A UNITED STATES
DEPARTMENT OF
COMMERCE
CATION

ESSA WBTM TDL 36

# ESSA Technical Memorandum WBTM TDL 36

**U.S. DEPARTMENT OF COMMERCE**
ENVIRONMENTAL SCIENCE SERVICES ADMINISTRATION
Weather Bureau

# Current High Altitude Observations–Investigation and Possible Improvement

M.A. ALAKA AND R.C. ELVANDER

Systems
Development
Office
SILVER SPRING
MARYLAND
July 1970

# ESSA TECHNICAL MEMORANDA

## Weather Bureau, Techniques Development Laboratory Series

The primary purpose of the Techniques Development Laboratory of the Office of Systems Development is to translate increases in basic knowledge in meteorology and allied disciplines into improved operating techniques and procedures. To achieve this goal, TDL conducts and sponsors applied research and development aimed at improvement of diagnostic and prognostic methods for producing weather information. The laboratory carries out studies both for the general improvement of prediction methodology used in the National Meteorological Service System and for more effective utilization of weather forecasts by the ultimate user.

ESSA Technical Memoranda in the Weather Bureau Techniques Development Laboratory series facilitate rapid distribution of material which may be preliminary in nature and which may be published formally elsewhere at a later date. The first five papers in the TDL series are part of the former Weather Bureau Technical Notes series.

Papers listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151. Price: $3.00 hard copy; $0.65 microfiche. Order by accession number shown in parentheses at end of each entry.

TN 10 TDL 1 Objective Prediction of Daily Surface Temperature. William H. Klein, Curtis W. Crockett, and Carlos R. Dunn, October 1965. (PB-168 590)

TN 11 TDL 2 Hurricane Cindy Galveston Bay Tides. N. A. Pore, A. T. Angelo, and J. G. Taylor, September 1965. (PB-168 608)

TN 29 TDL 3 Atmospheric Effects on Re-Entry Vehicle Dispersions. Karl R. Johannessen, December 1965. (PB-169 381)

TN 45 TDL 4 A Synoptic Climatology of Winter Precipitation from 700-mb. Lows for the Intermountain Areas of the West. D. L. Jorgensen, W. H. Klein, and A. F. Korte, May 1966. (PB-170 635)

TN 47 TDL 5 Hemispheric Specification of Sea Level Pressure from Numerical 700-mb. Height Forecasts. William H. Klein and Billy M. Lewis, June 1966. (PB-173 091)

WBTM TDL 6 A Fortran Program for the Calculation of Hourly Values of Astronomical Tide and Time and Height of High and Low Water. N. A. Pore and R. A. Cummings, January 1967. (PB-174 660)

WBTM TDL 7 Numerical Experiments Leading to the Design of Optimum Global Meteorological Networks. M. A. Alaka and F. Lewis, February 1967. (PB-174 497)

WBTM TDL 8 An Experiment in the Use of the Balance Equation in the Tropics. M. A. Alaka, D. T. Rubsam, and G. E. Fisher, March 1967. (PB-174 501)

WBTM TDL 9 A Survey of Studies of Aerological Network Requirements. M. A. Alaka, May 1967. (PB-174 984)

WBTM TDL 10 Objective Determination of Sea Level Pressure from Upper Level Heights. William Klein, Frank Lewis, and John Stackpole, May 1967. (PB-179 949)

WBTM TDL 11 Short Range, Subsynoptic Surface Weather Prediction. H. R. Glahn and D. A. Lowry, July 1967. (PB-175 772)

WBTM TDL 12 Charts Giving Station Precipitation in the Plateau States from 700-Mb. Lows During Winter. D. L. Jorgensen, A. F. Korte, and J. A. Bunce, Jr., October 1967. (PB-176 742)

WBTM TDL 13 Interim Report on Sea and Swell Forecasting. N. A. Pore and W. S. Richardson, December 1967. (PB-177 038)

WBTM TDL 14 Meteorological Analysis of 1964-65 ICAO Turbulence Data. DeVer Colson, September 1968. (PB-180 268)

WBTM TDL 15 Prediction of Temperature and Dew Point by Three-Dimensional Trajectories. Ronald M. Reap, September 1968. (PB-180 727)

(Continued on inside back cover)

U.S. DEPARTMENT OF COMMERCE
Environmental Science Services Administration
Weather Bureau

ESSA Technical Memorandum WBTM TDL 36

# CURRENT HIGH ALTITUDE OBSERVATIONS--INVESTIGATION AND POSSIBLE IMPROVEMENT

M. A. Alaka and R. C. Elvander

Systems Development Office
Techniques Development Laboratory

SILVER SPRING, MD.
July 1970

UDC 551.501.7.001.7:551.501.75.001.4:551.508.57

| | |
|---|---|
| 551.5 | Meteorology |
| .501.7 | Methods of observation and computation of upper air |
| .001.7 | Improvements |
| .501.75 | Methods of observation and computation of wind |
| .001.4 | Appraisal |
| .508.57 | Rawin equipment |

## CONTENTS

# CURRENT HIGH ALTITUDE OBSERVATIONS--
# INVESTIGATION AND POSSIBLE IMPROVEMENT

M. A. Alaka and R. C. Elvander

## ABSTRACT

On the basis of the available statistics of the atmosphere and of the random errors of observations, an estimate is made of the capability of current Rawin observations at 20 mb to measure differences in the strength of the zonal wind between two points separated by a given time or space interval. It is found that the random errors associated with these observations are high in comparison with the variability of the wind, especially during the summer months. In July, the error in the measured difference between two consecutive observations, taken 12 hours apart, exceeds the total true difference two out of five times; the error exceeds one-half the true difference three out of five times. In January, the chances are 80 percent that the measured difference between observations, taken 12 hours apart, is less than the total true difference. Considering that the average station separation of the current high-altitude program is 1000 km, the probability is only slightly more than 60 percent, in July, that the error in the measured difference between two simultaneous observations at adjacent stations is less than half the difference. In January, there is slightly less than 80 percent chance that the error is less than the total true difference and 70 percent chance that it is less than half the true difference.

The RMS random errors may be considerably reduced by proper processing. A procedure for "optimum" smoothing is introduced whereby the weights given to the different observations used in the smoothing are determined by climatology, by the relative location of observations, and by their RMS random errors. The reduction by this procedure of the initial RMS errors of observations is determined for different network configurations. It is found that with a station separation of 1000 km, optimum smoothing reduces the RMS random error of the 20 mb zonal winds from their current level of 4.5 kt to 4.0 kt in January and 3.4 kt in July. The improvement increases with network density.

The method of optimum interpolation is then used to determine the accuracy with which the field of zonal wind may be reconstructed from discrete observations. It is shown that in July, if the station separation is 850 km, the RMS error in reconstructing the zonal wind field by optimum interpolation from discrete observations is nowhere greater than the RMS error of the observations themselves. In January, the station separation must be no greater than 350 km to achieve comparable accuracy. Because of the slow time variation of the wind at 20 mb, it is possible, by using observations which are 24 hours old, to reduce the number of daily observations by about 30 percent without sacrificing any accuracy.

A high-altitude climatology is valuable both for use in lieu of daily observations, especially in summer, and for effective processing of daily observations. To compile a reliable climatology, it is not sufficient to mount a high resolution data acquisition program over a relatively short period. Two satisfactory data acquisition schemes are described.

## INTRODUCTION

To evaluate a meteorological data acquisition system, one must take into consideration the use made of the data and the processing which the data undergoes before it is used. We may conceive of two broad alternatives:

a. The observations are used "in the raw" without advantage of suitable processing to check their consistency or to compensate for the random errors associated with them.

b. The observations are analyzed in some "optimal" manner so that their information content is maximum.

Observations falling in one or the other of the above two categories may be evaluated by analysis of a set of related factors which include the temporal and spatial scale of the observed element, and the density, frequency, and accuracy of the observations made.

In the present report, we shall formulate some relevant relationships between the above factors and apply the relationships to evaluate current high altitude observational programs and, hence, to determine how best to improve these programs should improvement appear to be needed. We shall also attempt to assess the usefulness of a special high resolution data-acquisition experiment designed to permit a better estimate of observational requirements, at levels between 50 and 10 mb, than is possible from currently available information.

Because the procedure we have established depends on statistical information which is not readily available, the present study is limited to an evaluation of current observations of the zonal wind component at 20 mb. The same procedure can easily be extended to other meteorological elements and other levels once the necessary statistics have been compiled.

## THE CONCEPT OF OBSERVATIONAL EFFICIENCY

Consider a system of discrete observations which are separated by an interval $\tau$ in time or space. The efficiency of these observations (E) is defined (Bessemoulin, et. al., 1960; Alaka, 1967):

$$E = 1 - \frac{\sigma_\epsilon}{\sigma_\tau} \qquad (2.1)$$

where $\sigma_\epsilon$ is the root-mean-square (RMS) error of observations and $\sigma_\tau$ is the standard deviation of the true time or space variation of the measured meteorological element in the interval $\tau$. If $\tau$ is a space interval, $\sigma_\tau$ is related to the autocorrelation coefficient by the following relation (c. f. Gandin, 1963, Chapter 2):

$$\sigma_\tau^2 = \sigma_A^2 + \sigma_B^2 - 2\sigma_A\sigma_B\,\mu_{AB} \qquad (2.2)$$

where $\sigma_A^2$, $\sigma_B^2$ denote the variance of the element at locations A and B respectively, and $\sigma_\tau$ is the autocorrelation coefficient between the values of the element at these locations.

If the variance is homogeneous over the region of interest,

$$\sigma_\tau^2 = 2\sigma^2(1 - \mu_{AB}) \qquad (2.3)$$

If, furthermore, the autocorrelation coefficient is assumed to be both homogeneous and isotropic, $\mu_{AB}$ becomes $\mu(\tau)$, a function of distance alone and equation (2.1) becomes

$$E = 1 - \frac{\lambda}{\left\{2\left[1-\mu(\tau)\right]\right\}^{1/2}} \qquad (2.4)$$

where $\lambda = \frac{\sigma_\epsilon}{\sigma}$, the ratio of the RMS random errors to the standard deviation of the measured element.

If $\tau$ is a time interval, $\mu(\tau)$ denotes a time correlation coefficient with lag $\tau$.

From the above relations, it is seen that, for a given value of the random errors ($\sigma_\epsilon$), the efficiency increases as the time or space separation ($\tau$) between observations is increased. To appreciate the implications of this formulation, we may express the efficiency in terms of the probability that the random errors will not exceed a given fraction of the true variation of the measured element. The governing equation (Bessemoulin, 1960) is:

$$P = 1 - \frac{2}{\pi}\tan^{-1}\frac{\sigma_\epsilon}{K\sigma_\tau} \qquad (2.5)$$

Or, in terms of $\mu(\tau)$:

$$P = 1 - \frac{2}{\pi}\tan^{-1}\frac{\lambda}{K}\left\{2\left[1-\mu(\tau)\right]\right\}^{-1/2} \qquad (2.6)$$

The above equations express the probability P that $\epsilon_{\tau}$, the error in the measured difference between two observations separated by time or space interval $\tau$ does not exceed a fraction $k$ $(0 < k < 1)$ of the true difference $\Delta\tau$. Figure 1 illustrates this relation for different values of the efficiency E. As an illustration if E = .9, the probability is 94 percent that the error in the measured difference is less than the true difference, and 70 percent that the error is less than 1/5 the true difference. If E = .5, the above probabilities are reduced to 73 and 26.5 percent respectively.

The concept of efficiency is useful in evaluating the information content of observations "in the raw," i.e., those which are used without benefit of suitable processing to compensate for the random errors of observation. An efficiency of .9 or at least .8 (wide shaded band in Figure 1) is recommended for such observations.

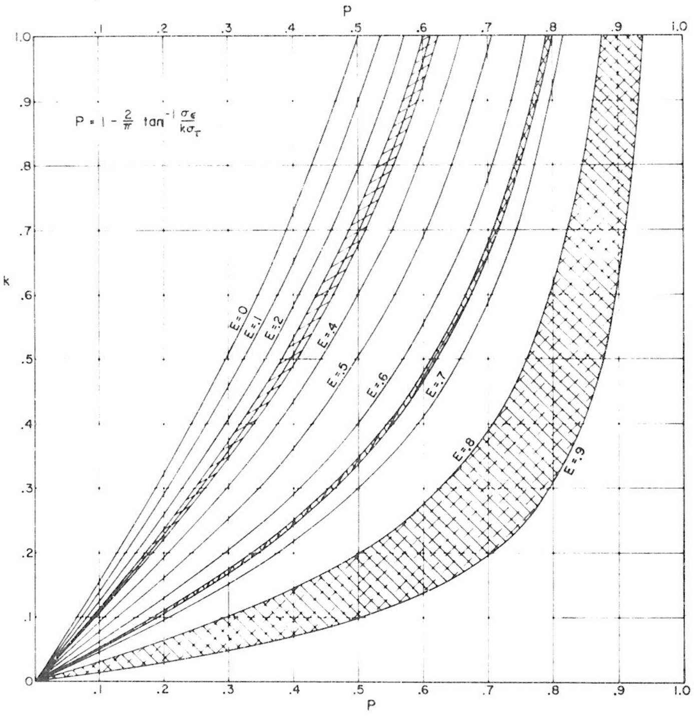


Figure 1--The probability, P, that the error in the measured difference between two observations separated by a space or time interval, $\tau$, does not exceed k times the true difference, as a function of the efficiency, E, of the observing system.

## THE EFFICIENCY OF RAWINSONDE OBSERVATIONS AT 20 mb

### a. Temporal Efficiency

If $\hat{\sigma}_\tau$ denotes the standard deviation of the difference in the measured value of a series of observations taken in the same locality at time intervals $\tau$ and if $\sigma_\epsilon$ denotes the RMS random errors[1], then the standard deviation of the true difference in the value of the element, $\sigma_\tau$, for the same interval is given by the following relation (Gandin, 1963, Chapter 2):

$$\sigma_\tau^2 = \hat{\sigma}_\tau^2 - 2\sigma_\epsilon^2 \qquad (3.1)$$

The above relation was used to determine the efficiency of 12- and 24-hour observations of the zonal wind component at 20 mb. Table 1 lists the 12- and 24-hour efficiencies for January and July and the values of $\hat{\sigma}_\tau$ on which they are based. The latter were adopted from recent computations by Colson, Greene, and Lewis (1968). A mean-square random error of 20 $kt^2$ was used[2] to obtain $\sigma_\tau$ in accordance with equation (3.1) and to compute the efficiency in accordance with equation (2.1). As expected, Table 1 indicates that the efficiency of rawin observations at 20 mb is much lower in July than in January.

Table 1--Efficiency of rawin observations of the zonal wind component at 20 mb in January and July

| Month | $\tau$ = 12 hours | | | $\tau$ = 24 hours | | |
|---|---|---|---|---|---|---|
| | $\hat{\sigma}_\tau$ (kt) | $\sigma_\tau$ (kt) | E | $\hat{\sigma}_\tau$ (kt) | $\sigma_\tau$ (kt) | E |
| January | 13.78 | 12.23 | .654 | 14.94 | 13.52 | .670 |
| July | 8.85 | 6.19 | .276 | 9.16 | 6.62 | .325 |

[1] The random errors include measurement errors plus errors due to features which are not resolved by the sampling network.

[2] This value was estimated from the computed structure function as proposed by Gandin (1963, Chapter 2).

The left and middle shaded bands in Figure 1 are those for July and January respectively. Each band spans the range between the efficiency of observations taken every 12 hours (left edge) and those taken every 24 hours. By reference to the figure, we find that in July, two out of five times, the error in the measured difference between two consecutive observations, taken 12 hours apart, exceeds the total true difference, and that the chances are better than three to five that the errors exceed one-half the true difference. The uncertainty in the measured 24-hour variation is not much better. The inevitable conclusion is that in July, unprocessed 12- or 24-hour wind observations have very limited operational value and offer little advantage over climatology, unless the random errors can be reduced to a small fraction of their present value.

The situation is much better in January. Although still very short of the desirable level, the efficiency of 12-hourly observations during this month is about .65. This means that the chances are 80 percent that the errors in the measured difference between observations, taken 12 hours apart, are less than the total true differnce, and somewhat more than 60 percent that the errors are less than half the true difference. Surprisingly, according to the available statistics, very little efficiency is gained by spacing the observations 24 hours apart. Evidently, there is much room for improvement. However, 20 mb wind observations made at 24- or 12-hour intervals with currently available sondes are operationally useful even without the benefit of processing to minimize the effect of the random errors.

The fragmentary statistics available indicate that a similar conslusion is valid for temperature observations.

b. Spatial Efficiency

Equation (2.4) may be used to determine the efficiency of the spatial distribution of a network of stations (Alaka, 1969). The

Table 2--Values of correlation coefficient, $\mu(\tau)$, of the zonal wind at 20 mb as a function of distance in January and July

| Month | $\tau$ (km × $10^{-2}$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| January | .987 | .971 | .954 | .934 | .913 | .889 | .864 | .837 | .808 | .778 |
| July | ---- | .780 | .724 | .672 | .623 | .578 | .537 | .498 | .462 | .429 |

continuous curves in Figure 2 show the manner in which the efficiency of current unprocessed observations of the zonal wind varies with station separation. Value of $\lambda = .7$ and $.17$ were used for summer

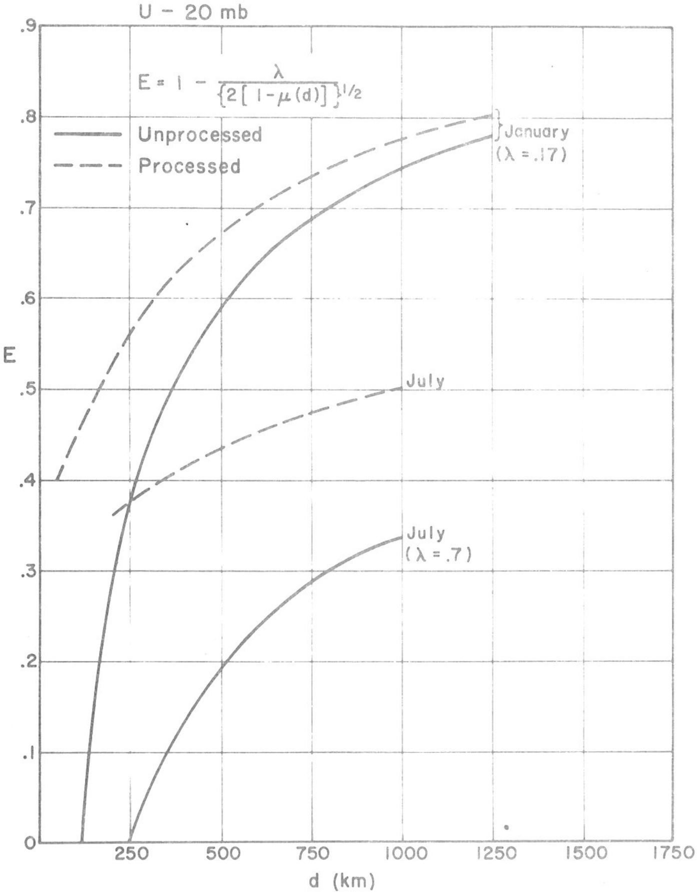


Figure 2--The efficiency, E, of radiosonde observations of the zonal wind at 20 mb in January, as a function of station separation.

and winter respectively. Corresponding values of $\mu(\tau)$ are given in Table 2. Considering that the average station separation of the current high alitude program is about 1000 km, the efficiency of these observations is about .75 in January and .35 in July. This means that in January (Figure 3) there is slightly less than 80 percent chance that the error in the measured difference, between two neighboring observations, is less than the total true difference and 70 percent chance that the error is less than half the true difference. In July (Figure 4) the probability is only slightly more than 60 percent that the error is less than the true difference and about 40 percent that it is less than half the difference.

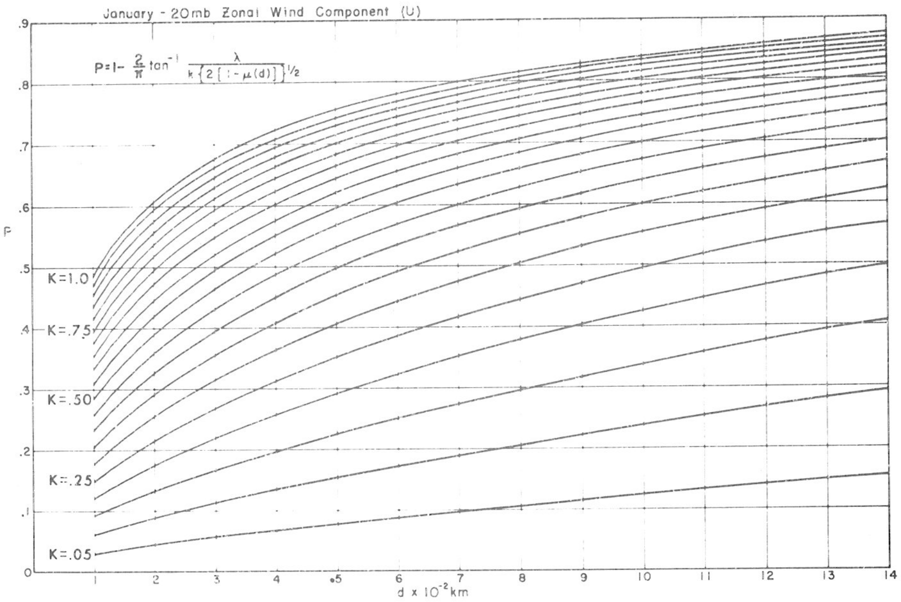


Figure 3--The probability, P, that the error in the measured difference of the 20 mb zonal wind in January, between two stations $\leq$ a fraction k of the true difference, as a function of the distance d between the two stations.

It should be noted that the spatial and temporal efficiencies of the current high-altitude program are not consistent. To make the spatial resolution consistent with one daily observation in January, the average distance between stations will need to be reduced by 50 percent. This means that the number of stations will need to be

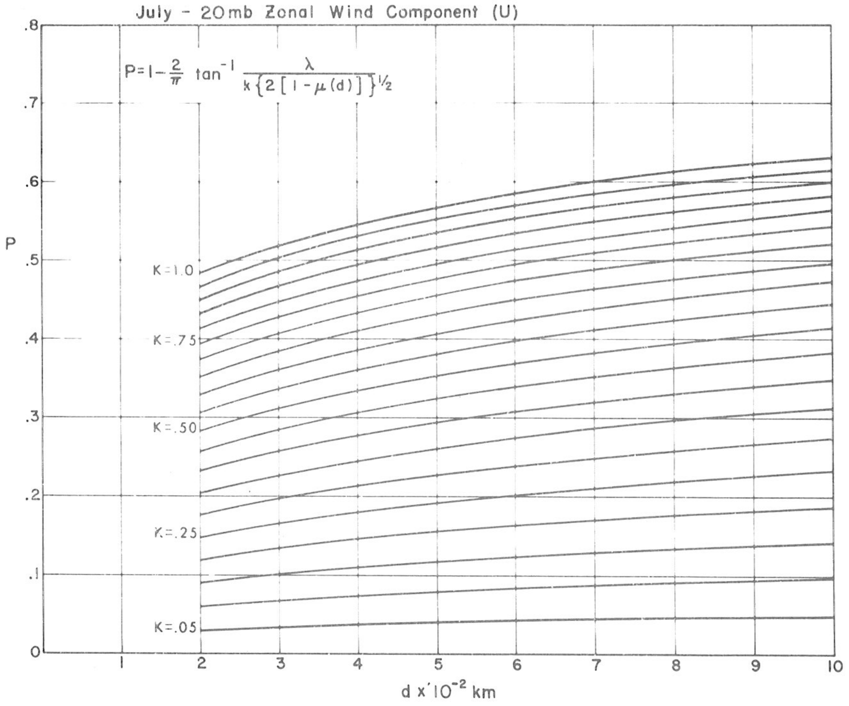


Figure 4--Same as Figure 3, for July

quadrupled. On the other hand, if the current spatial resolution is taken as a point of reference, the frequency of observations will need to be reduced to less than one per day in order to make the temporal and spatial efficiency equivalent.

The lack of balance between the spatial and temporal resolution is even more marked in July.

On the basis of the above analysis, we may draw the following conclusions for current unprocessed rawin observations of the zonal wind component at 20 mb:

1. The random errors associated with these observations are high in comparison with the variability of the wind especially during the summer months.

2. In winter, as typified by the month of January, observations made at intervals of 12 hours are operationally useful despite the comparatively high level of contamination by random errors of observation. The current distribution of high altitude stations in the U. S. is reasonably efficient and provides considerable information even without the benefit of "optimal" analysis. However, in the absence of such analysis a denser network is not recommended unless the magnitude of the random errors can be reduced.

3. In summer, the level of noise is so high that the observations offer little operational advantage over climatology unless they are suitably processed to minimize the effect of the random errors. In the absence of such a scheme, the purpose of observations during summer should be to collect data for a reliable climatology. This can best be achieved not by reducing the number of observing stations but rather by decreasing the frequency of observation, say, to once every two or three days, except at one or two "master" stations where observations continue to be made at 12- or 24-hour intervals or perhaps at more frequent intervals.

## THE INFORMATION CONTENT OF SUITABLY PROCESSED OBSERVATIONS

The conclusions of the previous section are not relevant if the observations are suitably processed. Let us assume that the main purpose of the observational program is to provide a continuous representation of a meteorological field from a number of discrete observations, with the help of some "analysis" technique. The accuracy with which this can be done from any given network of stations, depends on:

a. The temporal and spatial variability of the meteorological field in question;

b. The errors of observations; and

c. The characteristics of the analysis technique used.

Because of the dependence of the results on the analysis procedure and because of the large variety of current and potential analysis techniques, some standardization is necessary for our present purpose. A methodology based on "optimum interpolation" which minimizes RMS errors of interpolation has been proposed by a WMO Working Group on Networks (Bessemoulin, et. al., 1960) as an aid in the rational design of meteorological networks. This methodology has been elaborated by Gandin (1963) and forms the basis for operational numerical analysis in the USSR. Optimum interpolation also underlies numerical analysis schemes proposed in this country by Eddy (1967) and Petersen (1968), among others.

### a. Theory

Irrespective of its merit as a basis for numerical analysis, as compared with other schemes, optimum interpolation has certain properties which commend it for the present study. Among the most convenient of these properties is its ability to provide a measure of the interpolation error from the statistics of the atmosphere, thus obviating the necessity of handling numerous "actual" cases.

Let $\vec{r_i} = \vec{r_1}, \vec{r_2} \dots \vec{r_n}$ denote a set of independent variables defining the location of a point in the sampling space. Consider a function f whose measured values $\hat{f}_i$ at the points $\vec{r_i}$ have an error $\epsilon_i$, so that

$$\hat{f}_i = f_i + \epsilon_i \qquad (4.1)$$

We wish to determine the value $f_o$ at some point $\vec{r_o}$ from the measured values $\hat{f}_i$. If $f'_o$ and $\hat{f}'_i$ denote the deviation of $f_o$ and $\hat{f}_i$ from their mean values, we may express $f_o$ in term of the following linear combination

$$f'_o = \sum_{i=1}^{n} p_i \hat{f}'_i + I_o \qquad (4.2)$$

in which $p_i$ are weighting factors and $I_o$ is the error in determing $f'_o$.

If the errors are random and if, furthermore, the function under consideration is homogeneous and isotropic, it can be shown (Gandin, 1963, Chapter 2; Alaka, 1969) that for minimum mean-square value of the error, $I_o$, the weights $p_i$ may be determined from the following system of simultaneous linear equations.

$$\sum_{j=1}^{n} \mu_{ij} p_j + \frac{\sigma_{\epsilon_i}^2}{\sigma_i^2} p_i = \mu_{oi} \quad (i = 1, \dots n) \qquad (4.3)$$

in which $\sigma_i^2$ is the true variance of the elements at locations $\vec{r_i}$, $\sigma_{\epsilon_i}^2$ is the variance of the random errors of observation, $\mu_{oi}$ denotes the correlation coefficients between the values of the function at the locations $\vec{r_0}$ and $\vec{r_i}$, and $\mu_{ij}$ represents the corresponding correlation coefficients between all possible pairs of locations from which observations are used in the processing.

The mean-square error in determining $f_o$ can be shown to be:

$$\sigma_{I_o}^2 = \overline{I_o^2} = \sigma_o^2 \left( 1 - \sum_{i=1}^{n} \mu_{oi} \, p_i \right) \qquad (4.4)$$

Equation (4.3) reveals another advantage of this procedure; namely, its ability to take into account the effects of random errors. Equation (4.4) shows that, irrespective of the magnitude of these errors, the mean-square error of interpolation ($\sigma_{I_o}^2$) cannot exceed the variance of the element.

## b. Optimal Smoothing

We may enhance the efficiency of a set of observations by application of equations (4.3) and (4.4). Consider a set of regularly spaced observations with RMS random error $\sigma_\epsilon$. Let us attempt to find out to what extent we are able to reduce this error at a station x in the light of observations at the eight nearest stations. Equation (4.3) may be adapted to determine the appropriate weighting factors to be used for optimum results. Figure 5 shows the weights of the different observations of the zonal wind at 20 mb and the variation of their relative magnitudes with station separation for January. As would be expected, the weights are more nearly equal when the observations are close together than when they are far apart. As the spacing of observations increases, less and less weight is given to each of the eight outlying observations. The procedure is equivalent to a smoothing operation which is regulated on the basis of the random errors and of the statistics of the atmosphere.

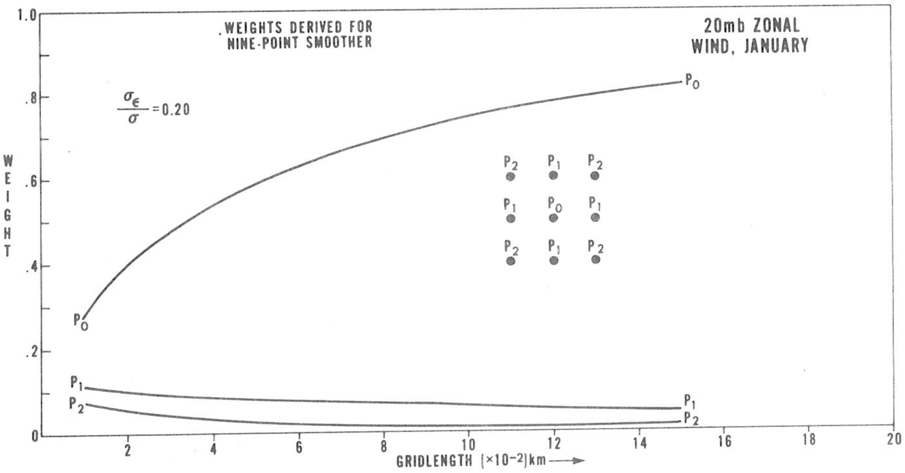
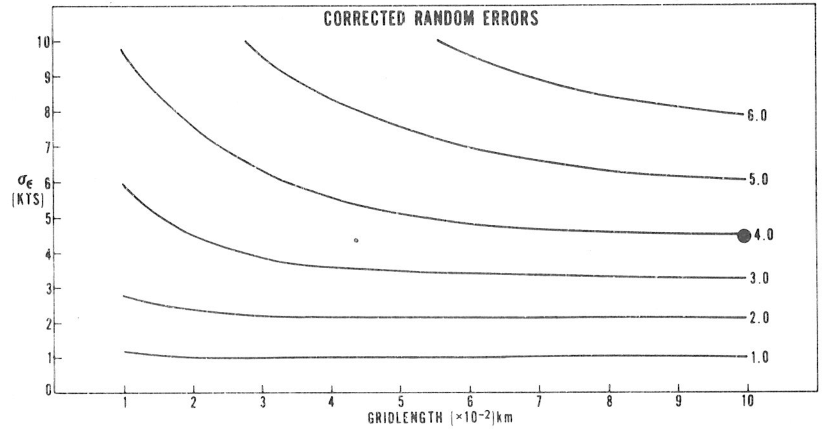


Figure 5--Optimum weights for a nine-point smoother for the 20 mb zonal wind in January, as a function of grid spacing.

Figure 6 shows the corrected RMS errors, in January and July respectively, as a function of the original random errors and of the distance between stations. We note that for a station separation of 1000 km, which is the average distance between current high-altitude stations in the U. S., the above operation reduces $\sigma_\epsilon$ from 4.5 kt to 4.0 kt in January and to about 3.4 kt in July. This improvement is by no means negligible and becomes even greater with decreasing station separation.

(a) 20mb, ZONAL WIND, JANUARY



(b) 20mb, ZONAL WIND, JULY

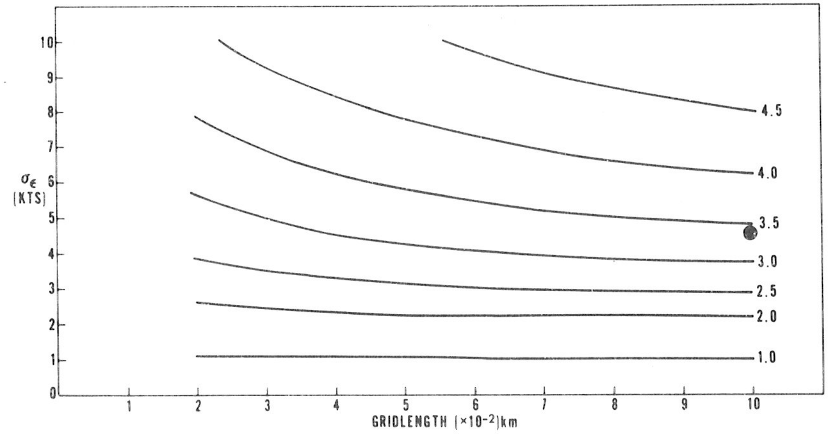


Figure 6--RMS random errors of the 20 mb zonal wind in (a) January and (b) July corrected by an optimum nine-point smoother, as a function of the original random errors and of grid spacing. The solid dot approximately represents the current situation over the United States.

The improved efficiency of the processed over the unprocessed observations may be appreciated by comparison of the appropriate curves in Figure 2. The improvement in July is particularly remarkable; it ranges from about 43 percent for a station separation of 1000 km to

200 percent when the spacing is reduced to 400 km. This is a good example of how otherwise useless observations may be redeemed by proper processing.

### c. Optimum Interpolation

Equations (4.3) and (4.4) also may be used to determine the accuracy with which the value of a meteorological field may be determined at points between observations. Let us suppose that we wish to determine the value of a field at a point 0 by optimum interpolation from the nearest 12 observations as shown in the upper left-hand corner of Figure 7. By symmetry, the observations are given one of two weights: $p_1$ for the four observations nearest the point 0, and $p_2$ for the eight outlying observations.

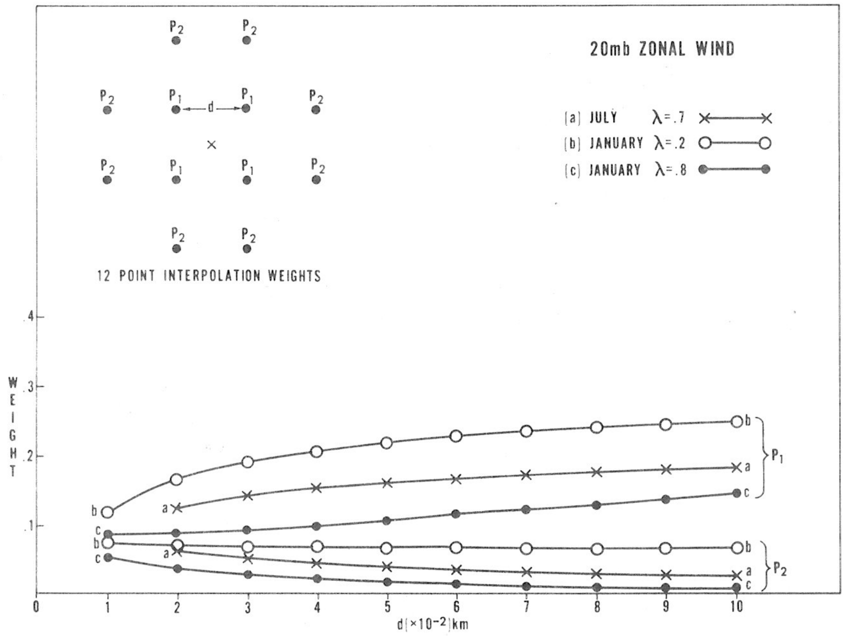
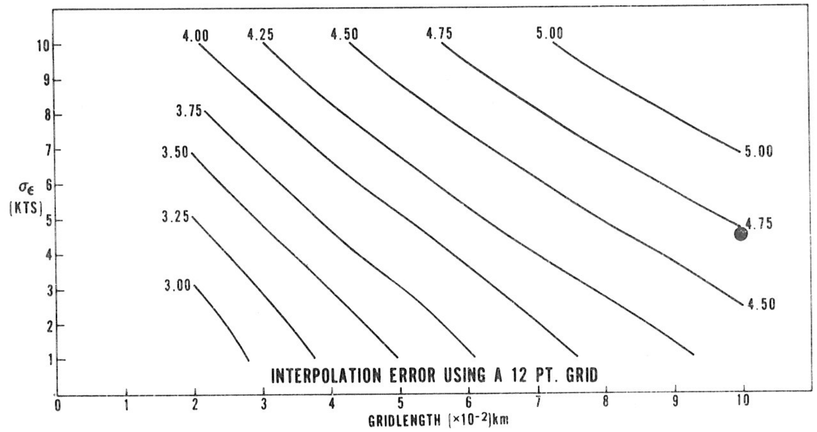


Figure 7--Optimum weights for interpolating the 20 mb **zonal wind** from 12 grid-point observations to a central point, as a function of grid spacing.

Figure 7 shows how these weights vary with the spacing of observations and with their accuracy. In particular, when the observations are relatively accurate ($\lambda = .2$), much greater weight is given to the central ring of observations than to the outlying observations, especially if the observations are widely spaced. On the other hand, if the random errors are comparatively large ($\lambda = .8$), $p_1$ and $p_2$ are more nearly equal. Figure 7 also shows the variation of the relative importance of $p_1$ and $p_2$ with season for the same value of $\sigma_\epsilon$. In January, when the variability is large, $p_1$ assumes greater relative importance than in July.

Figure 8 shows the variation of the RMS errors of interpolation, $\sigma_{I_0}$, at point O in January and July respectively, as a function of distance d between observations and of $\sigma_\epsilon$. It is seen that $\sigma_{I_0}$ may be improved by reducing d. Indeed, in July if the station separation is 850 km, $\sigma_{I_0}$ becomes equal to $\sigma_\epsilon$, i.e., the RMS error in determining the value of $u$ at the center of a square with sides 850 km long, and with the nearest observations at the corners of the square, is the same as the RMS error of the observations themselves. In January, the stations must be about 350 km apart to achieve the same order of accuracy.

(a) 20mb ZONAL WIND, JULY



(b) 20mb ZONAL WIND, JANUARY

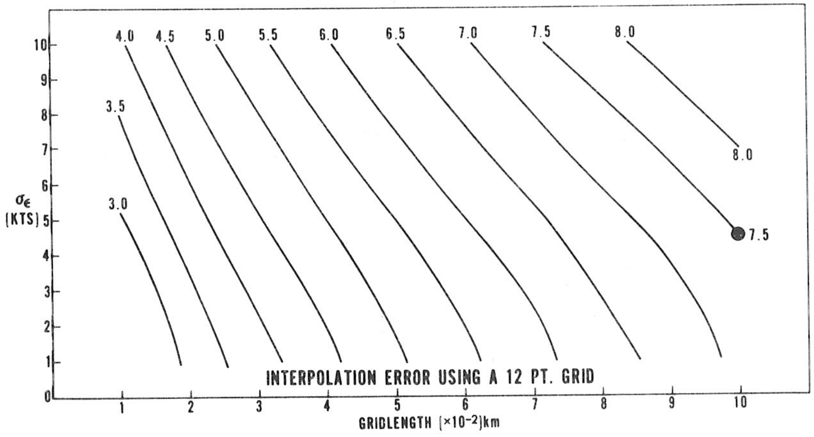


Figure 8--The RMS interpolation error, $\sigma_{I_0}$, of the 20 mb zonal wind in (a) January and (b) July as a function of the random error, $\sigma_\epsilon$, and the grid spacing, d, based on optimum interpolation from a 12-point grid. (The solid dot approximately represents the current situation over the United States.)

Figure 8 indicates that, for a given station separation, the RMS errors of interpolation ($\sigma_{I_0}$) are larger in January than in July. Thus, if the criterion in designing a network is to achieve a certain acceptable level of $\sigma_{I_0}$, a denser network would be required in winter than in summer. It must, however, be realized that the error in January is a much smaller fraction of the standard deviation of the element than in summer. If the quantity $Q = \frac{\sigma_{I_0}}{\sigma}$ is taken as a criterion, than a less dense network would be indicated for January (Figure 9).

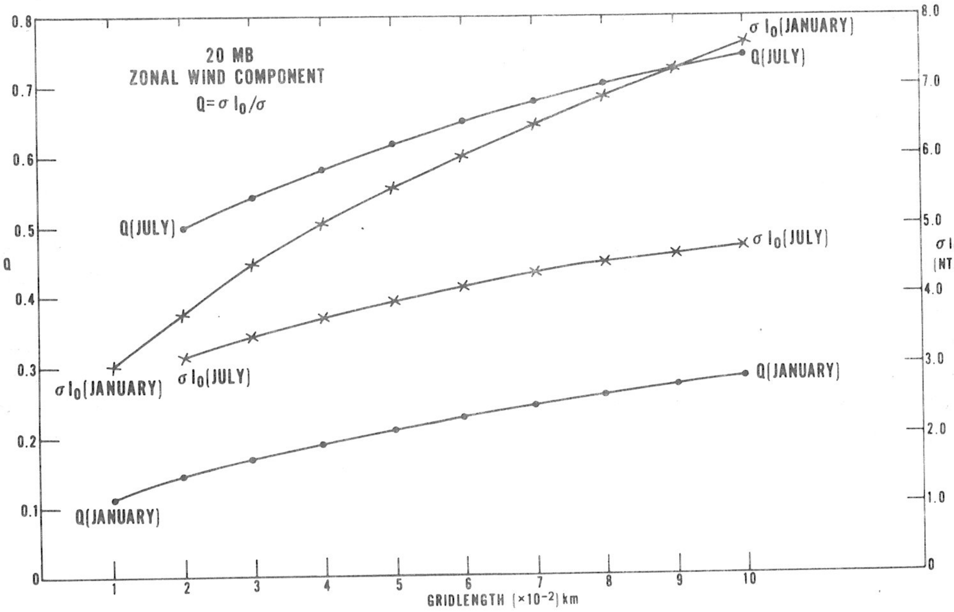


Figure 9--The variation with grid spacing of the RMS interpolation error, $\sigma_{I_0}$, of $Q = \frac{\sigma_{I_0}}{\sigma}$ of the 20 mb zonal wind in January.

Returning to Figure 8, we note that although $\sigma_{I_0}$ may be improved by reducing the distance between stations (d), the gain is comparatively small unless there is a corresponding reduction in $\sigma_\epsilon$. This is especially true in July when only a reduction of 1 kt in $\sigma_{I_0}$ can be achieved by reducing d from 1000 to 600 km, i. e., by almost tripling the number of stations!

The most rapid decrease of $\sigma_{I_0}$ may be achieved by reducing both the random errors and the station spacing in a combination which depends on the slopes of the isolines of $\sigma_{I_0}$ in Figure 8. Thus, in January, the most rapid improvement from present conditions, represented by the solid dot in Figure 8 may be achieved if $\sigma_\epsilon$ were reduced by 1 kt for every 70 km reduction in the spacing between observations. In July, the optimum combination is 1 kt to about 30 km.

## THE INFORMATION CONTENT OF OLD OBSERVATIONS

In view of the slow time variation of meteorological conditions at 20 mb, an attempt was made to determine the usefulness of observations 12 and 24 hours old as a supplement to current observations. The scheme followed was to alternate current observations with older observations in the manner shown in Figure 9. The older observations were, of course, given a smaller weight which was determined by considering the variance of the 12- or 24-hour changes as an additional source of random errors which was added to the variance of the random errors $\sigma_\epsilon^2$. Thus, in computing $\lambda$ in equation (4.3) an adjusted random error

$$\sigma_\epsilon' = (\sigma_\epsilon^2 + \sigma_\tau^2)^{1/2} \qquad (5.1)$$

was used for the older observations. Otherwise, they were treated as current observations. The relative weights which this procedure gives current and 24-hour old observations in July, and the variation of these weights with station separation are shown in Figure 10. The corresponding values for January are very similar. Note that when the station separation is large, old observations in the central ring are given more weight than current observations in the outer ring.

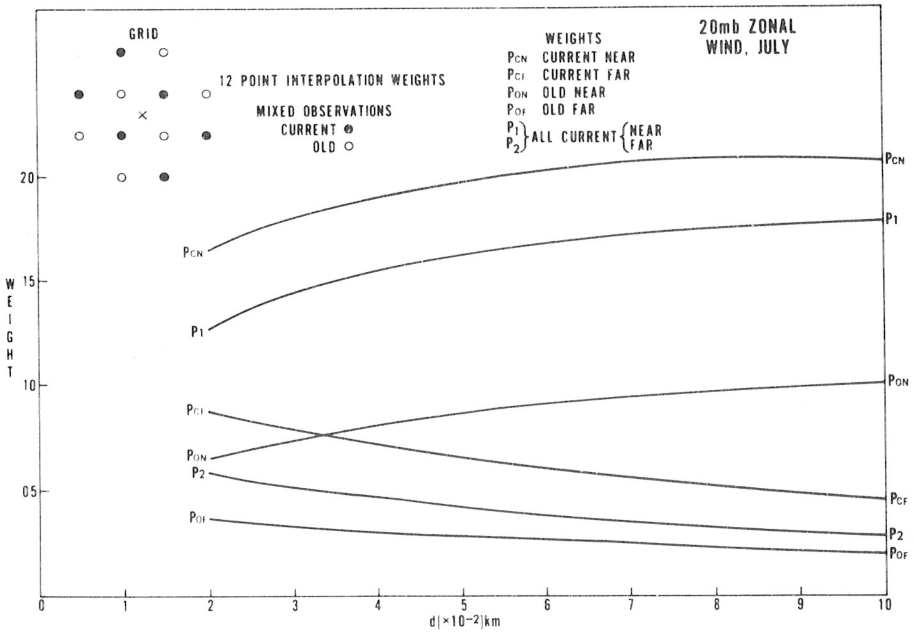


Figure 10--Optimum weights for a system of current and 24 hours old observations of 20 mb zonal wind in July.

Figure 11 compares the RMS error in determining the value of **u** at a central point from 12 outlying observations, when all the observations are current and when every other observation is 24 hours old. We note that little accuracy is lost by mixing old with new observations. In July, the difference in accuracy is almost negligible. From Figure 11, we note that in January the accuracy achieved by a network of current observations 1000 km apart is the same as that achieved by a network of mixed observations 830 km apart. This means that for every 70 current observations we require about 100 mixed observations of which only 50 are new. Thus, by using observations 24 hours old, we may reduce the number of daily observations by about 30 percent without sacrificing any accuracy. The percent savings in July are of the same order.

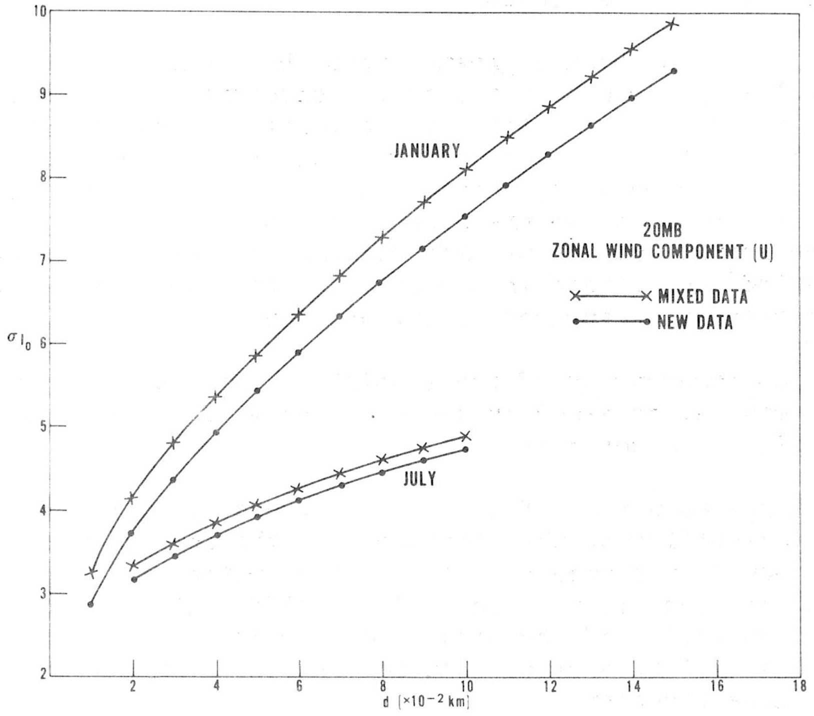


Figure 11--Comparison of RMS interpolation errors from current observations and from a mixture of current and 24-hour old observations, as a function of grid spacing.

## SOME CRITERIA FOR MEANINGFUL DATA ACQUISITION PROGRAMS

The above sections point to the usefulness of compiling a reliable high-altitude climatology both for direct daily operational use in lieu of unprocessed daily observations, particularly in summer, and for effective processing of these observations to maximize their usefulness. In particular, accurate values of the variance and space- and time-correlations of meteorological elements would be of extreme usefulness. In connection with the present study, an attempt was made to compute representative values of the parameters from existing data records. Difficulties were experienced because of:

a. The effect of the random errors in inflating the variances and the structure function and, hence, underestimating the correlation function.

b. The difficulty of correcting the variances and structure functions in the manner suggested by Gandin (1963) due to the lack of sufficient data from station pairs which are close together.

c. The marked non-homogeneity of the variances and the anisotropy of the covariances.

The possiblity has been suggested that the requisite data for a reliable climatology could be obtained by concentrating currently available resources and facilities over a relatively small area where a high resolution data acquisition program would be mounted for a relatively short period. In our opinion, such a project would be of limited value. In rendering this opinion, we are especially influenced by the serious effects on the results of the two key constraints; namely, "small area" and "short period" underlined in the previous sentence. We would advance the following reasons in support of our opinion:

a. The non-homogeneity of the statistical fields would make any results, obtained in one small area, of limited applicability in other areas.

b. The large magnitude of the random errors in comparison with the variability of the atmosphere at the levels under consideration would introduce errors in the computed variances and structure functions which would be difficult to correct unless the computation of the structure function includes data, over a reasonably long period, from several pairs of stations which are close together.

c. To avoid the distorting effect of truncated records, the data acquisition program should contain at least one cycle of the lower frequency--which contains appreciable energy. Thus, in the lower latitudes, the record should be long enough to accommodate the biennial cycle.

d. Aside from the above effect, the accuracy of a computed statistical function depends on the mass of data used in the computation. For instance, the RMS error of a correlation coefficient ($\sigma_\mu$) is given by

$$\sigma_\mu = \frac{1 - \mu^2}{\sqrt{n}} \qquad (6.1)$$

where $\mu$ is the value of the correlation coefficient and $n$ is the number of situations used in the computation. Equation (6.1) holds provided successive values of the quantities to be correlated are independent, i.e., about two-three days apart.

If m correlation coefficients are available to determine the correlation function over a given range (say, from $\mu = 1.0$ to 0.0) and the meteorological field is homogeneous and isotropic, then the RMS error in determining the correlation function over that range is

$$\sigma_{\bar{\mu}} = \frac{\sigma_{\mu}}{\sqrt{m}} = \frac{1-\mu^2}{\sqrt{N}} \qquad (6.2)$$

where N = m x n.

Table 3 (Gandin, Maskovich, Alaka, and Lewis, 1967) shows values of N for given values of $\mu$ and $\sigma_{\bar{\mu}}$.

Table 3--Values of N corresponding to different values of error and of the correlation coefficient.

| $\mu$ | | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_{\bar{\mu}}$ | 0.01 | 360 | 1,300 | 2,600 | 4,100 | 5,600 | 7,100 | 8,300 | 9,200 | 9,800 | 10,000 |
| | 0.02 | 90 | 300 | 650 | 1,023 | 1,400 | 1,800 | 2,100 | 2,300 | 2,400 | 2,500 |
| | 0.05 | 16 | 53 | 105 | 165 | 225 | 280 | 330 | 370 | 390 | 400 |

To appreciate the implications of Table 3, let us suppose that we have records for 60 independent situations, i.e., records of 120 to 180 days. To obtain the correlation function over the range 1.0 to 0.0 with a RMS error of 0.01, $360 \div 60 = 6$ individual correlation coefficients would be required for the interval near $\mu = 0.9$, $1300 \div 60 \approx 22$ in the interval near $\mu = 0.8$, etc. Assuming that 10 intervals in Table 3 will adequately describe the correlation function wanted, about 1,000 individual correlation coefficients (station pairs), i.e., about 45 stations would be needed.[3] If only 20 stations were available, 300 independent situations, or a record of 600-900 days in length would be required for comparable accuracy.

---

[3] The number of station pairs from a total of N stations is $\frac{N(N-1)}{2}$.

For time-lag correlations a mass of data $nxm^2 = 10^5$ would be adequate. Thus, if only one station is available (n=1) for computing such correlations and if four observations are made daily at this station, about 80 days of record would be required. If two stations are available and if these stations are far enough apart to produce independent records, and if the field is homogeneous, about 60 days would be sufficient.

## TWO POSSIBLE DATA ACQUISITION SCHEMES

We shall now apply the requirements detailed in the previous section to two possible data acquisition schemes:

### Scheme A

This scheme would involve essentially the same currently operating high-altitude stations plus one or two high-resolution "windows" as shown in Figure 12; the total would be about 25 stations. For the purpose of computing a representative space correlation function, it would be necessary and sufficient if observations were made at each of these stations once every three days for a period of 600 days.

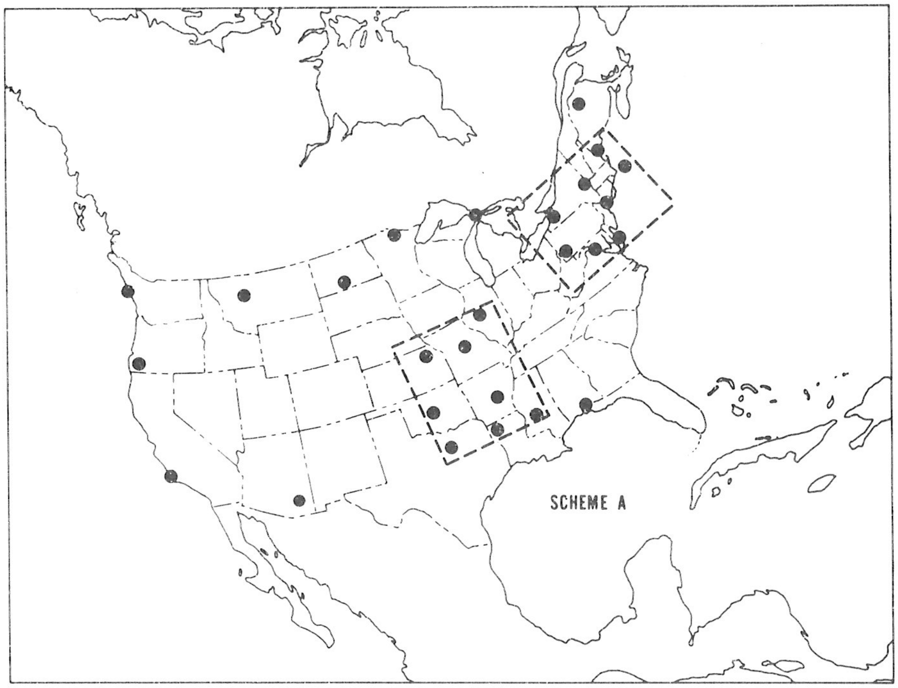


Figure 12--Scheme A for determining spatial and temporal statistics of meteorological fields at 20 mb.

For the purpose of time-lag correlations we would need two of the above stations, located far enough apart, say, one on the East Coast and the other on the West Coast, to carry out four or more observations per day, for periods of (two months) at a time, twice a year, throughout the period of the data-gathering program.

The above scheme assumes that the statistical fields are both isotropic and homogeneous. Since this assumption is not valid in general, the computations will result in a wide scatter of points which have to be averaged to obtain the required function.

Scheme B

Perhaps a better distribution of stations for climatological purposes is the one shown in Figure 13. The quasi-linear disposition of the stations is suitable for computing correlations in the meridional and zonal directions and does not, therefore, assume isotropy. However, this scheme requires a longer data-acquisition period. If the number of stations along each direction is 10, about 1,000 observations, taken three days apart, would be required at each station.

The requirements for the time-lag correlation would be the same as in Scheme A.

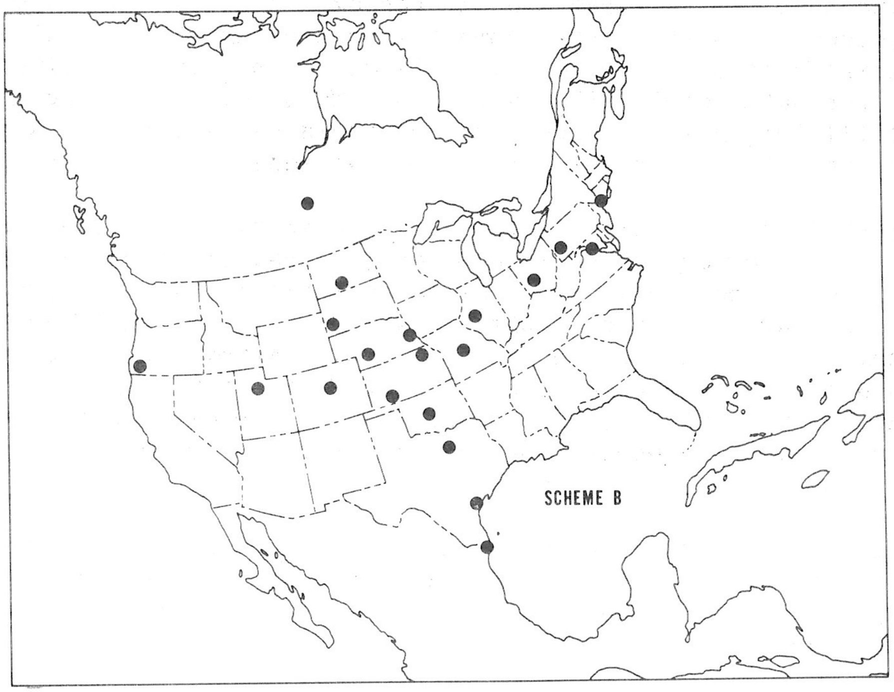


Figure 13--Scheme B for determining spatial and temporal statistics of meteorological fields at 20 mb.

## THE GAP BETWEEN PLANNING AND EXECUTION

The above discussion presupposes the ability to make high-altitude observations at will. That this is far from being the case is illustrated by Table 4, which shows the percentage of Rawins from high-altitude stations reaching 30 mb and 10 mb during 1968. The serious consequences of the failure of high-level observations may be appreciated from the fact that each failed observation reduces the computed covariances by the number of observations with which it can be paired. Thus, if there are 25 stations and one station fails to make an observation on a given day, the number of covariances available for computation would be reduced by 24.

Because of the large percentage of failures, the actual calendar periods required for the proper implementation of Schemes A and B would have to be much longer than the theoretical period postulated. The statistics which would enable determination of the actual calendar periods are not readily available. But it can be seen from Table 4 that for 10 mb these periods will have to be a large multiple of the theoretical periods which assume no failures.

However, the above consideration may not be the most serious obstacle to the compilation of a reliable high-altitude climatology. Table 4 shows that the failures are not random but differ with the seasons. Indeed it is known that more failures occur with low stratospheric temperatures and high winds. This means that irrespective of its length, the data sample will be biased. To obtain an unbiased sample, a still longer initial data record would be needed from which a sufficient density of representative observations could be extracted.

The inevitable conclusion is that so long as high-altitude observations continue to be made by balloon-borne sensors, it is extremely difficult, if not impossible, to obtain a sufficient mass of representative data to compute reliable statistics at 10 mb. This fact was forcefully brought to the authors when they initially attempted to compute correlation and structure functions at 10 mb from a ten-year record. The large number of missing observations reduced the data below the critical mass and, as a result, the computed points formed an amorphous cloud without any discernible trend.

At 20 mb the situation was much better. But even here, appreciable difficulty was encountered. This is the reason why the analysis in this study is confined to the zonal wind component for which the computed statistics appeared reasonable.

No doubt, at lower levels, the prospects of acquiring suitable data records become better as the probability of successful ascents increases.

Table 4--Percent of Rawins Reaching Indicated Level at Weather Bureau Hi-Altitude Stations.

| STATION | Jan-Mar '69 | | Apr-Jun '68 | | Jul-Sept '68 | | Oct-Dec '68 | |
|---|---|---|---|---|---|---|---|---|
| | 10 mb | 30 mb | 10 mb | 30 mb | 10 mb | 30 mb | 10 mb | 30 mb |
| Bismarck | 41.7 | 89.4 | 44.5 | 88.5 | 70.9 | 91.4 | 45.1 | 94.6 |
| Brownsville | 48.0 | 77.7 | 53.3 | 85.2 | 58.2 | 84.2 | 53.3 | 79.9 |
| Great Falls | 44.7 | 86.6 | 55.6 | 94.4 | 60.3 | 94.6 | 46.4 | 86.3 |
| International Falls | 58.1 | 93.3 | 65.9 | 95.6 | 55.5 | 91.2 | 54.4 | 95.6 |
| Medford | 52.2 | 83.3 | 35.9 | 87.8 | 65.0 | 89.1 | 50.3 | 86.9 |
| Nantucket | 35.0 | 79.7 | 34.5 | 87.6 | 38.8 | 79.2 | 36.2 | 79.1 |
| Peoria | 46.7 | 87.2 | 63.2 | 94.5 | 73.9 | 92.9 | 59.8 | 90.8 |
| Pittsburgh | 65.0 | 86.1 | 80.2 | 95.6 | 72.3 | 91.3 | 67.9 | 82.1 |
| Quillayute | 51.1 | 87.2 | 38.5 | 83.5 | 62.5 | 96.2 | 41.3 | 80.4 |
| Sault Ste. Marie | 62.2 | 95.0 | 44.0 | 91.8 | 50.0 | 92.9 | 51.1 | 92.9 |
| Shreveport | 49.4 | 73.9 | 58.8 | 94.5 | 59.3 | 89.6 | 59.8 | 85.3 |
| Tucson | 54.4 | 84.4 | 45.6 | 91.2 | 53.3 | 94.0 | 65.6 | 94.5 |
| Vandenberg AFB | 64.0 | 85.1 | 81.9 | 93.2 | 82.3 | 93.9 | 82.5 | 91.0 |

## ACKNOWLEDGMENTS

The authors acknowledge their indebtedness to Frank Lewis and Douglas Greene for their assistance in computing the statistics used in this study, and to Roy E. Wyett who contributed Table 4. The authors also extend their thanks to Willard V. Compton for assistance in drafting, and to Miss Catherine Ryan who typed the manuscript.

# REFERENCES

Alaka, M. A., "A Survey of Studies of Aerological Network Requirements," ESSA Technical Memorandum WBTM TDL-9, U. S. Weather Bureau, Silver Spring, Maryland, 1967, 20 pp.

Alaka, M. A., "Theoretical and Practical Considerations for Network Design," To be published in the Proceedings of the AMS Symposium on Meteorological Observations and Instrumentation, 1969.

Bessemoulin, J., et. al., "Contribution to the Study of Meteorological Networks," WMO Technical Note No. 30, Annex 9, World Meteorological Organization, Geneva, Switzerland, 1960, pp. 62-76.

Bessemoulin, J., et. al., "Preliminary Report on Research of the Working Group of the Commission on Synoptic Meteorology," WMO Technical Note No. 30, World Meteorological Organization, Geneva, Switzerland, 1960, 91 pp.

Colson, D., Greene, D., and Lewis, F., "Variability Studies of Temperature and Wind at SST Levels," Unpublished manuscript, Techniques Development Laboratory, ESSA, Weather Bureau, 1968.

Gandin, L. S., "Objective Analysis of Meteorological Fields," Hydrometeorological Publishing House, Leningrad, Translated into English Israel Program for Scientific Translations, 1963, 242 pp.

Gandin, L. S., Mashkovich, S. A., Alaka, M. A., and Lewis, F., "Design of Optimum Networks for Aerological Observing Stations," World Weather Watch Planning Report No. 21, World Meteorological Organization, Geneva, Switzerland, 1967, 58 pp.

Petersen, D. P., "On the Concept and Implementation of Sequential Analysis for Linear Random Fields," Tellus, 20, 1968, pp. 674-686.

(Continued from inside front cover)

WBTM TDL 16 Objective Visibility Forecasting Techniques Based on Surface and Tower Observations. Donald M. Gales, October 1968. (PB-180 479)

WBTM TDL 17 Second Interim Report on Sea and Swell Forecasting. N. A. Pore and Lt. W. S. Richardson, USESSA, January 1969. (PB-182 273)

WBTM TDL 18 Conditional Probabilities of Precipitation Amounts in the Conterminous United States. Donald L. Jorgensen, William H. Klein, and Charles F. Roberts, March 1969. (PB-183 144)

WBTM TDL 19 An Operationally Oriented Small-Scale 500-Millibar Height Analysis. Harry R. Glahn and George W. Hollenbaugh, March 1969. (PB-184 111)

WBTM TDL 20 A Comparison of Two Methods of Reducing Truncation Error. Robert J. Bermowitz, May 1969. (PB-184 741)

WBTM TDL 21 Automatic Decoding of Hourly Weather Reports. George W. Hollenbaugh, Harry R. Glahn, and Dale A. Lowry, July 1969. (PB-185 806)

WBTM TDL 22 An Operationally Oriented Objective Analysis Program. Harry R. Glahn, George W. Hollenbaugh, and Dale A. Lowry, July 1969. (PB-186 129)

WBTM TDL 23 An Operational Subsynoptic Advection Model. Harry R. Glahn, Dale A. Lowry, and George W. Hollenbaugh, July 1969. (PB-186 389)

WBTM TDL 24 A Lake Erie Storm Surge Forecasting Technique. William S. Richardson and N. Arthur Pore, August 1969. (PB-185 778)

WBTM TDL 25 Charts Giving Station Precipitation in the Plateau States From 850- and 500-Millibar Lows During Winter. August F. Korte, Donald L. Jorgensen, and William H. Klein, September 1969. (PB-187 476)

WBTM TDL 26 Computer Forecasts of Maximum and Minimum Surface Temperatures. William H. Klein, Frank Lewis, and George P. Casely, October 1969. (PB-189 105)

WBTM TDL 27 An Operational Method for Objectively Forecasting Probability of Precipitation. Harry R. Glahn and Dale A. Lowry, October 1969. (PB-188 660)

WBTM TDL 28 Techniques for Forecasting Low Water Occurrences at Baltimore and Norfolk. Lt. (jg) James M. McClelland, USESSA, March 1970. (PB-191 744)

WBTM TDL 29 A Method for Predicting Surface Winds. Harry R. Glahn, March 1970. (PB-191 745)

WBTM TDL 30 Summary of Selected Reference Material on the Oceanographic Phenomena of Tides, Storm Surges, Waves, and Breakers. Arthur N. Pore, May 1970. (PB-193 449)

WBTM TDL 31 Persistence of Precipitation at 108 Cities in the Conterminous United States. Donald L. Jorgensen and William H. Klein, May 1970. (PB-193 599)

WBTM TDL 32 Computer-Produced Worded Forecasts. Harry R. Glahn, June 1970.

WBTM TDL 33 Calculation of Precipitable Water. L. P. Harrison, June 1970. (PB-193 600)

WBTM TDL 34 An Objective Method for Forecasting Winds Over Lake Erie and Lake Ontario. Celso S. Barrientos, August 1970.

WBTM TDL 35 A Probabilistic Prediction in Meteorology: A Bibliography. A. H. Murphy and R. A. Allen, June 1970.