

A Skill Analysis of Soviet Seasonal Weather Forecasts¹

ROBERT E. LIVEZEY² AND SHERWIN W. JAMISON

University of Missouri, Columbia 65201

(Manuscript received 5 May 1977, in final form 1 September 1977)

ABSTRACT

Operational long-range weather prediction in the Soviet Union is reviewed. Methods for producing forecasts at the 5- and 10-day, monthly and seasonal range are described in terms of the synoptic, statistical and hydrodynamic tools available to Soviet forecasters. Skill scores for these forecasts published by the Soviets are summarized and examined.

Skill scores for Soviet operational forecasts of mean seasonal (about two months) temperature anomaly and precipitation category are computed separately for regions, seasons and years and compared to persistence skill scores. In addition, forecast-observation sets for the sign of the mean temperature anomaly are tested for "no skill." The forecasts for the sign of the mean temperature anomaly are found to be best by region for the Arctic and by season for March through April, but generally do not outperform persistence, exhibit demonstrable skill, or show an improvement trend over the verification period. Forecasts of the mean precipitation category are shown to be consistently better than persistence, but to have quite modest skill scores.

1. Introduction

The achievement of demonstrable and significant skill in producing long-range weather forecasts is an important accomplishment, both from a scientific and societal viewpoint. Consequently, schemes for making such forecasts should be critically examined for skill and, when demonstrated, their methodologies studied for potential exploitation.

Since 1968, forecasts of the mean temperature anomaly and mean precipitation category (below normal, near normal, above normal) for the forthcoming natural synoptic season (defined in Section 2b) for the region of the Soviet Union west of 90°E longitude have appeared in the Soviet journal *Meteorologiya i Gidrologiya* (Meteorology and Hydrology), until their recent cessation in mid-1975. Their availability presented a unique opportunity to examine the skill of an operational product issued by the world's largest long-range weather prediction complex, that of the Main Administration of the Hydrometeorological Service of the Soviet Union. Our resultant analysis is the primary topic of this article and is detailed in Section 4. However, a balanced perception of the results requires some understanding of the conceptual base and operational content of the forecasts, and for this reason a brief overview of Soviet philosophy and methods for long-range prediction is

presented in Section 2. In addition, for contrast and discussion, verification statistics compiled by Soviet authors for long-range products are summarized in Section 3.

2. Operational methods

The central long-range forecast activity in the Soviet Union that provides the background forecasts and methodological guidance to other institutes appears to be mostly concentrated at the Hydrometeorological Scientific Research Center (GMTs) in Moscow. For example, the official seasonal forecast is produced at GMTs and the selection of the analog used as the background for monthly forecasts is principally accomplished and supervised by that institute. Moreover, the methods employed at regional centers for 5- and 10-day forecasts, although they differ in some respects, all seem to be basically those developed by the late A. L. Kats at GMTs. Thus, because the forecasts discussed in Sections 3 and 4 were all made at the Moscow site, the focus here is principally (with one exception) on the methods used at GMTs. It should also be pointed out that, except for the longest range forecasts (seasonal and greater), all of the approaches outlined below represent substantial departures from the traditional synoptic schools of Soviet long-range forecasting.

a. Five- and ten-day synoptic-statistical-hydrodynamic method (Kats, 1973)

From 1965 to 1972, the operational method of forecasting 5- and 10-day mean temperature anomalies em-

¹ Contribution from the Missouri Agricultural Experiment Station, Journal Series No. 7835.

² Present affiliation: Development Division, National Meteorological Center, Washington, D. C. 20233.

ployed by the group headed by A. L. Kats at GMTs was the synoptic-statistical method. The method consisted of setting up a simple linear regression model for each forecast point and using as predictors a hemispheric zonal flow index, an Atlantic-Eurasian sector meridional index, and the daily mean temperature anomaly at the forecast point. The synoptic element of the method entered in a number of ways, including the use of the circulation indices. More importantly, separate sets of regression coefficients were calculated for different forms of "macroprocesses" as well as for different seasons, and the resulting equations were only applied on the first day of "natural synoptic periods."

The idea of a macroprocess, a kinematic type or class of large-scale circulation, and that of the natural synoptic period, a 5–7 day period in which on the average the macroprocesses retain uniformity, are at the heart of classical Soviet synoptic methods formulated by B. P. Multanovskiy.

One of the principal drawbacks of the method was the crucial necessity of applying it on the first day of a natural synoptic period. Hence, the forecasts were issued quite irregularly and were quite sensitive to the correct detection of the start of a new natural synoptic period. To overcome some of these difficulties and to exploit the advances in both short-range (≤ 72 h) numerical weather prediction and statistics, Kats developed his "synoptic-dynamical-statistical" procedure, which was later adopted for operational use.

The new method was still a linear regression model applied at a forecast point but otherwise bore little resemblance to the old method. First, the predictors consisted of 72 h dynamically forecast fields obtained from numerical models (for example, 500 mb heights, sea level pressure and 1000–500 mb thickness), and of various fields at the initial time, including the field of the predictand. The forecast fields characterized the average flow for the first five days and the initial-day fields helped to establish trends.

These fields contained much more information than the circulation indices. However, to emphasize only the synoptic or large-scale circulation characteristics, they were represented by expansion into series of empirical orthogonal functions, with higher order terms characterizing smaller scale processes deleted. After this was done for each field the various optimal sets of expansion coefficients were combined into one set, that was in turn expanded into a series of empirical orthogonal functions. The principal components of this final expansion were then selected as the compact set of predictors.

To obtain the regression coefficients, Kats used actual data rather than archived numerical-model output data in his dependent data set (often called the "perfect prog" method in the United States). For example, for a 10-day forecast he would build his regression model from fields of the predictand (the 10-day mean tem-

perature anomaly) associated with fields of the predictors from the zeroth and third day of the respective 10-day period. He argued that the regression coefficients would characterize the "ceiling" of efficiency and that "losses" resulting from systematic differences between the numerical forecasts and actual data would automatically decrease as the numerical models improved. Certainly this approach eliminates the necessity to regenerate the regression coefficients every time the numerical model is changed. Furthermore, because Soviet meteorologists have had to rely on the U. S. National Meteorological Center's six-layer primitive-equation model for their 72 h forecasts, it would have been a much more inconvenient, complicated and expensive task for them to build up a dependent data set from a foreign-based model.

Kats' synoptic-hydrodynamic-statistical method for 5- and 10-day weather forecasting is quite sophisticated and advanced and resembles in some respects approaches used by the U. S. National Weather Service in their forecasts for this time range.

b. Monthly method (Zverev, 1972b)

The methodology used to forecast the mean temperature anomaly or the precipitation class for a month was worked out in 1966 and 1967 and implemented at GMTs in 1968. The final forecast depends on the selection of an analog case extending over as much as four or five months that will encompass the following month and, after adjustments, allow the statistical specification of mean temperature anomaly and precipitation class fields.

The selection of the final analog is accomplished by the use of many tools, some quite conventional and some unique to the forecasting philosophy of Multanovskiy and his successor S. T. Pagava. Among the more conventional sources of information are hemispheric mean monthly charts of sea level pressure, surface temperature, 500 mb heights and 100 mb heights (to take into account surface conditions, tropospheric and stratospheric circulation, respectively). Also considered are the variation of the index of zonal flow, the state of the ice in the Barents Sea and the mean monthly sea surface temperature reported by two weather ships in the North Atlantic (to take into account the extra-atmospheric carriers of memory), and the positions and relative strength of semi-permanent features such as the Icelandic and Aleutian lows ("centers of action").

In addition, however, Soviet forecasters take a detailed look at the "history of macroprocesses" over two "natural synoptic seasons," that is, they note the frequency of occurrence of large-scale circulation types and the order of transformation from one type to another over a 3–5 month period. A natural synoptic season consists of 11–13 similar natural synoptic periods, and Soviet scientists recognize the existence of

six such seasons whose duration averages two months.

In their examination of the progression of the natural synoptic periods they look for "forecasting instructions," macroprocesses that they assess to be rhythmic with 45-, 75-, 90-, 120- or 150-day periods. The Multanovskiy school has been classifying, organizing and studying synoptic chart sequences since 1917.

Once a best analog case is selected the map sequence is often adjusted slightly to fit current major periodicities observed at key stations and sometimes to take into account Ye. N. Blinova's hydrodynamic forecast.³ In the spring, the map sequence is also adjusted toward M. T. Yudin's physical-statistical forecast (see Section 2c). Before the selection of the final analog, five regional centers also independently select an analog sequence case, and all choices are transmitted back and forth along with arguments until a consensus is reached.

c. Monthly physical-statistical method (Aristov and Bagrov, 1972)

In 1964 a group at the Main Geophysical Observatory in Leningrad under M. I. Yudin was charged with the task of developing a method for forecasting with a long lead time monthly mean precipitation amounts and temperature for May to July in the southern half of the European Territory of the Soviet Union and the Virgin Lands. Their efforts led to the development of the physical-statistical method, which, at the present time, is part of the monthly forecast guidance employed at GMTs.

The basic vehicle for the forecasts is a linear regression model relating coefficients of expansion into empirical orthogonal functions of the predictand field to coefficients of expansion of the predictors. The criteria for selection of predictors and the subsequent reduction of the size of the predictor set are what distinguish the physical-statistical method. The predictor set must satisfactorily distinguish the initial state of the physical system [atmosphere-ocean-(active) soil layer] from other states, include precursors of future states, contain extra- and intra-atmospheric carriers of memory but not meteorological noise, take into account the prehistory of the atmospheric circulation, and be relatively compact.

To describe the initial state of the atmosphere, complete fields of monthly means and their standard deviations are used, the former to isolate persistent processes and the latter to take into account the heterogeneity of processes and, implicitly, precursors of future events. The prehistory of the atmosphere is considered by including information for periods up to 28 months in advance of the initial month.

³ The operational procedure used is to run the latest version of Blinova's spectral model for a simulation time of 70 days, average the last 30 days of output, and provide this to the monthly forecast group 40 days in advance of the forecast month (Blinova, 1972).

Intra-atmospheric carriers of memory (i.e., large-scale processes) are simultaneously isolated from meteorological noise without the loss of significant information and reduced to a compact set of predictors by expansion of the fields into empirical orthogonal functions and retention of only the principal components. Certain spotty fields, such as precipitation and cloudiness, are spatially smoothed before expansion. Other fields used to describe the atmosphere are sea level pressure, 500 mb heights and surface temperature. In addition, a zonal circulation index, a geomagnetic index, and the number of days with one or another Vangengeym circulation type were used as predictors. Some of the extra-atmospheric carriers of memory included as predictors are the extent of ice in northern seas, sea surface temperatures, heat content of the oceans and the boundaries of snow cover.

Once the construction of the original set of predictors is complete, the reduction process can begin. First, predictors with either weak or unsteady correlations with the predictands are eliminated. Next, cases are noted where predictors have strong relationships to predictands in a certain range of values, but not outside that range. Those that can be identified as falling into this category and whose values are likely to occur outside of the allowable range are then dropped from the set. Finally, the remaining predictors in the set are examined by statistical methods to substantiate their degree of independence and after elimination of redundant information the remaining set is used to construct the forecast model. There is no close analog to the physical-statistical method in operational use by the U. S. National Weather Service.

d. Seasonal and longer period methods (Pokrovskaya, 1969)

Forecasts of the mean temperature anomaly and the mean precipitation category (above, near or below normal) for the forthcoming natural synoptic season for the area west of 90°E longitude have been produced by S. T. Pagava, N. A. Aristov, N. M. Zakharova, and others at GMTs since 1962. Publication of these forecasts in *Meteorologiya i Gidrologiya* began in 1968 and terminated with the forecast for the third natural synoptic season in 1975.

The method employs the same ideas of the classical Multanovskiy school that were encountered earlier in the discussions of the synoptic-statistical methods of 5- and 10-day and monthly forecasting in Sections 2a and 2b, respectively. The year is divided into six natural synoptic seasons which consist of 11-13 similar natural synoptic periods. The natural synoptic periods are defined as 5-7 day periods over which macroprocesses are uniform. Forecasting instructions are those macroprocesses that are expected to repeat after a certain number of days (rhythms).

An additional concept of importance is that of the

existence of "disturbance" natural synoptic periods, in which the macroprocesses making up a natural synoptic period differentiate it from all the others in the natural synoptic season. These disturbance periods are thought to be precursors and representative of the next season or the one thereafter.

Through the study of the frequency and forms of macroprocesses and the history of transformations from one natural synoptic period to another, the identification of forecasting instructions and disturbance macroprocesses, and the consideration of circulation indices, stratospheric circulation and sea surface temperatures, an analog year is selected and the forecast is made.

Another longer range forecast approach is the Vangengeym-Girs macrocirculation method developed at the Arctic and Antarctic Scientific-Research Institute in Leningrad. The method was originally intended for forecasting general meteorological and ice conditions in the Arctic, but is now also used to forecast hemispheric monthly mean temperature and pressure anomalies. Like most Soviet long-range forecasting methods it is principally synoptic in nature and depends on the selection of an analog.

The hemispheric circulation is broken down into nine basic types derived from classification of the Atlantic-Eurasian sector circulation into one of three types (W,C,E), and that of the Pacific-American sector into one of three types (Z,M₁,M₂). A circulation pattern is classified by first establishing whether it is predominantly zonal or meridional and, if the latter, then qualifying it further according to the positions of the major ridge and trough lines. A 70-year data set summarizing the number of days with each circulation type in each month was analyzed for 15-month periods extending from August of one year through October of the next. The frequency and sequence of types over 15 months turned out to have seven characteristic patterns.

The key to making a macrocirculation forecast is to fit actual data for the five-month period from August through December into one of the seven groups and use that group's characteristics to forecast the frequency and sequence of circulation types for the next ten months.

Actually, the process is more complicated in that groups are subdivided into subgroups and these and the climatic background are included in the refinement of the forecast. The author of the procedure, A. A. Girs, has identified five epochs during the last 90 years in which certain combinations of circulation types dominated others, and he has in turn broken these down into stages varying in duration from two to six years. Like many meteorologists in Leningrad (note Yudin's use of a geomagnetic index as a predictor in Subsection 2c), Girs also takes solar activity into account.

3. Reported skill

Before the presentation of the independent verification of seasonal forecasts in Section 4, it might prove useful to summarize the Soviets' assessment of their forecast skill. This is done below for each of the time ranges whose methodology was outlined in Section 2.

a. Skill measures

In this article frequent use is made of the skill score S , defined by

$$S = \frac{R - E_c}{T - E_c}, \quad (1)$$

where R is the number of correct forecasts, T the total number of forecasts and E_c the number expected to be correct based on climatology. For a multi-category set of forecasts (e.g., forecasts of above, near or below normal precipitation) where the climatological frequency p_i of observations that fall in each class i is known, the climatological expectancy can be obtained from

$$E_c = \sum_{i=1}^N p_i f_i, \quad (2)$$

where f_i is the number of forecasts in class i and N the total number of classes. If one subtracts from the total T that number E_c of forecasts that could have been made correctly from prior knowledge of the variable's behavior, then S measures the fraction of cases correctly forecast of those left over. Its maximum value is one, climatological skill is represented by zero, and it can take on negative values (Panofsky and Brier, 1958).

Because at many time scales the atmosphere is sometimes remarkably persistent (conditions fairly uniform or steady over time), a "persistence" forecast frequently has a positive S . For this reason forecasts are considered here to have real skill only when S is both positive and greater than the persistence skill S' . Both skills are normally reported in the discussion below.

Soviet meteorologists measure skill by the use of the parameter

$$\rho = \frac{n_+ - n_-}{T},$$

where n_+ and n_- are the number of correct and incorrect forecasts, respectively (Bagrov and Kats, 1973). It is a trivial exercise to show that for two-category forecasts (like the sign of the mean temperature anomaly)

$$\rho = S.$$

Because most of the Soviet skills discussed are for two-category forecasts S is employed exclusively. When the percentage P of correct forecasts is encountered it can be converted to S by use of

$$S = 2[(P/100) - 0.5].$$

TABLE 1. Skill of synoptic-statistical forecasts of the sign of the mean temperature anomaly for 1965-71.

Forecast	<i>S</i>	<i>S'</i>	<i>P</i> (%)
5-day	0.44	0.29	72
10-day	0.38	0.24	69

One final note is necessary. Many Soviet authors consider forecasts of anomalies (for instance temperature) as successful if *S* for the sign of the anomaly is greater than zero or if the mean absolute error of the forecast anomalies is less than the mean absolute value of the observed anomalies. However, both are *not* considered necessary for "success." Examples of this view of "success" are given in the following subsections. Here, only forecasts with $S > S' > 0$ are considered skillful.

b. Five- and ten-day forecast skill

Kats (1973) has compiled verification data on his synoptic-statistical method of 5- and 10-day mean temperature anomaly forecasting. The results of this compilation are shown in Table 1 for 339 forecasts for on the average 37 stations in the European Territory of the Soviet Union, Western Siberia and Kazakstan. These forecasts were made for the 10-day periods in each month beginning on days 1, 6, 11, 16, 21 and 26, *only* when the day was preceded by the start of a natural synoptic period. Thus out of 504 possible forecasts about one-third could not be issued.

For the period from February 1965 through August 1967, *S* for the 10-day forecasts was 0.43 (Kats *et al.*, 1968); hence there was a decrease in skill for the remainder of the period to 1971.

To illustrate the point made in the preceding subsection about alternative definitions of "success," we note that Zverev (1972a) reported 83 and 73% success for the 5- and 10-day forecasts, respectively, whose skill is summarized in Table 1. Thus, we conclude that he considered 11% (about 1400 station forecasts) successful when the 5-day mean temperature is forecast to be colder (warmer) than normal but turns out to be warmer (colder) than normal.

The skill of 135 forecasts of the sign of the 5-day mean temperature anomaly made between June 1972 and April 1973 by the synoptic-hydrodynamic-statistical scheme was given by Bagrov and Kats (1973) as

0.57 ($P = 79\%$). This statistic seems to indicate that the new scheme is superior to the old one.

c. Monthly forecast skill

N. I. Zverev, who has been responsible for the monthly forecasting operation, has summarized (1972b) the method's performance:

"In 1968, the forecasts of the mean monthly air temperature anomaly for the territory of the Soviet Union were correct for seven months, in 1969 for nine months, and in 1970 for ten months. . . . The accuracy of the forecasts for the European Territory and Western Siberia in 1968 and 1969 was 68%, and rose to 75% in 1970."

He has said elsewhere (Zverev, 1972a) that the correctness of monthly forecasts varies between 65 and 80% and that annually 8 or 9 out of 12 forecasts are correct. By inference from other material he has published, we believe his figures are representative of "success" (as defined in Section 2a), rather than of skill. For instance, Zverev (1972a) presents a table detailing the forecast and observed monthly mean temperature anomalies for Moscow from 1967 to 1972. These forecasts, according to Zverev, were accurate 75% of the time for the period 1967-71. Our examination of his table reveals, however, that only 65% ($S = 0.30$) of the forecasts correctly specified the sign of the mean temperature anomaly and that *S* was positive on the average in slightly less than 8 out of 12 months during the period.

d. Seasonal forecast skill

For the first three years of operation (1962-65) seasonal forecasts were reported to have an *S* for the sign of the mean temperature anomaly of 0.30 (Drogaytsev, 1966). Zverev (1972a) published the information in Table 2, which is compatible to other references, but did not specify whether one or several stations were included in the verification. This set of verification statistics will provide the principal point of contrast to the results presented in the next section.

4. Analysis of seasonal weather forecasting skill

The analysis of seasonal forecast skill consisted of a straightforward exploitation of Eqs. (1) and (2) and a simple statistical test. However, data reduction and processing in preparation for application of these

TABLE 2. Soviet skill of seasonal forecasts of sign of mean temperature anomaly for 1963-71.

	Synoptic season						Mean
	Winter	Spring	First half of summer	Second half of summer	Fall	Pre-winter	
<i>S</i>	0.32	0.49	0.33	0.42	0.33	0.22	0.35
<i>P</i> (%)	66	75	67	71	67	61	68

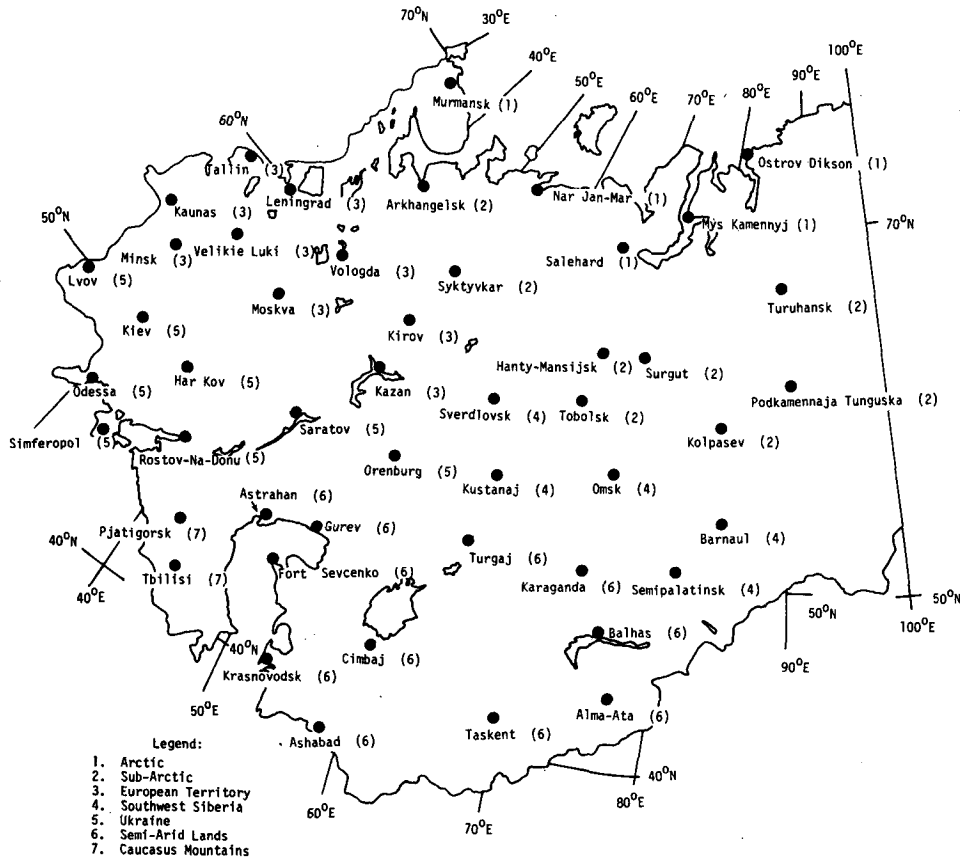


FIG. 1. Location and regional classification of cities used for verification of seasonal weather forecasts.

formulas were somewhat involved and require detailed description.

a. Procedure

Forty-eight Soviet cities somewhat uniformly distributed over the forecast region (Fig. 1) were chosen for the verification procedure because of accessibility of data for them. These data consisted of monthly means of temperature and precipitation obtained from the World Meteorological Organization data set tape provided by the National Climatic Center, and were augmented by use of the publication *Monthly Weather*

Reports of the USSR (translated title) when there was a significant amount missing for a particular month. Climatological means were not computed, but were obtained from *Monthly Climatic Data for the World* published by the National Climatic Center. The cities were further separated into seven regional groups as shown in Fig. 1.

The seasonal mean temperature anomaly forecasts for each city were converted into forecasts of the sign of the mean temperature anomaly, while the seasonal mean precipitation category forecasts were used directly. Forty-five sets of seasonal forecast charts were obtained for analysis from *Meteorologiya i Gidrologiya*.

TABLE 3. Skill by natural synoptic season summed over regions and years for 1968-74.

	Natural synoptic season						All
	Winter (Jan-Feb)	Spring (Mar-Apr)	First half of summer (May-June)	Second half of summer (Jul-Aug)	Fall (Sep-Oct)	Pre-winter (Nov-Dec)	
S	0.03*	0.33	-0.01*	-0.18	-0.03*	0.02*	0.03*
S'	0.04	0.49	0.14	0.23	0.28	0.06	0.20

* Did not pass no-skill test.

TABLE 4. Skill by region summed over natural synoptic seasons and years for 1968-74.

	Region							All
	Arctic	Sub-Arctic	European Territory	Southwest Siberia	Ukraine	Semi-arid Lands	Caucasus Mountains	
<i>S</i>	0.12*	0.03*	0.00*	0.12*	-0.04*	0.02*	0.01*	0.03*
<i>S'</i>	0.37	0.18	0.22	0.28	0.15	0.14	0.14	0.20

* Did not pass no-skill test.

In order to determine skill forecasts must be compared to corresponding verification data. This presented some difficulties in that the data set and forecasts described above were incompatible. Soviet forecasters make a seasonal prediction for the forthcoming natural synoptic season, and they recognize the existence of six such seasons per year. These average two months in length, but rarely, if ever, does a season's beginning and end both coincide with either the beginning or end of a calendar month. Furthermore, the bounds of a natural synoptic season can change from one year to the next and frequently do. Thus, to properly compare a forecast to actual conditions, the daily mean for every day in the natural synoptic season should be averaged to obtain the seasonal mean.

This was not practical here, so instead actual and climatic means for natural synoptic seasons were obtained by using the expression

$$\bar{\phi}^s = \frac{\sum_{i=1}^M d_i \bar{\phi}_i^m}{\sum_{i=1}^M d_i}$$

Here ϕ denotes a meteorological variable, the overbars with *m* and *s* the monthly and natural synoptic season means, respectively, d_i the number of days of the month *i* encompassed by the natural synoptic season, and *M* the total number of months or parts of months over which the natural synoptic season extends.

This procedure introduces errors for months that are not entirely included in a natural synoptic season, but the errors should be small when $d < 10$ or $d > 20$. Also, because of the importance placed on developing synoptic conditions at forecast time in the analog selection process, a bias toward the mean conditions for the entire month in which the natural synoptic season begins should more often reward than penalize the

forecaster. Rewards and penalties should be relatively equal in frequency for a bias toward the mean conditions over the entire month in which the season ends.

To ensure that the persistence forecasts did not have an unfair advantage they were generated from mean conditions for the first full month preceding the start of each natural synoptic season.

One final data problem stemmed from the fact that Soviet forecasters do not categorize precipitation in terms of climatological frequencies. Instead, with mean and actual precipitation denoted by \bar{r} and *r*, below, near and above normal precipitation are defined as $r < 0.8\bar{r}$, $0.8\bar{r} \leq r \leq 1.2\bar{r}$ and $1.2\bar{r} < r$, respectively. Thus, the climatological frequencies p_i for precipitation classes are not known *a priori* and must be estimated to obtain *S* values. Eighteen stations uniformly distributed over the forecast area had at least 30 years of data recorded on the World Meteorological Organization tape, so frequency distributions were compiled for each of these locations. The percent of cases falling into each of the precipitation categories was relatively uniform from station to station, and turned out to have mean values of 42, 26 and 32, respectively. These means, because of the spatial stability of the statistics, were used in the determination of E_c .

Separate two-by-two temperature anomaly sign and three-by-three precipitation category contingency tables for forecasts and observations were constructed for every region for every forecast period. Excluding 1975 forecasts, tables were separately summed over regions, seasons, years, regions and seasons, regions and years, seasons and years, and all regions, seasons and years together. Skills *S* and *S'* were then computed for all resulting tables and all tables were tested for no skill. These detailed procedures made it possible to identify seasons, regions or years in which forecasts consistently

TABLE 5. Skill by natural synoptic season.

	Natural synoptic season						All
	Winter (Jan-Feb)	Spring (Mar-Apr)	First half of summer (May-June)	Second half of summer (Jul-Aug)	Fall (Sep-Oct)	Pre-winter (Nov-Dec)	
<i>S</i> (Arctic, 1968-74)	-0.06	0.39	0.06	0.20	0.15	0.0	0.12
<i>S</i> (Table 2, 1963-71)	0.32	0.49	0.33	0.42	0.33	0.22	0.35

TABLE 6. Skill by year summed over regions and natural synoptic seasons.

	Year							All
	1968	1969	1970	1971	1972	1973	1974	
<i>S</i>	0.10	0.06*	0.03*	0.03*	-0.11*	-0.08	0.15*	0.03*
<i>S'</i>	0.34	0.18	0.05	0.38	0.10	0.28	0.13	0.20

* Did not pass no-skill test.

exhibited appreciable success or failure, to isolate combinations of seasons and regions that had extraordinary *S*'s, and to document the chronology or progression of skill over the 7-year period.

The test for no skill mentioned above consists of the determination of whether the hypothesis that a contingency table could have been generated with no skill can be rejected at the 95% confidence level. The test was useful for temperature anomaly sign tables, but not for precipitation category tables. To perform the test, a no-skill table must be constructed from the same data making up the actual table.

Let the subscripts *i* and *I* denote table rows, *j* and *J* table columns, and *n* and *n'* the number of cases for the actual and no-skill tables, respectively. The no-skill table can be constructed through use of

$$n'_{ij} = (\sum_i n_{iJ}) (\sum_j n_{Ij}) / \sum_i \sum_j n_{ij}$$

Chi-squared is then computed from

$$\chi^2 = \sum_i \sum_j [(n_{ij} - n'_{ij})^2 / n'_{ij}]$$

Because a two-by-two contingency table has one degree of freedom, the hypothesis that the table could have been generated with no skill can be rejected at the 95% confidence level if $\chi^2 > 3.84$ (Panofsky and Brier, 1958).

In order to check the representativeness of the forecast and observation data set and to validate the computational procedures, five-category monthly temperature anomaly forecasts and three-category monthly precipitation anomaly forecasts for the 1968-73 period made by the U. S. National Weather Service were obtained and verified. With the data set described above along with data from 54 U. S. stations, both two- and

five-category temperature anomaly skill scores and three-category precipitation anomaly skills were computed and compared to independently produced estimates. The latter include published skill (Namias, 1953, 1968; U. S. Committee for the Global Atmospheric Research Program, 1975) and unpublished estimates generated internally by the staff of the Long Range Prediction Group of the National Weather Service. The various skill score sets were entirely consistent.

This additional verification along with intensive quality control of man and machine operations and hand checks of selected calculations was intended to build confidence in the entire analytic procedure. The exercise was especially important in view of the results presented in the next subsection.

b. Skills for forecasts of the sign of the mean temperature anomaly

In Tables 3 and 4, the results of the skill analyses of seasonal forecasts of the mean temperature anomalies for individual natural synoptic seasons and regions are displayed.

First note that not one *S* in either table exceeds *S'* and that only two contingency tables passed the no-skill test (one of these has *S* = -0.18 and the other has the largest associated *S'*). Only for the spring is *S* substantial, and it is still 0.16 smaller than *S'*. For the second half of the summer the scores are particularly low. The cold part of the year (November-April) does have all positive skills.

A direct comparison of Table 3 to Table 2 (Zverev's statistics) is appropriate and reveals no apparent similarities. Note that there is an order-of-magnitude difference in all *S*'s except for that of spring and that no discernible correspondence exists in the relative magnitudes of *S* between seasons.

In Table 4 only two scores, those for the Arctic and Southwest Siberia, are substantially different from zero, and, once again, they are associated with the highest persistence scores in the table. The latter is principally the result of a very high score for March and April of 0.64. For the former the season-by-season skill is shown in Table 5 along with the top half of Table 2. If seasonal skills are ranked in descending order separately for each line of the table, their order is very

TABLE 7. Skill by natural synoptic season summed over regions for 1972 and 1974.

	Natural synoptic season						All
	Winter (Jan-Feb)	Spring (Mar-Apr)	First half of summer (May-June)	Second half of summer (Jul-Aug)	Fall (Sep-Oct)	Pre-winter (Nov-Dec)	
<i>S</i> (1972)	-0.62	-0.40	0.21	-0.72	0.29	0.38	-0.11
<i>S</i> (1974)	0.59	0.65	-0.28	-0.43	0.02	0.38	0.15

TABLE 8. Skill by natural synoptic season summed over regions and years for 1968–74.

	Natural synoptic season						All
	Winter (Jan–Feb)	Spring (Mar–Apr)	First half of summer (May–June)	Second half of summer (Jul–Aug)	Fall (Sep–Oct)	Pre-winter (Nov–Dec)	
S	0.06	0.08	0.08	0.04	0.08	0.06	0.07
S'	0.12	0.07	0.05	–0.07	0.01	–0.08	0.02

TABLE 9. Skill by region summed over natural synoptic seasons and years for 1968–1974.

	Region							All
	Arctic	Sub-Arctic	European Territory	Southwest Siberia	Ukraine	Semi-arid Lands	Caucasus Mountains	
S	0.09	0.13	0.14	0.03	0.05	0.01	0.03	0.07
S'	0.04	0.06	0.04	0.00	–0.02	0.00	–0.03	0.02

similar between lines—the four highest ranked seasons correspond exactly. This is true for only the Arctic skills, and it is significant that this region has the highest and most uniform (by season) skill. Further, the average skill for the Arctic for 1968–71 was 0.18, also higher than for any other region. All of these facts then suggest that the scores reported by Zverev, although somewhat high, may be more characteristic of a high-latitude forecast set than a regionally comprehensive set.

Table 6 is a presentation of the year-by-year skill. It may appear that the 1974 score represents a reversal of the general decline of seasonal forecast skill, so it should be pointed out that while an average skill of 0.32 was obtained for the first half of 1974, that obtained for the first half of 1975 was only –0.01. No trend can be confidently inferred from this short time sequence.

Note from Table 6 that two cases pass the no-skill test, but once again one is associated with a very high S' (1968) and the other has a negative skill (1973). Two other years in Table 6 deserve additional comment. First, 1974 has the only (S, S') pair presented in this section where S exceeds S' . Second, 1972, the year of the disastrous Soviet crop failures, was an equally disastrous year for seasonal forecasts as evaluated in this article.

TABLE 10. Skill by year summed over natural synoptic seasons and regions.

	Year							All
	1968	1969	1970	1971	1972	1973	1974	
S'	0.12	–0.01	0.02	0.07	0.14	0.06	0.08	0.07
S	–0.06	0.02	0.13	0.01	0.06	–0.03	–0.03	0.02

From Table 7 the main contributions to the negative skill in 1972 are apparent. Despite excellent scores for the last third of the year, the characteristically safe forecast period (Table 3) from January through April had extremely low scores, and the normally difficult July–August period had a forecast chart that was almost a virtual image of the actual temperature conditions. The success in 1974 was principally a result of consistently excellent scores for the cold half of the year.

Finally, the figures shown in the last column of Tables 3, 4 or 6 indicate that overall skill, as defined here, of Soviet seasonal forecasts of the sign of the mean temperature anomaly is insignificant and that the forecast set is indistinguishable from a set generated with no skill.

c. Skills for forecasts of precipitation category

Presented in Tables 8–10 are skills for seasonal forecasts of the mean precipitation category.

Note that for only three cases in all three tables was $S < S'$. Thus, these forecasts can be considered skillful. This is more a result of the degree of non-persistence of precipitation category rather than large skill scores, as evidenced by the very modest skills throughout.

In Table 8 the uniformity of skills by season is the most salient feature, while in Table 9 higher skills are noted for high-latitude areas and the European Territory. No trend of improvement is apparent in Table 10, but it is interesting that the year with the lowest skill for the sign of the mean temperature anomaly for the period 1972 (Table 6) is the year with the highest skill for the mean precipitation category. Outstanding forecast sets not shown that deserve mention are those for the European Territory in the spring ($S=0.32$) and

for the Ukraine in 1972 ($S=0.37$). Overall, seasonal forecasts of the mean precipitation category show a very small but real skill.

Acknowledgments. Dr. Donald L. Gilman, Chief of the Long Range Prediction Group, National Weather Service, contributed substantially to the content and coherence of this article, and the authors are grateful for his assistance. In addition, we would like to thank Dr. Sharon LeDuc for lending her statistical expertise, and Dr. John B. Hovermale for reviewing the manuscript. The report version and preliminary journal versions were typed by Mrs. Betty Allen, the figure was drafted by Mrs. Sally McConnel, and the final manuscript was typed by Mrs. Mary Daigle.

REFERENCES

- Aristov, N. A., and N. A. Bagrov, 1972: *Handbook on Monthly Weather Forecasts* (Chaps. 3-5). Foreign Tech. Div. Rep. No. FTD-MT-24-363-73 [NTIS AD 765 607/LL.]
- Bagrov, N. A., and A. L. Kats, 1973: Estimating the five- and ten-day mean temperature forecasts. *Meteor. Hydrol.*, No. 12, 10-21. Joint Publ. Res. Serv. Rep. No. 61632. [NTIS JPRS-61632/LL.]
- Blinova, Ye. N., 1972: State of the art and prospects for the development of hydrodynamic methods of long-range weather forecasting. *Meteor. Hydrol.*, No. 10, 24-34. Joint Publ. Res. Serv. Rep. No. 57652. [NTIS JPRS-57652/LL.]
- Drogaytsev, D. A., 1966: Book review: *Principles of the Synoptic Method of Seasonal Weather Forecasts* by S. T. Pagava, N. A. Aristov, L. I. Blyumina and Z. L. Turketti. *Meteor. Hydrol.*, No. 12, 66-69. Joint Publ. Res. Service Rep. No. 40218 [NTIS JPRS-40218/LL.]
- Kats, A. L., 1973: Three-ten day synoptic-hydrodynamical-statistical weather forecasting scheme. *Meteor. Hydrol.*, No. 6, 16-29. Joint Publ. Res. Ser. Rep. No. 59962. [NTIS JPRS-59962/LL.]
- , O. N. Khazova and N. M. Chapygina, 1968: Results of routine application of the ten-day air temperature forecasting methods. *Meteor. Hydrol.*, No. 2, 50-57. Joint Publ. Res. Ser. Rep. No. 45488. [NTIS JPRS-45488/LL.]
- Namias, J., 1953: Thirty-day forecasting—a review of a ten-year experiment. *Meteor. Monogr.*, No. 6, Amer. Meteor. Soc., 83 pp.
- , 1968: Long-range weather forecasting—history, current status and outlook. *Bull. Amer. Meteor. Soc.*, 49, 438-470.
- Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press, 224 pp.
- Pokrovskaya, T. V., 1969: *Synoptico-Climatological and Heliogeophysical Long-Term Weather Forecasting*. Foreign Tech. Div. Rep. No. FTD-MT-24-139-70. [NTIS AD 719 760/LL.]
- U. S. Committee, for GARP, 1975: *Understanding Climatic Change—A Program for Action*. U. S. Nat. Acad. Sci., 239 pp.
- Zverev, N. I., 1972a: Development of synoptic and statistical long-range weather forecasts. *Meteor. Hydrol.*, No. 10, 35-41. Joint Publ. Res. Ser. Rep. No. 57652. [NTIS JPRS-57652/LL.]
- , 1972b: Prospects for the development of the synoptic method of monthly weather forecasting. *Meteor. Hydrol.*, No. 1, 27-30. Joint Publ. Res. Ser. Rep. No. 55351. [NTIS JPRS-55351/LL.]