Optimum Smoothing of Two-Dimensional Fields

Philip Duncan Thompson

Major, U. S. Air Force

U.S. Joint Numerical Weather Prediction Unit

26 April 1955

National Oceanic and Atmospheric Administration

U.S. Joint Numerical Weather Prediction Unit

ERRATA NOTICE

One or more conditions of the original document may affect the quality of the image, such as:

Discolored pages
Faded or light ink
Binding intrudes into the text

This has been a co-operative project between the NOAA Central Library, National Center for Environmental Prediction and the U.S. Air Force. This project includes the imaging of the full text of each document. To view the original documents, please contact the NOAA Central Library in Silver Spring, MD at (301) 713-2607 x124 or www.reference@nodc.noaa.gov.

LASON
Imaging Contractor
12200 Kiln Court
Beltsville, MD 20704-1387
April 13, 2004

Abstract

The problem of smoothing out nonsystematic errors in a 2-dimensional field of measurements has been studied from the standpoint of minimizing the RMS difference between the true field and an arbitrarily weighted area average of the field of observations. For fields whose space-autocorrelation functions are invariant with rotation and have a simple and rather typical form, the optimum weighting function is a linear combination of Bessel functions, whose nondimensional rate of decay depends partially on the so-called "signal-to-noise" ratio, but primarily on the ratio of the scales of the true field and error field. A comparison of optimum averaging with the analyst's subjective process of smoothing indicates that the former is significantly superior in its ability to distinguish between random small-scale fluctuations and minor synoptic features of only slightly greater scale. Finally, the minimum RMS error of linearly smoothed fields is expressed in terms of the statistical properties of the true fields and the observing system itself.

1. Introduction

One of the main functions of synoptic map analysis is that of smoothing out nonsystematic errors in observations taken at a more or less dense network of discrete points or, possibly(in the view of some meteorologists), of smoothing out real fluctuations whose scale is so small that they cannot be reconstructed in detail.* At the present time, the analyst's process of smoothing is rather subjective and ill-defined ---- generally being carried out "by eye" and with no definite weight assigned to each piece of data. Where data coverage is adequate, this process usually consists in drawing continuous isopleths on a horizontal projection (e.g., contours of an isobaric surface), in such a way that they fluctuate as little as possible, but "fit" the data within the tolerances of the probable nonsystematic error of reported measurements. Loosely speaking, the isopleths "fit" the data if the reported measurements are interpolated roughly linearly between the two nearest isopleths in the direction normal to the isopleths. The fact that the spacing of the isopleths depends on their local direction means that the isopleths are not positioned by two-point linear interpolation in any fixed direction and, accordingly, implies that the value of the smoothed distribution at any single point depends on reported measurements at a number of neighboring points.

In an earlier paper, the author (1954) has pointed out that the analyst's smoothed distribution is, in the above and other respects, very similar to a running area average of the field of reported measurements, in which each piece of data is weighted less and less with increasing distance from the point where the average applies.

*In fact, if the data were to be used exclusively for purposes of numerical weather prediction, and if the observations were taken at a sufficiently dense network of evenly spaced points, this would be the only function of synoptic analysis. Under these conditions, the function of interpolation would be unnecessary.

The mathematical properties of a special type of weighted area average were developed in considerable detail, and it was shown that such area averages have certain advantages (inherent in the form of the physical equations) over time averages with analogous properties. The whole question of smoothing out nonsystematic error, however, was passed over with the remark that "...to minimize the error... without sacrificing what little resolving power we do possess, the observed state must be averaged in such a way as to suppress fluctuations whose scale is comparable with the distance between neighboring observation stations, but leaving disturbances of much larger scale essentially intact". Put more precisely, the real problem is to choose the weighting factors so that they bring about the best compromise between obliterating the small-scale fluctuations completely and partially removing the large-scale disturbances of somewhat greater amplitude. The remaining questions center around the criterion for the "best" compromise.

The purpose of this paper, to state it briefly, is to find the weighting function which is "best" in the sense that the mean square error of the smoothed distribution is the least. To do so, we shall extend Wiener's methods for smoothing functions of one variable to "statistically isotropic" fields or functions of two variables, whose auto- and cross- correlation functions do not depend on the direction in which the fields are shifted. The condition that the RMS difference between the true field and the mean observed field be minimized is expressed in an integral equation, which relates the optimum weighting function to the autocorrelation functions for the true field and the "error" distribution. For Gaussian autocorrelation functions, the weighting function is found to be a linear combination of Bessel functions. The optimum nondimensional radius of the effective domain of averaging is determined by the relative wavelengths and amplitudes of the "errors" and the large-scale disturbances, and its absolute size is fixed by the distance between observation points.

The optimum averaging process is compared with the "smoothing" process of a skilled synoptic analyst in a case when the spectrum of disturbances was fairly broad. Finally, the minimum RMS error of the smoothed field is expressed as a function of the RMS error of reported measurements and the distance between neighboring observation stations.

2. The Condition for Optimum Smoothing

The problem of averaging or smoothing out nonsystematic error and small-scale fluctuations on the threshold of detectability will be recognized as essentially equivalent to that of filtering the "noise" from the output of a communications system ---- a problem that is familiar to the electrical engineer and which has been treated extensively by Wiener (1949) in his theory of stationary time series. Following Wiener, we shall take the minimization of the RMS difference between the true field and the smoothed field of observations as the criterion for optimizing the smoothing process.

We begin by considering a "true" field $g(x,y)$ and the corresponding field of observations $f(x,y)$, which consists partly of $g(x,y)$ and partly of a superimposed "error" $\epsilon(x,y)$. The latter need not be interpreted as genuine error, but may be thought of as any small-scale fluctuation that is random with respect to disturbances of larger scale. We shall also consider a weighted area average $\bar{f}(x,y)$, constructed by applying an integral operator to the observed field $f(x,y)$. That is,

$$\bar{f}(x,y) = \iint K(\rho)\ f(x-\xi, y-\eta)\ d\xi\, d\eta$$

in which x and y are cartesian horizontal coordinates, ξ and η are the corresponding dummy variables of integration, and ρ is the distance from the origin to the variable point (ξ, η) ---- i.e., $\rho^2 = \xi^2 + \eta^2$. The indicated area integration generally extends over the entire (ξ, η) plane. The weighting function K is unspecified, except that its area integral is unity. Under some conditions, for example, K may turn out to be a delta-function, in which case the integration actually extends over only one point in the (ξ, η) plane.

The mean square difference I between the true field $g(x,y)$ and the smoothed field of observations $\bar{f}(x,y)$, when taken over the entire (x,y) plane, is

$$I(K) = \mathrm{l.i.m.} \iint \left[g(x,y) - \iint K(\rho) f(x-\xi, y-\eta)\, d\xi\, d\eta \right]^2 dx\, dy$$

where the symbol "$\mathrm{l.i.m.}$" stands for the limit of the area average as the periphery of the area approaches infinity in all directions. Expanding the square of ($g - \bar{f}$) and inverting the order of integration, we may rewrite the equation above as

$$I(K) = \gamma(0,0) - 2 \iint K(\xi,\eta)\, \chi(\xi,\eta)\, d\xi\, d\eta$$

$$+ \iint K(X,Y)\, dX\, dY \iint K(\xi,\eta)\, \phi(\xi - X, \eta - Y)\, d\xi\, d\eta \qquad (1)$$

where X and Y are dummy variables of integration corresponding to ξ and η, and where*

$$\gamma(\xi,\eta) = \mathrm{l.i.m.} \iint g(x,y)\, g(x-\xi, y-\eta)\, dx\, dy$$

$$\chi(\xi,\eta) = \mathrm{l.i.m.} \iint g(x,y)\, f(x-\xi, y-\eta)\, dx\, dy$$

$$\phi(\xi,\eta) = \mathrm{l.i.m.} \iint f(x,y)\, f(x-\xi, y-\eta)\, dx\, dy$$

The functions γ and ϕ will be recognized as the autocorrelation functions for g and f, respectively, and χ is the cross-correlation between g and f. For simplicity, we shall now suppose that the correlation functions depend only on the magnitude of the shift from $(0,0)$ to (ξ,η), and not on its direction ---- an assumption that is very nearly fulfilled in reality. With this simplification, $\gamma(\xi,\eta) = \gamma(\rho)$, etc., and Eq. (1) reduces to:

*Each of the fields in question may be regarded as the deviation of the given field from a linear function of X and Y, whose derivatives are the unweighted average derivatives of the given field. Since the smoothing operation, when applied to a linear function, yields the same linear function, this interpretation involves no loss of generality.

$$I(K) = \gamma(0) - 4\pi \int_0^\infty K(\rho)\, \chi(\rho)\, \rho\, d\rho \qquad (2)$$
$$+\, 2\pi \int_0^\infty K(\rho)\, \rho\, d\rho \int_0^\infty K(R)\, R\, dR \int_0^{2\pi} \phi(r)\, d\theta$$

in which R is the distance from the origin to the variable point (X, Y), r is the distance between the variable points (X, Y) and (ξ, η) ---- i.e., $r^2 = R^2 + \rho^2 - 2R\rho \cos\theta$ — and θ is the angle between radii from the origin to the points (X, Y) and (ξ, η).

It should be noted that, according to Eq. (2), the mean square difference between the true field and the smoothed field of observations does not depend on the details of those fields, but is determined entirely by their correlation functions. As will be seen later, these statistics can be expressed in terms of the autocorrelation functions for the true field and the "error" field ---- the former of which probably does not change markedly from one day to the next, and the latter of which (in the case of nonsystematic errors of measurement) depends only on the characteristics of the observing system.

Stated concisely, the mathematical problem is to find the function K for which $I(K)$ is the least, given the correlation functions $\gamma(\rho)$, $\chi(\rho)$, and $\phi(\rho)$ and the condition that the area integral of K is unity. Let us suppose that $K(\rho)$ is, in fact, the weighting function for which $I(K)$ is the least. Then $I(K + \delta M)$ must be greater than $I(K)$, <u>for any choice of the constant δ and for any unspecified function $M(\rho)$</u>. That is to say, substituting $K + \delta M$ for K in Eq. (2), transposing terms and factoring,

$$I(K+\delta M)-I(K) = 2\pi\delta^2\int_0^\infty M(\rho)\,\rho\,d\rho\int_0^\infty M(R)\,R\,dR\int_0^{2\pi}\phi(\lambda)\,d\theta$$

$$+\,4\pi\delta\int_0^\infty M(\rho)\,\rho\,d\rho\left[\int_0^\infty K(R)\,R\,dR\int_0^{2\pi}\phi(\lambda)\,d\theta\;-\;\chi(\rho)\right] > 0$$

We next assume tentatively that the expression in square brackets is different from zero. If this is the case, then the integral over ρ in the second term on the left hand side of the inequality is different from zero. Suppose it is greater than zero. Now, it can be easily shown that, owing to the nature of an autocorrelation function, the first term on the left hand side is always greater than zero. Thus, under the conditions assumed above, there exists a sufficiently small negative value of δ for which the inequality is not satisfied. Conversely, if the integral over ρ in the second term on the left hand side of the inequality is less than zero, there must exist a sufficiently small positive value of δ for which the inequality is violated. Accordingly, it must be concluded that the quantity in square brackets is zero ---- i.e.,

$$\int_0^\infty K(R)\,R\,dR\int_0^{2\pi}\phi(\lambda)\,d\theta = \chi(\rho) \qquad (3)$$

This equation expresses the necessary condition which K must satisfy, in order that the mean square difference between the true field and the smoothed field of observations be minimized. It is apparent that this condition, taken together with the given correlation functions χ and ϕ , completely determines the optimum weighting function if it exists at all.

3. Solution of the Equation for the Optimum Weighting Function

The remaining problem is to solve the integral equation (3) for $K(R)$, regarding $\phi(r)$ and $\chi(\rho)$ as known. It is convenient to begin by considering $F(u)$, the Fourier-Bessel transform of $\phi(r)$, which is defined as

$$F(u) = \int_0^\infty J_0(ru)\,\phi(r)\,r\,dr$$

where J_0 is the zero-order Bessel function of the first kind with real argument. Now, according to the Fourier-Bessel theorem, the inverse transform is

$$\phi(r) = \int_0^\infty J_0(ru)\,F(u)\,u\,du$$

for all ordinary functions ϕ and their transforms F . Thus, substituting the transform of $F(u)$ for $\phi(r)$ and inverting the order of integration,

$$\int_0^{2\pi} \phi(r)\,d\theta = \int_0^\infty F(u)\,u\,du \int_0^{2\pi} J_0\left(u\sqrt{R^2+\rho^2-2R\rho\cos\theta}\right)d\theta$$

The integral over θ is discussed by Watson (1922). It is equal to

$2\pi J_0(R\mu) J_0(\rho\mu)$. Introducing these results into Eq. (3), and again inverting the order of integration,

$$2\pi \int_0^\infty J_0(\rho\mu) F(\mu) \mu \, d\mu \int_0^\infty J_0(R\mu) K(R) R \, dR = \chi(\rho)$$

The integral over R will be recognized as $\bar{K}(\mu)$, the Fourier-Bessel transform of $K(R)$. Thus, the integral equation above reduces to:

$$\chi(\rho) = 2\pi \int_0^\infty J_0(\rho\mu) F(\mu) \bar{K}(\mu) \mu \, d\mu$$

But this equation simply states that $\chi(\rho)$ is the Fourier-Bessel transform of $2\pi F(\mu) \bar{K}(\mu)$, whence (in view of the Fourier-Bessel theorem) $2\pi F(\mu) \bar{K}(\mu)$ is the Fourier-Bessel transform of $\chi(\rho)$. In other words,

$$\bar{K}(\mu) = \frac{1}{2\pi} \frac{X(\mu)}{F(\mu)}$$

where $X(\mu)$ is the Fourier-Bessel transform of $\chi(\rho)$. Finally, we obtain $K(R)$ by applying the inverse transform to $\bar{K}(\mu)$, with the result that

$$K(R) = \frac{1}{2\pi} \int_0^\infty \frac{J_0(R\mu) X(\mu)}{F(\mu)} \mu \, d\mu \qquad (4)$$

in which, as noted before,

$$X(\omega) = \int_0^\infty J_0(\rho\omega)\ \chi(\rho)\ \rho\, d\rho$$

$$F(\omega) = \int_0^\infty J_0(\rho\omega)\ \phi(\rho)\ \rho\, d\rho$$

Formally, at least, Eq. (4) gives the optimum weighting function corresponding to the known correlation functions χ and ϕ .

A more revealing form of Eq. (4) may be obtained by normalizing the correlation functions χ and ϕ , and by introducing the fact that the true field $g(x,y)$ and the "error" field $\epsilon(x,y)$ are assumed to be uncorrelated. By definition,

$$\chi(\rho) = \text{l.i.m.} \iint g(x,y)\ g(x-\xi,\ y-\eta)\ dx\,dy$$

$$+ \text{l.i.m.} \iint g(x,y)\ \epsilon(x-\xi,\ y-\eta)\ dx\,dy$$

Since g and ϵ are uncorrelated, however,

$$\chi(\rho) = \gamma(\rho) = \gamma(0) \cdot \frac{\gamma(\rho)}{\gamma(0)}$$

Similarly,

$$\phi(\rho) = \text{l.i.m.} \iint g(x,y)\, g(x-\xi, y-\eta)\, dx\, dy$$

$$+ \text{l.i.m.}\ 2 \iint g(x,y)\, \epsilon(x-\xi, y-\eta)\, dx\, dy$$

$$+ \text{l.i.m.} \iint \epsilon(x,y)\, \epsilon(x-\xi, y-\eta)\, dx\, dy$$

Again, since g and ϵ are uncorrelated,

$$\phi(\rho) = \gamma(\rho) + E(\rho) = \gamma(o)\left[\frac{\gamma(\rho)}{\gamma(o)} + \frac{E(o)}{\gamma(o)}\,\frac{E(\rho)}{E(o)}\right]$$

where $E(\rho)$ is the autocorrelation function for the "error" field $\epsilon(x,y)$. Substituting these expressions for χ and ϕ into Eq. (4),

$$K(R) = \frac{1}{2\pi}\int_0^\infty J_0(Ru)\cdot\frac{1}{1 + k^2\,\dfrac{\bar{E}(u)}{\bar{\gamma}(u)}}\cdot u\, du \qquad (5)$$

in which $\bar{E}(u)$ and $\bar{\gamma}(u)$ are the Fourier-Bessel transforms of the normalized autocorrelation functions for ϵ and g, respectively, and k^2 is $E(o)/\gamma(o)$. The constant k is the ratio of the RMS amplitudes of the "error" field and the true field ---- or, in the language of the electrical engineer, the "noise-to-signal" ratio.

To place the general result stated in Eq. (5) in a familiar setting, we shall consider two special cases. First, if there is no error, $\mathcal{R}$ vanishes and the weighting function reduces to:

$$K(R) = \frac{1}{2\pi}\int_0^\infty J_0(R\mu)\,\mu\,d\mu$$

The integral above is zero except when $R = 0$, and becomes infinite as R approaches zero. Thus, since the area integral of K is unity, K behaves likes a delta-function. This implies that, if there is no error, unit weight should be given to the point where the average applies and none to any other point, verifying that an error-free field of observations should not be smoothed at all. A less obvious conclusion applies when the scale of the true field is the same as that of the error field. In this case, $\bar{E}(\mu)$ and $\bar{\gamma}(\mu)$ are sensibly the same, and $K(R)$ again becomes a delta-function. This means that "noise" cannot be filtered out of a "signal" of the same (or smaller) scale by <u>smoothing</u>, simply because the signal amplitude is reduced in the same proportion as the error amplitude. It may, of course, be possible to filter out the noise by differentiation ---- rather than by integration or averaging ---- provided the "noise-to-signal" ratio is small enough.

4. The Optimum Smoothing Process for "Gaussian" Correlation Functions.

Although Eq. (5) yields the optimum weighting function for any type of "true" field and any uncorrelated "error" field of smaller scale, it is readily appreciated that numerical calculations of the transforms of the correlation functions and the weighting function itself are extremely laborious, if the correlation functions do not have definite analytic forms. Accordingly, in order to gain a general understanding of the manner in which the optimum smoothing process depends on the statistical properties of the true field and error field, we shall assign simple and rather realistic analytic forms to the autocorrelation functions $E(\rho)$ and $\gamma(\rho)$. By its nature, the normalized autocorrelation function attains a maximum value of unity at $\rho = 0$, and has zero slope at $\rho = 0$. Moreover, since the fields under consideration are not periodic, one would expect that the correlation function decreases monotonically to zero with increasing ρ. The rate at which it decreases will depend, of course, on the "scale" of the field -- i.e., a typical distance over which the field has the same sign. These considerations suggest that the "shapes" of the normalized autocorrelation functions might be closely approximated by that of the Gaussian error function. That is,

$$\frac{E(\rho)}{E(0)} = e^{-a^2\rho^2}, \qquad \frac{\gamma(\rho)}{\gamma(0)} = e^{-b^2\rho^2}$$

where a and b are inverse measures of the scale of the fields $\epsilon(x,y)$ and $q(x,y)$, respectively. The corresponding Fourier-Bessel transforms are (Watson, p.394):

$$\bar{E}(u) = \frac{1}{2a^2} e^{-\frac{u^2}{4a^2}} \qquad \bar{\gamma}(u) = \frac{1}{2b^2} e^{-\frac{u^2}{4b^2}}$$

Substituting these expressions in Eq. (5),

$$K(R) = \frac{1}{2\pi} \int_0^\infty J_0(Ru) \cdot \frac{1}{1 + \lambda^2 \exp(p^2 u^2)} \cdot u\, du \tag{6}$$

in which $\lambda = kb/a$ and $p^2 = (a^2 - b^2)/4a^2b^2$. This equation defines the process of optimum smoothing under fairly general conditions, and provides the basis for most of the remaining discussion.

A simple and fairly accurate method for applying the smoothing process to values at discrete points consists in (1) regarding the entire plane as a nested set of ring-shaped regions, all centered on the origin, (2) averaging the values at all discrete points in each ring, (3) multiplying each average by the <u>integral</u> of the weighting function taken over the corresponding ring and (4) summing the weighted averages over all rings. Since the weighting function depends only on the distance from the origin, step (3) of this procedure leads one to consider an integral J (related to Eq. 6) of the form:

$$J = \int_0^R K(r)\, r\, dr = \frac{1}{2\pi} \int_0^R r\, dr \int_0^\infty J_0(ru) \frac{u\, du}{1 + \lambda^2 \exp(p^2 u^2)}$$

$$= \frac{1}{2\pi} \int_0^\infty \frac{u\, du}{1 + \lambda^2 \exp(p^2 u^2)} \int_0^R r\, J_0(ru)\, dr$$

The integration with respect to r can be carried out by making use of the recursion formulae for the Bessel functions, with the result that

$$J = \frac{1}{2\pi}\int_0^\infty J_1(q)\,\frac{dq}{1+\lambda^2\exp\left(\frac{p^2q^2}{R^2}\right)} \tag{7}$$

where $q = Ru$.

We now consider the behavior of the function $\left[1+\lambda^2\exp\left(p^2q^2/R^2\right)\right]^{-1}$. Under ordinary conditions, the "noise-to-signal" ratio is small and the ratio of the scales of the true field to error field is large ---- i.e., $k \ll 1$ and $a \gg b$. Thus, since $\lambda^2 \ll 1$ and $p^2/R^2 \gg 1$ the function $\left[1+\lambda^2\exp\left(p^2q^2/R^2\right)\right]^{-1}$ is nearly unity at $q=0$, remains near unity until q approaches $R\left(\ln 1/\lambda^2\right)^{\frac{1}{2}}/p$ (the inflection point for small λ), then decreases sharply, and rapidly approaches zero beyond the inflection point. In these circumstances, therefore, the integral (7) may be approximated by

$$J = \frac{1}{2\pi}\int_0^{\frac{R}{p}\left(\ln\frac{1}{\lambda^2}\right)^{\frac{1}{2}}} J_1(q)\,dq$$

The integral above is readily evaluated. According to another well-known recursion formula, it is simply

$$J = \int_0^R K(r)\,r\,dr = \frac{1}{2\pi}\left[1 - J_0\left(\frac{R}{p}\sqrt{\ln\frac{1}{\lambda^2}}\right)\right] \tag{8}$$

The weighting function $K(R)$ can now be found by differentiating Eq. (8) with respect to R.

$$K(R) = \left(-\frac{\alpha^2}{2\pi}\right)\cdot\frac{1}{\alpha R}J_1(\alpha R) = \frac{\alpha^2}{4\pi}\left[J_0(\alpha R) + J_2(\alpha R)\right] \quad (9)$$

in which $\alpha = (\ln 1/\lambda^2)^{\frac{1}{2}}/p$. This formula gives the approximate weighting function for the types of fields normally encountered. It does not deviate from exact results obtained from Eq. (7) by more than a few percent in the neighborhood of $\lambda = .4$ and $a/b = 4$, except when R is large ---- in which case the weighting function has already become negligibly small.* It also approaches the correct limiting form when the scales of the true field and the error field become comparable.

*The approximation discussed above is not, however, valid if λ approaches or exceeds unity. When the "noise-to-signal" ratio is very large, Eq. (6) degenerates to

$$K(R) = \frac{1}{2\pi\lambda^2}\int_0^\infty J_0(Ru)\, e^{-p^2u^2}\, u\, du$$

This type of integral has been mentioned previously (Watson, p.394). When properly normalized, the weighting function for high noise-level takes the form

$$K(R) = \frac{1}{4\pi p^2}\, e^{-\frac{R^2}{4p^2}}$$

The result above shows that the optimum degree of smoothing is independent of noise level when the "noise-to-signal" ratio is very large. As is intuitively evident, the best strategy in this case is to smooth very strongly---provided, of course, that the scale of the true field is greater than that of the error field.

5. Dependence of the Degree of Smoothing on Scale and "Noise-to-Signal" Ratio.

Since the precise form of the weighting function is probably not critical, the greatest interest centers on the optimum degree of smoothing. The latter depends on the scale parameter α , which is inversely proportional to the effective radius of the domain of averaging. The manner in which the degree of smoothing depends on the characteristics of the true field and error field is shown in Fig. 1, on which $5\alpha d$ (as computed from Eq. 9) is plotted as a function of the signal-to-noise ratio and the ratio of the scales of the true field and error field. (The length d is a typical distance between adjacent observing stations). This diagram illustrates the points that were made earlier ---- namely, that the field of observations should not be smoothed if the signal-to-noise ratio is extremely large (i.e., if the error is very small), or when the scales of the true field and the error field are comparable. Otherwise, the most striking feature of these results is that the optimum degree of smoothing depends very strongly on the scale ratio, the effective radius of the domain of averaging being 4 times as great for a scale ratio of 10 as it is for a scale ratio of 2.

With reference to meteorological variables, the ratio of the scale of the true field to that of the error field probably ranges from 3 to about 6 over areas of good data coverage, and from 1 to 3 over the oceans. The signal-to-noise ratio probably varies from about 10 for the long planetary waves to about 3 for minor synoptic features of shorter wavelength. The region of relative scale and

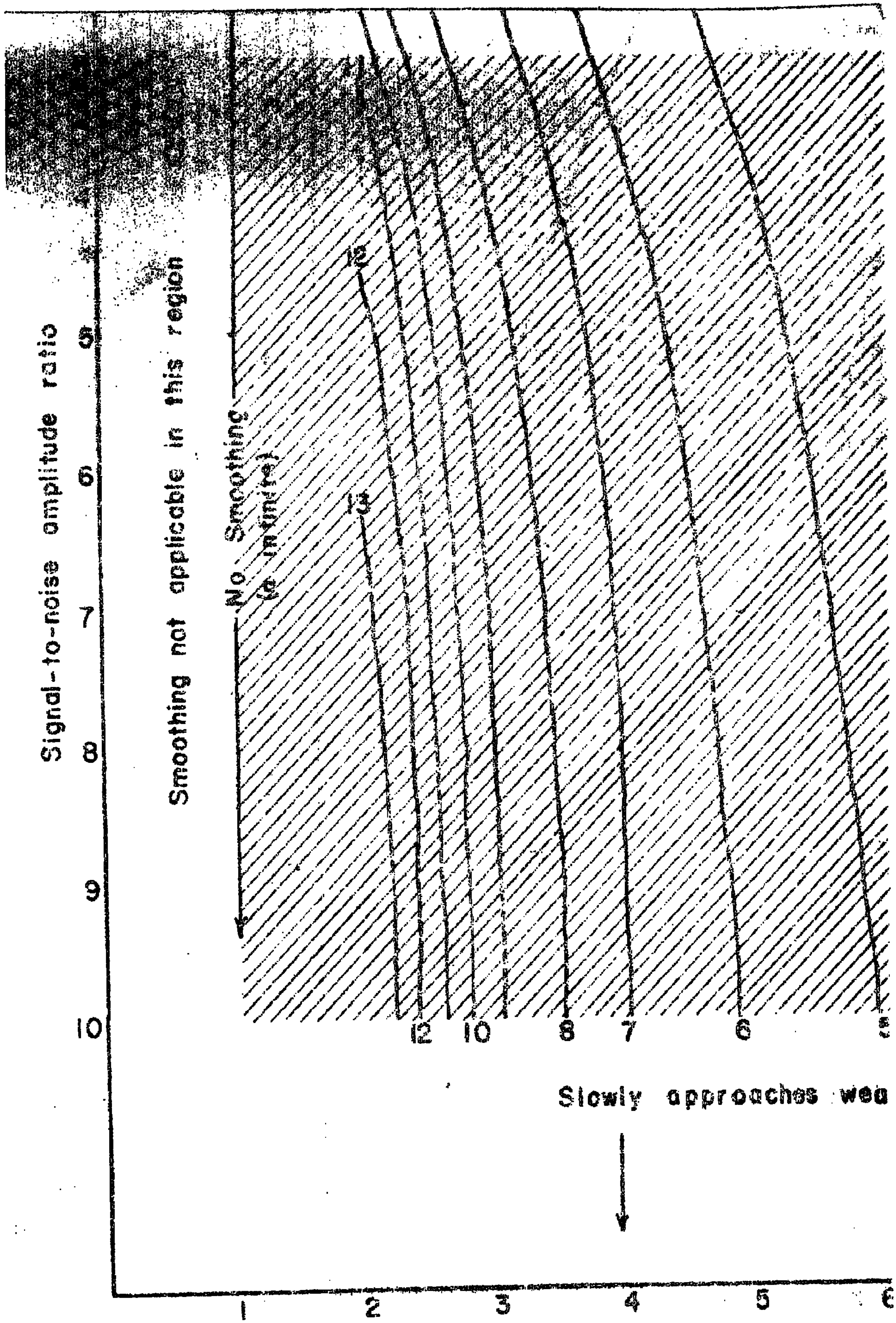

Signal-to-noise amplitude ratio
Smoothing not applicable in this region
No Smoothing
(n infinite)
5
6
7
8
9
10
15
13
12
10
8
7
6
5
Slowly approaches wea
1
2
3
4
5
6
Signal-to-noise scale
Ratio

Fig 1

noise level delineated by these ranges is the shaded area in Fig. I. It is readily seen that the optimum effective radius of the domain of averaging may vary by a factor of 3, a fact that should be taken into account in the analysis of meteorological data. It is also likely, of course, that a skilled analyst introduces qualitative considerations of a similar nature in deciding how much weight is to be assigned to any single piece of data in an irregularly spaced network of observations. It is a matter of some interest, therefore, to compare the results of optimum smoothing with those of the subjective smoothing practiced by an experienced analyst.

6. A Comparison of Optimum Smoothing With Subjective Smoothing.

In order to compare the subjective and optimum processes of smoothing from the standpoint of minimizing the RMS difference between the true field and the smoothed data field, both processes were applied to the same synthetic field of "observations", consisting of an artificially constructed "error" field superimposed on a specified true field. For purposes of isolating differences due to method alone, it is sufficient to deal with an assumed true field that is more or less typical of real meteorological conditions. Accordingly, a set of height values at the points of a square grid, interpolated to the nearest 10 feet between the contours shown in Fig. 2(a), were regarded as correct. The contours had been drawn to fit real data for 1500Z, 22 April 55 ---- a case that was deliberately selected as one in which the spectrum of wave components was fairly broad.

The "error" field was constructed by matching a random sequence

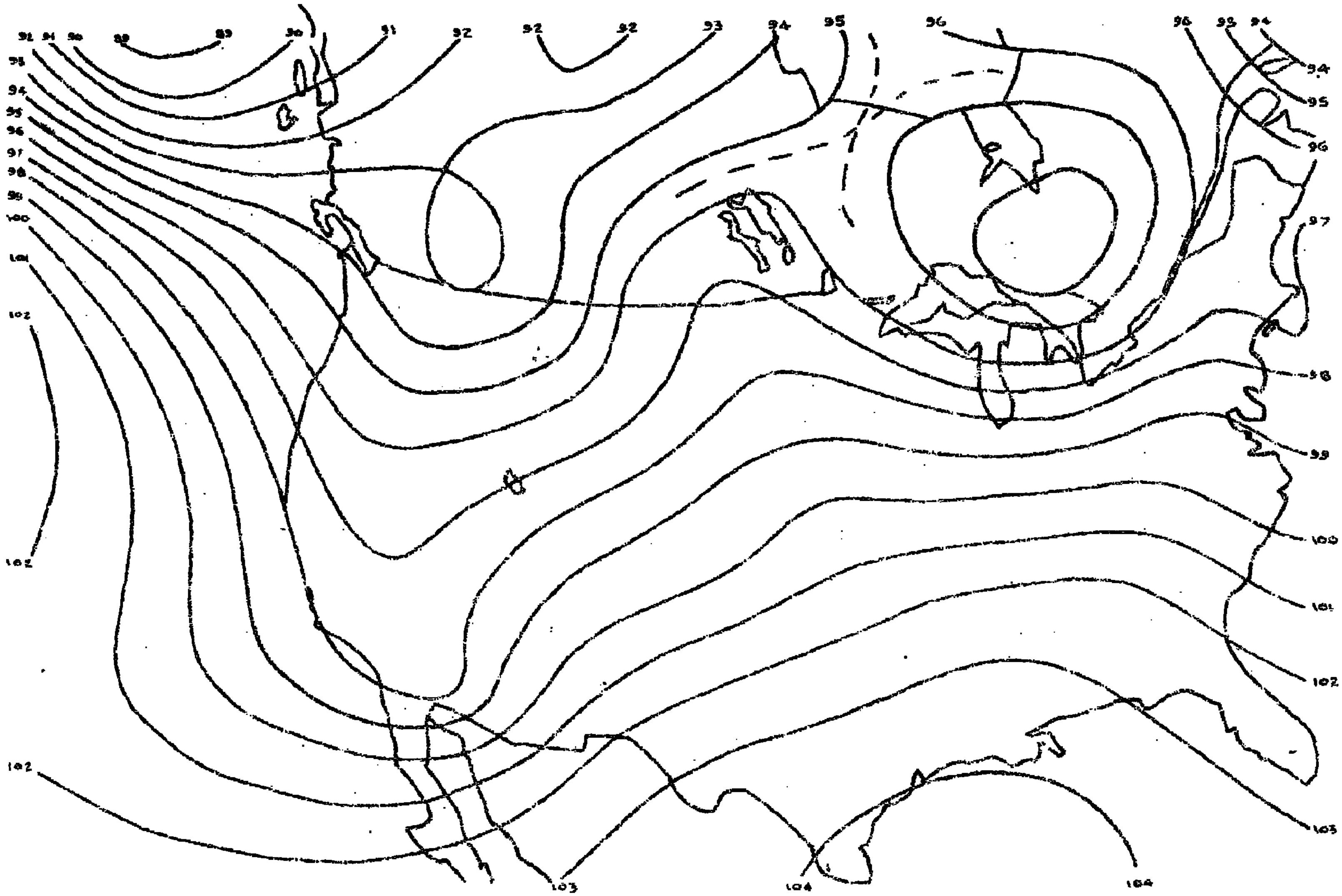


Fig 2(a)

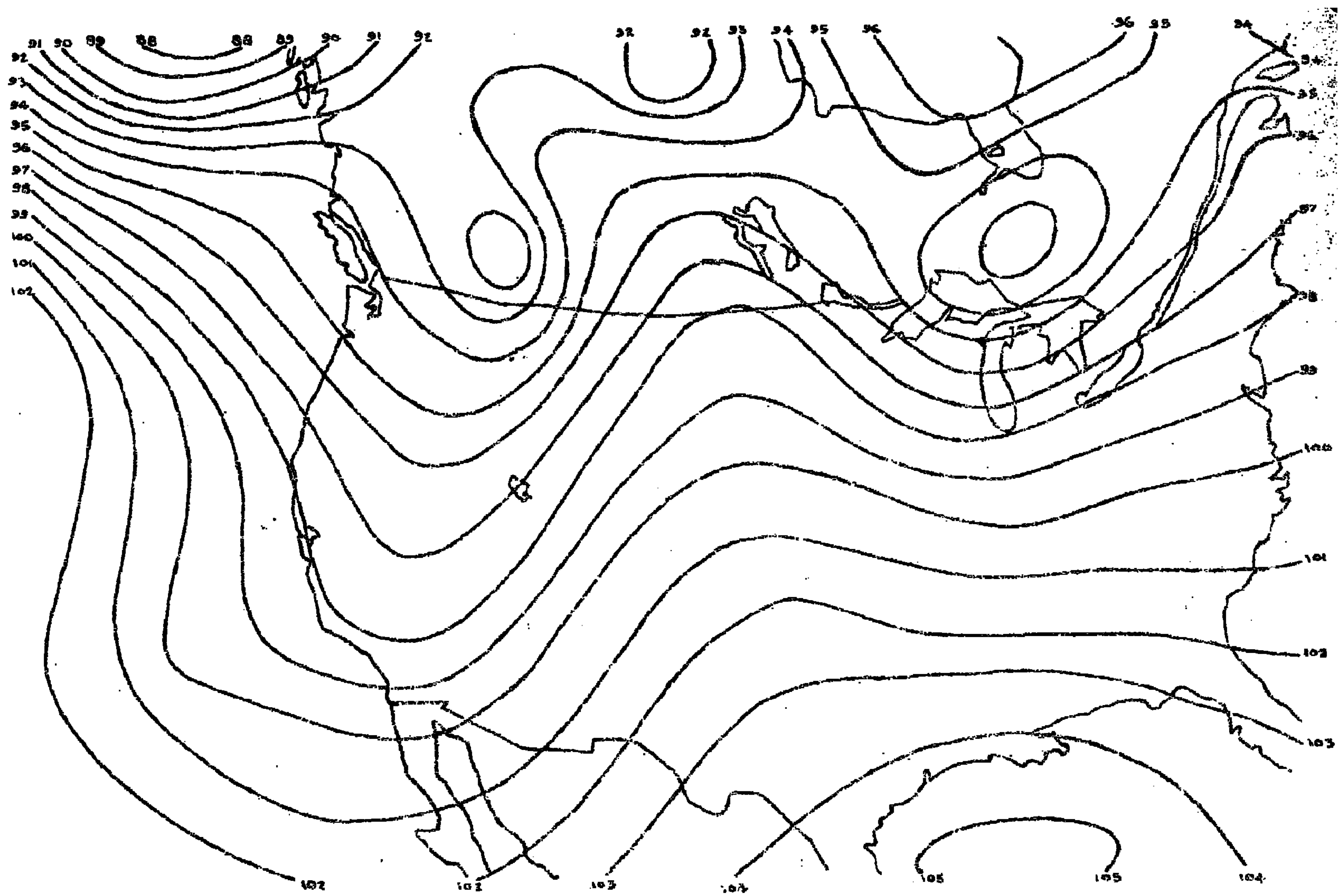
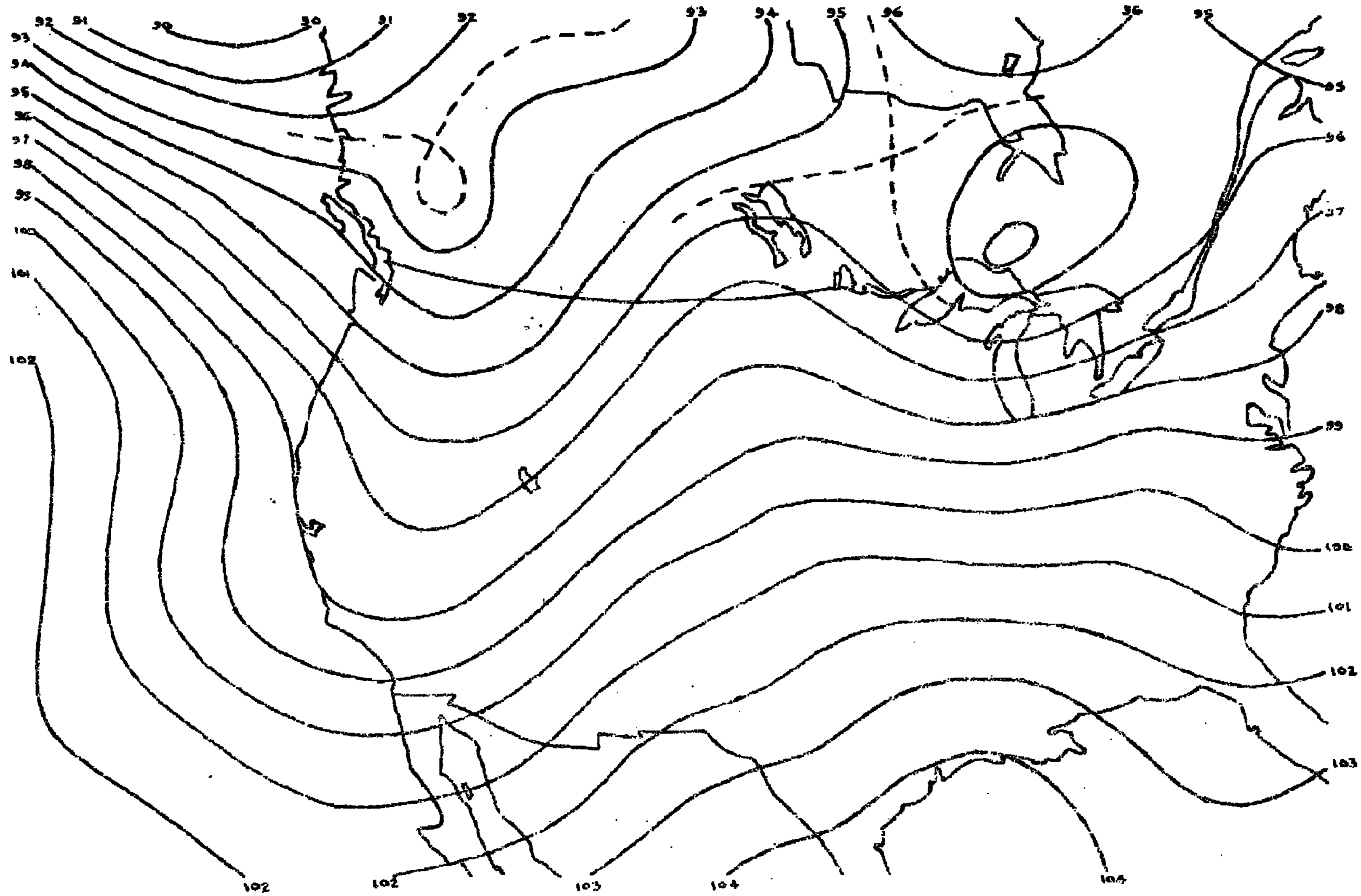


Fig. 2(b)



92
91
90
90
91
92
93
94
95
96
96
95
93
94
95
96
97
98
95
100
101
102
95
96
97
98
99
100
101
102
103
102
102
103
104
104

Fig. 2(c)

of errors to a predetermined sequence of gridpoints, spaced about 200 miles apart. The sequence of errors was formed by letting each two-digit group in a list of random numbers determine the magnitude and sign of each error in the sequence. The correspondence between the sequences of errors and random numbers was arranged in such a way that the frequency distribution of errors is normal, with an RMS error of 100 feet. The "error" field so formed is fairly typical of the combined fields of instrumental, reading, height evaluation, and roundoff errors, although somewhat exaggerated to accentuate differences of method.*

The "observations" at the gridpoints were obtained simply by adding the "error" field to the assumed true field. The resulting field of observations was then analyzed by an experienced meteorologist, who was informed that the field contained random errors with a normal frequency distribution and RMS error of 100 feet, but who did not have access to the true field. The same field of observations was also subjected to the process of optimum smoothing, by computing the weighted average for each gridpoint as a weighted sum of the "observed" heights at that point and at the 20 points nearest to it. The weighting factors were calculated from Eq. (7), the computed autocorrelation functions for the assumed true field and the artificial error field being used to estimate the scale ratio and signal-to-noise ratio.

*The mean square total error is approximately the sum of the mean squares of the individual types of error, provided different types of error are not highly correlated.

The results of this experiment are summarized in Figures 2(b) and (c), which show the subjectively smoothed and "optimum" averaged fields, respectively. These are to be compared with Figure 2(a), which shows the corresponding "true" field. In general, it appears that the subjective process of "analysis" and the optimum averaging process perform the function of smoothing in approximately the same way, since the subjectively smoothed and optimum averaged fields are more similar to each other than either is to the true field. Each of them removes fluctuations whose wave lengths are comparable with twice the distance between observation points (e.g., the component with wavelength about 900 kilometers, associated with the rather "square" trough and ridge over the northwestern U.S.), but retains the large-scale features of greater amplitude. Both produce about the same degree of smoothness.

Over most of the area considered, there is no significant and systematic difference between the two smoothed fields. Over the southeastern quadrant, however, the subjectively smoothed field is noticeably smoother than the optimum average ---- too smooth, in fact, to reproduce the very weak trough in the southwest flow over western U.S. (where it actually shows anticyclonic contour curvature), or the shallow trough along the east coast. The optimum average, on the other hand, does reflect these minor synoptic features, though with somewhat reduced amplitude. The most obvious reason for this difference is that the scale of the features mentioned above is considerably less than that of the large-scale pattern, but is still somewhat greater than twice the distance between observation points. Even a skilled analyst probably finds it difficult to distinguish qualitatively between random small-scale fluctuations and real disturbances of only slightly larger scale and, for this reason, thinks in terms of the scale of the predominant wave band in the spectrum. The optimum smoothing process, however, does distinguish between random and real fluctuations ---- provided the scale of the latter is greater than that of the error field, and if the existence of small-scale disturbances is reflected in the auto-correlation function for the true field. In short, whatever difference there is between the subjective and optimum processes of smoothing is probably due to differences in the fidelity with which they reproduce fluctuations whose scale is intermediate between that of the error field and the major synoptic features of the true pattern. That there is a significant difference in performance is revealed in the RMS errors of the smoothed fields ---- 52 feet in the case of subjective smoothing, and 42 feet for optimum smoothing.

7. The Information Value of a Network of Data in Smoothing Out Non-systematic Error.

It was emphasized earlier that the mean square error of the smoothed field is expressible in terms of the autocorrelation functions ---- which, in the cases studied here, are assumed to have the form of the Gaussian error function. Since the optimum weighting function for this type of autocorrelation function is now known, it is possible to calculate the minimum RMS error attainable by "linear" smoothing, which is a measure of the maximum amount of information about the true field that can be extracted from an error-contaminated field of observations. Substituting from Eq. (9) into Eq. (2), we may write the minimum mean square error I_{min} of the smoothed field as:

$$I_{min} = \gamma(0) \Big\{ 1 - 2\alpha \int_0^\infty J_1(\alpha\rho)\, e^{-b^2\rho^2}\, d\rho \tag{10}$$

$$+ \frac{\alpha^2}{2\pi} \int_0^\infty J_1(\alpha\rho)\, d\rho \int_0^\infty J_1(\alpha R)\, dR \int_0^{2\pi} \left(e^{-b^2 r^2} + k^2 e^{-a^2 r^2} \right) d\theta \Big\}$$

The first integral in the curly brackets is one of the types mentioned before (Watson, p. 394). The triple integral can be simplified considerably by noting that:

$$\int_0^{2\pi} \phi(r)\, d\theta = 2\pi \int_0^\infty J_0(Ru)\, J_0(\rho u)\, F(u)\, u\, du$$

where $F(u)$ is the Fourier-Bessel transform of $\phi(r)$. Introducing this identity and inverting the order of integration, we find that the third term in the curly brackets of Eq. (10) takes the form:

$$\alpha^2 \int_0^\infty \left[\int_0^\infty J_1(\alpha\rho)\, J_0(\rho u)\, d\rho \right]^2 F(u)\, u\, du = \int_0^\alpha F(u)\, u\, du$$

The integral in square brackets above is one of the so-called discontinuous integrals (Watson, p. 406), Finally, we substitute all of these results into Eq. (10) to obtain the mean square error of the smoothed field as a function of the scale ratio S and the noise-to-signal ratio k .

$$I_{min} = \gamma(0)\left\{\left(\frac{k}{S}\right)^{\frac{2S^2}{S^2-1}} + k^2\left[1-\left(\frac{k}{S}\right)^{\frac{2}{S^2-1}}\right]\right\} \tag{11}$$

in which $S = a/b$. It has been shown previously that there are two cases in which the data field should not be smoothed ---- namely, when there is no error, or when the true field and the error field are of the same scale. In the former case $(k = 0)$, Eq. (11) verifies that the mean square error of the smoothed field (or, more precisely, the unsmoothed field) is zero. In the latter case $(S = 1)$, Eq. (11) reduces to:

$$I_{min} = k^2\gamma(0) = \frac{E(0)\,\gamma(0)}{\gamma(0)} = E(0)$$

This simply shows that the minimum mean square error of the "smoothed" field in the case when $S = 1$ is the mean square error of the reported measurements.

The general way in which the minimum RMS error of the smoothed field depends on the spacing of observation stations is illustrated in Figure 3, on which the RMS error (expressed as a fraction of the RMS error of reported measurements) is plotted as a function of the

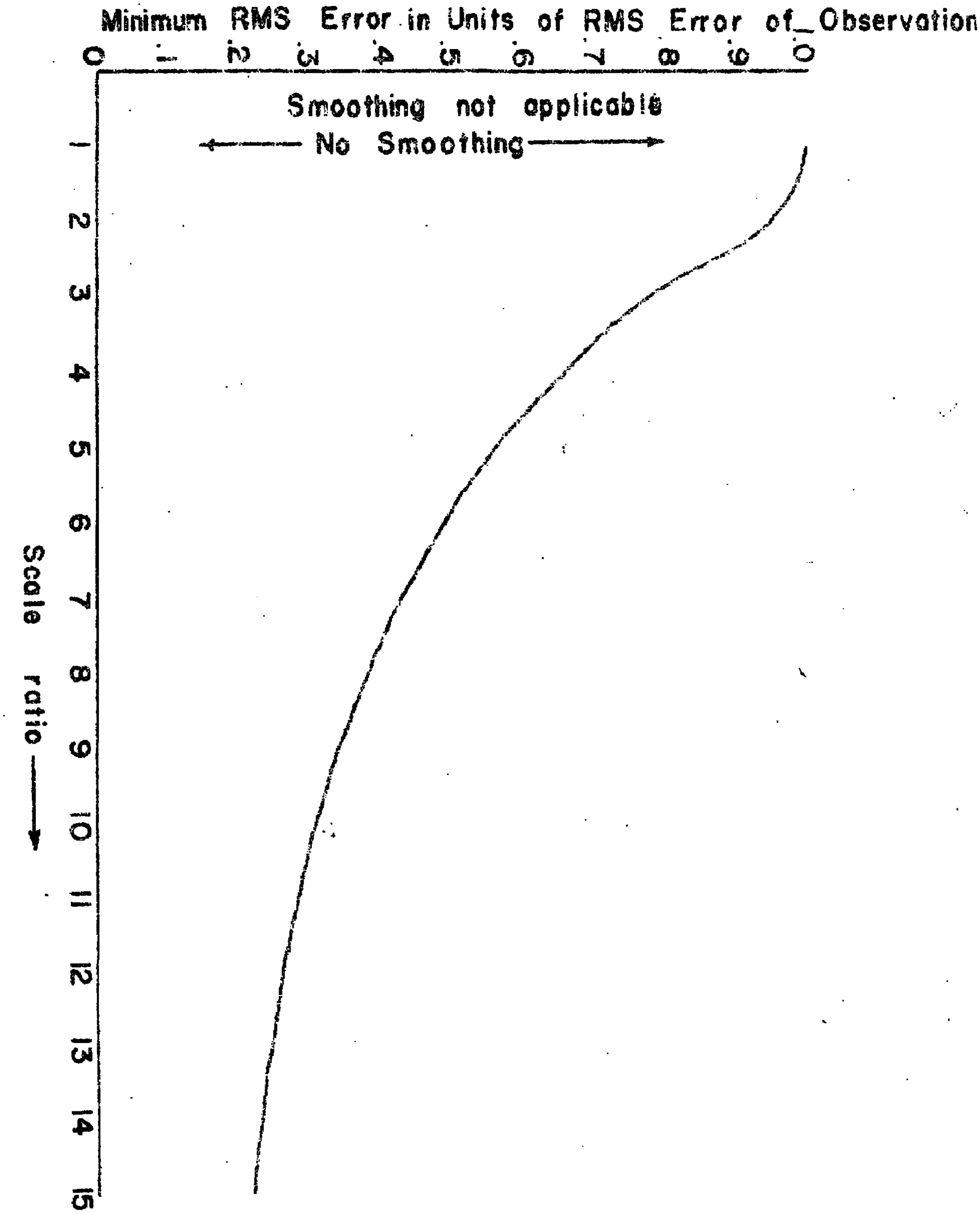

Minimum RMS Error in Units of RMS Error of Observation
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
Smoothing not applicable
No Smoothing
Scale ratio
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Fig. 3

scale ratio. Since the optimum weighting function does not change markedly with varying noise-to-signal ratio, the calculations were carried out for a single typical value of $\mathcal{R}$, equal to one-tenth. Figure 3 shows that the minimum RMS error is reduced only slightly by increasing the density of the observing stations if the scale ratio is greater than 6, but indicates that a considerable gain in accuracy is to be made by increasing the scale ratio from 2 to about 5.

In interpreting these results, it should be borne in mind that the foregoing analysis has dealt entirely with a *continuous* true field and an error field that is *simulated* by a continuous field, whose scale is determined by the distance between *discrete* observing points. As pointed out earlier, application of the smoothing operation to data at a network of discrete points requires that the weighted average of *continuous* observations over a ring-shaped region of finite width be approximated by the unweighted average of the observations at *discrete* points in the ring, multiplied by the integral of the weighting function taken over the ring. This approximation is clearly best when the number of observations in the effective domain of averaging is large enough to constitute a representative sample of nonsystematic errors, and when the width of the ring is much less than the scale of the true field. Both of these conditions, of course, are enhanced by a large scale ratio. Moreover, it is evident that the discrete point approximation is most nearly valid when applied at the observation points, where bona fide data figure most heavily in the smoothed field. This simply means that the use of the smoothing process as a combined operation of smoothing and interpolation does not result in any more real information about the detail of the true field than is revealed in the original observations.* In short, one must distinguish between the value of a network of observations as a means of resolving the detailed structure of the true field, and its value as a means of reducing nonsystematic error at each point of the network. At present, of course, we are concerned solely with the latter.

Despite these cautionary remarks, it turns out that the results of an analysis of continuous smoothing may be used to estimate the gain of information (reduction of error) that can be expected from averaging data at discrete points. According to Figure 3, a good compromise between accuracy and economy is attained when S is about 5, which corresponds to a linear density of about 15 stations spread over a characteristic wave length of the true field. For typical wavelengths of disturbances in the flow aloft, a reasonable spacing of stations is around 200 miles. It should be pointed out, however, that it is probably more economical to reduce the errors of the smoothed field by reducing the errors of reported measurement, rather than by smoothing over a very dense network.

*This is especially obvious in the case of no error, when the weighting function is zero everywhere, except at the origin.

The remaining problem ---- that of estimating the information value of a network of essentially error-free observations ---- clearly involves the details of the use to which the data are put. For example, if the observations are to be used as initial data for a numerical forecast, the information value of the network depends on the errors of approximating derivatives by finite differences over a typical distance between neighboring observation stations. Such considerations are beyond the scope of the present paper.

8. Summary and Conclusions.

This paper has dealt exclusively with the problem of smoothing out nonsystematic errors of reported measurements or other small-scale fluctuations that are random with respect to fields of larger scale. By extending Wiener's methods for smoothing stationary time series to two-dimensional fields whose autocorrelation functions are invariant with rotation, it has been found possible to define and analyze an "optimum" linear smoothing process ---- "optimum" in the sense that the RMS difference between the true and smoothed fields is the least. The optimum smoothed field is essentially a weighted area average, whose weighting function is determined by the "signal-to-noise" ratio and the autocorrelation functions for the true field and "error" field.

The form of the optimum weighting function has been investigated in detail for fields whose autocorrelation functions have the shape of the Gaussian error function ---- a type that is more or less characteristic of aperiodic fields. In this case, the optimum weighting function is a linear combination of Bessel functions, whose rate of decrease away from the origin depends on the signal-to-noise ratio, the ratio of the scales of the true field and "error" field, and the distance between neighboring observation stations. The manner in which the radius of the effective domain of averaging depends on the scale ratio and the signal-to-noise ratio substantiates two conclusions that were previously established under more general conditions ---- namely, that the field of observations should not be smoothed at all if the noise-to-signal ratio is very small (in which case there is obviously nothing to gain by smoothing) or if the scale of the true field is the same or less than that of the "error" field (when reduction of error cannot be attained by smoothing, without simultaneously reducing the amplitude of the true field in the same or greater degree). When the signal-to-noise ratio is very small, the optimum degree of smoothing approaches a limit that is independent of noise level; in this case, the best strategy is to average over a very large region, whose effective radius is determined by the relative scales of the true field and error field. In the normal range of signal-to-noise ratio, the optimum radius of the effective domain of averaging depends most strongly on the scale ratio, varying by a factor of 3 or 4 from one extreme to another. Although a skilled analyst probably introduces qualitative considerations of a similar nature in smoothing observations over an irregularly spaced network of stations, it is suggested that the objective and quantitative character of the optimum linear smoothing process may prove advantageous from the standpoint of maintaining consistently high performance over a variety of ordinary situations.

The subjective and "optimum" processes of smoothing have been applied to the same synthetic field of "observations", consisting of an artificial field of random errors superimposed on an assumed (but fairly typical) "true" field. Comparisons between the two smoothed fields, and between each smoothed field and the "true" field show that the two different processes accomplish about the same type and degree of smoothing. In the one case presented here, the only significant difference was a reflection of the fact that the optimum smoothing process is more capable of distinguishing small-scale random fluctuations from minor synoptic features of only slightly greater scale. The RMS errors of the subjectively smoothed field and the optimum average were 52 and 42 feet, respectively, indicating that the optimum linear smoothing process can probably match the performance of a skilled analyst, and may exceed it.

Finally, the minimum RMS error of linearly smoothed fields has been expressed in terms of the signal-to-noise ratio, the standard deviation of the "true" field, and the scale ratio (which, in turn, depends on the spacing of observation stations). This result shows that the information value of a network of observations can be enhanced considerably by increasing the linear density of observing stations from 6 to about 15 per wavelength, but is not substantially increased beyond a linear density of 18 stations per wavelength. It is to be emphasized that this conclusion applies only to the information value of a network from the standpoint of smoothing out nonsystematic error, and does not bear directly on the problem os interpolation or of reducing truncation error.

Acknowledgement

The author gratefully acknowledges his indebtedness to Dr. H. M. Mott-Smith of Norden Laboratories for several valuable and illuminating discussions of this problem, and particularly for pointing out a number of general properties of the Fourier-Bessel (or Hankel) transforms of correlation functions.

Thanks are also due to Mr. E. B. Fawcett of the Joint Numerical Weather Prediction Unit for his careful subjective analysis of the synthetic field of observations referred to in Section 6.

References

Watson, G.N., 1948: Theory of Bessel Functions (2nd Ed.). Cambridge University Press, London. 804 pp.

Wiener, N., 1949: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Technology Press of Massachusetts Institute of Technology, Cambridge, Mass. 163 pp.

information value of a network from the standpoint of smoothing out nonsystematic error, and does not bear directly on the problem os interpolation or of reducing truncation error.

Acknowledgement

The author gratefully acknowledges his indebtedness to Dr. H. M. Mott-Smith of Norden Laboratories for several valuable and illuminating discussions of this problem, and particularly for pointing out a number of general properties of the Fourier-Bessel (or Hankel) transforms of correlation functions.

Thanks are also due to Mr. E. B. Fawcett of the Joint Numerical Weather Prediction Unit for his careful subjective analysis of the synthetic field of observations referred to in Section 6.

References

Watson, G.N., 1948: Theory of Bessel Functions (2nd Ed.). Cambridge University Press, London. 804 pp.

Wiener, N., 1949: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Technology Press of Massachusetts Institute of Technology, Cambridge, Mass. 163 pp.