HEMISPHERICAL PREDICTION BY STATISTICAL TECHNIQUES

By Robert M. White, Royce C. Derby, Duane S. Cooley and Florence A. Seaver

Geophysics Research Directorate, Air Force Cambridge Research Center

(Original manuscript received 7 May 1957)

ABSTRACT

A statistical procedure suitable for forecasting the hemispherical midtropospheric contour-height distribution is developed. Least-squares empirical filters are designed to yield the 24-hour 500-mb contour heights as functions of the preceding two-day sequence of contour heights at the same level. The results indicate that the statistical procedure may provide useful forecasts for twenty-four hours over regions of the hemisphere with adequate data coverage. The utility of the system for extended-range forecasting is also discussed.

1. Introduction

All meteorological forecasting techniques imply the application of filters to sets of observed data. In synoptic-subjective forecasting, the individual may be considered as a human filter who rejects or accepts certain pieces of meteorological information which he uses in accordance with his experience to prepare a forecast. In dynamical forecasting, a theoretical filter known as the Green's function, is applied to a set of dynamically specified variables to yield a tendency of some meteorological parameter. By an iterative process, this theoretical filter is successively applied to yield a forecast for a finite time period. In statistical forecasting, empirical filters are constructed on the basis of the historical record of specific events.

Whether the filter design is the result of theoretical dynamical reasoning or based on statistical empirical investigations, meteorological forecasting can be considered as a problem in optimum filter design. A good case can be made for the view that weather prediction problems are not fundamentally different from those encountered in other fields such as communications, where optimum filter design is of great importance (Wiener, 1954). In the report that follows, we shall discuss the design of certain statistical prediction filters suitable for hemispherical forecasting of the midtropospheric contour-height distribution and which can also be used to obtain extended forecasts by an iterative process.

2. Physical considerations

Filters operate on signals. In communications, the signal may be a current, a voltage or a radar return. In meteorology, we have an infinite variety of possible signals on which a filter might conceivably operate. The time series of any meteorological parameter at any point in the atmosphere can be considered as a signal. The key problem in successful and useful meteorological filter design for prediction purposes is in the selection of the signals on which the filters are to operate. In this report, we will use the terms “statistical prediction filters” and “statistical operators” interchangeably.

In the present investigation, we will consider statistical filters of the type

$$\hat{D}_i = \sum_{j=1}^{n} G_{ij} F_j, \quad i = 1 \cdots n. \quad (1)$$

In this equation, the filter comprises the set of weighting coefficients indicated by $G_{ij}$ where there exists a different filter for every point $i$. The signals on which the filter operates are expressed by $F_j$ which represent $j$ functions of the initial and past meteorological conditions. For convenience, we consider the predictand as a deviation from its mean. Thus we define $\hat{D}_i$ as the predicted deviation of the 500-mb contour height $z$.

The set of $n$ equations (1) can be conveniently written in matrix notation as

$$\hat{D} = GF,$$  (2)

where $D$ is the vector of the predictands, $G$ is the filter matrix in which each row corresponds to a filter for a point $i$, and $F$ is the vector of the predictors $F_i$.

The usefulness of the resulting filter matrix $G$ depends importantly upon the manner in which $F$ is selected, for this function determines the nature of the physical ingredients which are incorporated in the statistical procedure. In a loose sense, $F$ defines a statistical analogue of a dynamical system as pointed out by White and Palfson (1955). Because of the large amount of computations involved, we desire a function $\hat{F}$ which is simple and economical to process but which still retains substantial physical information.

In this study, the predictands $D$ are the 24-hour 500-mb contour-height deviations, and the $F_i$ are chosen to be the 500-mb contour heights at initial time, $t = 0$ and the preceding observation time, $t = -24$ hours at a grid of points. This selection of $F_i$ should result in the inclusion of many of the complex
large-scale nonlinear atmospheric physical processes in a rather simple manner.

The use of the two-day sequence of contour heights is statistically equivalent to using the initial contour-height distribution and the field of the 24-hour height changes to predict the 24-hour contour-height distribution. Because the height change at a point in the atmosphere represents the integrated effect of all the physical processes in the atmosphere over a finite time period, the two-day sequence of the contour heights should contain an integrated measure of the physical processes occurring in the most recent past (see Appendix).

The statistical procedure which is used in this study does not permit us to specify which physical processes have been included. Only the performance of the filters can tell us whether they successfully incorporate sufficient physical information. Many alternative selections of the function \( F \) might have been made. We could have made explicit use of the various nonlinear terms representing various physical processes as specified by the dynamical equations. Such a procedure was used by the third author of this paper in another study, Cooley (1957).

3. Mathematical procedures

The filters, \( G_{ij} \), of equation (1) are obtained by a least-squares procedure. We seek to minimize the sums of the squares of the forecast errors given by

\[
S_t = \sum \left[ D_t - \hat{D}_t \right]^2 = \sum \left[ D_t - \sum G_{ij} F_j \right]^2,
\]

where \( \sum \) indicates a summation in time for all of the forecasts. It can be shown (Anderson, R. L., and T. A. Bancroft, 1952) that the set of coefficients corresponding to \( G_{ij} \) can be obtained by the solution of the matrix equation

\[
G_t R = A_t,
\]

where \( G_t \) is the row vector of the filter coefficients and \( A_t \) is the row vector of the covariances between \( F_j \) and \( D_t \). In equation (4), the matrix \( R \) is defined as the symmetrical covariance matrix

\[
R = \begin{pmatrix}
    r_{11} & r_{12} & \cdots & r_{1m} \\
    r_{21} & r_{22} & \cdots & r_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    r_{m1} & r_{m2} & \cdots & r_{mm}
\end{pmatrix}
\]

whose elements \( r_{jk} \) are the covariances of the independent variables \( F_j \) given by

\[
r_{jk} = \bar{F}_j \bar{F}_k - \bar{F}_j \bar{F}_k, \quad j = 1 \cdots m, \quad k = 1 \cdots m.
\]

The complete filter matrix \( G \) and the complete predictor-predictand covariance matrix \( A \) for all points \( i \) are readily given by composition.

\[
GR = A.
\]

The solution to the matrix equation (7a) for the filter matrix \( G \) may be obtained in principle by the inversion of the matrix \( R \).

\[
G = AR^{-1}.
\]

In practice, many difficulties are encountered. The order of the matrix \( R \) is equal to the number \( m \) of independent variables \( F_j \). While high-speed electronic computers are very efficient in the inversion of matrices, even their capacities are limited and the order of the matrix \( R \) must be restricted. In the present study, the hemispherical contour-height distribution at 500 mb is defined at a grid of 288 points between 25 and 90N. Since a two-day sequence of the contour heights at these grid points is used, there is a total of 576 independent variables to consider if the straightforward mathematical approach indicated by (7) is used.

To restrict the order of the matrix, the hemisphere is considered as a series of five overlapping sectors. Each predictor sector extends over 100 degrees of longitude, and contains 80 points arranged in a diamond grid as shown in fig. 1. Filters operating on the two-day sequence of contour heights within each predictor sector were designed to yield forecasts for points in the easternmost regions of these sectors as indicated in table 1 and fig. 1. The overlapping, as shown in fig. 1, is introduced to insure that sufficient upstream boundary information is available to yield reasonable forecasts in the western predictand region.

Formidable mathematical difficulties still remain. The two-day sequence of contour heights within each sector yields a total of 160 independent variables. Aside from the resulting high order of the matrix \( R \), the independent variables at these grid points are highly correlated both in space and in time tending to make the matrix singular. This difficulty may be overcome by replacing the original grid-point variables by a reduced set of transformed variables, which retain almost all of the pertinent information. Such a reduction in the number of independent variables is a statistical necessity if filters obtained in

<table>
<thead>
<tr>
<th>Sector number</th>
<th>Predictor sector</th>
<th>Number of grid points</th>
<th>Predictand sector</th>
<th>Number of grid points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60W–155W</td>
<td>80</td>
<td>60W–135W</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>10E–85W</td>
<td>80</td>
<td>10E–55W</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>80E–15W</td>
<td>80</td>
<td>80E–15E</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>150E–65E</td>
<td>80</td>
<td>150E–85E</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>140W–125E</td>
<td>80</td>
<td>140W–155E</td>
<td>56</td>
</tr>
</tbody>
</table>
the sample are to approximate the population filters closely. Lorenz (1956) has shown that the sample and population reduction of predictand variance are approximately related by the expression

\[ S_0 \sim S' - \frac{M}{N - 1} R_0, \quad R_0 + S_0 = 1, \]  

where \( S_0 \) is the population per cent reduction of variance, \( S' \) is the per cent reduction of variance obtained in a given sample, \( M \) is the total number of independent variables, \( N \) is the sample size and \( R_0 \) is the unexplained per cent of the population variance. It is clear that if the ratio of the number of predictors to the sample size is large, then the sample variance per cent reduction will exceed significantly the population per cent reduction of variance, and the application of the filters to a new sample will yield erratic results.

The prediction operators were developed from a sample which consisted of the daily January and February 500-mb grid-point data prepared by the Navy project AROWA for the years 1946-1952. A total sample of 413 cases was available.

There are many possible linear transformations of the 500-mb contour-height distribution which might have been suitable. We chose to represent the contour-height distribution by means of Tchebyscheff polynomials. These polynomials have been used extensively in meteorological studies, notably by G. P. Wadsworth and J. G. Bryan (1948) and R. G. Miller and T. F. Malone (1954). Such a transformation may be represented in matrix notation as

\[ TZ = U, \]  

where each row of the matrix \( T \) consists of the values of the orthonormal Tchebyscheff functions of some degree, \( Z \) is the vector of 500 mb contour heights arranged in appropriate order, and \( U \) is the vector of the transformed variables. Here, since the grid points cover an area, the \( T \) matrix actually represents a set of two-dimensional orthogonal polynomials. Polynomials of the fifth degree in the east-west direction and fourth degree in the north-south direction were used. Including the cross-product terms, there are 30 possible polynomials of which only 23 were used to specify the 500-mb pressure-height distribution. The 23 polynomial coefficients were chosen to yield the maximum specification while those which were highly related to other coefficients of the set were eliminated. Those coefficients selected are shown above the solid line dividing the matrix in table 2. Here the row identification stands for the degree of the polynomial in the north-south direction and the column identification for the degree in the east-west direction. It was found that on the average 97 per cent of the variance of the 500-mb contour-height distribution could be specified by these 23 polynomial coefficients.

<table>
<thead>
<tr>
<th>Sector</th>
<th>( T_{10} )</th>
<th>( T_{11} )</th>
<th>( T_{12} )</th>
<th>( T_{13} )</th>
<th>( T_{14} )</th>
<th>( T_{15} )</th>
<th>( T_{16} )</th>
<th>( T_{17} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>25°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>30°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>35°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>40°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>45°N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Array of Tchebyscheff coefficients used to represent the 500-mb contour-height distribution for each sector. Row and column identification indicate degree of polynomial in north-south and east-west direction, respectively.
The number of independent variables in the statistical prediction scheme is thus reduced from 160 to 46 by expressing the two-day sequence of contour-height distribution by a Tchebyscheff polynomial representation. The vector $F$ of equation (2) may now be written as

$$ F = \begin{pmatrix} U_0 \\ U_{-24} \end{pmatrix}, $$

(10)

where $U_0$ represents the 23 coefficients at time $t = 0$ and $U_{-24}$ represents the 23 coefficients at time $t = -24$ hours. The prediction equation thus becomes

$$ \hat{D} = G \begin{pmatrix} U_0 \\ U_{-24} \end{pmatrix} $$

(11)

which upon expansion for a given point $i$ is equivalent to

$$ \hat{D_i} = G_{11} U_1 + G_{12} U_2 + \cdots + G_{46} U_{46}. $$

(12)

All of the mathematical computations on which the results of this paper are based were performed on the International Business Machines model 701 EDPM located at Suitland, Maryland.

4. Results

The utility of the forecast operators is indicated by the per cent of the 24-hour 500-mb contour-height variance which can be explained by the statistical system. This quantity is called the per cent reduction of the variance and is denoted by $R^2$ since it is equal to the square of the multiple correlation coefficient between the predictors and the predictand. The geographical distribution of $R^2$ for the North American predictand sector extending between 60W longitude and 125W longitude is shown in fig. 2.

A consistently high $R^2$ is obtained over the region

---

**Fig. 2.** Per cent of the variance of the 24-hour 500-mb contour height over North America explained by statistical operators based on 100-degree sector.
averaging about 74 per cent, which suggests that the prediction filters may have considerable utility over North America. A certain "shrinkage" in $R^2$ can be expected when these filters are applied to another sample. According to approximate formulae given by Lorenz (1957), a four to seven per cent shrinkage can be expected.

The maximum reductions of variance, about 80 per cent, are found in the southeastern United States. It is in this region that the accuracy of dynamical forecasts is greatest, according to Gates et al (1956). If this comparison is valid, it suggests that the statistical procedure incorporates much the same dynamics as the mathematical model on which the forecasts made by Gates are based.

Other features of the distribution of $R^2$ indicate further similarity with the results of dynamical forecasts. For example, there is a decrease in the field of $R^2$ in the lee of the Rocky Mountains. The lee effect is very pronounced in most dynamical forecasts and it had been hoped that the statistical forecasts might perform better in these lee regions.

Several experiments on the effects of various predictor sector sizes were carried out for the North American region. The predictor sector size was reduced from 100 to 70 degrees and $R^2$'s were obtained for the North American region. The results are shown in fig. 3. The $R^2$'s in the southeastern United States are slightly increased. Although such increases are
probably without statistical significance, they may be attributed to the fact that the field of contour heights was more accurately specified for the smaller sector with the fixed set of polynomial coefficients which was used. In the western regions, however, the $R^2$'s are substantially lower than in the first case. This decrease is attributed to the lack of upstream information along the western border in the case of the smaller predictor sector size.

It is not possible to display the filters for each of the 288 grid points. The general features of the filters are similar for all points not located close to boundaries. To illustrate these general characteristics, the 70-degree sector filter corresponding to grid-point 40N 70W, near Nantucket, Massachusetts, is shown in fig. 4. The distribution of the regression coefficients comprising the filter for the initial time is shown in fig. 4a, while a similar distribution of these coefficients for time $t = -24$ hours is shown in fig. 4b.

The numerical value of the 24-hour 500-mb contour-height forecast at Nantucket is obtained in practice by computing the product of this filter function and the two-day sequence of the contour-height distribution. These filters can be thought of as codifying and systematizing a set of synoptic forecasting rules of which they are a graphical representation. The general oscillatory character and scale of fluctuations of the filters is not unlike that of a 500-mb contour-height...
chart. When the two-day sequence of 500-mb contour-height patterns is highly correlated with the patterns of this filter, then the contour height at Nantucket can be expected to undergo large changes. The filters possess their maxima approximately ten degrees to the west of the forecast points, and their minima are located about 40 degrees to the west of the forecast points. The amplitude of the filter fluctuations decreases with distance from the maximum point. The oscillatory character of the filter is evident in the time domain also. The absolute magnitudes of the filter coefficients are about three times larger at initial time than they are at the time 24 hours earlier and the coefficients are generally of opposite sign.

Both the space and time variations of the filter are not unexpected. The filter “wave length” is about 60 degrees of longitude, which is approximately the same scale as transient disturbances at 500 mb (White and Cooley, 1956). The displacement of the region of maximum coefficients to the west of the forecast point reflects the general eastward progression of weather systems, and the decrease in the magnitudes of the coefficients in time reflects the importance of the latest information in forecasting.

Fig. 4a, b. Graphical illustrations of forecasting filter for predicting 24-hour 500-mb contour height at latitude 40N and longitude 70W near Nantucket, Massachusetts. This filter operates on the field of contour heights at time $t = 0$ (Fig. 4a) and at time $t = -24$ hrs (Fig. 4b) covering 70-degree sector.
The distribution of $R^2$ for the entire hemisphere is shown in fig. 5. It is apparent that the quality of the results is highly influenced by the availability of data. Over most of the western hemisphere, where the data are known to be relatively plentiful, the variance reductions generally exceed 60 per cent. However, over regions such as China and the western Pacific where the data coverage is poor, the per cent reductions can reach very low levels not greatly exceeding those which might be expected by chance.

The per cent reductions reach their maxima of about 80 per cent in the Aleutians, the southeastern United States, and over the coast of Western Europe. Except for the southeastern United States, the areas of maximum per cent reduction are closely associated with areas of very large pressure-height variance. The magnitude of the forecast errors implied by the statistical operators is shown in fig. 6. This is a chart of the square root of the unexplained variance, and is equivalent to the standard deviation of the errors of the within-sample forecasts. Assuming that these residuals are normally distributed and that there is no "shrinkage" on an independent sample, 67 per cent of all forecasts can be expected to have errors smaller than the values indicated. There is a marked and systematic increase in the values of the residual errors from low to high latitudes. This is opposite, in general, to the latitudinal variation of the per cent reduction and is accounted for by the strong increase in the pressure-height variance with latitude.

5. Extended range forecasts

It is the intention of the authors to make application of these statistical operators for extended-range forecasting. The extended range forecasts will be ob-
Fig. 6. Standard error in feet of the 24-hour forecast 500-mb contour height for the Northern Hemisphere.

tained by an iterative procedure. This procedure would involve the following steps:

a. Apply the operators to obtain a 24-hour 500-mb forecast for each of the five sectors.
b. Apply operators to the predicted 24-hour contour-height field and obtain another 24-hour 500-mb forecast for each of the five sectors. This is the 48-hour predicted contour-height field.
c. Iterate successively as indicated in a and b by 24-hour time steps to obtain a predicted contour-height field at any time in the future.

The implication of the results shown in figs. 5 and 6 is that useful forecasts for periods of two or three days may be possible for regions of the eastern United States, the Atlantic and Western Europe. It is the authors’ feeling that a 60 per cent reduction of variance is the limit below which the forecasting utility for even one day is questionable. Beyond this time period, the areas in which, the statistical operators yield poor results will soon contaminate the predictions over the remainder of the hemisphere.

As of the writing of this paper, no test of the practical utility of these operators had been completed on independent data. A series of five-day forecasts of the hemispherical 500-mb contour-height distribution is presently being programmed for the International Business Machines model 704 EDPM. The results of this feasibility test will be reported subsequently.

APPENDIX

Let the height tendency at a point in the atmosphere be specified by some mathematical model of the
atmosphere whose equation is

$$L \left( \frac{\partial z}{\partial t} \right) = F(x, y, z), \quad (1)$$

where \(L\) is some time independent linear differential operator, \(\partial z/\partial t\) is the height tendency determined by (1) and \(F(x, y, z)\) some function of the initial conditions. Because the mathematical model is an approximation to reality, the computed height tendency will deviate from the true height tendency \(\partial z/\partial t\). Thus there will exist a residual

$$E(x, y, z) = L \left( \frac{\partial z}{\partial t} \right) - F(x, y, z), \quad (2)$$

where \(E(x, y, z)\) now is some function of the meteorological conditions which express all those physical processes neglected in equation (1) and may also include observational and computational errors. Equation (2) may alternatively be written as

$$\frac{\partial}{\partial t} L(z) = F(x, y, z) + E(x, y, z). \quad (3)$$

In finite difference form, the left-hand side of (3) represents the tendency of some specific linear combination of the true contour heights. Equation (3) states that this tendency is a measure of the physical processes as represented by \(F(x, y, z)\) and \(E(x, y, z)\).

Because instantaneous height tendencies are not available to us from observations, it is necessary to consider the change in \(L(z)\) over a finite time period. In this particular study, the time interval considered is 24 hours. We now approximate equation (3) by

$$[L(z)]_{t=0} - [L(z)]_{t=-24} = [F(x, y, z) + E(x, y, z)] \Delta t, \quad (4)$$

where \(F(x, y, z)\) and \(E(x, y, z)\) are now considered to be taken as the average values of \(F\) and \(E\) during the time interval.

However, rather than use specific linear combinations of \(z\) at time \(t = 0\) and time \(t = -24\) hours as indicated in equation (4), we can seek the linear combination of \(z\) at time \(t = 0\) and time \(t = -24\) hours which will yield the best forecast. If the specific linear combination, \(L(z)\), is actually the best, then our mathematical procedure will indicate this. If there is a better one, which implies that other physical processes such as \(E(x, y, z)\) are important, then it will specify such linear combinations.

REFERENCES