UDC 551.509.313/.314

SPECTRAL MODIFICATION BY OBJECTIVE ANALYSIS [1]

J. J. STEPHENS and A. L. POLAN
The Florida State University, Tallahassee, Fla.

ABSTRACT

The power transfer function for the objective analysis of scalar fields by linear interpolation is determined with a simple statistical model of the data distribution. The method of successive corrections is used in a computational verification of the result.

It is shown that the spectral filtering by interpolation is due to distance-dependent weighting and discrete sampling. True spectra can be inferred from interpolated waves that are long with respect to average sampling interval.

1. INTRODUCTION

The interpolation phase of objective analysis usually provides field estimates on a regular net through linear combinations of station observations at irregularly distributed locations. These estimates may be adjusted in an initialization for a prognostic model, or they may be used directly in a diagnostic analysis of the data. In either case, the relation between the observed and interpolated spectra is of interest.

Spectral modifications associated with linear operations in a data continuum or in regularly spaced discrete arrays are easily determined. However, the analysis is more difficult when the observation sites are irregularly distributed, and results derived for continuous fields have been used to infer the discrete response (Stephens 1967). Stephens and Stitt (1970) have introduced a simple statistical model of the distribution of observing sites that can be used to find the average response for a linear arithmetic operation on randomly distributed data in the plane. It is modified here to find the power transfer functions for interpolation with a simple scan technique and for the method of successive corrections introduced by Cressman (1959). The Cressman weight function is used in a computational verification of the formulation.

2. AVERAGED ESTIMATES

In application to a scalar field, successive correction estimates $Z^e_{j,k}$ at the grid point $(j\Delta x, k\Delta y)$ are generated from samples of the observed field Z^o and a guess field Z^g by a correction formula of the type

$$Z^e_{j,k}=Z^g_{j,k}+\sum_{m=1}^{M} w_n(r_m,R,M)\{Z^o(j\Delta x+r_m\cos\theta_m,\ k\Delta y+r_m\sin\theta_m)-Z^g(j\Delta x+r_m\cos\theta_m,\ k\Delta y+r_m\sin\theta_m)\} \quad (1)$$

where M is the number of stations within the influence radius R. The weight function $w_n(r_m,R,M)$ is normalized to unity by

$$w_n(r_m,R,M)=\frac{W(r_m,R)}{\sum_{m=1}^{M}W(r_m,R)}. \quad (2)$$

Here, $W(r_m,R)$ is the weight associated with the mth ordered station in which the coordinates are (r_m,θ_m) in a local polar coordinate system.

Station positions vary with each grid point, so that analysis of eq (1) is difficult. However, if the locations are randomly distributed over the domain, then the average response can be obtained. The analysis given below is patterned after that shown by Stephens and Stitt (1970).

Under the random distribution assumption, r_m and θ_m are random variables with associated probability densities $p_m(r_m,d)$ and $f_m(\theta_m)$, respectively. Here, d is the average station separation defined such that, if there are N stations in the total area A, $Nd^2=A$. The average weight for each station is then determined by averaging its location over all possible values.

The Fourier representations

$$Z^o(x,y)=\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}O_{l\kappa}e^{iK_x lx+iK_y\kappa y} \quad (3)$$

and

$$Z^g(x,y)=\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}G_{l\kappa}e^{iK_x lx+iK_y\kappa y}, \quad (4)$$

where $K_x=2\pi/L_x$ and $K_y=2\pi/L_y$, are introduced in eq (1) for interpolation in a rectangular domain of dimensions $L_x\times L_y$:

$$Z^e_{j,k}=Z^g_{j,k}+\sum_{m=1}^{N}w_n(r_m,R,N)\left\{\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}(O_{l\kappa}-G_{l\kappa})\times e^{iK_x lj\Delta x+iK_y\kappa k\Delta y}e^{iK_x lr_m\cos\theta_m+iK_y\kappa r_m\sin\theta_m}\right\}. \quad (5)$$

Since all stations in the domain may appear within the influence area for a particular realization of the data array, all are included in eq (5). A field estimate averaged over all possible station distributions and including all stations is then obtained by

$$\overline{Z^{\theta_m,r_m}_{j,k}}=\int_0^{\infty}\int_{-\pi}^{\pi}Z^e_{j,k}p_m(r_m,d)f_m(\theta_m)d\theta_m dr_m. \quad (6)$$

Since all values of θ_m in $(-\pi,\pi)$ are equally likely,

$$\int_{-\pi}^{\pi}f_m(\theta_m)e^{iK_x lr_m\cos\theta_m+iK_y\kappa r_m\sin\theta_m}d\theta_m=J_0(r_mK_{l\kappa}) \quad (7)$$

[1]Research supported by the U.S. Army Electronics Command, Contract No. DAA07-69-C-0062 under Project THEMIS.

where

$$K_{l\kappa}=\sqrt{l^2K_x^2+\kappa^2K_y^2}$$

and J_0 is the Bessel function (Oliver 1964) of the first kind and order zero. Then, with suppressed average notation, the grid point estimate can be written as

$$Z^e_{j,k}=Z^g_{j,k}+\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}(O_{l\kappa}-G_{l\kappa})e^{iK_x lj\Delta x+iK_y\kappa k\Delta y}I_{l\kappa}(N,\ R) \tag{8}$$

where

$$I_{l\kappa}(\beta,\ R)=\sum_{m=1}^{\beta}\int_0^{\infty}p_m(r_m,\ d)w_n(r_m,\ R,\ \beta)J_0(r_mK_{l\kappa})dr_m. \tag{9}$$

In general, all N stations will not be in the influence area. There is a probability $P_N\{r_N<R\}$ that the Nth ordered station is within the influence area and a probability $(1-P_N)$ that it is not. Accordingly,

$$Z^e_{j,k}=(1-P_N)Z^e_{j,k}(N-1)+P_N\{Z^g_{j,k}+A_{j,k}(N)\} \tag{10}$$

where $Z^e_{j,k}\ (N-1)$ is the estimate when the Nth station is outside the influence radius and the notation

$$A_{j,k}(N)=\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}(O_{l\kappa}-G_{l\kappa})e^{iK_x lj\Delta x+iK_y\kappa k\Delta y}I_{l\kappa}(N,\ R) \tag{11}$$

has been introduced. Similarly, there is a probability P_{N-1} that $r_{N-1}<R$ when $r_N>R$ so that

$$Z^e_{j,k}(N-1)=(1-P_{N-1})Z^e_{j,k}(N-2)+P_{N-1}\{Z^g_{j,k}+A_{j,k}(N-1)\}. \tag{12}$$

This reasoning can be continued until only the closest station is left for consideration. If it is not within R, then the guess field is used as the estimate. Thus,

$$Z^e_{j,k}(1)=(1-P_1)Z^g_{j,k}+P_1\{Z^g_{j,k}+A_{j,k}(1)\}. \tag{13}$$

These results can be combined by induction to write

$$Z^e_{j,k}=Z^g_{j,k}+\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}H_{l\kappa}(O_{l\kappa}-G_{l\kappa})e^{iK_x lj\Delta x+iK_y\kappa k\Delta y} \tag{14}$$

where

$$H_{l\kappa}=\sum_{\beta=1}^{N}h_\beta I_{l\kappa}(\beta,R), \tag{15}$$

$$h_N=P_N$$

and

$$h_\beta=P_\beta\prod_{\alpha=\beta+1}^{N}(1-P_\alpha). \tag{16}$$

Properly, P_β is the probability that the βth station is within R when the $(\beta+1)$ is not. However, in the computations shown below, the empirically determined $p_\beta(r_\beta,d)$ from which the P_β are derived are independent rather than joint distributions.

This analysis differs from that of Stephens and Stitt (1970) only in that they normalized the weights under the assumption that all N stations were always present. That approximation simplifies the form of $H_{l\kappa}$, but the results differ appreciably from those presented here only for wavelengths much greater than the data separation and for influence radii less than the minimum definable wavelength. The relative behavior with respect to changes in (R/d) for a given wavelength is not changed appreciably by this more complete formulation.

The simple scan technique chosen for examination here generates grid point estimates by means of the linear combination

$$Z^e_{j,k}=\begin{cases}\sum_{m=1}^{M}w_n(r_m,\ R,\ M)Z^o(j\Delta x+r_m\cos\theta_m,\\ \qquad\qquad k\Delta y+r_m\sin\theta_m) & M\neq 0\\ Z^g_{j,k} & M=0.\end{cases} \tag{17}$$

This differs from the method of successive corrections in that a guess field is used only if there are no stations present. The averaged counterpart of eq (17) is

$$Z^e_{j,k}=\sum_{l=-\infty}^{\infty}\sum_{\kappa=-\infty}^{\infty}O_{l\kappa}H_{l\kappa}e^{iK_x lj\Delta x+iK_y\kappa k\Delta y}+Z^g_{j,k}\prod_{m=1}^{N}(1-P_m). \tag{18}$$

3. TRANSFER FUNCTIONS

The Fourier coefficients $\{E_{l\kappa}\}$ for $Z^e(x,y)$ can be introduced in eq (14) to conclude that

$$E_{l\kappa}=H_{l\kappa}O_{l\kappa}+(1-H_{l\kappa})G_{l\kappa}. \tag{19}$$

The real quantity $H_{l\kappa}$ will be termed the distance influence function. It can be interpreted as the transfer or response function for the method of successive corrections only when the guess field is zero.

For assessing the relative roles of the observed and guess fields in generating the estimates, the distance influence function $H_{l\kappa}$ was calculated with the numerical techniques and probability densities used by Stephens and Stitt (1970). The results are shown in figure 1 for various combinations of influence radii, data separations, and wavelengths for $l=\kappa$. The distance influence or (amplitude) transfer function for a continuous distribution of data is taken from Stephens (1967). Both the continuous and discrete responses are based on the Cressman (1959) weight function

$$W(r_m,\ R)=\frac{R^2-r_m^2}{R^2+r_m^2}. \tag{20}$$

The results will vary with different weight functions.

If the guess field is disregarded for the moment, it can be concluded that the average effect of discrete sampling is to filter somewhat more than if the weight were applied to a continuous field. While the H_{ll} curve is nearly universal for long waves, there are systematic variations. For a given separation, an increase in influence radius leads to greater filtering. For a given ratio of influence radius to wavelength, a data separation increase decreases H_{ll}.

If attention is now restricted to waves in which lengths are greater than twice the influence radius, eq (19) can be used to infer that the estimated field is determined by

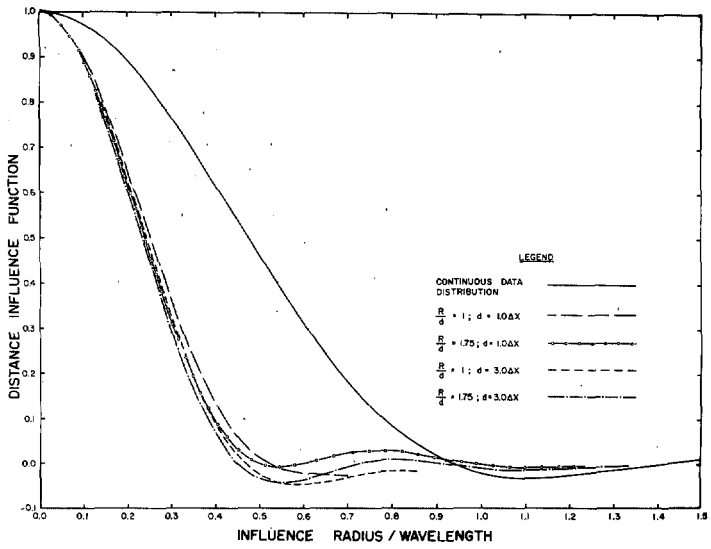

FIGURE 1.—Distance influence function for various influence radii and data separations as a function of the ratio of influence radius to wavelength.

the observations for long waves and by the guess field at shorter wavelengths. The latter is also true for wavelengths less than twice the influence radius. Clearly, any wave for which G_{lk} is an adequate representation of O_{lk} will be effectively reconstituted at the grid points.

The Fourier coefficients $\{E_{lk}^{(M)}\}$ for the Mth scan with the estimated field always used as the guess field for the subsequent pass are

$$E_{lk}^{(M)}=\left\{H_{lk}^{(1)}+\sum_{m=2}^{M} H_{lk}^{(m)} \prod_{k=1}^{m-1}(1-H_{lk}^{(k)})\right\} O_{lk} + \left\{\prod_{k=1}^{M}(1-H_{lk}^{(k)})\right\} G_{lk} \quad (21)$$

where $H_{lk}^{(k)}$ is the distance influence function evaluated with the influence radius used on the kth pass. In the special case where the same influence radius is used for each scan, eq (21) reduces to

$$E_{lk}^{(M)}=\left\{H_{lk}\left[1+\sum_{m=1}^{M-1}(1-H_{lk})^m\right]\right\} O_{lk}+(1-H_{lk})^M G_{lk}. \quad (22)$$

For all H_{lk} such that $(1-H_{lk})^2<1$,

$$\lim_{M\to\infty} E_{lk}^{(M)}=\frac{H_{lk}}{1-(1-H_{lk})} O_{lk}=O_{lk}. \quad (23)$$

The limit diverges for $H_{lk}<0$. This suggests that repeated scans coupled with selective low-pass filtering would yield an effective longwave analysis scheme.

Except for influence radii comparable to data separations where the guess field alternative is more probable, the distance influence function is the amplitude transfer function for the scan method. This corresponds to the case of a null guess field for the method of successive corrections.

4. COMPUTATIONAL VERIFICATION

A computational check of the formulation was made with the set of observed fields

$$Z_k^o(x,y)=\sin\left(\frac{2\pi x}{L}k\right)\sin\left(\frac{2\pi y}{L}k\right) \quad (24)$$

where $k=1, 2, \ldots, 10$; $L=21\Delta x$; and $\Delta x=1$. Random station locations were generated with an average separation $d=1\Delta x$, and eq (24) was evaluated at each site to form a set of observations. After removing the average from each observation, interpolated values were found with the method of successive corrections with the Cressman weight on a 21×21 net. Two scans were used with an initial guess-field of zero and $R=1.75\,d$. Discrete spectra were calculated for the interpolated array as well as for eq (24) evaluated at the grid points. Experimental power transfer functions were then calculated as the ratio of the interpolated variance to the true variance for each wave number. The process was repeated four times for the first four wave numbers and three times for $k=5$, 10. A different station array was generated for each of these realizations. The power transfer function for each experiment is shown in figure 2 with the average responses connected by dashed lines.

Theoretical power transfer functions for the two passes were determined from eq (22):

$$|E_{ll}|^2=H_{ll}^2|O_{ll}|^2 \quad (25)$$

for the first scan and

$$|E_{ll}|^2=\{2H_{ll}-H_{ll}^2\}^2|O_{ll}|^2 \quad (26)$$

for the second scan. These constitute the theoretical model to be used for comparison.

As shown in figure 2, there is substantial agreement between the experiments and the model summarized in (25) and (26). The experimental response for long waves is in good quantitative agreement, although rather consistently less than the theory.

The agreement for waves shorter than twice the influence radius is only qualitative, and the experimental values are consistently greater than the theory. The second-pass discrepancy is not due to the interpolation of the guess field to the stations. When following Stephens and Stitt (1970), it can be shown that eq (26) would be modified to

$$|E_{ll}|^2=\left\{2H_{ll}-H_{ll}^2\cos\left(K_x l\frac{\Delta x}{2}\right)\cos\left(K_y l\frac{\Delta y}{2}\right)\right\}^2|O_{ll}|^2 \quad (27)$$

if that secondary interpolation is included in the analysis. Since $H_{ll}<0$ in this shortwave region, the actual discrepancy is larger than that indicated. The agreement would be better for the long waves.

Aliasing is apparently responsible for the discrepancy. Spectra determined from the grid are necessarily band-limited; yet, because of the variability of station locations, waves of all lengths would be implied on a continuous

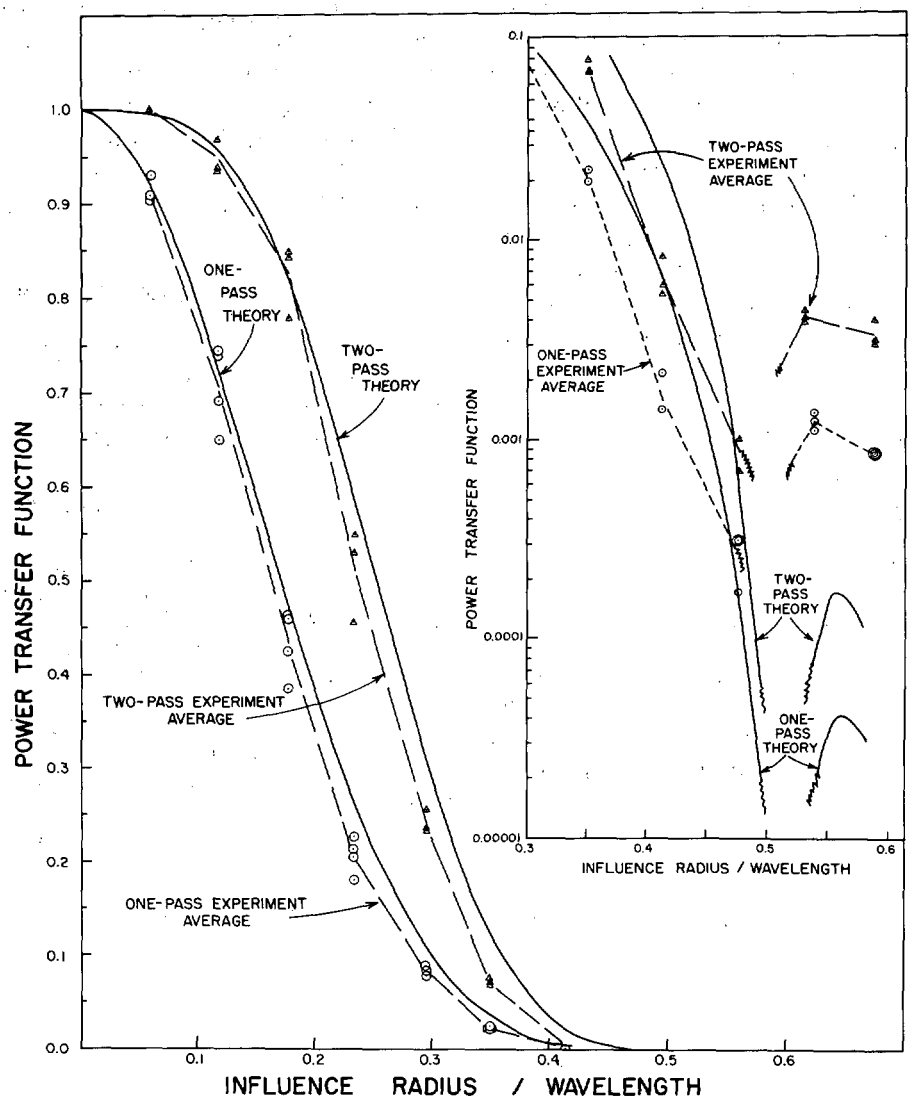

FIGURE 2.—Power transfer functions for one- and two-pass applications of the method of successive corrections with a zero guess field for $(R/d)=1.75$ and $d=\Delta x$.

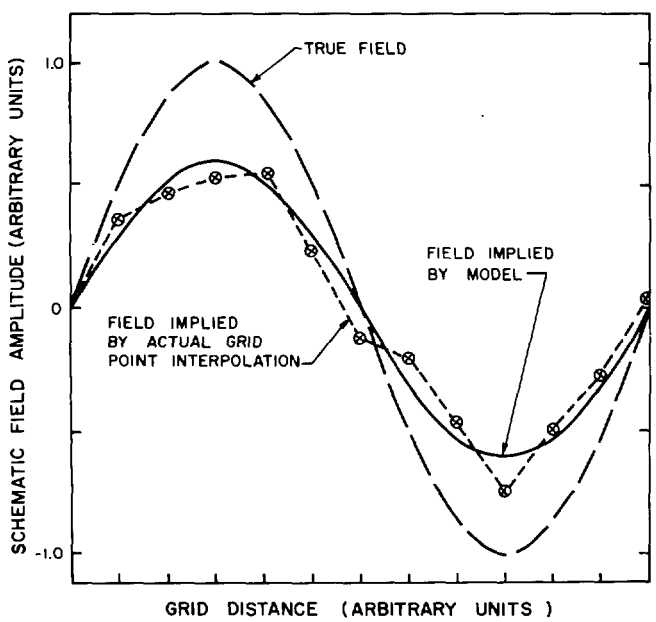

FIGURE 3.—Schematic illustration of the difference between the model and actual interpolations.

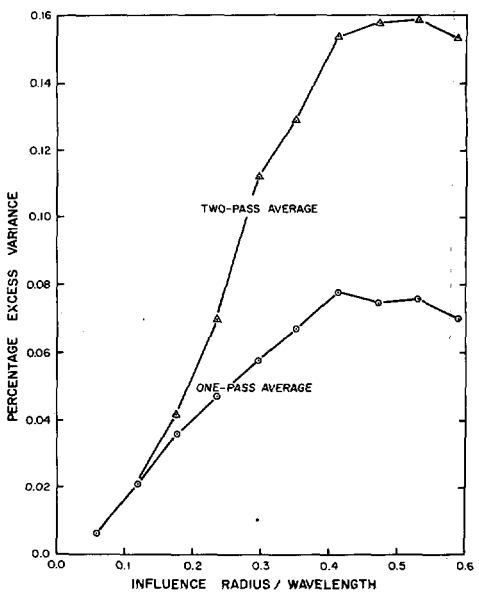

FIGURE 4.—Percentage excess variance for one- and two-pass applications of the method of successive corrections with a zero guess field for $(R/d)=1.75$ and $d=\Delta x$.

distribution of analyzed points. Waves less than twice the grid spacing are aliased into larger waves. An example of this is shown schematically in figure 3. The analyzed field would be that indicated by the dashed line. While its spectrum would reflect a principal contribution at the appropriate wave number, other wavelengths would also be represented. Further, the percentage contribution by aliasing to wavelengths other than the nominal one would be expected to increase with wave number. This is shown in figure 4 for both passes for $R=1.75\ d$. It was found that the excess variance decreased with increasing influence radius.

The results shown in figure 2 are for each of the $Z_k^o(x,y)$ evaluated separately. Computations for

$$Z^o(x, y)=\sum_{k=1}^{10} \frac{1}{k} Z_k^o(x, y) \tag{28}$$

show no significant differences.

5. CONCLUSIONS

Linear interpolation to a grid significantly modifies observed field spectra in the absence of an adequate guess field. The filtering derives from the use of a distance-dependent weighting and from discrete sampling. In general, the continuum response will not give an adequate measure of the filtering.

The average response model introduced here can be used to infer the true (observed) spectrum from that of the interpolated values, except for waves comparable to the shortest definable one in length.

While the data distributions used here were random, the results shown by Stephens and Stitt (1970) suggest that the model can be applied to actual distributions with success.

ACKNOWLEDGMENTS

The authors are indebted to Dr. O. Talagrand of the Geophysical Fluid Dynamics Laboratory, NOAA, at Princeton, N.J., for his constructive and complete review and to Mrs. Janina Richards for her care in typing the manuscript.

REFERENCES

Cressman, George P., "An Operational Objective Analysis System," *Monthly Weather Review*, Vol. 87, No. 10, Oct. **1959**, pp. 367–374.

Oliver, F. W. J., "Bessel Functions of Integer Order," *Applied Mathematics Series* 55, National Bureau of Standards, Washington, D.C., **1964**, 1046 pp. (see pp. 925–995).

Stephens, J. J., "Filtering Responses of Selected Distance-Dependent Weight Functions," *Monthly Weather Review*, Vol. 95, No. 1, Jan. **1967**, pp. 45–46.

Stephens, J. J., and Stitt, J. M., "Optimum Influence Radii for Interpolation With the Method of Successive Corrections," *Monthly Weather Review*, Vol. 98, No. 9, Sept. **1970**, pp. 680–687.

[*Received May 5, 1970; revised August 12, 1970*]