Theses and Dissertations   1. Thesis and Dissertation Collection, all items

1963

A statistical experiment in dynamical prediction of the 24-hour sea-level pressure change

Reilly, Donald J.; Ward, Compton E.

Monterey, California: U.S. Naval Postgraduate School

A STATISTICAL EXPERIMENT IN DYNAMICAL PREDICTION OF THE 24-HOUR SEA-LEVEL PRESSURE CHANGE

DONALD J. REILLY AND COMPTON E. WARD

A STATISTICAL EXPERIMENT IN DYNAMICAL

PREDICTION OF THE 24-HOUR SEA-LEVEL PRESSURE CHANGE

* * * *

Donald J. Reilly

and

Compton E. Ward

A STATISTICAL EXPERIMENT IN DYNAMICAL

PREDICTION OF THE 24-HOUR SEA-LEVEL PRESSURE CHANGE

by

Donald J. Reilly

Captain, United States Marine Corps

and

Compton E. Ward

Lieutenant, United States Navy

A STATISTICAL EXPERIMENT IN DYNAMICAL

PREDICTION OF THE 24-HOUR SEA-LEVEL PRESSURE CHANGE

by

Donald J. Reilly

and

Compton E. Ward

This work is accepted as fulfilling

the thesis requirements for the degree of

MASTER OF SCIENCE

IN

METEOROLOGY

from the

United States Naval Postgraduate School

ABSTRACT

Two statistical prediction models for the 24-hour surface pressure change are developed. One model employs the terms in a dynamic model as the independent variables in a linear regression equation. The other model combines these variables with parameters capable of reflecting the long-wave, long-term influences in a multivariate discriminate analysis. The regression equations were developed from data taken from the month of November 1962 at 50N latitude. A discussion of the results of both methods is presented along with a critique of the procedures used in obtaining the data.

TABLE OF CONTENTS

TABLE OF SYMBOLS AND ABBREVIATIONS

A   Coefficient of independent variable

A_o Constant

CDC   Control Data Corporation

F   A statistic relating the variances of chi-square distributed variables

L   Wavelength

N   Wave number

PRV   Per-cent reduction in variance

R   Multiple linear correlation coefficient

RMSE   Root-mean-square error

S_y Standard error of estimate

Y   Dependent variable

Y' Predicted value of dependent variable

a   Radius of the earth

d   Grid distance

f   Coriolis parameter

f_m Mean value of the coriolis parameter in mid-latitudes

g   Acceleration of gravity

h   1000-500 mb thickness

m   Map factor

mb   Millibar(s)

n	Number of independent variables
p	Atmospheric pressure
p_o	Sea-level pressure
p_G	Surface pressure
s_y	Standard error
t	Time
z_p	Height of isobaric level p
$\propto$	Specific volume
χ^2	Chi-square statistical frequency
∇	Two-dimensional del operator on a constant pressure surface
η	Absolute vorticity
ω	Vertical velocity (dp/dt)
Ω	Angular velocity of the earth
ϕ	Latitude
θ	Potential temperature
γ	Vertical component of relative vorticity
V_g	Geostrophic wind
V_p	Vector horizontal wind at isobaric level p

1. Introduction

Numerical techniques for the sea-level pressure prognosis have
not shared the same success as those used for forecasting the 500-mb
contour map. Methods presently in operational use consist in general
of a 500-mb prognosis coupled with prognosis of thickness using some
form of the first law of thermodynamics. Methods similar in nature
have been suggested by Haltiner and Hesse [1958] and Reed [1956] .
In addition, it has become standard practice with such units as the
U. S. Navy Fleet Numerical Weather Facility, Monterey, California,
(FNWF) to superimpose certain empirical corrections on the location
and intensity of cyclones and anticyclones. . The inadequacy of
numerical methods to predict cyclogenesis and anticyclogenesis is
perhaps the greatest deficiency of the presently operational numerical
techniques. This study was undertaken to help eliminate some of
these problems.

Two models were developed: (1) a model using solely dynamic
parameters in the form of a multiple linear regression equation; and
(2) a stepwise multivariate regression analysis using the "dynamic"
predictors as well as certain other parameters selected to reveal
long-term and long-wave influences on the surface pressure change.
For convenience the two models are hereafter referred to as "the
dynamic model" and "the statistical model", respectively. Separate
and detailed descriptions of the two models are presented in the
following sections.

Latitude 50N in early winter months was chosen as a latitude offering typical, if not difficult, forecast problems for the investigation. The dependent data were taken, insofar as possible, from 30 consecutive days beginning in late October, 1962. Early December of the same year provided five days of independent data. Twenty-four geographical locations, at intervals of 15 degrees of longitude around the latitude belt were chosen as "stations". The "stations" were arbitrarily numbered from 1 to 24, starting at ocean station "Papa" in the Gulf of Alaska, in an eastward direction.

The data used in both models were taken from numerical computations and printout charts prepared by use of the CDC 1604 electronic digital computer. The programs and data tapes necessary to compute and print the charts were supplied by FNWF. All statistical computations were made on the same computer using selected programs from the BIMD[1] Fortran library.

The techniques employed in this work were not designed to supplant the numerical methods now in use. This effort was conducted so as to illuminate some of the factors not now considered that may be influential in causing a sea-level pressure change. Time steps of twenty-four hours were attempted in connection with both models in

---

[1] A compilation of statistical electronic computer programs that were compiled and edited by the Biology and Medical Department of the University of California, Los Angeles. This manual is distributed through the UCLA campus bookstore.

order to test the efficacy of this longer-than-normal time step in
conjunction with statistical methods.

2. A derivation of the dynamical prediction equation

The following development is taken, after Reed [1962] , with some modifications and simplifications in the later stages of the derivation. Pertinent remarks will indicate where these modifications are applicable.

The frictionless vorticity equation for the 1000-mb level may be well approximated by

$$\frac{\partial}{\partial t}(\mathcal{S}_o + f) = -V_o \cdot \nabla(\mathcal{S}_o + f) + f\left(\frac{\partial \omega}{\partial p}\right)_o .$$ (1)

If we assume a parabolic vertical velocity profile between the surface and 500 mb (subscript 5) of the form

$$\omega = \omega_o + (\omega_5 - \omega_o)\left[1 - \left(\frac{p - p_5}{p_o - p_5}\right)^2\right]$$ (2)

and substitute for $\left(\frac{\partial \omega}{\partial p}\right)_o$ in (1), we obtain

$$\frac{\partial}{\partial t}(\mathcal{S}_o + f) = -V_o \cdot \nabla(\mathcal{S}_o + f) - \frac{2f}{p_o - p_5}(\omega_5 - \omega_o) .$$ (3)

The geostrophic wind is used to approximate the vorticity and, following Reed, the 1000-mb contour pattern may be regarded as consisting of a set of equally-spaced circular highs and lows of the form

$$Z_o = -\frac{fm}{g}Uy + B\sin\frac{2\pi}{L}x \sin\frac{2\pi}{L}y$$ (4)

superimposed on a constant zonal current U. Here x and y are the eastward and northward distance elements, respectively, and the

4

absolute constant resulting from the integration of U along the meridian

is taken as zero. Then the 1000-mb relative vorticity becomes

$$\zeta_o = \frac{g}{f} \nabla^2 Z_o = -\frac{8\pi^2}{f L^2} g \left[Z_o + \frac{f_m}{g} U y \right] . \tag{5}$$

Substituting (5) into (3) yields the equation

$$\frac{\partial}{\partial t} \left(-Z_o + G' - \frac{f_m}{g} U y \right) = -V_o \cdot \nabla \left(-Z_o + G' - \frac{f_m}{g} U y \right) - \frac{f^2 L^2}{4\pi^2 g} \left(\frac{\omega_5 - \omega_o}{P_o - P_5} \right) \tag{6}$$

$$G' = \frac{f^2 L^2}{8\pi^2 g} , \qquad H' = G' - \frac{f_m}{g} U y .$$

The particular form of the adiabatic thermodynamic equation

used here is

$$\frac{\partial}{\partial t} \left(\frac{\partial z}{\partial p} \right) = -V \cdot \nabla \frac{\partial z}{\partial p} - \sigma \omega, \qquad \sigma = -\frac{\alpha}{g\theta} \frac{\partial \theta}{\partial p} . \tag{7}$$

With the assumption of a linear geostrophic-wind hodograph in the

layer between 1000 mb and 500 mb, equation (7) may be integrated

with respect to p from 1000 to 500 mb to give

$$\frac{\partial}{\partial t} \left(Z_5 - Z_o \right) = -V_o \cdot \nabla \left(Z_5 - Z_o \right) + \sigma \left(\frac{P_o - P_5}{3} \right) \left(2\omega_5 + \omega_o \right) . \tag{8}$$

Next multiply (8) by

$$k' = \frac{3}{8\pi^2} \frac{f^2 L^2}{g\sigma (P_o - P_5)^2} , \tag{9}$$

a slowly varying parameter which will be regarded here as constant, and add the modified version of (8) to (6) giving

$$\frac{\partial}{\partial t}\left[-Z_0 + H' + k'(Z_5 - Z_0)\right] = -V_0 \cdot \nabla\left[-Z_0 + H' + k'(Z_5 - Z_0)\right] + \frac{f^2 L^2 \omega_0}{8\pi^2 g(p_0 - p_5)} . \quad (10)$$

Use of the kinematic boundary condition $\omega_0 = V_{g_0} \cdot \nabla p_g$ allows the vertical velocity ω_0 to be included within both brackets of (10) as a terrain effect. This last effect is not considered in this paper, since it is not one of the dynamic factors explicitly selected for the statistical regression employed in this study. Hence we are left with the result

$$\frac{\partial}{\partial t}\left[-Z_0 + H' + k'(Z_5 - Z_0)\right] = -V_0 \cdot \nabla\left[-Z_0 + H' + k'(Z_5 - Z_0)\right] \quad (11)$$

as the prediction equation.

Equation (11) is now rearranged to give

$$\frac{\partial}{\partial t}\left[k'Z_5 - (1 + k')Z_0 + H'\right] = -V_0 \cdot \nabla\left[k'Z_5 - (1 + k')Z_0 + H'\right] \quad (12)$$

or

$$\frac{\partial}{\partial t}\left[kZ_5 - Z_0 + H\right] = -V_0 \cdot \nabla\left[kZ_5 - Z_0 + H\right]$$

$$k = k'/(1 + k'), \quad H = H'/(1 + k'). \quad (13)$$

We next make use of the principle, first introduced by Fjortoft [1952] and extended by Reed [1962], of employing an equivalent advecting wind which gives the same instantaneous advection as V_0,

6

but which has the property of changing more slowly with time. This concept has proved to be particularly valuable in graphical integration of the dynamical equations where long time steps are employed.

Thus (13) may be written in the equivalent form

$$\frac{\partial}{\partial t}\left[k Z_5 - Z_o + H\right] = -V_E \cdot \nabla\left[k Z_5 - Z_o + H\right] \tag{14}$$

or, alternatively as

$$\frac{\partial Z_o}{\partial t} = V_E \cdot \nabla\left[k Z_5 - Z_o + H\right] + \frac{\partial}{\partial t}\left(k Z_5\right), \tag{15}$$

where

$$V_o = k \times \frac{g}{f}\nabla Z_o, \quad V_E = k \times \frac{g}{f}\nabla Z_E, \quad Z_E = \left[k Z_5 + H\right]. \tag{16}$$

7

3. Simplification of the dynamic model

In Reed's model the scalar field of z_E whose gradient defines V_E has a term $(-M)$ representative of terrain effects in addition to those in equation (16). As already noted, in approaching the statistical application of the dynamic model we have set $M=0$.

A word of discussion regarding Reed's function G and our function H is appropriate. From equations (6) and (9), G is expressible as

$$G = G'/(1+k') = \frac{f^2 L^2}{8\pi^2 g}/(1+k') , \quad L = \frac{2\pi a \cos\phi}{N} .$$

Reed uses a mean value of $k=0.55$ or $k=1.22$ for wave number $N=6$ in connection with his 12-hr Lagrangian prediction technique. It may be shown that $G = \left[\frac{\Omega^2 a^2}{2 N g (1+k')}\right](\sin 2\phi)^2$ and hence G is an increasing function of latitude up to $\phi = 45N$. For $45 < \phi < 55$ degrees which is typical of the range of latitude encountered in this study, G is a slowly decreasing function. The mean northward gradient of the G-field centered at latitude 50N over a 10-degree span, is such that $V_G \doteq 3$ knots in an eastward direction.

Recall that our function H is given by $G - \frac{f_m}{g} \frac{U_y}{1+k'}$, and using $(1+k')=2.22$ it follows that the term involving U is equivalent to a mean maximum zonal wind (according to Namias and Clapp [1951]) of 2 knots. In lower latitudes the effects of the G and U terms are of opposite algebraic sign whereas in the latitude belt of this discussion

the gradient of H is equivalent to an eastward wind of 5 knots. Consequently the H-field in (16) has been neglected relative to kz_5 in the z_E field. With this simplification the prediction equation (15) becomes

$$\frac{\partial z_o}{\partial t} = V_E \cdot \nabla h + \frac{\partial}{\partial t}(k Z_5) \tag{17}$$

since $h \approx z_5 - z_o$.

The last term $\frac{\partial}{\partial t}(k Z_5)$ in (17) may be obtained by employing the graphical prediction technique for the barotropic model as described in Haltiner and Martin [1957, pp. 395-398]. If $\bar{z}_5$ represents the space-mean 500-mb height, the Fjortoft method leads to the result

$$\frac{\partial}{\partial t}(\bar{Z}_5 - Z_5 + J) = - V_{\bar{Z}+J} \cdot \nabla (\bar{Z}_5 - Z_5 + J) \tag{18}$$

at the level of nondivergence (assumed to be 500 mb). Fjortoft's space-mean advecting wind $V_{\bar{Z}+J}$ is given by

$$V_{\bar{Z}+J} = \frac{g}{f} k \times \nabla (\bar{Z}+J) . \tag{19}$$

We therefore have the familiar result

$$\frac{\partial Z_5}{\partial t} = V_{\bar{Z}+J} \cdot \nabla (\bar{Z}_5 - Z_5 + J) - \frac{\partial \bar{Z}_5}{\partial t} . \tag{20}$$

The Fjortoft graphical treatment makes use of the following function for J:

$$J = \int_0^\phi \frac{\Omega^2 d^2 \sin\phi \cos\phi \, d\phi}{m^2 g} .$$

Petterssen [1956, pp. 392] gives the distribution of the J-field

based upon a grid mesh d=1000 km. At latitude 50N, the gradient of

J which he shows is equivalent to an easterly geostrophic wind of

8.5 knots. In regions of maximum advection of $(\bar{z}-z+J)$, it appears

reasonable on the basis of a statistical approach to approximate the

500-mb advecting space-mean wind $V_{\bar{z}+J}$ by $V_{\bar{z}}$, where $V_{\bar{z}}$ is

the space-mean geostrophic wind at 500 mb. Moreover, since our

"stations" are far apart and confined to the 50N latitude circle, the

working hypothesis made in the following statistical analysis is that

$\dfrac{\partial \bar{\bar{z}}}{\partial t}$ gives only "feedback" contributions to the term $V_{\bar{z}} \cdot \nabla(\bar{z}_5 - z_5 + J)$

For simplicity the 1000-mb height change resulting from (15) and

(18) becomes

$$\frac{\partial \bar{z}_0}{\partial t} = V_E \cdot \nabla h + k\, V_{\bar{z}} \cdot \nabla(\bar{z}_5 - z_5 + J).$$ (21)

Note that since $V_E \doteq k\, V_{\bar{z}}$ it follows that for 24-hr advection it is

appropriate to consider the first term on the right side of (21) as a

reduced advection. Furthermore the vector $k\, V_{\bar{z}}$ in (21) has also

been treated as the reduced advecting wind V_E , since an inter-

pretation of this kind has been found useful in "advecting" the movement

of rain areas associated with progressive sea-level cyclones by

Renard [1959] . We have noted that Reed gives the value k=0.55

as appropriate to a Lagrangian forecast technique with 12-hr time

increments. In our computations k was rounded off to 0.5 in view of

10

the subjectivity of hand measurements of the space-mean geostrophic wind $V_{\bar{z}}$. Hence with these simplifications our prediction model now becomes

$$\frac{\partial Z_o}{\partial t} = A_o + A_1 \left[-V_E \cdot \nabla h \right] + A_2 \left[-V_E \cdot \nabla (\bar{Z}_6 - Z_5 + J) \right]. \tag{22}$$

Here the regression coefficients A_o, A_1, A_2 are introduced as unknowns in order to absorb any statistically-determined feedback relationships contained in the preceeding analysis, such as for example, the assumption $k V_{\bar{Z}+J} \cdot \nabla (\bar{Z}_5 - Z_5 + J) = V_E \cdot \nabla (\bar{Z}_5 - Z_5 + J)$.

In equation (22) the best-fit determination of A_o, A_1, and A_2 will be made by least squares. In essence, however, this equation presents a dynamically formulated problem similar to that of Reed [1956] and Haltiner and Hesse [1958] .

4. Computational procedures for the dynamic model

The local change of the 1000-mb height in equation (22) is written for a 24-hr time step and is considered to be strictly proportional to the 24-hr change of sea-level pressure, Δp_o, which serves as the dependent variable. The advecting terms in the working equation (22) are then used as the independent variables in a multiple linear regression equation of the form

$$Y = A_o + A_1 X_1 + A_2 X_2$$

$$\Delta p_o = A_o + A_1 \left[-V_E \cdot \nabla h \right] + A_2 \left[-V_E \cdot \nabla \left(\bar{Z}_5 - Z_5 + J \right) \right]. \tag{23}$$

The advecting terms were evaluated by a quasi-Lagrangian technique using fixed-point computations at points determined by trajectory tracing. At each of the 24 stations for 30 days an upwind point was determined. This was accomplished by specifying a geostrophic wind from the 500-mb space-mean height field terminating at the station in question and then tracing the upwind trajectory in the contour channel for a distance corresponding to 24 hours. In the cases of difluent and confluent contours 6-hour steps on the initial chart were used so that more representative space-mean winds were available at each step.

The parameters $\bar{z}$ and J as used in the vorticity term were computed using a square grid distance of 782 km. The space-mean 500-mb height field used to determine the advecting wind was obtained by subjecting

the 500-mb heights to four scans using a "smoother" of the type

$$(A_{ij})_S = \frac{1}{8}(A_{ij} + \nabla^2 A_{ij})_I \qquad (24)$$

where subscripts S and I refer to smoothed and initial values, respectively.

As previously noted, k in $V_E = k V_{\frac{z}{2}}$, was taken to be 0.5 and the wind speed in all cases was reduced by this factor. With this choice of advecting wind, the upwind point thus determined for fields of both h and $(\bar{z}_5 - z_5 + J)$ was the same for both advection computations. The difference of the values of h and $(\bar{z}_5 - z_5 + J)$ from their initial values over the station was recorded as the 24-hr advective change.

FNWF data tapes served to give printouts of the entire fields for all 24-hr forecast periods under consideration. Values of Δp_0 at the stations along 50N were then obtained by bilinear interpolation.

5. Results for the dynamic model

The BIMD 06 program of the BIMD library was used to perform the least-squares regression analysis based upon the use of equation (23) for the dynamic model. This was accomplished for each of the three data stratifications shown in Table 1. Relevant statistics obtained by this analysis are displayed in Table 2. Here PRV stands for per cent reduction of variance, while R is the multiple correlation coefficient.

After analyzing the results obtained by stratifying the data it was decided that due to the greater percent reduction in variance given by data from the fifteen stations over or near the land areas that this group would be used to develop the final regression equation. The resulting equation was (land areas only):

$$\Delta p_o = 6.59 - .04 \left(-V_E \cdot \nabla \eta \right) - 1.8 \left(-V_E \cdot \nabla h \right). \tag{25}$$

Note that the regression coefficients are negative indicating that advection of higher values of η and of h each cause a negative contribution to the pressure change. This agrees with usual synoptic observations.

Since the assumption of a linear relationship between the predictors and the predictand was not necessarily valid, scattergrams of Δp_o versus both independent variables were examined in order to investigate possible indications of a preferred relationship. The BIMD 27 program

14

as adapted to the CDC 1604 computer was employed to plot these scatter-grams. The results using thickness and absolute vorticity advection, respectively, are shown in figures 1 and 2. Examination of these figures shows that while a linear regression between Δp_o and $-\mathbb{V}_E \cdot \nabla h$ is reasonably valid, the relationship between Δp_o and $-\mathbb{V}_E \cdot \nabla \eta$ appears[2] to be non-linear with no obvious correlation.

The simple correlation coefficients found between Δp_o and $-\mathbb{V}_E \cdot \nabla h$ and $-\mathbb{V}_E \cdot \nabla \eta$ were -.41 and -.04 respectively (see also Table 2a). Note that the percent reduction in variance attributed to the partial correlation of the vorticity advection is insignificant, indicating that the thickness-advection parameter will give equally good results when used alone as a predictor. This was not due to a significant correlation between the variables $-\mathbb{V}_E \cdot \nabla h$ and $-\mathbb{V}_E \cdot \nabla \eta$, since their simple correlation coefficient was only 0.17.

The regression equation developed from these parameters was tested on an independent sample of 75 cases drawn from data gathered in the month of December with the results shown in Table 2b.

The apparent lack of success of the 24-hour absolute vorticity advection, as measured by the method here employed, to furnish significant predictability for the subsequent 24-hour pressure change

[2]The symbol η has been introduced to represent $(\bar{z}_5 - z_5 + G)$.

was surprising, particularly in view of the comparative usefulness of the thickness advection. Some of the primary reasons for these differences in the statistical behavior of these two predictors were revealed upon a critical re-examination of the initial-data fields. These are discussed in subsections (a), (b) and (c), below, with some resulting conclusions in (d).

(a) Advection of 500 mb absolute vorticity. The values of this variable are sensitively dependent upon the field of $\overline{z}_5 - z_5 + J$. This field had the characteristic of exhibiting unusually strong gradients over short distances near the vorticity centers, with relatively weak gradients elsewhere. Thus any cumulative error in constructing a 24-hour upwind trajectory, based on the use of the equivalent-advecting wind V_E can give rise to sizeable differences in the value of $\overline{z}_5 - z_5 + J$ to be advected. Reed [1962] has also referred to the question of trajectory accuracy.

(b) Advection of thickness. The printed fields of $h = z_5 - z_0$ displayed a smaller degree of non-linearity over short distances and were subject to considerably smaller upwind-point error. The difference in appearance of the $\overline{z}_5 - z_5 + J$ and $z_5 - z_0$ fields may be attributable to the smoothing process used in obtaining $\overline{z}_5$ whereas no smoothing was employed in obtaining the thickness field.

(c) Approximation $V_E = k V_5$ for advection of absolute vorticity
In a small percentage of cases of 24-hour 500-mb advection the value

16

of $\frac{1}{2}\left[V_{\bar{z}} \cdot \nabla\left(\bar{z}_5 - z_5 + J\right)\right]$ was not equal to that using the

advecting wind $V_E = \frac{1}{2} V_{\bar{z}}$. This occurred when a 24-hour

trajectory passed over an extreme value of $\bar{z}_5 - z_5 + J$.

(d) It must be concluded that the statistical Lagrangian technique

employed here suffers from the defect of employing excessive time-

steps. While the procedure bears some similarity to the Fjortoft

technique, it does not have the advantage of scanning comparative data

from all latitudes within the grid map. Hence the smoothing capability

usually available in most prognostic procedures (and in analysis in

general) could not be used as a prognostic aid here.

6. A multivariate linear regression analysis for 24-hour prediction
of sea-level pressure

In this phase of the investigation a total of 28 independent variables
were tested as possible predictors in a purely statistical approach.
Recent investigations by Ostby-Viegas [1960] , Miller [1962] , and
others employing statistical reduction methods have utilized the
advantages of a computer to reduce large numbers of possible predictors
to a significant few. Such methods may use either objectively-determined
data with no immediate rationale for the relationship, or dynamically-
based data to select variables. Such significant multiple linear re-
gressions which exist are found by a statistical screening process.
Once an objectively chosen variable has been selected through the
screening process it is usually possible to find a causal relationship
between it and the predictand on the basis of synoptic-dynamic
considerations

In this section, the intent was to utilize the two predictors already
employed in the dynamic model (see equation (23)). Since the predictors
already chosen rest heavily on 500-mb parameters it was decided to
choose, as far as possible, additional objective parameters from this
level. Furthermore, since the month of November 1962 was charac-
terized by contrasting mid-latitude regimes in the Pacific and Atlantic
Oceans, with higher than normal mid-latitude zonal flow in the Pacific
and anomalous blocking action in the Atlantic, it was felt that the

18

24-hour values of the dynamic (advective) parameters might contain considerable scatter at stations influenced by such contrasting meteorological activity. Consequently it was decided to select a limited number of 500-mb parameters indicative of the long wave features and, to some degree, of the extended-period circulation anomalies. Recourse is made here to some of the extended-forecast concepts of Namias [1951] . At each station the latest 500-mb height z_c and those for each of the three preceeding 24-hour map times were read off or interpolated from the contoured printout charts. From these data the two sets of parameters, the two-day trend T_i and three-day mean $\widetilde{Z}_i$ were computed for each station, and a regression equation of the form

$$\Delta P_{o,i} = B_{io} + B_{i_1}\left(-V_E\cdot\nabla \eta\right) + B_{i_2}\left(-V_E\cdot\nabla h\right) \qquad (26)$$
$$+ \sum_{j=o}^{11} B_{i-2j} T_{i-2j} + \sum_{j=o}^{11} C_{i-2j} \overline{Z}_{i-2j} + Du_i + Ev_i$$

is sought, where j is a 15° longitude interval, considered positive eastward. The subscript i-2j , indicates that we are examining data at all upwind points 30° longitude apart. Note that the two-day trend at station i is given by

$$T_i = \left(z_c - z_{-2}\right), \qquad (27)$$

and $\overline{Z}_i$ for simplicity has been defined here as the simple arithmetic mean $$\overline{Z}_i = \left(z_c + z_{-1} + z_{-2} + z_{-3}\right)/4 . \qquad (28)$$

The implication of the summation sign before the T_{i-z_j} and $\bar{Z}_{i-z_j}$ terms is that we are considering all of the large-scale upwind and downwind rates of change and mean heights, in contributing predictability to Δp_o at the site denoted by the subscript i .

The variables u_i and v_i are the 850-mb zonal and meridional geostrophic-wind components at station i . They are presented in equation (23), firstly, in order to include possible relevant low-level effects, and secondly, in order that a physically important factor such as $\overline{uv}$ (where the superior bar indicates a zonal mean) may at least be implicit within the set of independent variables. The significance of this cross-covariance is that it suggests effects of the zonal-index cycle (see Haltiner and Martin, [1957, pp. 446-448].).

With 28 possible predictors appearing in (23), the BIMD 09 program was used to perform the regression analysis of the statistical model. This program is a modification of one originally written by M. A. Efroymson [1955] of the Esso Research and Engineering Company. The important features of the program include a stepwise screening of the predictors using arbitrary upper and lower critical F-values as cutoff limits for inclusion or rejection of variables.

The F statistic as employed in this test is the ratio of the mean squares explained by the regression to the residual or unexplained mean squares. According to Anderson [1960] the F-ratio is the ratio of two chi-square distributed variables with k and n-k-1 degrees of

20

freedom, where n is the number of cases in the data sample. Miller [1962] suggests values for the critical or cutoff F-values for introducing predictors into (23). If P is the total number of possible predictors (28 in this investigation) and k is the number of predictors already selected, his critical F-value is given as

$$F_{\alpha *} \equiv \frac{F_\alpha}{P-k+1}$$

(29)

where $\alpha = \dfrac{\alpha *}{P-k+1}$. The value $\alpha * = .05$ is the usual critical significance level in the selection test. It is apparent that the level imposed by this method of determining a critical F-value will decrease as more predictors are chosen.

Inasmuch as the BIMD 09 screening program uses a fixed F-level throughout the screening process, it seems desirable to perform several regression analyses of the data with differing F-levels but retaining one coinciding with Miller's (F = 10.0). In a recent analysis (Martin et al., [1963]) the recommendation is made that a lower F-level for rejection of a previously selected variable should be taken as zero when using the BIMD 09 program with Miller's selection criterion. Accordingly, predictors were arbitrarily selected in this analysis with upper critical F-levels of 10.0 and 5.0, and a lower limit in each case of zero.

Selected parameters are shown in Table 3 in the order in which they were chosen by the analysis of the dependent data. The form of the F-test used to compute the significance of the final regression equation

21

is given (after Anderson, [1960, p.89]) as below:

$$F(k, n-k-1) = \left(\frac{R^2}{1-R^2}\right)\left(\frac{n-k-1}{k}\right) \qquad (30)$$

where R is the multiple correlation coefficient. The percent reduction in variance, R^2, is given by the formula

$$R^2 = 1 - \left(\frac{S_y}{\Delta_y}\right)^2 \qquad (31)$$

where S_y the reduced standard error, and Δ_y, the total standard error, are available at each step of the BIMD 09 program printout.

The cumulative percent reduction in variance was computed for each step of the regression (see Table 3). Note that using Miller's suggested initial F-level of 10.0 only two parameters are chosen by the screening process. These significant predictors are $-V_E \cdot \nabla h$, the advection of thickness as employed in the dynamic model, and T_i the 2-day 500-mb height change over the station. Together, these parameters give a cumulative percent reduction in variance of 18.43. When the F-level is lowered to 5.0 four additional parameters are selected, each of which contributes a relatively small gain in PRV. The four additional parameters selected (in the order chosen) were $-V_E \cdot \nabla h$ or the advection of z_5-z_5+ J as employed in the dynamic model; U_i, the geostrophic component of the 850-mb wind over the station; T_{i-6}, the 2-day 500-mb height trend 90° upstream;

and $\overline{Z}_{i-10}$, the 3-day mean 500-mb height 150° upstream. The resulting regression equation which includes the six selected variables (see Table 3) is

$$\Delta p_o = 116.1 - 1.6(-V_E \cdot \nabla h) - .18(-V_E \cdot \nabla h)$$
$$- .28\,\overline{T}_i - .54 u_i - .24\,\overline{T}_{i-6} - .14\,\overline{Z}_{i-10}. \tag{32}$$

While the computed F-levels indicate that these last predictors are statistically significant it is possible that such correlations as are indicated arise from "noise" and/or erroneous data. Under such circumstances the regression equation may tend to "overfit" the sample. A discussion of this effect is given by Panofsky and Brier $\left[1958,\ \mathrm{p.176}\right]$.

The existence of "overfitting" of the dependent sample is demonstrated by the instability of the regression equation when applied to the independent data. This phenomenon may evidence itself by a substantial decrease ("shrinkage") in the percent reduction of variance explained with the independent sample. For example, when equation (32) was tested on an independent sample of 75 cases, a PRV of 10.6 percent occurred, compared to 22.04 percent for the dependent sample. Although the "shrinkage" here is large, some stability of the final equation is indicated and an improvement shown over the dynamic model.

From practical considerations we wish to use only the most efficient predictors. Those which offer little improvement in predictability (by the added percent reduction in variance criterion) are chiefly of theoretical interest. Undoubtedly the most practical prediction

equation giving the least overfit would involve the two parameters

selected in the statistical model with $F \geqslant 10$, namely, advection of

thickness $-V_E \cdot \nabla h$ and the 2 -day height difference T_t . The

usefulness of those predictors with lower F-levels is doubtful, especially

when the small percent reduction in variance achieved by their use is

considered.

Errors in data sampling have already been referred to in section 6,

particularly in reference to the dynamic predictors. Other errors are

map-scale error and interpolation errors, both of which are considered

to be negligible in this instance.

7. Conclusions

The prediction equations developed by this investigation do not offer a significant improvement to existing methods. This is felt to be due partly to the limitations inherent in applying the statistical technique as a linear operator and partly attributable to the fact that single-point observations fail to capture gradient effects in the manner of a closely spaced grid. Considering that the chosen latitude and season indicate that baroclinic development can normally be expected, any useful filter should predict non-linear effects in a consistent fashion. The results obtained by applying the developed regression equations to the independent samples indicate that this is not being done and that we must recognize some of the shortcomings of the measurement methods. The most immediate improvements suggested from the results are: (1) decreasing the Lagrangian time step, and (2) employing an entire grid map to verify actual Δp_o patterns.

Table 1. Stratification of data for analysis of the dynamic model

Grouping	Number of stations in sample	Population size
Individual stations	1	30
Continental stations	15	450
Ocean stations	9	270
Total of stations	24.	720

Table 2. Summary of pertinent statistics for various sample stratifications for the dynamic model

(a) Dependent data

Grouping	Sample size	Variable	PRV	R	Final F value
Total stations	720	$-V_E \cdot \nabla \eta$ $-V_E \cdot \nabla h$ combined	.002 .144 .146	.382	61.31
Land stations	450	$-V_E \cdot \nabla \eta$ $-V_E \cdot \nabla h$ combined	.002 .179 .181	.426	49.64
Ocean stations	270	$-V_E \cdot \nabla \eta$ $-V_E \cdot \nabla h$ combined	.004 .130 .134	.366	20.62

(b) Independent data

Grouping	Sample size	Variable	PRV	R	Final F value
Land stations	75	combined	.01	.10	< 1

27

Table 3. Summary of results obtained from dependent data for the
statistical model

Predictor	F level on entry	Percent reduction in variance	Coefficient in regression equation ($F \gg 5.0$)
$-V_E \cdot \nabla h$	90.6	16.65	-1.6
T_i	10.8	1.78	-.28
$-V_E \cdot \nabla \eta$	6.9	1.06	-.18
u_i	6.5	.98	-.54
T_{i-6} ($90°$ upstream)	5.9	.86	-.24
$\bar{\bar{Z}}_{i-10}$ ($150°$ upstream)	5.0	.71	-.14
Constant term	-	-	116.1

Standard deviation of Δp_c, s_y = 7.484 mb

$$R^2 = 0.22$$
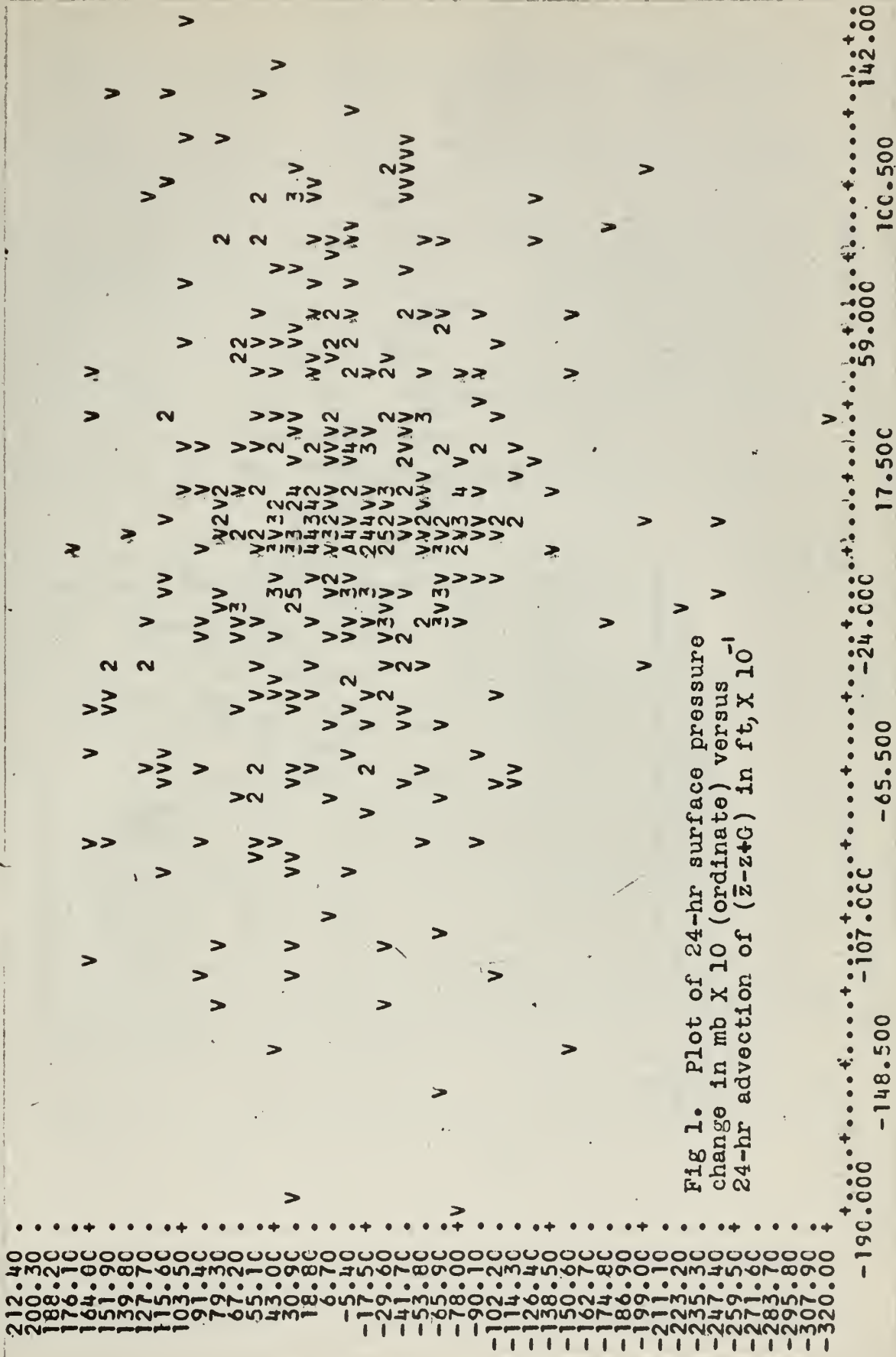
$$R = 0.48$$

$$F(6,443) = 20.9$$

$$F_c(.99) = 2.85$$

Fig l. Plot of 24-hr surface pressure change in mb X 10 (ordinate) versus 24-hr advection of $(\bar{z}-z+G)$ in ft, X 10^{-1}

-190.000   -148.500   -107.000   -65.500   -24.000   17.500   59.000   1CC.500   142.00

212.40
208.30
2C6.2C
176.1C
164.0C
151.90
139.80
127.70
115.60
1C3.50
91.4C
79.30
67.2C
55.1C
43.0C
30.9C
18.8C
6.70
-5.40
-17.5C
-29.6C
-41.7C
-53.8C
-65.9C
-78.00
-90.1C
-1C2.2C
-114.3C
-126.4C
-138.50
-15C.6C
-162.7C
-174.8C
-186.90
-199.1C
-211.20
-223.3C
-235.3C
-247.4C
-259.5C
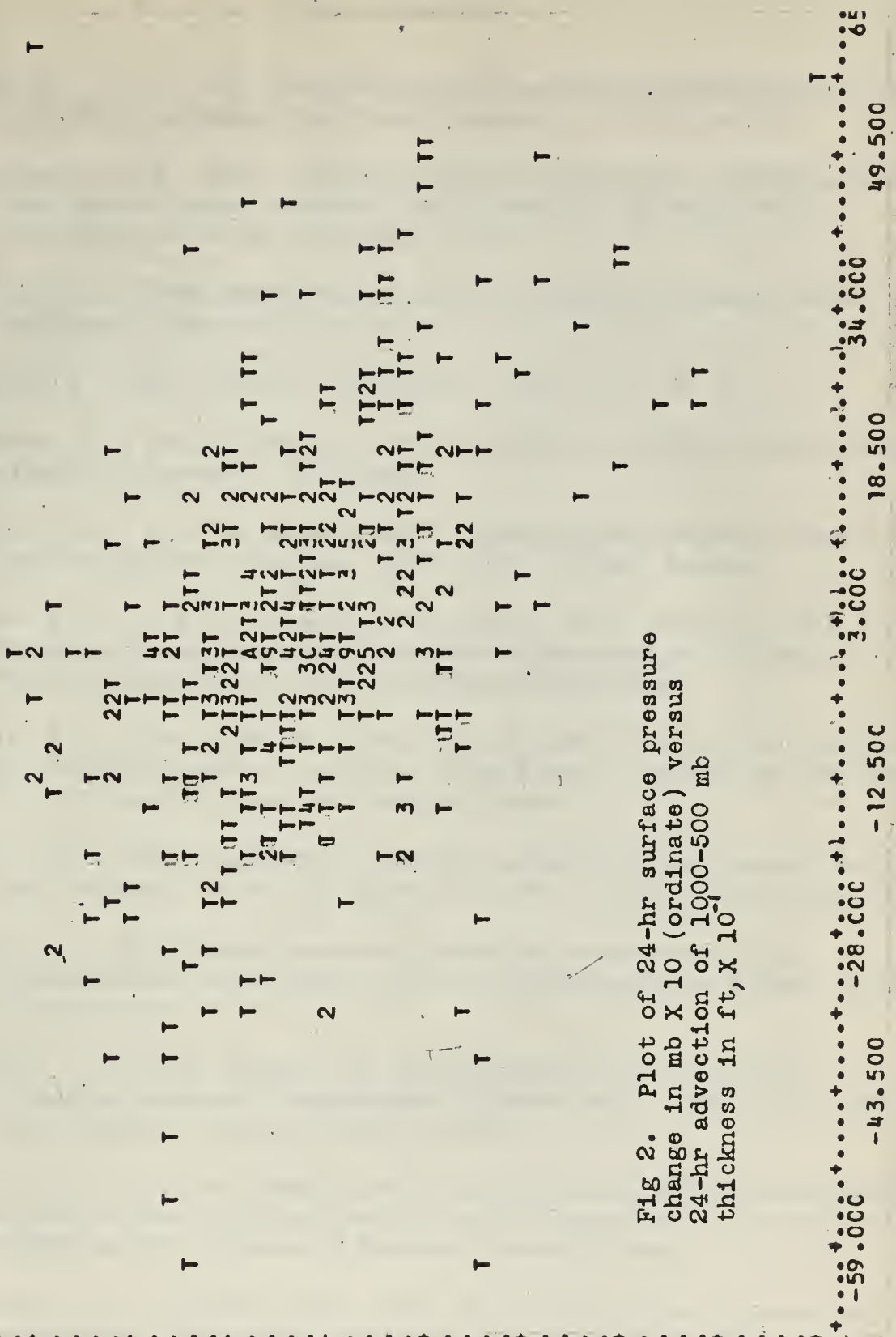-271.6C
-283.70
-295.80
-307.90
-320.00

29

Fig 2. Plot of 24-hr surface pressure
change in mb X 10 (ordinate) versus
24-hr advection of 1000-500 mb
thickness in ft, X 10^{-1}

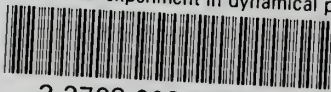# BIBLIOGRAPHY

Anderson, T.W., 1960, An Introduction to multivariate statistical analysis, John Wiley and Sons, New York, London

Efroymson, M.A., 1960, "Multiple regression analysis", in Mathematical methods for digital computers (A. Ralston & H.S. Wilfe, ed.), John Wiley and Sons, New York, London

Fisher, R.A., 1948, Statistical methods for research workers, Oliver and Boyd, Edinburgh, London

Fjortoft, R., 1952, Graphical techniques, Tellus, vol. 4, no. 3

Haltiner, G.J. and T.S. Hesse, 1958, Graphical prognosis including terrain effects, J. Meteor., 15, 103-107

Haltiner, G.J. and F.L. Martin, 1957, Dynamical and physical meteorology, McGraw-Hill Book Co. Inc., New York, Toronto, London

Martin, F.L., J.R. Borsting, F.L. Steckbeck, A.H. Manhard, 1963, Statistical prediction methods for North American anticyclones, U. S. Naval Postgraduate School, unpublished manuscript

Miller, R.G., 1958, "Statistics and the predictibility of weather", in Studies in statistical weather prediction, Final Report, contract No. AF 19(604) -1590, Travelers Weather Research Center

Miller, R.G., 1962, Statistical prediction methods by discriminant analysis, Met. Monogr., 4, no. 25, Amer. Met. Soc., Boston, Mass., pp.1-53

Namias, J., 1951, General aspects of extended-range forecasting, Compendium of Meteorology, American Meteorological Society, Boston, Massachusetts

Namias, J. and P.F. Clapp, 1951, Observational studies of general circulation patterns, Compendium of Meteorology, American Meteorological Society, Boston, Massachusetts

Ostby, F.P. and K.W. Veigas, 1960, A moving coordinate prediction model applied to east coast cyclones, Scientific Report No. 1, contract No. AF19(604)-5207, Travelers Weather Research Center

Panofsky, H.A. and G.W. Brier, 1958, Some applications of statistics to meteorology, Penn State University, University Park, Penn.

Petterssen, S. , 1956, Weather analysis and forecasting, volume 1, McGraw-Hill Book Co. Inc. , New York, Toronto, London

Reed, R.J. , 1956, A graphical method for preparing 1000-millibar prognostic charts, J. Meteor. , 14, 65-70

Reed, R.J. , 1962, Experiments in 1000-millibar prognosis, National Meteorological Center Technical Memorandum, Washington, D. C.

Renard, R.J. , 19 June 1956, An objective method for quantitative precipitation forecasting at Monterey, California, paper presented at 177th National Meeting of American Meteorological Society, San Diego, California

Spiegel, M.R. , 1961, Theory and problems of statistics, Schaum Publishing Co. , New York