Barnston, 1988: An operational multifield antilog prediction system for United States seasonal temperatures. Part 1: System design and winter experiments. *J. Geophys. Res.,* **93,** 10 953–10 974.

Lorenz, E. N., 1969a: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26,** 636–646.

——, 1969b: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc.,* **50,** 345–349.

——, 1969c: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.

Madden, R. A., and D. J. Shea, 1978: Estimates of the natural variability of time-averaged temperatures over the United States. *Mon. Wea. Rev.,* **106,** 1695–1703.

Namias, J., 1951: General aspects of extended-range forecasting. *Compendium of Meteorology,* T. Malone, Ed., Amer. Meteor. Soc., 802–813.

——, 1968: Long-range weather forecasting—history, current status and outlook. *Bull. Amer. Meteor. Soc.,* **49,** 438–470.

Nicholls, N., 1980: Long-range weather forecasting: value, status and prospects. *Rev. Geophys. Space Phys.,* **18,** 771–788.

Radinovic, D., 1975: An analog method for weather forecasting using the 500–1000 mb relative topography. *Mon. Wea. Rev.,* **103,** 639–649.

Ratcliff, R. A. S., 1974: The use of 500 mb anomalies in long-range forecasting. *Quart. J. Roy. Meteor. Soc.,* **100,** 234–244.

Roads, J. O., and T. P. Barnett, 1984: Forecasts of the 500 mb height using a dynamically oriented statistical model. *Mon. Wea. Rev.,* **112,** 1354–1369.

Ruosteenoja, K., 1988: Factors affecting the occurrence and lifetime of 500 mb height analogs: a study based on a large amount of data. *Mon. Wea. Rev.,* **116,** 368–376.

Sawyer, J. S., 1970: Observational characteristics of atmospheric fluctuations with a time scale of a month. *Quart. J. Roy. Meteor. Soc.,* **96,** 610–625.

Schuurmans, C. J. E., 1973: A 4-year experiment in long-range weather forecasting, using circulation analogues. *Meteor. Rundsch.,* **26,** 2–4.

Shabbar, A., and J. L. Knox, 1986: Monthly prediction by the analogue method. *Proc. First WMO Workshop on the Diagnosis and Prediction of Monthly and Seasonal Atmospheric Variations over the Globe,* College Park, Maryland, World Meteor. Org. WMO-TD-No. 87, 672–681.

Stefanick, M., 1981: Space and time scales of atmospheric variability. *J. Atmos. Sci.,* **38,** 988–1002.

Titkos, E., 1981: A method of long-range forecasting based on simulation of large-scale atmospheric processes. *Proc. Symposium on Current Problems of Weather-prediction,* Vienna, Central Institute of Meteorology and Geodynamics, 173–176.

Toth, Z., 1988: Dependency of long-range temperature predictability in Hungary on circulation weather type and forecast class. *Proc. 13th Annual Climate Diagnostics Workshop,* Cambridge, NOAA (in press), 497–500.

Treidl, R. A., E. C. Birch and P. Sajecki, 1981: Blocking action in the Northern Hemisphere: a climatological study. *Atmos.-Ocean,* **19,** 1–23.

van den Dool, H., W. H. Klein, and J. E. Walsh, 1986: The geographical distribution and seasonality of persistence in monthly mean air temperatures over the United States. *Mon. Wea. Rev.,* **114,** 546–560.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.