A Technique for Maximizing Details in Numerical Weather Map Analysis[1]

STANLEY L. BARNES

Atmospheric Research Laboratory, University of Oklahoma Research Institute, Norman, Okla.

(Manuscript received 13 January 1964, in revised form 16 April 1964)

ABSTRACT

This paper summarizes the development of a convergent weighted-averaging interpolation scheme which can be used to obtain any desired amount of detail in the analysis of a set of randomly spaced data. The scheme is based on the supposition that the two-dimensional distribution of an atmospheric variable can be represented by the summation of an infinite number of independent waves, i.e., a Fourier integral representation. The practical limitations of the scheme are that the data distribution be reasonably uniform and that the data be accurate. However, the effect of inaccuracies can be controlled by stopping the convergence scheme before the data errors are greatly amplified. The scheme has been tested in the analysis of 500-mb height data over the United States producing a result with details comparable to those obtainable by careful manual analysis. A test analysis of sea level pressure based on the data obtained at only the upper air network stations produced results with essentially the same features as the analysis produced at the National Meteorological Center. Further tests based on a regional sampling of stations reporting airways data demonstrate the applicability of the scheme to mesoscale wavelengths.

1. Introduction

In recent years there has been considerable effort put forth toward improving the interpolation phase of objective analysis schemes so that more detail can be represented in the analyses. The type of scheme which has been most successful at doing this is a surface-fitting scheme, that is, the method of fitting a geometrical surface to the reported data and calculating the values determined by that surface at any other points of interest, specifically, the grid points. The works of Dellert,[2] of Pfeffer *et al.*,[3] and of Penn, Kunkel and Mount (1963) are all based on that method. It is the author's opinion that such methods suffer from three major disadvantages: the calculations are complicated and require considerable time to complete; the data to which the surface is fitted are chosen in a rather artificial manner (that which produces the best results); the effect of erroneous data can be disastrous since each datum is given equal ranking in determining the shape of the surface.

In order to avoid the undesirable effects which erroneous data produce, a smoothing process has been recommended[4] employing a least-squares fit of the surface to the data with the influence of each datum weighted according to its distance from the grid point. This latter method is actually a combination of a surface-fitting method and the weighted-averaging method of interpolation.

The weighted-averaging method determines the values of the variable at grid points as the sum of weighted values of the individual data. The closer a data point to the grid point in question, the greater influence the datum at that point exerts. Such schemes, variously refined, are now being used for determining the broad features of atmospheric distributions (Bergthórrson and Döös, 1955; Cressman, 1959; McDonell[5]). The major disadvantage of such schemes has been their tendency to smooth out all small variations in the field, whether caused by data errors or actual atmospheric disturbances.[6] This paper describes a technique for re-

[1] The research in this report has been sponsored by the National Science Foundation under Grant NSF G-19688.

[2] Dellert, G. T., Jr., 1962: The triangle method—an objective analysis technique. Unpublished manuscript of the U. S. Weather Bureau, presented at the 43rd Annual Meeting of the American Meteorological Society in New York City, 21–24 January 1963, 18 pp.

[3] Pfeffer, R. L., *et al.*, 1963: Objective analysis by polynomial fitting. Unpublished manuscript of the Lamont Geological Laboratory, Columbia University, presented at the 43rd Annual Meeting of the American Meteorological Society in New York City, 21–24 January 1963, 13 pp.

[4] Pfeffer *et al.*, *ibid.*

[5] McDonell, J. E., 1962: On the objective analysis system used at the National Meteorological Center. Tech. Memo. No. 23, U. S. Weather Bureau, 31 pp.

[6] Cressman (1959) has succeeded in depicting smaller-scale components by the application of weight factors using successively smaller influence radii which are expressed in terms of grid length. The detail of the analysis is, therefore, limited by one's choice of grid length rather than by the data distribution. Also, a smoothing operator (Shuman, 1957) is applied to eliminate wavelengths of two grid units or less. Although this technique may be adequate for large-scale analyses, it does not insure the retention of the smallest components which can be adequately described from a given set of data.

gaining the details lost in applying such a weighted-averaging scheme.

The weight factor used in this objective analysis has been developed from the fundamental premise that the two-dimensional distribution of an atmospheric variable can be represented by the summation of an infinite number of *independent harmonic waves*, that is, by a Fourier integral representation.[7] In its current form,[8,9] the weight factor is explicitly related to the density of observations which, after all, determines the ultimate resolution obtainable in any strictly scalar analysis. Because of the data density dependency, applications of the scheme are restricted to regions in which the distribution of data is reasonably uniform, although this does not exclude the possibility that means can be found to tie together the analyses in regions of vastly differing data densities. As will be shown, the technique of regaining lost details is also consistent with the fundamental premise.

2. Development of the interpolation scheme

Under the assumption that the distribution of an atmospheric quantity, $f(x,y)$, can be depicted by a Fourier integral representation, we may define a corresponding smoothed function, $g(x,y)$, which is obtained by applying a filter to the original function as follows:

$$g(x,y)=\int_0^{2\pi}\int_0^{\infty} f(x+r\cos\theta,\ y+r\sin\theta)wrdrd\theta, \quad (1)$$

where the weight factor, or filter, is

$$w=(1/4\pi k)\exp(-r^2/4k), \quad (2)$$

with r and θ as polar coordinates, the origin being at the point (x,y), and k being a parameter determining the shape of the weight factor which is to be related to the density of the observed data concerning $f(x,y)$.

Rearranging (1), we may express the weight factor in an alternate form:

$$g(x,y)=\int_0^{2\pi}\int_0^{\infty} f(x+r\cos\theta,\ y+r\sin\theta) \times(\eta/2\pi)d(r^2/4k)d\theta, \quad (3)$$

where

$$\eta=\exp(-r^2/4k) \quad (4)$$

is the new weight factor. This form (3) is preferred over the form given by (1), since in (1) the maximum weight is not applied at $r=0$.

Interpolation by (3) is not practical for two reasons: first, we don't know the analytical form of $f(x,y)$ (indeed, that is what we are trying to represent based on a few random pieces of information), and second, we cannot integrate the function to infinity. Therefore, we must approximate $g(x,y)$ by placing a finite limit on the region of influence of any datum concerning $f(x,y)$ and by taking a weighted average of only those M number of data within that region. In a practical form,

$$g(x,y)=\frac{\sum_j^M \eta(r_j)\cdot f_j}{\sum_j^M \eta(r_j)}. \quad (5)$$

It is obvious from (4) that

$$\int_0^{2\pi}\int_0^{\infty}(1/2\pi)\eta d(r^2/4k)d\theta=1.$$

Integrating with respect to θ and then to some distance R, we may write the above integral

$$\int_0^R \eta d(r^2/4k)+\int_R^{\infty}\eta d(r^2/4k)=1.$$

We define the second of these integrals as ϵ. Thus,

$$\int_0^R \exp(-r^2/4k)d(r^2/4k)=1-\epsilon.$$

Carrying out the integration, we see that

$$\exp(-R^2/4k)=\epsilon$$

or

$$R^2/4k=-\ln\epsilon\equiv E. \quad (6)$$

If ϵ is small enough, we can represent the weighted influence of any datum with sufficient accuracy. For example, choosing $E=4$ means we have represented 98 per cent of the influence of any datum within the circular region whose radius is R. We define R as the "radius of influence" of the weight factor η. The relationship (6) will be used later to form a link between the data distribution and the choice of weight factor.

Since the component waves which make up the distribution $f(x,y)$ are independently related, we shall in the following refer to only a single harmonic wave, keeping in mind that the discussion applies equally well to all component waves present in $f(x,y)$. To further simplify the discussion, we shall consider only a one-dimensional wave,

$$f(x)=A\sin ax \quad (7)$$

---

[7] Sasaki, Y., 1960: An objective analysis for determining initial conditions for the primitive equations. Tech. Rep. (Ref. 60–16T), Dept. of Oceanography and Meteorology, Texas A & M University, 22 pp.

[8] Barnes, S. L., 1960: An objective analysis of a pre-frontal squall line. Tech. Rep. (Ref. 60–17T), Dept. of Oceanography and Meteorology, Texas A & M University, 44 pp.

[9] Barnes, S. L., 1962: An improved interpolation technique for numerical weather map analysis. Tech. Rep. (Ref. ARL–1296–3), Atmospheric Research Laboratory, University of Oklahoma Research Institute, 14 pp.

where $a=\pi/L$, L being the half wavelength. Substitution of (7) into (3) yields

$$g(x)=D(A\sin ax)=Df(x) \tag{8}$$

where

$$D=\exp(-ka^2), \tag{9}$$

D being the "accuracy index" relating to the fractional amount of the original wave amplitude retained in the smoothed interpolated function $g(x)$.

In (9), D is a function of two lengths, k, which is related to the weight factor and determines its shape, and a, which is related to the wavelength of the disturbance. It is preferable to express those two lengths, and hence the accuracy index D, in terms of the one length which limits the detail and accuracy of the interpolated result, that is, the distance between data points. This can be done by first substituting (6) and the definition of a into (9), the result being

$$D=\exp[-(\pi^2/4E)\ (R/L)^2], \tag{10}$$

and then dividing the numerator and denominator in the exponent by $(\bar{d})^2$ so that

$$D=\exp[-(\pi^2/4E)\ [(R/\bar{d})/(L/\bar{d})]^2]. \tag{11}$$

The parameter $\bar{d}$ is defined as the average distance between data points, a necessity arising from the fact that our observation sites are not uniformly spaced. In the following discussion, $\bar{d}$ is used in two ways, 1) meaning the average for the whole set of observation sites, and 2) meaning the average for the group of sites immediately surrounding any given site.

Eq (11) is presented in graphical form in Fig. 1 with $E=4$. Note that the value of the ordinate, $R/\bar{d}$, is dependent upon the choice of weight factor. The region between 0 and 1 is not shown, since if the weight factor has a region of influence which does not include at least two data points, the resultant analysis will have first order discontinuities, possibly even zero order. In general, R should always be chosen greater than $\bar{d}$; $(R/\bar{d}>1)$.

Along the abscissa, waves which fall in the region $0\leq L/\bar{d}<1$ should be ignored since it is impossible to describe any feature of such short waves. Analysis of waves whose dimensions are such that $L/\bar{d}=1$ is possible only when the data points lie at the maximum and minimum points of the wave. It is important to note that random errors in the observed data generate fictitious waves of dimension $L/\bar{d}=1$ (using the second meaning of $\bar{d}$). In general then, waves whose dimensions are such that $0\leq L/\bar{d}\leq 1$ should be filtered from the analysis as much as possible. We can be reasonably confident that waves of dimensions $L/\bar{d}>2$ are accurately described by the interpolated result, except that they suffer a loss in the magnitude of their amplitude. For example, suppose we have chosen a weight factor such that $R/\bar{d}=1.8$ and the resultant interpolated pattern shows a predominant wave of length $L/\bar{d}=3$. Eq (11) (see also Fig. 1) shows that we have represented only 80 per cent of the amplitude of this wave, 20 per cent of it having been lost through the smoothing inherent in the scheme.

In searching for a technique which might correct the analysis by regaining some of that lost amplitude, the following experiment was designed:

1) Choose a weight factor appropriate to the data distribution and the scale of disturbance to be analyzed, and perform an initial interpolation over the grid. This is a "first-guess" analysis.
2) At each data point, subtract from the reported value the value obtained in the first-guess analysis at that point. (The data point is regarded as being in the center of the square formed by the four nearest grid points, regardless of its actual position within that area. The interpolated value at the data point is determined by the arithmetic average of the interpolated values at the four nearest grid points.)
3) Determine the smoothed field of difference values by interpolation of the information obtained in Step 2 using the *same* weight factor as was used initially.
4) Add the field obtained in Step 3 to the first-guess analysis, producing a second-guess analysis.
5) Repeat Steps 2 through 4 replacing the first-guess analysis by the second-guess, then replacing the second by the third, etc. Continue iterating until the differences have been diminished to tolerable amounts.

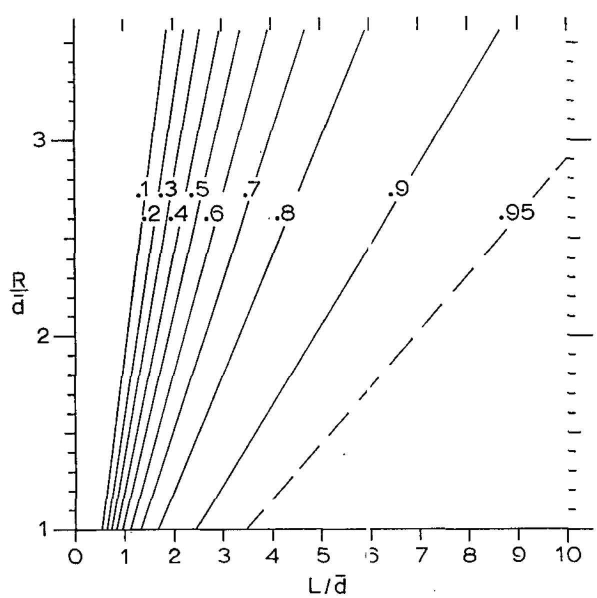

FIG. 1. Initial accuracy index D as a function of half wavelength L, average distance between data points $\bar{d}$, and radius of influence R of the weight factor η with $E=4$.

A discussion of the experimental results using that technique will be found in the next section. It suffices to say here that the results were better than anticipated, for they showed an improvement in the representation of not only the longer waves, but of all details of the pattern. It was found afterward that the convergence of each succeeding "guess" toward the "real" field of values was inherent in the iteration scheme. The reason is explained in the following paragraphs.

Using the simple one-dimensional example again, we wish to find out if

$$\lim_{N\to\infty}[f(x)-g_N(x)]=0 \tag{12}$$

where N is the number of iterations, $f(x)(=A\sin ax)$ is the "real" distribution, and $g_N(x)$ is the interpolated field obtained after N iterations. As before, the first interpolation results in the initial guess

$$g_0(x)=Df(x). \tag{13}$$

Steps 2, 3, and 4 of the process produce

$$g_1(x)=g_0(x)+[f(x)-g_0(x)]D. \tag{14}$$

Substitution of (13) into (14) yields

$$g_1(x)=Df(x)[1+(1-D)].$$

Similarly, it can be shown that the N-th iteration yields

$$g_N(x)=f(x)D\sum_{n=0}^{N}(1-D)^n, \tag{15}$$

and (12) may then be written

$$\lim_{N\to\infty}(f(x)[1-D\sum_{n=0}^{N}(1-D)^n)]=0. \tag{16}$$

The question thus reduces to the following: we must determine whether or not

$$\lim_{N\to\infty}[\sum_{n=0}^{N}(1-D)]^n=1/D. \tag{17}$$

The ratio test of the power series in (17) shows that

$$\frac{(1-D)^{n+1}}{(1-D)^n}=1-D. \tag{18}$$

Recalling that D is constant for any given wavelength and for any choice of weight factor (and the weight factor was not changed during the iteration process), then $1-D$ is also constant. Furthermore, from (11) we see that $0\leq D\leq 1$. From the practical viewpoint, we can remove from consideration the limiting cases $D=0$ and $D=1$. Therefore, the right-hand side of (18) is a constant less than one, the requirement for convergence of the series in (17). That series is, in fact, a geometric series converging to the value

$$\frac{1}{1-(1-D)}=\frac{1}{D}. \tag{19}$$

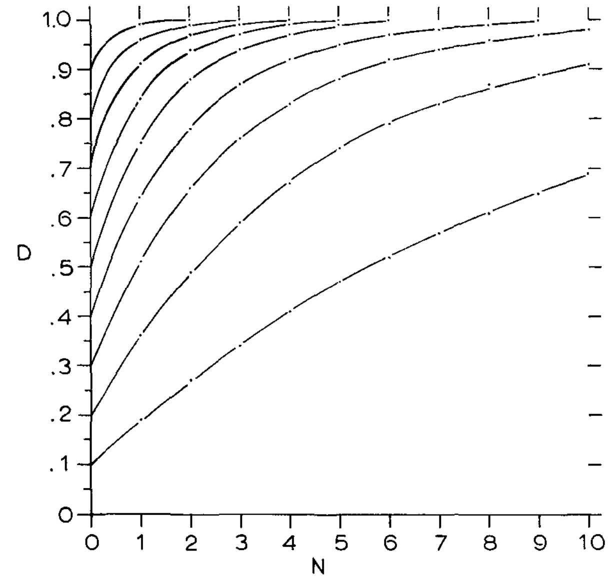

Fig. 2. Final accuracy index D' as a function of initial D and number of iterations N.

The question on (12) may now be resolved. Since the premise of the interpolation scheme is that the waves which make up the distribution $f(x,y)$ are *independent*, the argument applies individually to each component wave. Thus, the iteration scheme regains the amplitude of not only the longer waves, but of all waves represented in the data.

We may now define a new accuracy index from (15) as that obtained after N iterations, that is,

$$D'=D\sum_{n=0}^{N}(1-D)^n. \tag{20}$$

Table 1 shows calculated values of D' for various initial D and up to 10 iterations. Fig. 2 shows the same relationships in graphical form. It will be noted from Fig. 1 that once a weight factor is chosen (fixing $R/\bar{d}$), a direct correspondence exists between initial D and $L/\bar{d}$. In other words, since the weight factor is held constant, we have a means of tracing the final accuracy of our result back through the initial accuracy index D to the wavelength of a particular disturbance. In Table 1, this relationship is shown for a weight factor choice such that $R/\bar{d}=1.6$. For example, if in the final analysis we note a wave of a length such that $L/\bar{d}=1.5$ and we have performed 3 iterations of the interpolation process using a weight factor such that $R/\bar{d}=1.6$, we will have increased the amplitude accuracy of that wave from an initial 50 per cent to 94 per cent.

The question then arises, "If we can make the final analysis fit the data as exactly as desired through repeated iterations of the interpolation, how many iterations are necessary or desirable?" The answer to this question necessarily depends upon the type of data input. If the data are accurate, or if the errors involved comprise but a small portion of the total variations observed over the field, and if one desires to depict the maximum detail exhibited by the data, then one is justified in "forcing" the analysis to converge to fit the data nearly exactly. On the other hand, if data errors are known to be of fairly large amplitude, or if one wishes to see only the broad features of the field, then a fewer number of iterations are called for. In particular, when data errors are a problem, as with most upper air analyses, then a table such as Table 2 should be prepared. Table 2 shows the amount of amplitude accuracy gained in each iteration as a function of the initial accuracy. The values were obtained from Table 1 by the following relation:

$$\Delta D' = D'_{n+1} - D'_{n}. \qquad (21)$$

As in Table 1, the relations between initial D and $L/\bar{d}$ are shown for $R/\bar{d}=1.6$. Note in Table 2 that the maximum increase in amplitude during the first iteration occurs at $D=0.5$; $[(L/\bar{d})=1.5]$. During the second iteration, the maximum increase moves to the $D=0.3$ column; $[(L/\bar{d})=1.1]$. And so on with each iteration the maximum increase moves toward lower values of D (shorter waves). As mentioned before, random errors in the data field will produce fictitious waves of a length $L/\bar{d}=1$. It makes sense, therefore, to stop the iterations just before waves of this size are increased in amplitude by an amount exceeding the increase in the longer wave amplitudes. In the case under discussion $(R/\bar{d}=1.6)$, Table 2 shows that N should not be allowed to exceed 3 since with the 4th iteration the column $L/\bar{d}=1.0$ is increased by 8 per cent, surpassing the increase in any other column.

Of course, in any practical application of the iteration scheme, the economic factor of machine time enters the picture placing further limitation upon the number of iterations performed.

3. Experimental results

The initial tests of the iteration-interpolation scheme were performed on an IBM 7090 using 500-mb height data as reported over the United States and part of Canada. A 27×18 grid was used with a 100-n mi mesh size. Seventy-seven data points were located within the boundaries of the grid. Fig. 3 shows the grid and distribution of data points. In this case, $\bar{d}=300$ n mi was applicable, and a weight factor was chosen such that $R/\bar{d}=1.6$, which is to say that the radius of influence of any datum is 480 n mi.

The number of data affecting each grid point is used as a reliability check with a minimum of two pieces of data required for interpolation. A smoothed distribution of the reliability grid for this case of 500-mb heights is shown in Fig. 4. There were only a few points along the southeastern boundary of the grid which could not be interpolated. Within the major portion of the United States, no fewer than 12 pieces of data, and in a few cases as many as 20, were considered in determining the interpolated value at the grid points.

Fig. 5 shows the results of the "first-guess" interpolation of 500-mb heights at 1200 GCT 15 March 1962. The method of obtaining this and all subsequent analyses was as follows: the interpolated values at each grid point are printed out, plotted on a reproduction of the grid, and analyzed manually by linear interpolation between grid points. Automated means of performing this step can be used if adequate facilities are available. Since this particular program was not meant to be operational, the method used seemed most expedient,

TABLE 1. Values of final accuracy index D' as a function of initial D and number of iterations N from equation (20). Relationships between $L/\bar{d}$ and D are for $R/\bar{d}=1.6$.

$L/\bar{d}$	0.8	1.0	1.1	1.3	1.5	1.8	2.1	2.7	3.9
N \ D	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1	0.19	0.36	0.51	0.64	0.75	0.84	0.91	0.96	0.99
2	0.27	0.49	0.66	0.78	0.88	0.94	0.97	0.99	1.00
3	0.34	0.59	0.76	0.87	0.94	0.97	0.99	1.00	
4	0.41	0.67	0.83	0.92	0.97	0.99	1.00		
5	0.47	0.74	0.88	0.95	0.99	1.00			
6	0.52	0.79	0.92	0.97	1.00				
7	0.57	0.83	0.94	0.98					
8	0.61	0.87	0.96	0.99					
9	0.65	0.89	0.97	1.00					
10	0.69	0.91	0.98						

TABLE 2. $\Delta D'$ as a function of initial D and number of iterations N from equation (21). Relationships between $L/\bar{d}$ and D are for $R/\bar{d}=1.6$.

$L/\bar{d}$	0.8	1.0	1.1	1.3	1.5	1.8	2.1	2.7	3.9
N \ D	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.09	0.16	0.21	0.24	0.25	0.24	0.21	0.16	0.09
2	0.08	0.13	0.15	0.14	0.13	0.10	0.06	0.03	0.01
3	0.07	0.10	0.10	0.09	0.06	0.03	0.02	0.01	0.00
4	0.07	0.08	0.07	0.05	0.03	0.02	0.01	0.00	
5	0.06	0.07	0.05	0.03	0.02	0.01	0.00		
6	0.05	0.05	0.04	0.02	0.01	0.00			
7	0.05	0.04	0.02	0.01	0.00				
8	0.04	0.04	0.02	0.01					
9	0.04	0.02	0.01	0.01					
10	0.04	0.02	0.01	0.00					

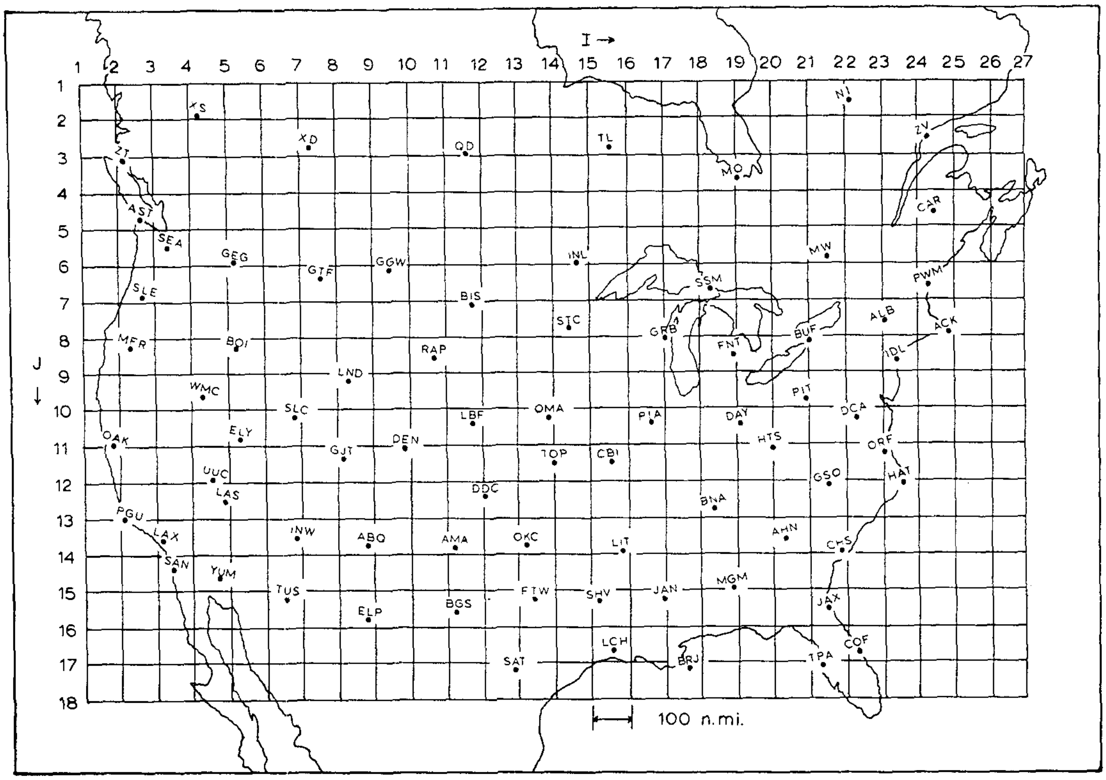

FIG. 3. Analysis grid and data distribution for 500-mb height tests.

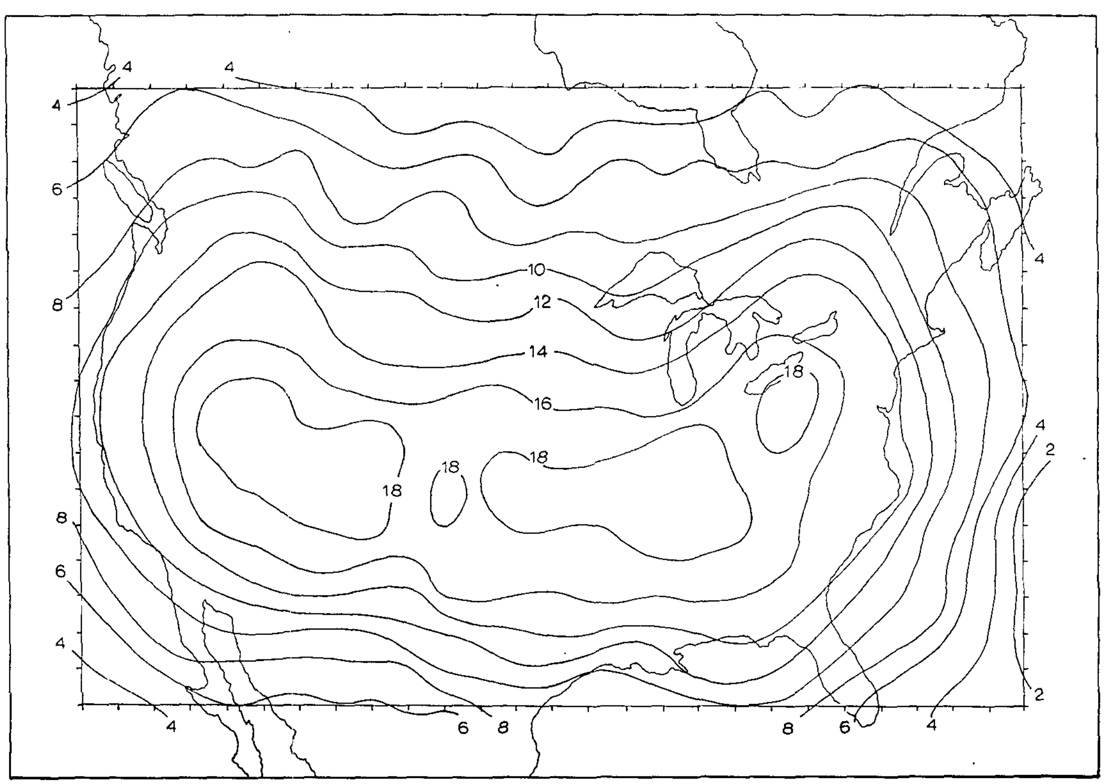

FIG. 4. Smoothed distribution of the number of data affecting the calculation of the interpolated value at each grid point for the 500-mb height tests.

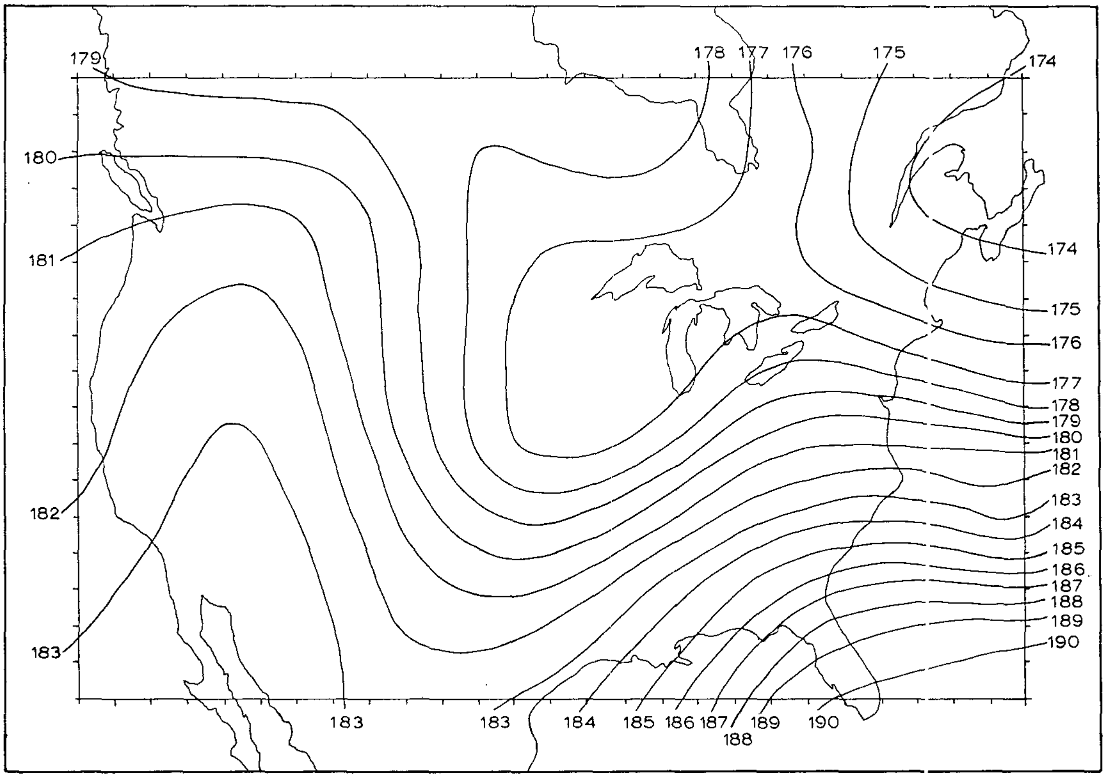

FIG. 5. Result of initial interpolation ($N=0$) of 500-mb contours for 1200 GCT 15 March 1962. Labels are in hundreds of feet.

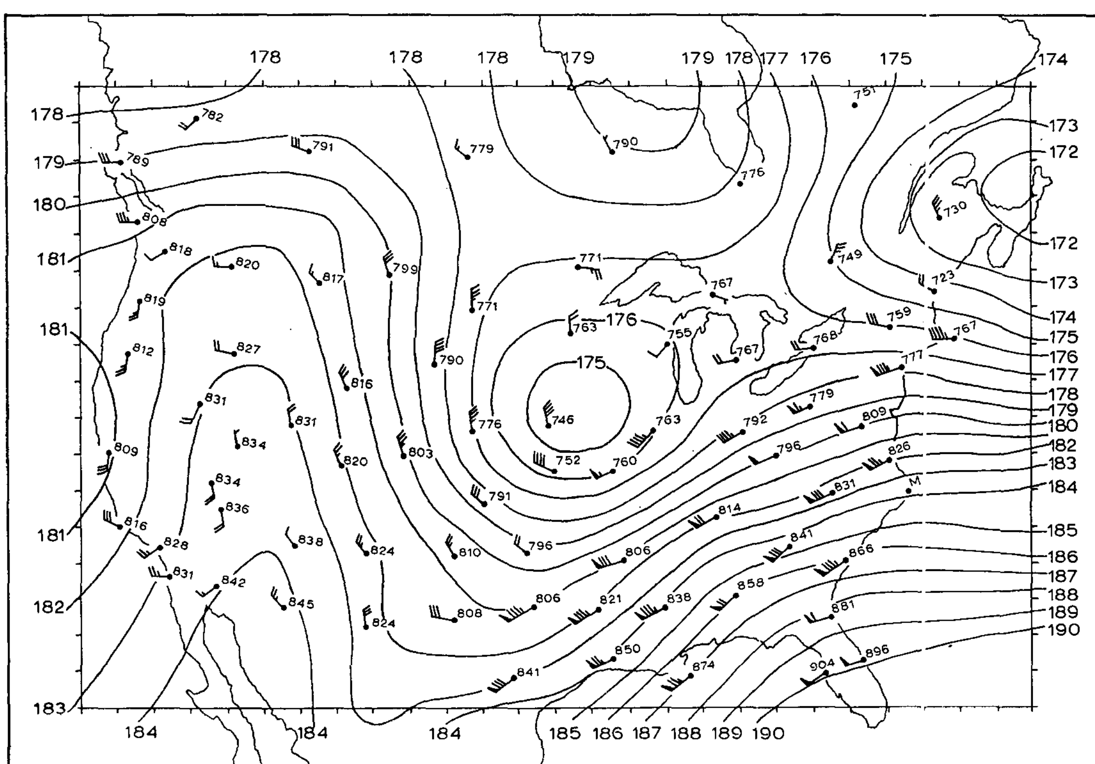

FIG. 6. Result of interpolation of 500-mb contours for 1200 GCT 15 March 1962, after three iterations ($N=3$).

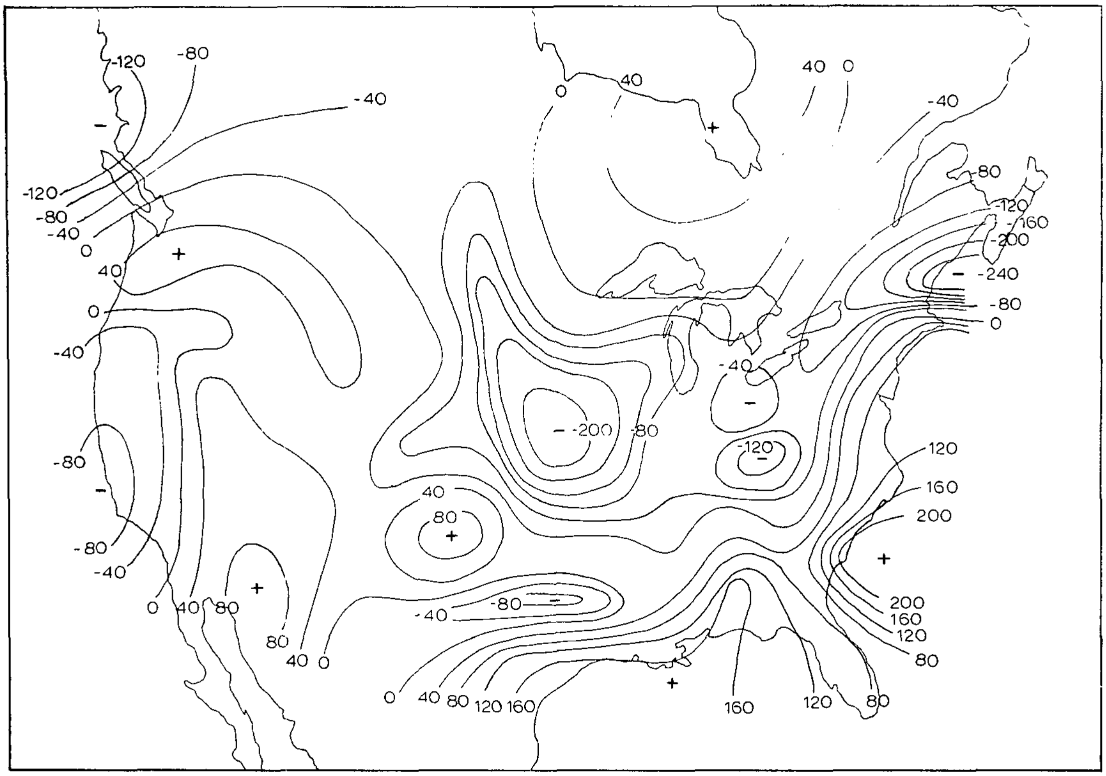

FIG. 7. Manual analysis of difference field generated by subtracting first interpolated values (Fig. 5) from reported 500-mb heights at each data point. Contours are labelled in feet.

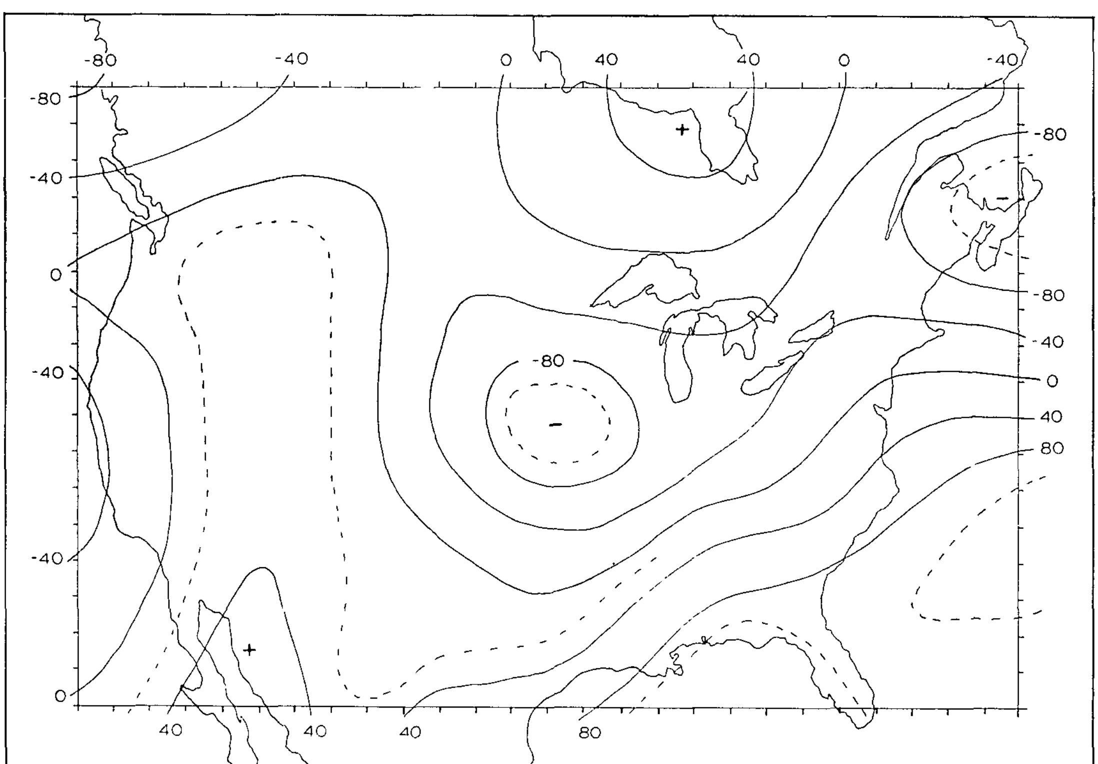

FIG. 8. Machine analysis of same difference data shown in Fig. 7. Contours are labelled in feet.

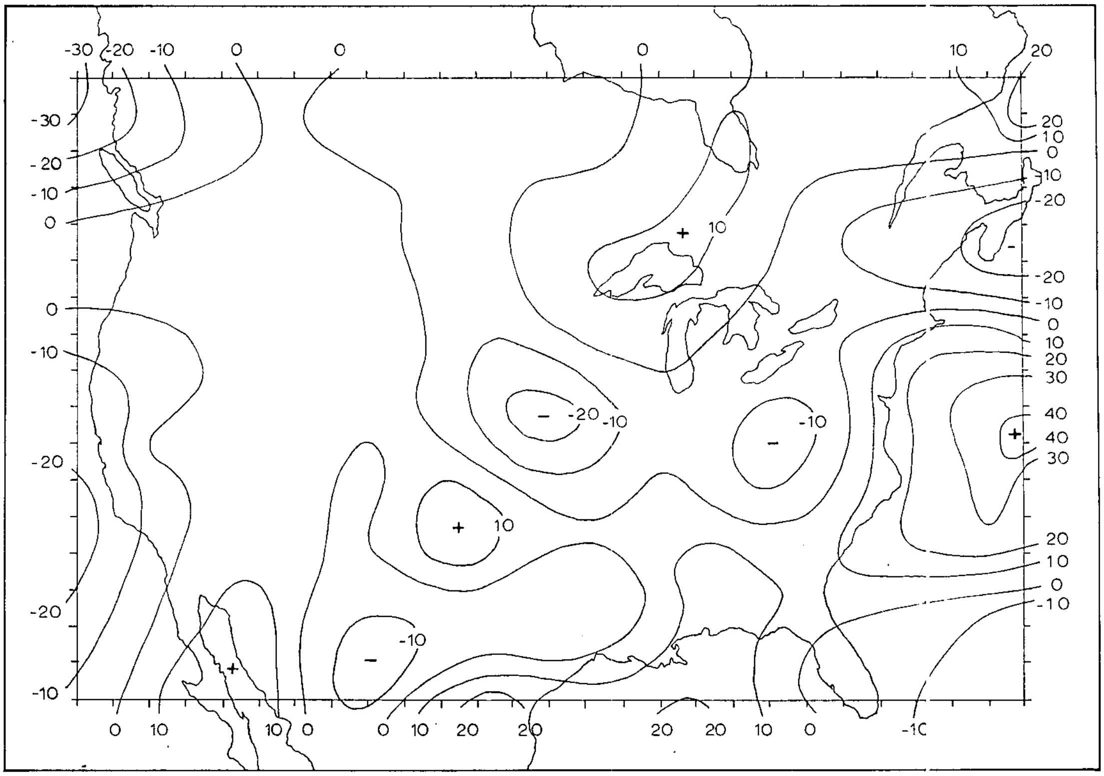

Fig. 9. Machine analysis of 500-mb difference data after three iterations.

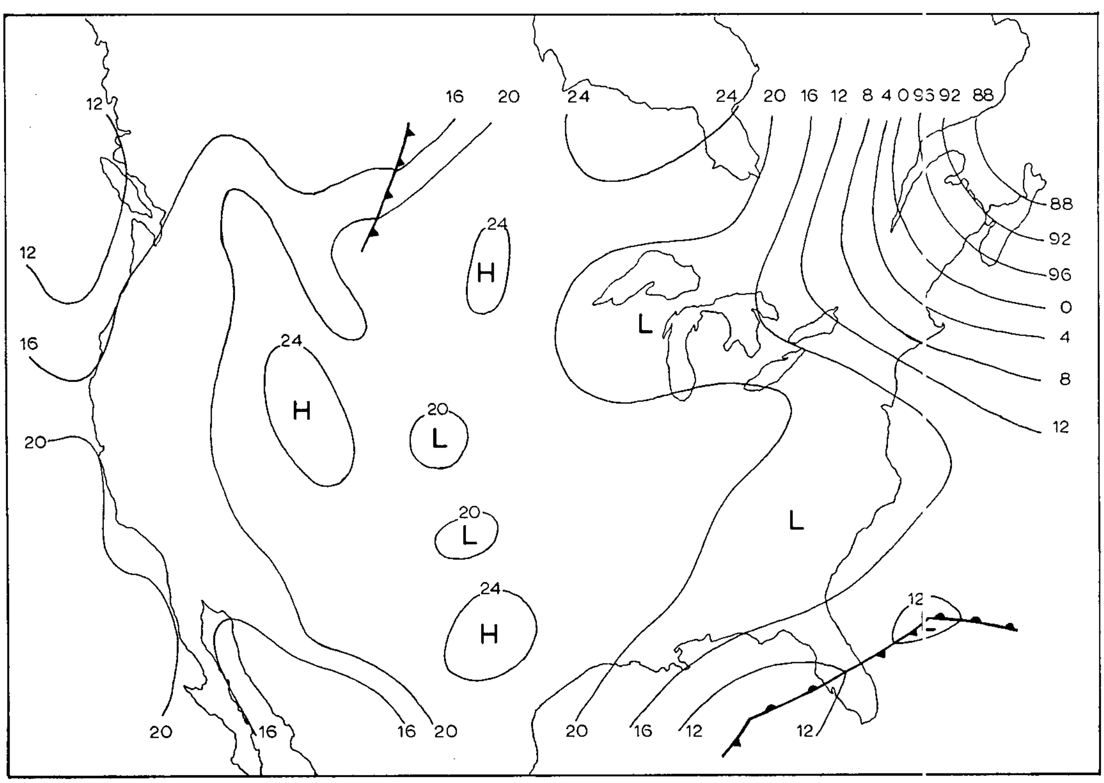

Fig. 10. Manual analysis of sea level pressure in millibars for 1200 GCT 15 March 1962. (After NMC.)

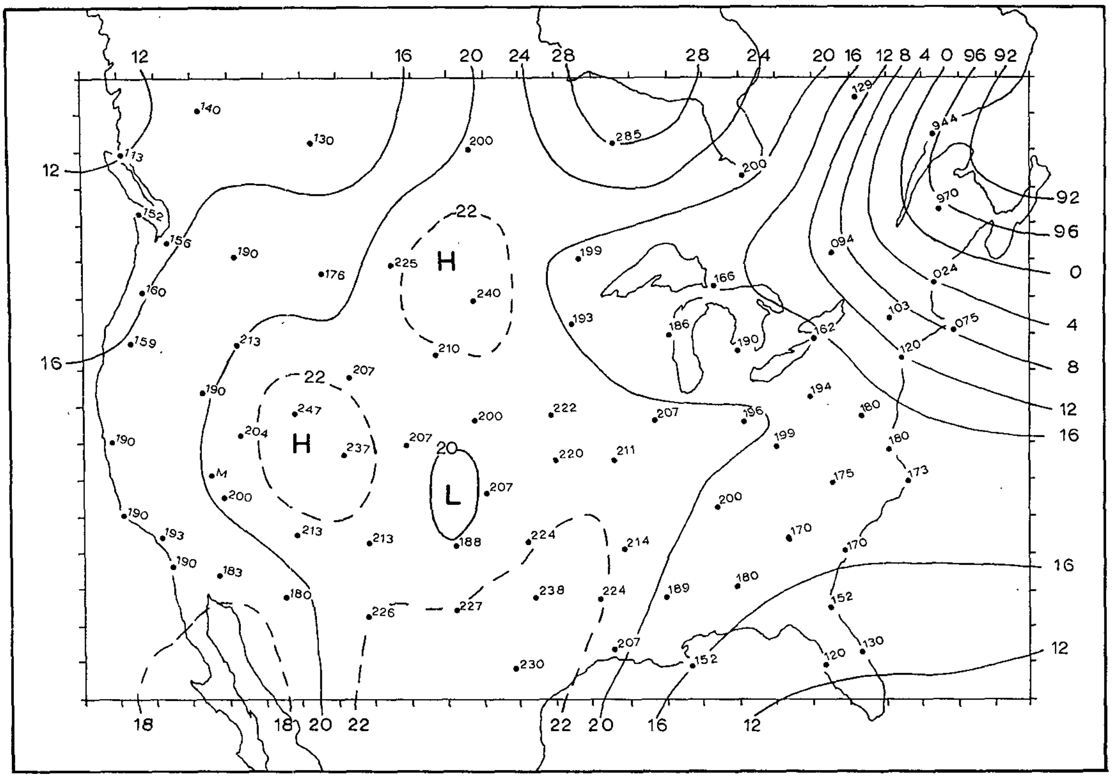

FIG. 11. Machine analysis of sea level pressure for 1200 GCT 15 March 1962 based on data shown.

even if not completely objective. Fig. 5 is also a prime example of the problem of over-smoothing in a weighted-averaging scheme. There existed a 17,400-ft low center just east of Omaha, Nebraska (OMA), and heights above 18,400 ft were reported in Arizona.

Fig. 6 shows the interpolated results after 3 iterations. The reported height values and winds are plotted, although it should be mentioned that *wind information was not used in the analysis*. The results within the interior of the grid agree quite well with the reported heights. Around the boundaries, the fit of the heights is adequate, but the apparent sharp trough off the California coast was not represented. It is particularly interesting to see how the scheme reacted to the obviously erroneous height reported at Portland, Maine (PWM), 17,230 ft. A recheck of the reported sounding failed to show any calculation error, but even the National Meteorological Center (NMC) analysis disregarded the reported value. Likewise, the IBM 7090 disregarded the report. The explanation is fairly simple, however. The distance between PWM and the adjacent stations is on the order of 150 n mi. The error then amounts to a wave of length $L/\bar{d}\sim 0.5$ since $\bar{d}=300$ n mi. Checking again with Fig. 1, since $R/\bar{d}=1.6$, the amplitude of a wave of that dimension would have initially been represented by something less than 10 per cent. Fig. 2 shows that even after 3 iterations, still less than 30 per cent of this amplitude has been retained.

Figs. 7, 8 and 9 are examples of the difference fields generated in the iteration steps, Fig. 8 and 9 showing the convergence of scheme. Fig. 7 is a manual analysis of the differences between the data point values and the first-guess interpolation values (Fig. 5). Note the error at PWM showing up as the large negative value off the coast of Maine, and also note the low center in the middle section of the grid with the short wave pattern to the southwest. (The short wave was associated with a surface system as will be shown later.) Fig. 8 is the result of applying the weight factor to the difference data used to obtain Fig. 7. Fig. 9 shows the final smoothed difference field after 3 iterations. Note that over the majority of the United States the differences have been reduced to generally less than 20 ft, a value approaching the normal computational accuracy of the radiosonde method. This indicates that further iterations would not be called for.

Since at 500 mb the most predominant waves are long waves (long with respect to the spacing between stations), it was decided to test the resolution of the scheme by using sea level pressure data which generally contain short waves of larger amplitude than found at 500 mb. Fig. 10 shows the surface analysis after NMC for 1200 GCT 15 March 1962. That analysis is based on the dense network of surface stations in the United States and Canada. Fig. 11 is the machine analysis of the same map as shown in Fig. 10, but it is based on *only*

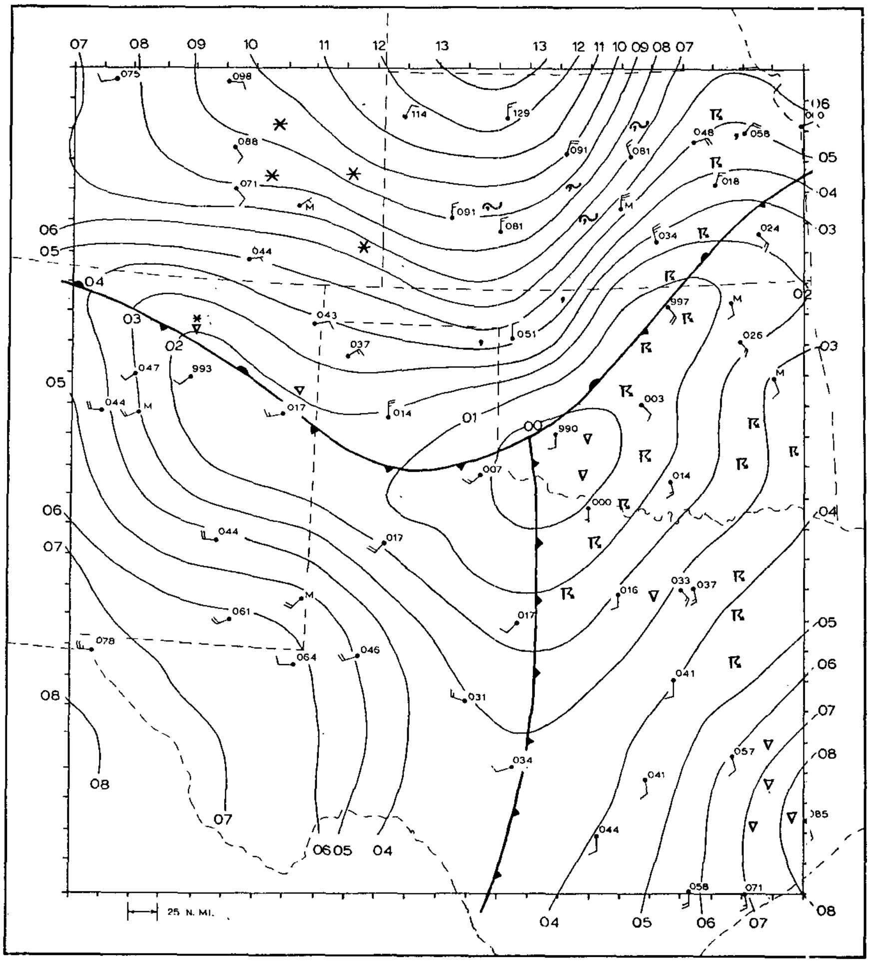

FIG. 12. Machine analysis of sea level pressure in millibars for 0000 GCT 18 February 1961. Fronts and weather determined from manual analysis.

those pressures shown plotted, i.e., at the upper air reporting stations. Three iterations were carried out.

Although the resolution and amplitude of the small highs and lows over the west central part of the map are not as good as the manual analysis, the essential features are shown in the machine result. Of course, one cannot hope to depict the frontal discontinuities on the basis of scalar analysis of a coarse data distribution. The surface lows in the vicinity of western Kansas are associated with the short wave trough at 500 mb.

In an effort to see just how much detail could be obtained from the interpolation scheme, another test case was assembled, this time dealing with the surface analyses of pressure, temperature, and dew point on a regional scale. A 25×28 grid of mesh size 25 n mi was placed over the portions of Texas, Oklahoma, New Mexico, Colorado and Kansas shown in Fig. 12. In this case, $\bar{d}=100$ n mi and $R/\bar{d}=1.6$. All analyses were carried to 3 iterations. Interpolated values in the extreme southwest corner of the grid were not obtainable for lack of data. The subjectively determined surface fronts and the major weather types are shown superimposed upon the machine analysis of pressure in Fig. 12 (again, winds are shown but were not used in the analysis). A squall line with accompanying hail and a few tornadoes was moving through central Oklahoma at the time. Although the analysis differs from the data in several respects, the essential features of the pressure distribution are represented.

Fig. 13 shows the interpolated pattern of surface temperatures. One could not justify placing the polar front at the warm edge of the large temperature gradient since the other data show it to be slightly farther north. But then, we are not attempting to position discontinuities by this method.

The dew point analysis in Fig 14 shows two interesting features aside from the gradients associated with the fronts. The first is the relatively moist area in eastern New Mexico associated with the shower area there. The size of this system is of the order of 100 n mi in half

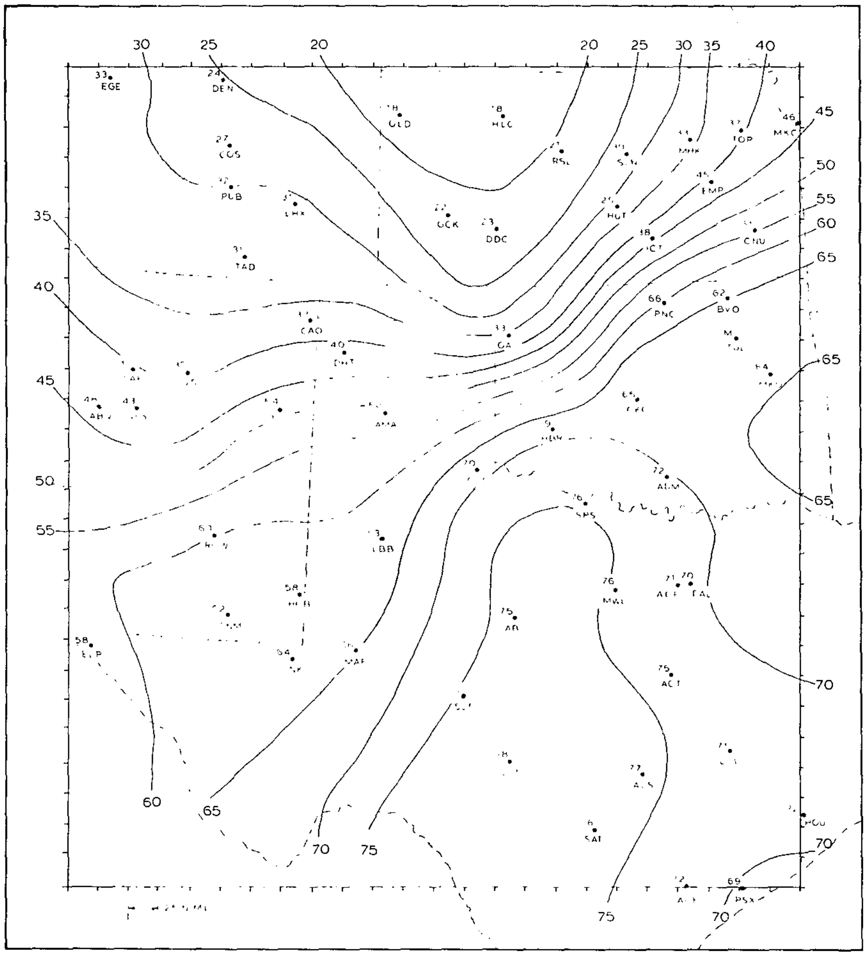

FIG. 13. Machine analysis of surface temperature for 0000 GCT 18 February 1961.

wavelength. This indicates the analysis scheme is applicable to mesoscale phenomenon provided one has sufficient information concerning them.

The other interesting result is the region of low dew points in the Big Bend region of Texas. The lowest reported dew point was 9F at Wink, Texas (INK), yet the analysis scheme extrapolated a 2F value in the Big Bend region. This occurred because the only two stations affecting that region were Wink and Midland, Texas (MAF), with the two lowest reported dew points. During the iteration process, the smoothed differences at the points in the Big Bend were always negative, thus diminishing the first-guess value of 9.7F with each iteration. However, we cannot dispose of the extrapolation ability of the scheme merely on the basis that it happens only when too few stations enter the calculation. The 65F value in the northeastern part of Oklahoma was the result calculated on the basis of 8 stations, not one of which reported a value greater than 62F. The fact that the scheme does extrapolate is quite clear; the validity of those extrapolations is still open to question.

Again, in Figs. 12, 13 and 14, only the general features have been analyzed with any great detail. It is conceded that a skilled analyst could get more detail out of the data by placing more weight on each report. But to do so is to say the reports are accurate and representative of the smallest scale of disturbance visible in the distribution of data. If that is in fact the case, then the same result can be obtained objectively merely by carrying out the iterations a greater number of times, or by using a weight factor with slightly smaller radius of influence. The convergence of the interpolated result to the desired fit of the data is insured in either case.

4. Conclusions

In summary, we list the advantages of the convergent weighted-averaging interpolation scheme described in this paper.

1) It is based on the concept that the distribution of an atmospheric parameter at any given time can

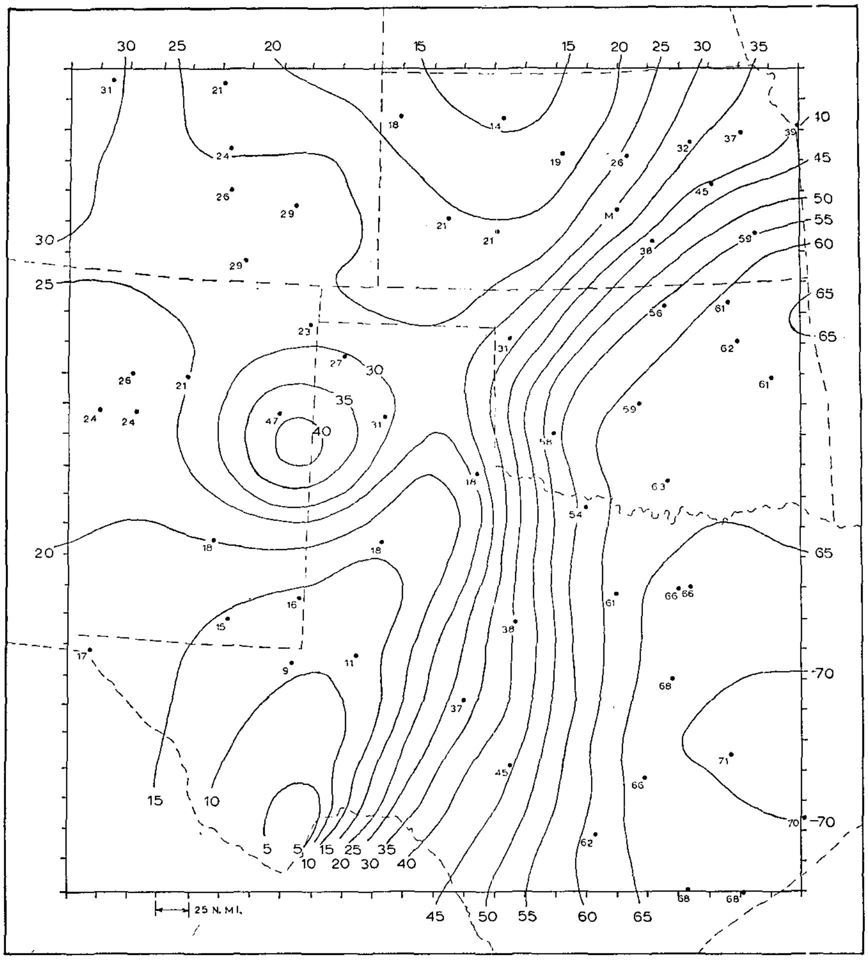

FIG. 14. Machine analysis of surface dew points for 0000 GCT 18 February 1961.

be represented as the sum of an infinite number of independent component waves.

2) The theoretical limit of the detail obtainable is governed only by the density of the data distribution; the practical limit is additionally governed by the accuracy of the input data and the purpose of the analysis.
3) The scheme is computationally simple; therefore, it is quickly executed and economical. (The three maps shown in Fig. 12, 13 and 14 involving 57 stations on a 25×28 grid were executed in 2 min 43 sec on the IBM 7090. The entire program including input of program and data, execution, and printing of results on the IBM 1401 was accomplished in about 10 min, two-thirds of which was printing time.)

Although the scheme is not limited in theory to any particular distribution of data, from a practical viewpoint, applications should be made to reasonably uniform data distributions. This is mainly for economical reasons. If the data spacings are fairly uniform so that the actual distance between individual stations does not deviate greatly from the average distance between the entire group of stations, then a relatively small radius of influence may be chosen for the weight factor and still insure continuous, well-behaved results. Of course, the smaller the radius of influence of the weight factor, the faster the scheme converges to show the shorter wavelengths.

On the other hand, should the scheme be applied to, say, the data distribution found in an area including North America and the northern Atlantic Ocean, the radius of influence of the weight factor would have to be very large to insure continuity of the results. Although no tests were performed on such a distribution of data, experience suggests that the resulting smoothing of the shorter wavelength disturbances appearing in the data over North America would be so great that many tens of iterations would have to be performed to regain those details. At present, extending the iterations much beyond ten does not appear to be economically justified. Furthermore, it should be remembered that while the shorter wavelength variations are being amplified, the deviations caused by data errors at the the more widely-spaced stations are being amplified to a higher degree. For these reasons, direct application of the scheme to obtain maximum detail in regions wherein the data densities vary considerably is not recommended.

A solution to the problem of applying the scheme to non-uniform data distributions is beyond the scope of this paper. However, two possibilities are suggested. One might attempt to fill in the data distribution by employment of nonconventional data as is currently being done at the NMC. Or, one might analyze the various regions of different data densities separately, employing the proper weight factor and number of iterations individually to each region, and then tie the analyses together in some proper manner.

As for better representation of discontinuities between air masses, again nonconventional data might be employed. For surface analyses, the most promising source of such data appears to be autographic records, converting the time ordinate to a space ordinate.

Acknowledgment. The author extends thanks to Mr. Basil Taylor, Central Computer Department, Continental Oil Company, Ponca City, Oklahoma, for handling the liaison concerning the computer program, and to Mr. Lawrence Hagen for his assistance in analyzing the machine results. Also, the author is grateful to Professor Y. Sasaki for his encouragements during the study and for his critical review of this paper.

REFERENCES

Berghórrson, P., and B. Döös, 1955: Numerical weather map analysis. *Tellus*, **7**, 329–340.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.

Penn, S., B. Kunkel and W. D. Mount, 1963: On objective analysis of continuous and discontinuous parameters. *J. appl. Meteor.*, **2**, 345–350.

Shuman, F. G., 1957: Numerical methods in weather prediction: II. Smoothing and filtering. *Mon. Wea. Rev.*, **85**, 357–361.