Interpolation and Filtering of Spatial Observations Using Successive Corrections and Gaussian Filters

M. A. PEDDER

Department of Meteorology, University of Reading, Reading, United Kingdom

(Manuscript received 10 August 1992, in final form 25 March 1993)

ABSTRACT

This paper describes a simple empirical analysis system based on Bratseth's method of successive corrections applied to detrended field data, which approximates an optimal interpolation of fields with a spatially variable mean sampled within a limited domain by scattered observations. As in other empirical interpolation schemes, the influence function that determines the weights applied to increment variables is chosen such that unresolvable scales tend to be strongly damped, even if the contribution from observation error is not represented in a formally equivalent correlation model for observed increment variables. Unlike most empirical successive correction schemes, the number of iterations is not necessarily considered as a prescribed analysis parameter. Instead, the number of iterations can be chosen on a judgemental, posterior basis such that the analysis approximates the observed field to within some acceptable limit.

The analysis generated by this form of successive corrections can be represented by a continuous function of location variables. Consequently, it is possible to express the result of posterior filtering of the analysis field as a weighted sum of "filtered" influence functions, each of which is defined by the convolution of an increment autocorrelation function with a continuous linear filter. If both the autocorrelation model and filter are based on a simple Gaussian function of spatial lag, then these convolution integrals can be solved analytically. This leads to a numerically inexpensive method of scale selection analysis that is in some ways less ambiguous than methods based on applying filters directly to the observed data.

The performance of the analysis system is demonstrated by applying it to simulated observations sampling two-dimensional fields.

1. Introduction

Following the pioneering work of Cressman (1959), empirically formulated successive correction schemes have been widely used in meteorology for interpolating and smoothing spatial fields sampled by scattered observations. All such schemes are based on the idea of correcting a prior estimate of the field at a particular location by adding to it a weighted mean of the differences between surrounding observations and their prior estimates, the weights applied to these increment data being calculated using a prescribed "influence" function of spatial separation.

A common approach to the design of empirical successive correction schemes is to consider the analysis system as a spatial filter acting on the observed field; analysis objectives are then expressed in terms of a theoretical wavenumber response that is valid in the limit of continuous spatial sampling. Barnes (1964) was the first worker to apply this idea to the objective design of an empirical successive correction scheme. His work was later extended by Maddox (1980), who developed a method of spatial filtering (generally referred to as "scale selection") based on a modified version of the Barnes analysis algorithm. In a recent review of methods for atmospheric data analysis, Daley (1991) notes that the Barnes algorithm is one of the most successful successive correction algorithms and that it is still widely applied.

Despite their advantages, the original Barnes method and the related method of scale selection analysis due to Maddox (1980) do not always perform well when applied to irregularly spaced observations on a spatial domain. In a recent study, Buzzi et al. (1991; henceforth BGPA) showed that in the case of inhomogeneous spatial sampling, a Barnes analysis could produce an unrealistic interpolation of the sampled field even when this is reasonably well resolved by error-free observations. They showed that it is possible to correct for this effect by comparing the analysis obtained directly from observing site data with that obtained via an interpolation of observing site data onto a uniform spatial grid. However, they also found that for the purposes of scale separation, even better results can be obtained if the Barnes algorithm is applied to uniform gridpoint data that have previously been interpolated from observing sites using surface splines. Unfortunately, an approach to scale separation based on surface spline

Corresponding author address: Dr. M. A. Pedder, Department of Meteorology, University of Reading, 2 Earley Gate, Whiteknights, P.O. Box 239, Reading, RG6 2AU, England.

interpolation followed by gridpoint filtering lacks the obvious advantages of the more direct successive correction method, namely, of being simple to formulate and numerically inexpensive (in relative terms) when applied to a very large database.

An alternative approach to improving the performance of analysis by successive corrections, without losing some of its advantages, is to replace the low-order iteration method introduced by Barnes with the multiple iteration scheme suggested by Bratseth (1986). This offers the theoretical advantage of converging toward the same "optimal" weighting scheme as a statistical interpolation algorithm, and can indeed be regarded as an iterative approach to statistical interpolation of background error variables (Lorenc 1986, 1992). Practical tests of Bratseth's method have been reported by Franke (1988) and Seaman (1988). Both of these studies considered the problem of correcting a background field, which is assumed to be known at both analysis and observing sites, and in effect, treated Bratseth's method primarily as an approximation to statistical interpolation of background error variables with known (or assumed) covariance and observation error statistics. Here, we consider the application of Bratseth's approach to the problem of interpolating observations in the absence of such useful prior information: we shall suggest that in such an application it is possible to use ideas similar to those used by Barnes as a basis for selecting suitable values for the prior parameters of the analysis system so as to combine good data resolution with effective suppression of measurement and representativity errors.

When a conventional successive correction approach is used as a basis of scale selection analysis, its net theoretical response is normally considered to predict the spectral relationship between the observed (input) field and the filtered analysis (output) field. However, in the case of inhomogeneous spatial sampling the actual (discrete) amplitude response of the analysis system cannot be considered as a spatially invariant property, and may differ substantially from its theoretical response (see, for example, BGPA). A similar problem arises due to the effect of domain boundaries, even when spatial sampling is relatively uniform (Pauley and Wu 1990). Consequently, the predicted theoretical response function is often an ambiguous and unreliable measure of scale selection properties. Here, we suggest an alternative approach to defining scale selection objectives, based on the idea of applying a continuous, spatially invariant filter to analysis rather than observed variables. The associated theoretical response is in this case also spatially invariant, and accurately describes the spectral relationship between the input field, which is a continuous approximation to the observed field, and an output analysis field. We show that it is easy to incorporate this form of spatial filtering into a successive correction analysis system based on Bratseth's method when both the weights applied to observed variables and the filter are derived from a simple Gaussian function of spatial separation.

2. Formulation

a. Basic formulation

We consider a successive correction analysis scheme (henceforth abbreviated to SC) based on the following recursion formulas:

$$\mathbf{f}_g^k = \mathbf{f}_g^{k-1} + \mathbf{W}_g(\mathbf{f}_o - \mathbf{f}_o^{k-1}) \tag{1a}$$

$$f_o^k = \mathbf{f}_o^{k-1} + \mathbf{W}_o(\mathbf{f}_o - \mathbf{f}_o^{k-1}), \tag{1b}$$

where $\mathbf{f}_g^k$ and $\mathbf{f}_o^k$ are data vectors representing the analysis following k iterations evaluated on analysis and observation points, respectively, $\mathbf{f}_o$ is a vector of observations, and $\mathbf{W}_g$ and $\mathbf{W}_o$ are matrices containing weighting factors that interpolate increment variables onto the analysis and observation points, respectively.

The iteration cycle is initialized using

$$\mathbf{f}_g^0 = \bar{\mathbf{f}}_g$$

$$\mathbf{f}_o^0 = \bar{\mathbf{f}}_o, \tag{2}$$

where $\bar{\mathbf{f}}_g$ and $\bar{\mathbf{f}}_o$ represent a "mean" field, evaluated on both analysis and observation points, which is estimated as a linear function of the observations. Although the mean field could be calculated simply as the arithmetic mean of the observations, it is often more appropriate to assume that it can be represented by a low-order polynomial function of spatial coordinates, such as a linear trend surface, and to estimate this using ordinary least-squares methods (see, for example, Thiebaux and Pedder 1987): the idea here is that the mean field component of the analysis can account for variations associated with wavelength scales larger than the scale of the sampled domain but not contribute significantly to any harmonic in a Fourier representation of the analysis on the finite analysis domain (see section 2b).

We later regard (1) as an approximation to statistical interpolation, with the mean field playing the role of a "background" or "first-guess" field. In this sense our approach is analogous to optimal analysis of fields with a spatially variable mean (Bretherton et al. 1976; Le Traon 1990). However, for reasons of numerical simplicity we assume that an ordinary least-squares estimator of the mean field provides an adequate approximation to the formally optimal weighted least-squares estimator that should in principle be preferred in the presence of correlated residuals. Theory applying to the estimation of trends in time series data suggest that the asymptotic properties of ordinary least-squares estimation provide some justification for this assumption, provided that the problem is sufficiently well overdetermined (Grenander and Rosenblatt 1957).

Apart from the method of initializing (1), the foregoing scheme is similar to that described by Achtemeier

(1989) in that the weighting factors in $\mathbf{W}_g$ and $\mathbf{W}_o$ are not varied between iterations. As in Achtemeier's scheme (which was based on the method of Barnes 1964), we assume a simple normal-error influence function model as a basis for generating these weighting factors, as represented by

$$\Omega(x_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{x_{ij}}{L}\right)^2\right], \quad (3)$$

where $x_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ is the distance between two points on the observed domain and L is a user-prescribed influence scale.

In conventional distance weighting schemes, the weight matrices in (1) are calculated as $\mathbf{W}_g = \mathbf{C}_g^{-1}\mathbf{\Omega}_g$ and $\mathbf{W}_o = \mathbf{C}_o^{-1}\mathbf{\Omega}_o$, where $\mathbf{\Omega}$ denotes a matrix of influence function values and $\mathbf{C}$ is an associated diagonal matrix of normalization constants, with kkth entry given by

$$C_{kk} = \sum_{j=1}^{n} \Omega(x_{kj}), \quad (4)$$

where n is the number of observations. The row vectors in $\mathbf{W}_g$ and $\mathbf{W}_o$ are thus normalized so that their elements sum to unity. [Lorenc (1992) refers to this normalization as being in "model space."] It is then not difficult to show (by back substitution) that (1a) is equivalent to an analysis given by

$$\mathbf{f}_g^k = \bar{\mathbf{f}}_g + \mathbf{C}_g^{-1}\mathbf{\Omega}_g \sum_{t=0}^{k-1} (\mathbf{I} - \mathbf{C}_o^{-1}\mathbf{\Omega}_o)^t(\mathbf{f}_o - \bar{\mathbf{f}}_o), \quad (5a)$$

where $\mathbf{I}$ is the identity matrix. If the iteration given by (1b) converges, then this leads to

$$\mathbf{f}_g^\infty = \bar{\mathbf{f}}_g + \mathbf{C}_g^{-1}\mathbf{\Omega}_g\mathbf{\Omega}_o^{-1}\mathbf{C}_o(\mathbf{f}_o - \bar{\mathbf{f}}_o). \quad (5b)$$

Following the ideas of Bratseth (1986), an alternative normalization is with respect to observation space (cf. Lorenc 1992). This means applying a normalization so that $\mathbf{W}_g = \mathbf{\Omega}_g\mathbf{C}_o^{-1}$ and $\mathbf{W}_o = \mathbf{\Omega}_o\mathbf{C}_o^{-1}$. Using scheme (1a) is then equivalent to an analysis given by

$$\mathbf{f}_g^k = \bar{\mathbf{f}}_g + \mathbf{\Omega}_g\mathbf{C}_o^{-1} \sum_{t=0}^{k-1} (\mathbf{I} - \mathbf{\Omega}_o\mathbf{C}_o^{-1})^t(\mathbf{f}_o - \bar{\mathbf{f}}_o). \quad (6a)$$

Because (3) generates weights that are positive for all x_{ij}, (6a) should theoretically always converge (Franke 1988), leading to

$$\mathbf{f}_g^\infty = \bar{\mathbf{f}}_g + \mathbf{\Omega}_g\mathbf{\Omega}_o^{-1}(\mathbf{f}_o - \bar{\mathbf{f}}_o). \quad (6b)$$

Bratseth's point was that this limiting solution is mathematically identical to statistical interpolation (henceforth abbreviated to SI), and theoretically implies an optimal (minimum variance) estimate of $\mathbf{f}_g$ (when linear in observed variables) for a suitable choice of $\mathbf{\Omega}_o$ and $\mathbf{\Omega}_g$. In this sense, a scheme based on (6a) provides a better approximation to an optimal weighting scheme than one based on (5a). Actually, the intention here is not to arrive at an analysis that necessarily satisfies (6b) as closely as possible but rather to use (6a) in combination with a suitable influence function scale as a means of approximating a prescribed bandwidth of observed scales. In this sense our method is more closely related to the ideas of Barnes (1964) and Maddox (1980) than to the formal objective of SI.

An important practical consequence of using Bratseth's iteration scheme is that because the weighting factors in (1) are not varied between iterations, the analysis following k iterations can be represented by

$$\mathbf{f}_g^k = \bar{\mathbf{f}}_g + \mathbf{\Omega}_g\beta^k, \quad (7)$$

where β^k is a parameter vector that depends only on observed variables, the number of iterations, and the weighting matrix $\mathbf{\Omega}_o$. An algorithm for updating the parameter vector β^k and estimating residual increments is presented in appendix A. A useful feature of this algorithm is that it is easy to approximate $\mathbf{\Omega}_o$ in such a way that the numerical overheads involved increase only in proportion to the number of observations n rather than in proportion to the order of at least n^2, as is commonly assumed to apply to SI methods that involve solving a linear system of the form $\mathbf{\Omega}_o\beta = (\mathbf{f}_o - \bar{\mathbf{f}}_o)$ [see (6b)]. Apart from its algorithmic simplicity, the SC method thus offers obvious advantages if computational resources are limited or the number of observations is very large. A further advantage of (7) is that it is easily modified to represent the effect of applying a spatial filter to the analysis field (see section 2c).

In general, the estimators given by (5a) and (6a) are not equivalent. [Equation (5a) is certainly not guaranteed to converge to the SI estimator (6b) for any arbitrary distribution of observing sites.] However, (5a) and (6a) are formally equivalent in the limit of continuous spatial sampling on an infinite domain; in simple terms, all the diagonal entries in $\mathbf{C}_g$ and $\mathbf{C}_o$ are then equal to the same normalization constant, which therefore cancels through (5b) to give result (6b). This formal equivalence in the limit of continuous sampling is useful, because it allows us to treat the two schemes as identical when referring to a continuous approximation to their spectral response properties.

b. *Wavenumber response and design factors*

A common approach to the description and design of empirical SC analysis systems is to refer to the amplitude response, considered as a function of spatial wavenumber, of a continuous linear filter that approximates the behavior of the a posteriori discrete weighting process associated with the actual analysis system. For this purpose, it is often sufficient to consider one-dimensional sampling, and here, we therefore treat location and frequency domain variables as scalars.

In common with earlier studies, we express the total effective theoretical response in terms of a basic am-

plitude response function, which is given by the Fourier transform of a suitably normalized influence function $\Omega(x)$. In the case of the normal-error influence function, this is given by

$$w(x) = \frac{1}{(2\pi)^{1/2}L}\exp\left[-\frac{1}{2}\left(\frac{x}{L}\right)^2\right], \qquad (8a)$$

which has the well-known Fourier transform

$$R(\mu) = \exp\left(-\frac{\mu^2 L^2}{2}\right), \qquad (8b)$$

where $\mu = 2\pi/\lambda$, λ being spatial wavelength. In the limit of continuous sampling the theoretical amplitude response of either (5a) or (6a) is given by

$$R^k(\mu) = \bar{R}(\mu) + [1 - \bar{R}(\mu)]R(\mu)\sum_{t=0}^{k-1}[1 - R(\mu)]^t,$$

where $\bar{R}(\mu)$ is the response associated with the linear process estimating the mean field. However, if the mean field takes the form of a low-order polynomial trend, then $\bar{R}(\mu)$ can be taken as zero for all $\mu > 0$, and equal to unity at $\mu = 0$. Hence, the theoretical response of SC is effectively represented by

$$R^k(\mu) = R(\mu)\sum_{t=0}^{k-1}[1 - R(\mu)]^t. \qquad (9)$$

Figure 1 shows curves of theoretical amplitude response versus dimensionless wavelength λ/L for various values of k when $R(\mu)$ is given by (8b). As in the method of Barnes (1964), we use this as a basis for choosing a suitable value of L in (3b). Our primary objective in applying (1) will be to resolve spatial features in the observed field associated with wavelength scales appreciably larger than some minimum resolvable wavelength λ_N, but suppress the effect of wavelength scales of the same order or less than λ_N. The latter condition is clearly satisfied if $L \sim \lambda_N/2$, provided that k is not too large. If we consider λ_N to be of the order twice the observed mean separation scale $\bar{x}_o$, then that implies $L \sim \bar{x}_o$. On a two-dimensional domain, a common measure of mean separation scale is $(A/n)^{1/2}$, where A is the area of the region enclosing n observations, and it is this measure that is referred to in later sections. (Considerations applying to the case of anisotropic sampling are discussed in appendix B.)

In some applications, the theoretical amplitude response of an analysis system might be considered less important than its ability to approximate observations to within some subjectively prescribed limit. The latter might typically be defined in terms of the root-mean-square (rms) perceived error of the analysis evaluated on the observing sites. (In fact this may be the only quantitative method of evaluating an analysis in the absence of useful prior information regarding its spectral content.) A reasonable strategy is then to regard L as a prescribed parameter and, subsequently, use as many iterations as are necessary to produce an "acceptable" fit to the observed data. When used in this way, the rate of convergence may be increased by incorporating an overrelaxation parameter in the recursive algorithm estimating β^k in (7) (see appendix A). Consequently, an excessive number of iterations need not be required to approximate observations closely, provided that the observed field does not contain a large contribution from relatively small wavelength scales.

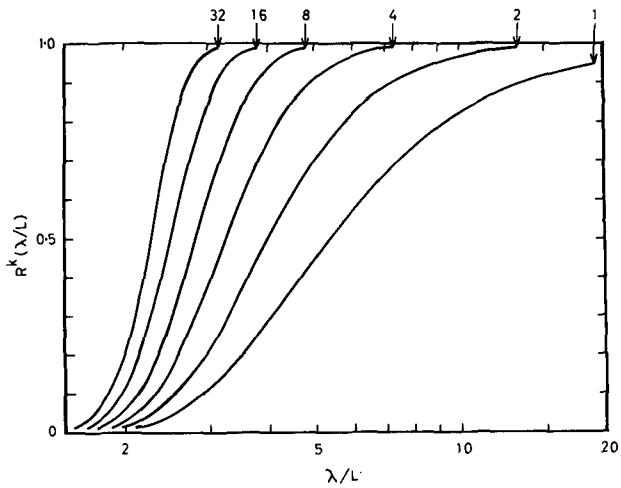

FIG. 1. Theoretical amplitude response versus dimensionless wavelength scale corresponding to a successive correction analysis system using a Gaussian influence function with fixed influence scale L. Numbers along the top of the diagram identify the number of iterations.

c. SC and SI noise response

At this point, it is worth considering how the foregoing approach to SC analysis compares with an (formally) optimal SI approach when the latter also assumes a simple influence function model for increment correlations. To this end, let us define a field $\tilde{f}(\mathbf{x})$ that corresponds to a prescribed subset of observed scales (which we wish to analyze), and represent the observed field by $f(\mathbf{x}) = \tilde{f}(\mathbf{x}) + f'(\mathbf{x})$, where $f'(\mathbf{x})$ is the error of representativity on $f(\mathbf{x})$ considered as an estimate of $\tilde{f}(\mathbf{x})$. According to this specification, the gjth element of $\mathbf{\Omega}_g$ in (6b) should correspond to the correlation between $[\tilde{f}(\mathbf{x}_g) - \bar{\tilde{f}}(\mathbf{x}_g)]$ and $[f(\mathbf{x}_j) - \bar{f}(\mathbf{x}_j)]$, which we assume equal to $\Omega(x_{gj})$. On the other hand, the ijth element of $\mathbf{\Omega}_o$ should correspond to the correlation between $[f(\mathbf{x}_i) - \bar{\tilde{f}}(\mathbf{x}_i)]$ and $[f(\mathbf{x}_j) - \bar{f}(\mathbf{x}_j)]$. Clearly, this cannot be equal to $\Omega(x_{ij})$ unless the observed field has the same spectral distribution as that associated with $\tilde{f}(\mathbf{x})$ (which is unlikely). If we consider f' and $(\tilde{f} - \bar{\tilde{f}})$ as independent random variables, then in the absence of measurement error, the optimal SI estimator of $\tilde{f}(\mathbf{x})$ should be of the form

$$\mathbf{f}_g = \bar{\tilde{\mathbf{f}}}_g + \mathbf{\Omega}_g(\mathbf{\Omega} + \gamma\mathbf{\Omega}')^{-1}(\mathbf{f}_o - \bar{\tilde{\mathbf{f}}}_o), \qquad (10)$$

where $\mathbf{\Omega}$ is derived from the same influence function model as $\mathbf{\Omega}_g$, $\mathbf{\Omega}'$ corresponds to the two-point correlations between representativity errors, and γ is the ratio of representativity error variance to the variance of $(\tilde{f} - \bar{f})$. It follows that the SC analysis described in the previous subsection, which uses the same influence function to model both $\mathbf{\Omega}_g$ and $\mathbf{\Omega}_o$, must be suboptimal in the sense that it cannot generally converge to (10) in the limit as $k \to \infty$. Fortunately, this deficiency may be partially offset in practice by the fact that the effect of limiting the number of iterations of (1) can be rather similar to that of including a noise term in the SI analysis represented by (10). This can be understood in qualitative terms by noting that, following derivations given in Daley (1991), the theoretical amplitude response of the increment weighting process in (10) can be written

$$R^{\infty}(\mu) = \frac{R(\mu)}{R(\mu) + \sigma R'(\mu)}, \tag{11}$$

where $R'(\mu)$ is the Fourier transform of the representativity error autocorrelation function $\Omega'(x)$, and $\sigma = \gamma(L_e/L)$, where L_e is a characteristic length scale associated with $\Omega'(x)$. In general, we expect $R'(\mu)$ to decay more slowly with increasing μ than $R(\mu)$ for $L_e < L$. Hence, (11) predicts that for $\gamma > 0$, small wavelength features should be damped more strongly than large wavelength features in the optimally formulated SI scheme. Incorrectly specifying $\gamma = 0$ in (10) would imply $R^{\infty}(\mu) = 1$ for all $\mu > 0$, which could clearly lead to spurious emphasis of small wavelength scales. However, such a specification error may be less serious in the case of the SC approximation to (10), since for realistic values of k Fig. 1 shows that $R^k(\mu)$ decreases to very small values for $\lambda \sim 2L$ (subject to the assumption the high-wavenumber amplitude response of the discrete weighting scheme is not too different from its theoretical value).

If purely random measurement errors affect observations, then their effect can be allowed for by adding a constant noise term γ_o to the diagonal elements of $\mathbf{\Omega}_o$ in either the SI analysis system represented by (10) or its approximation represented by (6a). In many applications of SI, the effects of both measurement error and representativity error have been approximated in this way. Though questionable in some applications (see Lorenc 1986), such an approximation seems reasonable provided that representativity error does not contribute significantly to the observed increment autocorrelation at spatial lags greater than $\bar{x}_o$ [so that we can approximate $(\mathbf{\Omega} + \gamma\mathbf{\Omega}')$ in (10) by $(\mathbf{\Omega} + \gamma_o\mathbf{I})$]. However, there is again the problem of choosing a suitable value for the noise parameter γ_o in the absence of useful prior information, though we might reasonably expect the "slow convergence" property of the SC iterations to compensate to some extent for the absence of a realistic noise model in the specification of the observed increment correlation matrix $\mathbf{\Omega}_o$. The following considers only the simplest SC scheme in which no noise term is included in its weight formulation.

d. Scale selection analysis

The term *scale selection* is occasionally used in the meteorological analysis literature to imply a process of spatial filtering designed to isolate a relatively narrow bandwidth of wavelength scales. Ooyama (1987) has suggested that some form of explicit low-pass filtering is often desirable, even in optimally formulated interpolation schemes. In SC schemes such as that described by Maddox (1980), different degrees of selective filtering are obtained by varying the scale parameters of the influence functions used to generate the a posteriori weights applied to observations following a fixed number of iterations. Here we suggest an alternative approach to scale selection that does not involve repeating the iterations represented explicitly by (1) and that does not suffer from some of the ambiguity associated with the foregoing approach.

Because the weighting factors in (1) are calculated from a continuous function of spatial lag, the analysis field itself can be represented as a continuous function of location. Thus, following k iterations, the corresponding analysis field can be evaluated at any point $\mathbf{x}_g$ using

$$\hat{f}(\mathbf{x}_g) = \bar{f}(\mathbf{x}_g) + \sum_{j=1}^{n} \hat{\beta}_j \Omega(\mathbf{x}_g - \mathbf{x}_j), \tag{12}$$

where $\hat{\beta}_j$ is the jth element of a parameter vector given by

$$\hat{\boldsymbol{\beta}} = \mathbf{C}_o^{-1} \sum_{t=0}^{k-1} (\mathbf{I} - \mathbf{\Omega}_o\mathbf{C}_o^{-1})^t(\mathbf{f}_o - \bar{\mathbf{f}}_o),$$

which may be regarded as an approximation to an SI parameter vector represented by

$$\boldsymbol{\beta} = \mathbf{\Omega}_o^{-1}(\mathbf{f}_o - \bar{\mathbf{f}}_o).$$

(Note absence of superscript symbols on $\boldsymbol{\beta}$ throughout this section.)

Now consider a continuous low-pass filter $g(\mathbf{x}')$ applied to $\hat{f}(\mathbf{x}_g)$ so that a filtered analysis field is defined by

$$\hat{f}^1(\mathbf{x}_g) = \int_{-\infty}^{+\infty} g(\mathbf{x}')\hat{f}(\mathbf{x}_g + \mathbf{x}')d\mathbf{x}'.$$

We assume that the low-pass filter has no effect on the mean field, which is reasonable if $\bar{f}(\mathbf{x})$ takes the form of a low-order polynomial in location variables. Using (12), the filtered analysis variable at location $\mathbf{x}_g$ can therefore be calculated as

$$\hat{f}^1(\mathbf{x}_g) = \bar{f}(\mathbf{x}_g) + \sum_{j=1}^{n} \hat{\beta}_j \int_{-\infty}^{+\infty} g(\mathbf{x}')\Omega(\mathbf{x}_g - \mathbf{x}_j + \mathbf{x}')d\mathbf{x}'. \tag{13}$$

Suppose we base $g(\mathbf{x}')$ on the normal-error function (8a): in the one-dimensional case, $g(\mathbf{x}')$ is thus defined by

$$g(x') = \frac{1}{(2\pi)^{1/2}L_g}\exp\left[-\frac{1}{2}\left(\frac{x'}{L_g}\right)^2\right],$$

where L_g is the filter scale parameter. If $\Omega(\mathbf{x}_g - \mathbf{x}_j)$ is based on (3), then in the one-dimensional case the jth convolution integral in (13) has an analytical solution given by

$$\tilde{\Omega}^1(x_g - x_j) = \frac{L}{L_1}\exp\left[-\frac{1}{2}\frac{(x_g - x_j)^2}{L_1^2}\right],$$

where $L_1 = (L_g^2 + L^2)^{1/2}$. In the d-dimensional case, we can specify $g(\mathbf{x}')$ as the product of d one-dimensional normal-error filters, in which case it is easy to show that $\tilde{\Omega}^1(\mathbf{x}_g - \mathbf{x}_j)$ is given by

$$\tilde{\Omega}^1(\mathbf{x}_g - \mathbf{x}_j) = \left(\frac{L}{L_1}\right)^d \exp\left(-\frac{1}{2}\frac{|\mathbf{x}_g - \mathbf{x}_j|^2}{L_1^2}\right). \quad (14)$$

The effect of applying a normal-error filter to the output of either SI or its SC approximation is thus obtained simply by replacing the influence function $\Omega(\mathbf{x}_g - \mathbf{x}_j)$ in (12) by the "filtered" influence function given by (14). The same principle is easily extended to the case where the method is required to effect high-pass filtering {in which case $\Omega(x)$ is replaced by $[\delta(x) - \tilde{\Omega}^1(x)]$, where $\delta(x) = 1$ if $x = 0$ and 0 otherwise} or bandpass filtering [in which case $\Omega(x)$ is replaced by the difference between two low-pass-filtered influence functions].

The normal-error filter response corresponding to a given value of the scale parameter L_g is obtained by replacing L by L_g in (8b), and is represented by the curve labeled 1 in Fig. 1. Such a filter clearly offers rather poor scale selectivity in the sense that its taper, for example, as measured by $L_g dR/d\lambda$ at the half-amplitude point $R = 0.5$, is relatively small. However, greater scale selectivity can be achieved by applying the aforementioned principle to the case where $g(x')$ is applied recursively. By analogy with (1a), we express this by a recursion relationship of the form

$$\hat{f}^K(\mathbf{x}_g) = \hat{f}^{K-1}(\mathbf{x}_g) + \int_{-\infty}^{+\infty} g(\mathbf{x}')[\hat{f}_g(\mathbf{x} + \mathbf{x}') - \hat{f}_g^{K-1}(\mathbf{x} + \mathbf{x}')]d\mathbf{x}', \quad (15)$$

with $\hat{f}^0(\mathbf{x}_g) = 0$. If $\hat{f}(\mathbf{x}_g)$ is given by (12), then the output obtained from (15) can be written in a similar way; that is,

$$\hat{f}^K(\mathbf{x}_g) = \bar{f}(\mathbf{x}_g) + \sum_{j=1}^{n} \beta_j \tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j), \quad (16)$$

where $\tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j)$ is a filtered influence function. To relate $\tilde{\Omega}^K(\mathbf{x})$ to $\Omega(\mathbf{x})$, we define two continuous linear operators $\mathcal{L}$ and $\mathcal{I}$ such that

$$\mathcal{L}\xi(\mathbf{x}) \to \int_{-\infty}^{+\infty} g(\mathbf{x}')\xi(\mathbf{x} + \mathbf{x}')d\mathbf{x}'; \quad \mathcal{I}\xi(\mathbf{x}) \to \xi(\mathbf{x})$$

and

$$\mathcal{L}^l\xi(\mathbf{x}) \to \mathcal{L}\mathcal{L}^{l-1}\xi(\mathbf{x}).$$

We can then express (15) as a sequence of linear operators acting on $\hat{f}(\mathbf{x}_g)$, and by substituting (12), thus relate the filtered influence function in (16) to $\Omega(x)$. The result [which is the continuous spatial analog of (9)] is expressed by

$$\tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j) = \mathcal{L}\sum_{l=0}^{K-1}(\mathcal{I} - \mathcal{L})^l\Omega(\mathbf{x}_g - \mathbf{x}_j); \quad K > 0$$

or, as an equivalent power series in $\mathcal{L}$, by

$$\tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j) = \sum_{t=1}^{K} a_t\mathcal{L}^t\Omega(\mathbf{x}_g - \mathbf{x}_j), \quad (17)$$

where a_t is an expansion coefficient. Noting that

$$\mathcal{L}^t\Omega(\mathbf{x}_g - \mathbf{x}_j) \to \mathcal{L}\tilde{\Omega}^{t-1}(\mathbf{x}_g - \mathbf{x}_j),$$

it is then not difficult to show that in the case of the normal-error filter applied to a normal-error-type influence function, (17) gives

$$\tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j) = \sum_{t=1}^{K} a_t\left(\frac{L}{L_t}\right)^d \exp\left(-\frac{1}{2}\frac{|\mathbf{x}_g - \mathbf{x}_j|^2}{L_t^2}\right), \quad (18)$$

where

$$L_t = (tL_g^2 + L^2)^{1/2}.$$

The effect of repeated filtering applied to the output of the original SC or SI analysis system is thus obtained by simply replacing $\Omega(\mathbf{x}_g - \mathbf{x}_j)$ in (12) by a filtered influence function calculated using (18).

In practice, there are obvious practical limitations to this method of spatial filtering, since the number of exponentials that need to be evaluated at each analysis location increases in proportion to K, while the equivalent filter taper increases only slowly with increasing K for $K > 2$, (as can be inferred from Fig. 1). Nevertheless, it provides a relatively simple approach to isolating spatial features that are not associated with a sharply defined or particularly narrow bandwidth of wavelength scales. It is also easily modified to deal with the requirement of applying different degrees of scale selection along different directions (see appendix B).

It is worth noting that the foregoing approach to spatial filtering of analysis field variables can be considered as equivalent to applying a discrete symmetric filter to data sampled on a uniform spatial grid of infinite extent in the limit where the grid space tends to zero. Such a filter is spatially invariant, implying a less ambiguous relationship between theoretical and actual wavenumber response than in the case of a filter applied directly to observations, which is the principle used in

other scale separation schemes based on distance weighting methods.

3. Tests with simulated data

This section describes results obtained from applying the SC analysis method to the problem of estimating a simulated scalar field sampled by scattered observations.

The two-dimensional field simulation (of 850-hPa height) was based on the method described in Gomis et al. (1990). In BGPA, tests of a scale selection analysis system considered sampling this field on one particular arrangement of 79 station sites lying within a rectangular domain measuring approximately 2800 km × 2400 km, corresponding to a mean station separation scale of about 290 km. In the study reported here, tests were carried out by applying an analysis to the same number of observing sites scattered over the same area but using a random sampling method to simulate multiple sets of observations associated with different station distributions. Analysis errors were evaluated on a 23-point × 21-point regular grid covering an inner verification domain measuring 2200 km × 2000 km. In the following, error statistics associated with a particular analysis formulation were obtained by averaging squared-error variables derived from 100 independent realizations of the random sampling pattern. The effects of random observing error were also simulated by adding Gaussian noise increments to the simulated observations.

a. Convergence properties

Figure 2 illustrates the convergence characteristics of the SC method applied to observations sampling the simulated field with and without added noise increments. The SC analysis was in all cases based on linear detrending followed by iterations applied to detrended observations, using a fixed influence scale of 300 km (chosen empirically to be roughly equal to the average observed separation scale) and an overrelaxation constant $\alpha = 2$ (see appendix A). For comparison, Fig. 3 shows rms perceived error and rms analysis error versus the assumed value for the noise-to-signal ratio parameter γ_o in an SI estimation based on solving (6b) applied to detrended data, using the same influence function model as in the SC scheme to represent increment correlations.

In the absence of noise on observations, the SI analysis using a fixed influence scale would be expected to provide the most accurate analysis when $\gamma_o = 0$, which

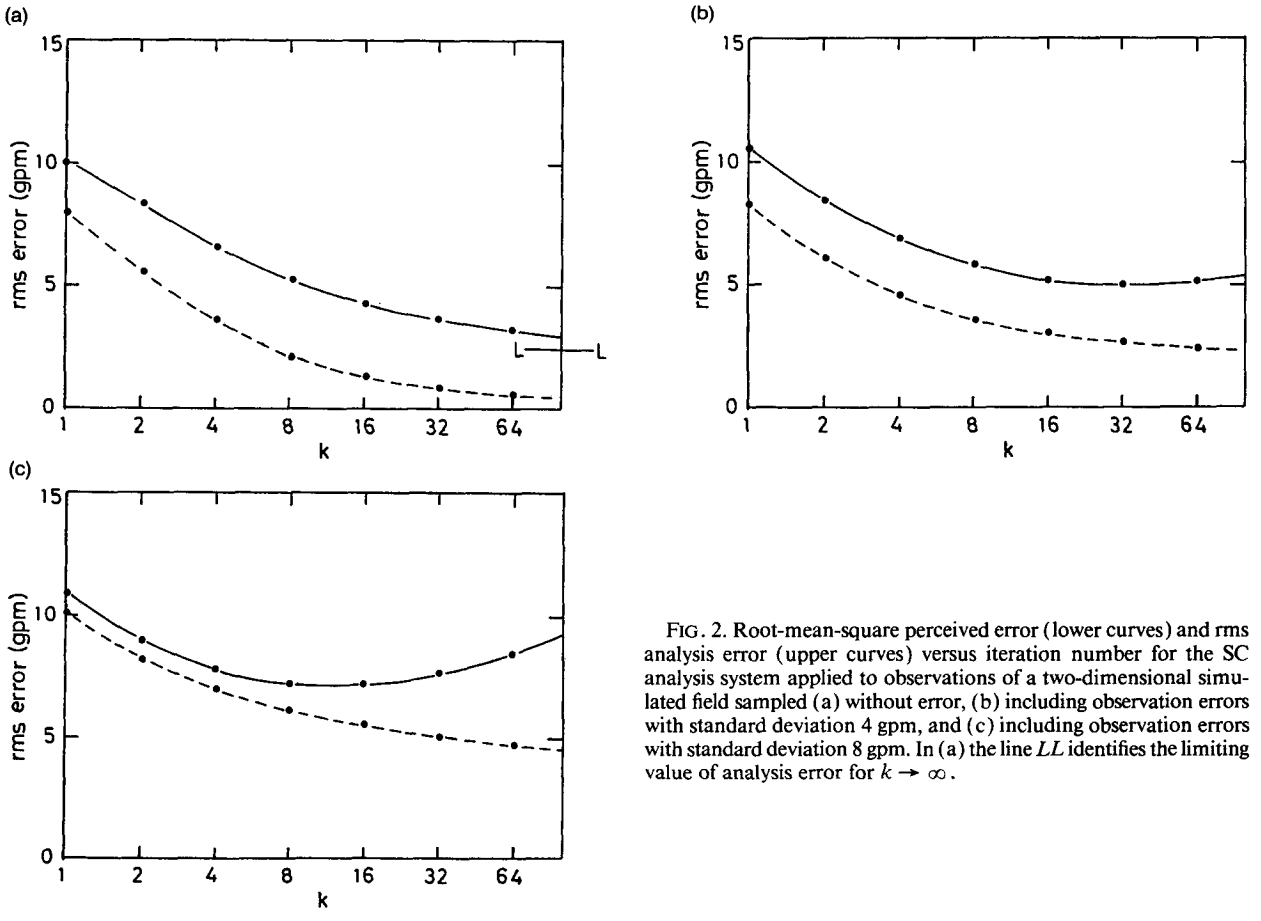

FIG. 2. Root-mean-square perceived error (lower curves) and rms analysis error (upper curves) versus iteration number for the SC analysis system applied to observations of a two-dimensional simulated field sampled (a) without error, (b) including observation errors with standard deviation 4 gpm, and (c) including observation errors with standard deviation 8 gpm. In (a) the line LL identifies the limiting value of analysis error for $k \to \infty$.

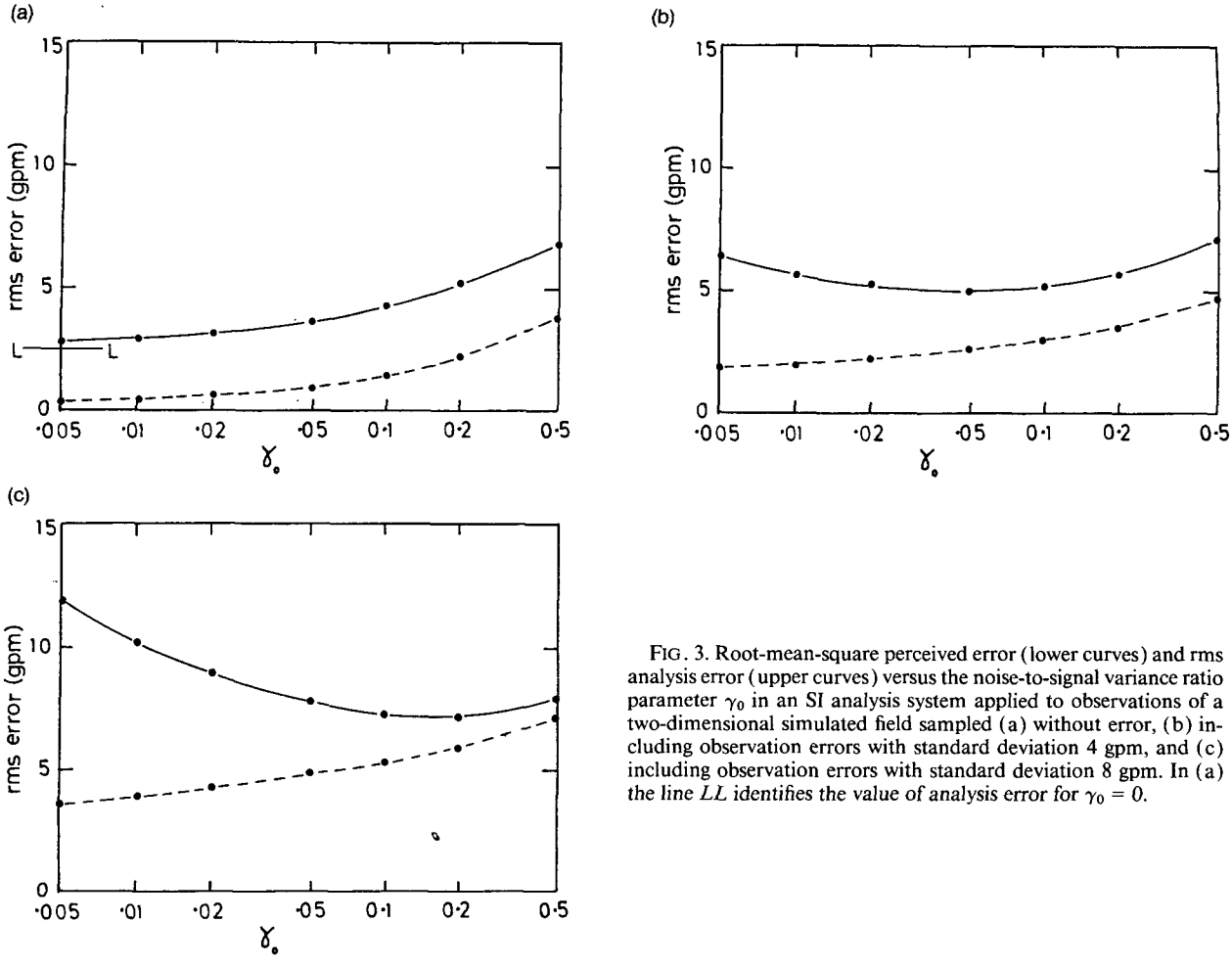

FIG. 3. Root-mean-square perceived error (lower curves) and rms analysis error (upper curves) versus the noise-to-signal variance ratio parameter γ_0 in an SI analysis system applied to observations of a two-dimensional simulated field sampled (a) without error, (b) including observation errors with standard deviation 4 gpm, and (c) including observation errors with standard deviation 8 gpm. In (a) the line *LL* identifies the value of analysis error for $\gamma_0 = 0$.

is formally equivalent to the limit $k \to \infty$ in its SC approximation based on (6a). The results suggest that the SC analysis does indeed converge toward that limit, though the rate of convergence decreases rapidly with increasing k. In general, the rate of convergence toward the observations also decreases as the noise level increases, which is consistent with the tendency for the high-wavenumber response to remain small even following a large number of iterations (see Fig. 1). Symptoms of overfitting become evident when the number of iterations exceeds some "optimal" value that minimizes rms analysis error. Clearly, the effect of seriously *overprescribing* the number of iterations in the SC scheme is similar to that of seriously *underprescribing* the value of the smoothing parameter γ_o in the SI scheme. For the SC method, serious overfitting arises only when the number of iterations corresponds to a relatively large-amplitude response at wavelength scales close to the minimum resolvable scale inferred from the mean observed separations scale; in the foregoing formulation this corresponds to the order of 50 iterations.

As can be seen in Figs. 2 and 3, the rms analysis error associated with a particular value of rms perceived error is virtually the same for the two schemes, which is consistent with the idea that the effect of limiting the number of iterations in the SC scheme is rather similar to assuming a nonzero value for observing error variance in the increment covariance model of an SI scheme (see section 1c). In some applications, the analysis might be targeted toward a certain level of data resolution as measured by an empirically prescribed rms perceived error. The SC algorithm provides a simple method of generating an analysis with an rms perceived error that approximates the required value; since residuals on observing sites are calculated as one part of the iteration cycle (see appendix A), it is a trivial matter to identify the number of iterations required to reduce the rms perceived error to or below the required value.

b. SC versus Barnes-type and surface spline methods

Figure 4 compares analysis fields generated from three types of analysis system applied to noise-free observations of the simulated field sampled on the spatial pattern of station sites used in BGPA. The SC analysis follows 32 iterations applied to detrended observations,

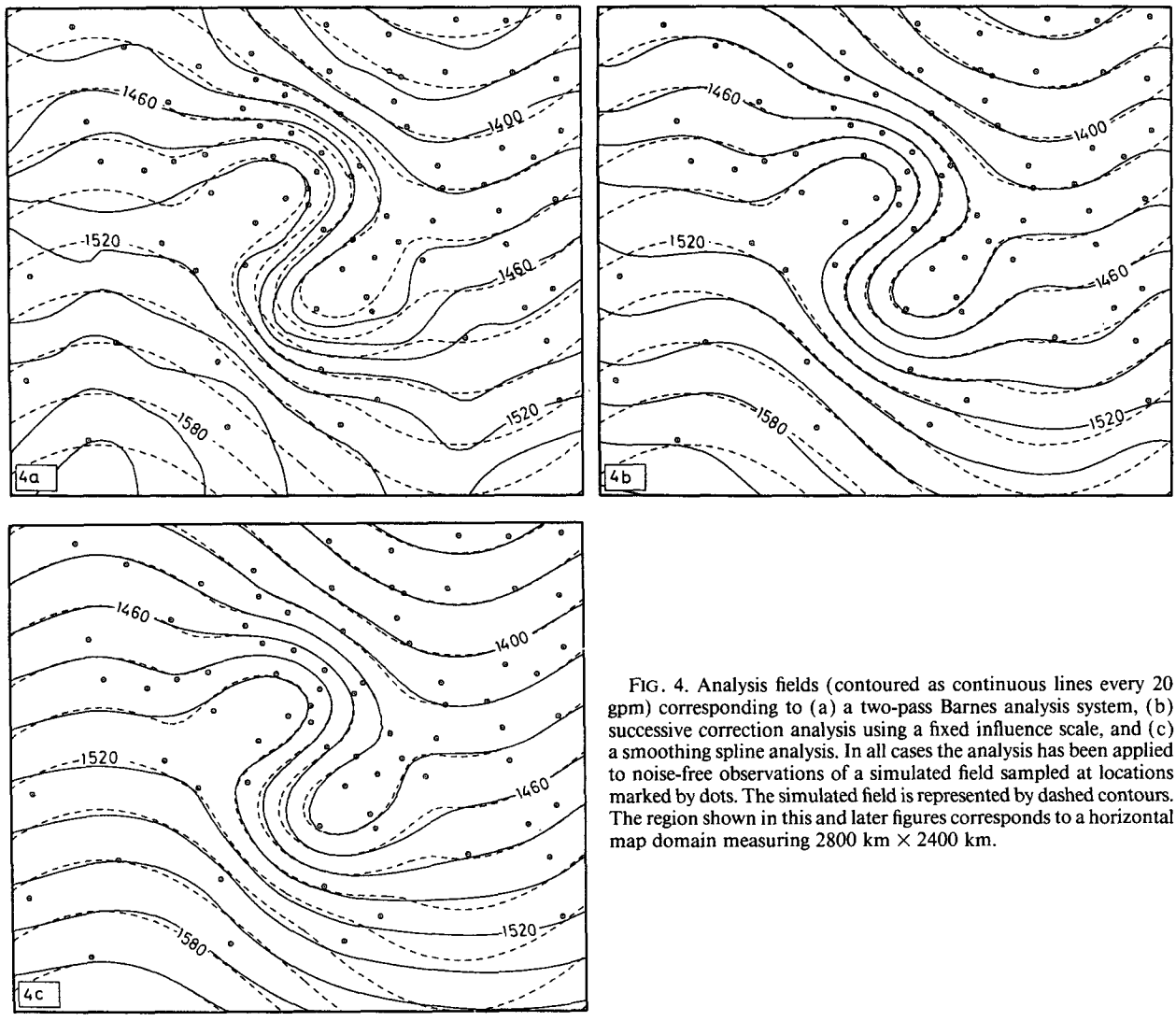

FIG. 4. Analysis fields (contoured as continuous lines every 20 gpm) corresponding to (a) a two-pass Barnes analysis system, (b) successive correction analysis using a fixed influence scale, and (c) a smoothing spline analysis. In all cases the analysis has been applied to noise-free observations of a simulated field sampled at locations marked by dots. The simulated field is represented by dashed contours. The region shown in this and later figures corresponds to a horizontal map domain measuring 2800 km × 2400 km.

with other parameters specified as above, and approximates the observations with an rms residual of 0.6 gpm. The Barnes-type ("two pass") analysis is as detailed in BGPA, and was designed to have a theoretical response rising from 0.5 at $\lambda = 500$ km to near unity for $\lambda > 1000$ km. This approximates the observations with an rms residual of 3.4 gpm. The spline analysis corresponds to an order 2 smoothing spline, with a smoothing parameter estimated using generalized cross validation (GCV) as detailed in Wahba and Wendelberger (1980). This results in an rms perceived error of only 0.01 gpm, and thus, nearly interpolates the observations.

Both the spline and SC analysis clearly provide a close approximation to the simulated field over most of the observed domain; the spline analysis is slightly more accurate near the boundaries, whereas the SC analysis is slightly more accurate near the center. Within the inner verification region, the rms analysis error of the SC analysis is 2.1 gpm, while that for spline analysis is 2.5 gpm. Thus, even though the SC analysis has a slightly lower data resolution than the spline analysis, it is nevertheless comparable in analysis accuracy. In contrast, the Barnes analysis provides a much less "smooth" interpretation of the observed field than either the spline or SC analysis—despite its much lower data resolution—and results in a much larger rms analysis error of 4.6 gpm within the inner verification region.

The foregoing comparisons are obviously valid only for one particular arrangement of observing sites. Table 1 presents error statistics associated with each method applied to multiple realizations of a random sampling pattern and for various values of rms random error on observations. The GCV spline analysis on average provides a slightly more accurate estimate of the simulated field than the SC analysis based on a fixed number of iterations, and always gives better data resolution. Both

TABLE 1. Root-mean-square analysis errors evaluated on grid points and rms perceived errors evaluated on observing points (in brackets), following analysis of simulated observations with and without added noise increments. Units are geopotential meters. The Barnes two-pass analysis is as detailed in the text. The SC analysis is based on 32 iterations using an influence scale of 300 km. The spline analysis is based on an order 2 smoothing spline, with the smoothing parameter estimated using generalized cross validation. The noise-to-signal variance ratio for observations was estimated as the ratio of noise variance to mean-square variation of simulated field data on the verification grid following linear detrending.

	rms observation error (noise/signal variance ratio as percentage)			
	0 (0)	2 (0.35)	4 (5.6)	8 (22.3)
Barnes	6.58 [3.00]	6.79 [3.18]	7.36 [3.66]	9.33 [5.16]
SC	3.79 [0.91]	4.14 [1.56]	5.00 [2.71]	7.46 [5.19]
Spline	3.14 [0.01]	3.86 [0.61]	5.00 [1.76]	7.46 [5.19]

the Barnes and SC schemes were designed to have a rather similar theoretical amplitude response for wavelength scales greater than about 500 km. Nevertheless, the SC scheme appears to result in a much more accurate resolution of resolvable features in the observed field, especially at low noise levels.

4. Actual amplitude response and scale selection

a. Actual (discrete) amplitude response

Estimates of actual (discrete) amplitude response produced by the SC scheme have been obtained using a method similar to that described in BGPA. For this purpose, the simulated field consisted of plane zonal wave disturbances of known amplitude and spatial phase sampled without added noise by random arrangements of observing sites (the latter as detailed in section 3). Estimates of actual amplitude and phase response were derived from a two-dimensional Fourier analysis of fields sampled on a regular grid, following the method described by Errico (1985). For a prescribed wavelength λ, the zonal extent of this grid was chosen to equal an integer number of wavelengths m, subject to the restriction $m\lambda \leq X_a < (m+1)\lambda$, where $X_a = 2400$ km in these experiments.

Table 2 summarizes results obtained for selected zonal wavelengths, based on analysis of 100 realizations of the random spatial sampling pattern by 79 observations on the outer domain specified as in section 3. Columns headed "unfiltered" refer to the effect of approximating the observed field using the SC method based on an influence scale of 300 km, 32 iterations, and an overrelaxation parameter $\alpha = 1$. Columns headed "filtered" refer to the effect of solving (16), using a filtered influence function based on a filter scale parameter $L_g = 300$ km and $K = 2$ [see (18)]: this should have the effect of a low-pass filter acting on the unfiltered field with theoretical amplitude response of 0.5 at $\lambda = 1200$ km.

According to Fig. 1, the theoretical response of the unfiltered SC analysis system should be of order unity for wavelengths greater than about 800 km. The somewhat lower values of actual response demonstrate that the theoretical response does not necessarily provide an accurate measure of amplitude gain associated with the discrete SC weighting scheme. There is also evidence of significant phase error at wavelengths not much greater than the smallest resolvable wavelength (considered to be of the order of 600 km in these experiments), which also reflects the influence of discrete sampling on irregularly spaced observation points. On the other hand, the difference between the response of the unfiltered and filtered analysis is consistent with an actual response close to that predicted from the theoretical response of a normal-error filter with scale $L_g = 300$ km, at least for wavelength scales appreciably larger than 600 km. Furthermore, the filter introduces no significant additional phase error at wavelengths that are reasonably well resolved by the unfiltered analysis.

b. Example of scale selection analysis

As explained in BGPA, the simulated field referred to in section 3a is intended to simulate a "mesoscale" anomaly field superimposed on a "macroscale" back-

TABLE 2. Empirical wavenumber response characteristics of the SC analysis before and after applying a normal-error filter to the output field estimated directly from observations of unit amplitude plane zonal waves, averaged over 100 realizations of a random sampling pattern. Shown are the mean and standard deviations of the amplitude and phase response, the latter being expressed as phase shift relative to the input wave. The mean actual filter gain has been calculated from the ratio of amplitude response values corresponding to the filtered and unfiltered analysis systems.

λ (km)	Unfiltered		Filtered		Filter gain	
	Amplitude	Phase (°)	Amplitude	Phase (°)	Actual	Theory
2400	0.96 ± 0.02	0.0 ± 1.4	0.88 ± 0.02	0.0 ± 1.1	0.92	0.93
2000	0.96 ± 0.02	0.0 ± 1.0	0.81 ± 0.02	0.0 ± 0.8	0.85	0.87
1600	0.95 ± 0.03	0.0 ± 1.0	0.70 ± 0.02	0.0 ± 0.8	0.73	0.75
1200	0.92 ± 0.04	0.0 ± 1.6	0.45 ± 0.02	0.0 ± 1.5	0.48	0.50
1000	0.86 ± 0.04	0.2 ± 1.9	0.27 ± 0.02	0.2 ± 1.9	0.32	0.30
800	0.62 ± 0.06	0.0 ± 3.8	0.12 ± 0.02	0.0 ± 3.9	0.19	0.13
600	0.17 ± 0.03	0.2 ± 6.3	0.02 ± 0.01	0.6 ± 8.6	0.10	0.01

ground field with a zonal wavelength of the order of 3000 km (see Fig. 5). Figures 6 and 7 show the results of applying SC and Barnes analysis schemes to the problem of isolating these components of the total disturbance field when the latter is sampled by noise-free observations. The Barnes scale selection analysis scheme is as specified in BGPA; it was designed to have a low-pass theoretical response of 0.5 at $\lambda = 1800$ km and a bandpass theoretical response that rises from 0.5 at $\lambda = 500$ km to the order of unity at $\lambda = 1000$ km before decreasing to 0.5 at $\lambda = 1800$ km. The SC analysis was derived from (a) a total field analysis based on the same parameter specification as in section 3a ($L = 300$ km; 32 iterations; $\alpha = 2$), and (b) a low-pass analysis based on a normal-error filter scale $L_g = 450$ km and iteration number $K = 2$. (This particular combination of filter scale and iteration number was chosen so that the SC scheme would have a theoretical low-pass response rather similar to that associated with the Barnes scheme.) The bandpass output of the SC analysis was obtained simply by subtracting the low-pass analysis data from the total field analysis data.

In the case of the macroscale field analysis (Figs. 6a and 7a), it is clear that neither scheme completely eliminates the contribution from the mesoscale disturbance field, though the SC analysis (rms gridpoint error of 7.5 gpm) appears to be rather less sensitive in this respect than the Barnes scheme (rms gridpoint error of 14.1 gpm). However, the most obvious difference between the two schemes is in their estimation of the mesoscale field structure (Figs. 6b and 7b); the SC analysis provides a much more realistic representation of the symmetric dipole structure than the Barnes analysis, and has less of a tendency to generate spurious mesoscale anomaly features near the domain boundaries. Within the inner verification region the rms gridpoint error of the Barnes bandpass analysis is 14.5 gpm, while that of the SC bandpass analysis is 6.7 gpm. To a large extent, the error in the SC analysis can be attributed to its underestimation of the amplitude of the mesoscale anomaly. The Barnes bandpass analysis also underestimates the mesoscale anomaly amplitude, but additional errors arise from its failure to capture the underlying symmetry of the anomaly pattern.

As demonstrated in BGPA, a Barnes-type scale selection analysis system can be very effective when applied to a field sampled without error on a regular spatial grid, but relatively large phase errors on the Fourier components of a sampled field can result when it is applied to an irregular spatial distribution of observations; the latter effect can explain the "distorted" version of the mesoscale anomaly field evident in Fig. 7b. In contrast, Fig. 6a provides strong evidence that such phase errors are largely absent when scale selection is effected using the filtered influence function approach in combination with a successive correction approximation of the observed data.

5. Summary and conclusions

A suitably formulated successive correction scheme based on multiple iterations using a constant influence scale can provide a more effective approach to estimating spatial fields from scattered observations than the more conventional Barnes method, which usually involves varying the influence scale between just two iterations. The method described here is easy to formulate and can be applied to large datasets without requiring substantial computing resources or expensive numerical techniques. For a suitable choice of influence

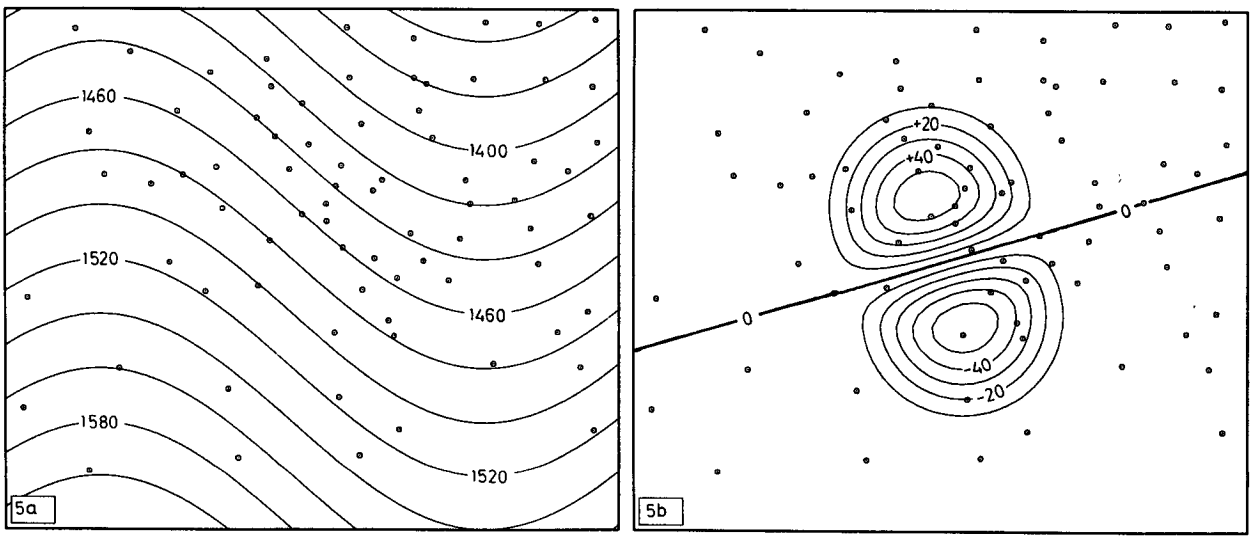

FIG. 5. (a) Macroscale (contoured every 20 gpm) and (b) mesoscale (contoured every 10 gpm) components of the simulated two-dimensional field represented by dashed contours in Fig. 4.

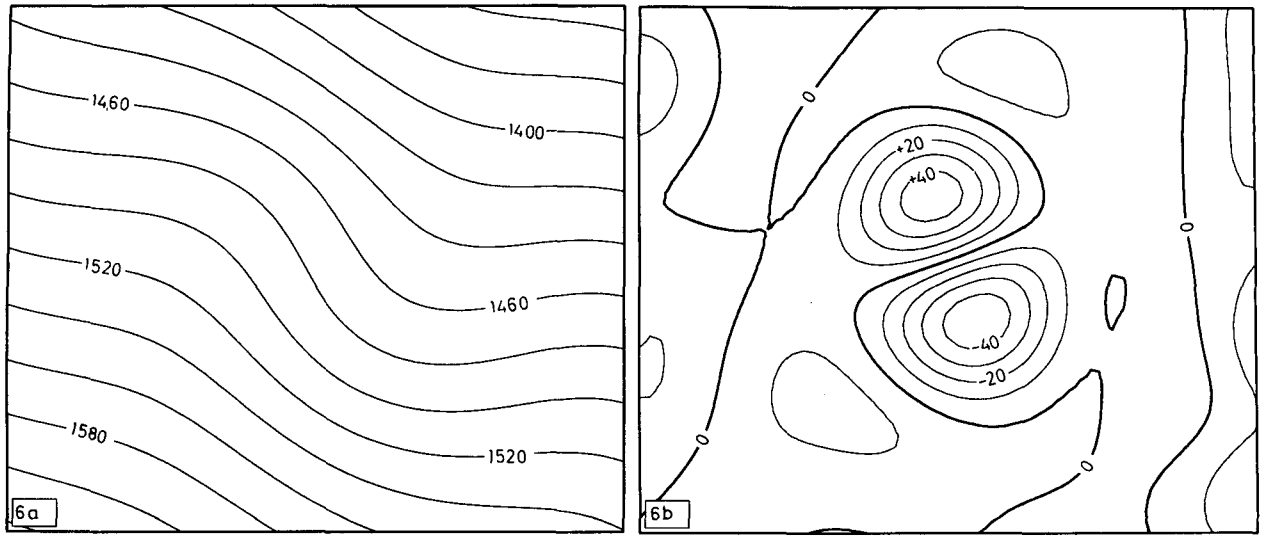

FIG. 6. (a) Macroscale and (b) mesoscale components of the simulated two-dimensional field estimated using the SC scale separation analysis system applied to noise-free observations at locations marked by dots in Fig. 5. Contouring convention as in Fig. 5.

scale, the successive correction algorithm converges only slowly toward the observed field at small wavelength scales, and this helps to suppress the influence of measurement and representativity errors on observations without the need to allow explicitly for these effects in the design of its lag-dependent weight functions.

When applied as a method of scale selection, the effect of selective filtering can be achieved through a simple modification to the weight functions that model the correlation between gridpoint and observed variables. The associated filter response can be considered as spatially invariant, and this property avoids some of the ambiguity of other methods based on applying filters directly to observations, especially when these are not sampled on a regular grid. In terms of analysis accuracy, we do not claim that this approach can necessarily compete with methods based on applying discrete filters to gridpoint data generated using optimally formulated surface spline or statistical interpolation methods but rather suggest that it is an attractive alternative when the computational resources required to solve very large linear systems might be a major consideration.

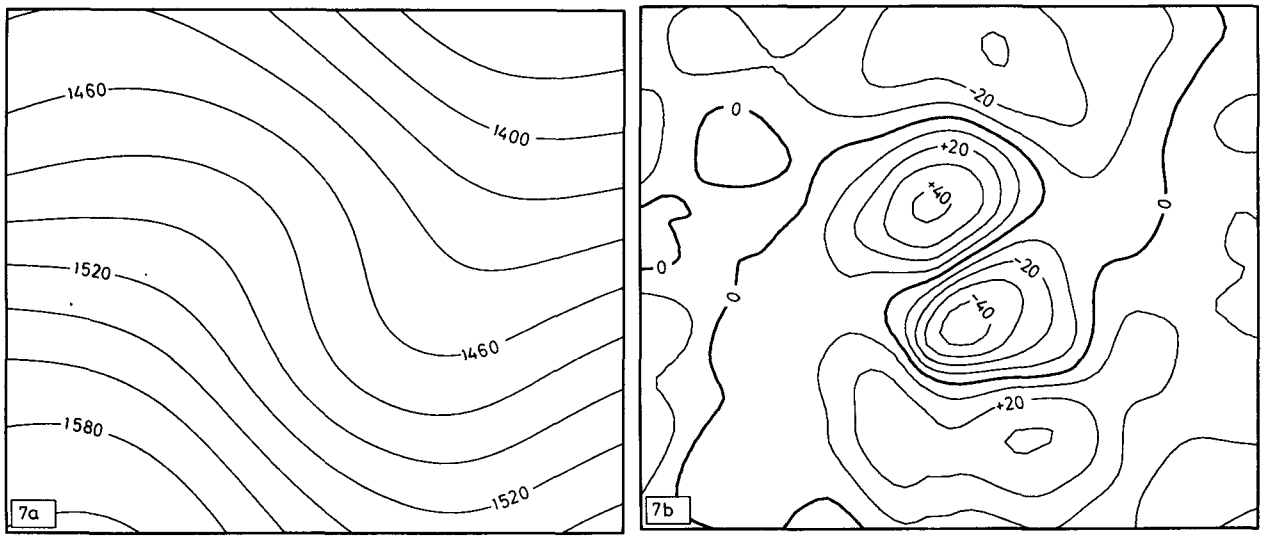

FIG. 7. As Fig. 6 but using the two-pass Barnes scale separation analysis system as detailed in Buzzi et al. (1991).

Acknowledgments. I am grateful to Damia Gomis for assistance with preparing simulated field data and some of the analysis, and Bryan Johns for advice on mathematical derivations.

APPENDIX A

Algorithm for Successive Corrections

Consider the iteration cycle represented by (1) applied to data sampled at n observing sites, with the weighting factors in (1a) calculated using Bratseth's formulation, so that $\mathbf{W}_o = \mathbf{\Omega}_o\mathbf{C}_o^{-1}$, where $\mathbf{C}_o$ is a diagonal matrix of normalization constants. If $\boldsymbol{\beta}^k$ is a parameter vector as in (7), and $\mathbf{r}^k$ a parameter vector that is related to the residual increments ("perceived errors") on observing locations following k iterations (see the following), then, by a small modification to the algorithm given in appendix F of Daley (1991), both $\boldsymbol{\beta}^k$ and $\mathbf{r}^k$ can be updated using

$$\boldsymbol{\beta}^k = \boldsymbol{\beta}^{k-1} + \alpha\mathbf{C}_o^{-1}\mathbf{r}^{k-1} \tag{A1}$$

$$\mathbf{r}^k = \mathbf{f}'_o - \mathbf{\Omega}_o\boldsymbol{\beta}^k, \tag{A2}$$

where α (≥ 1) is an optional overrelaxation parameter and $\mathbf{f}'_o = (\mathbf{f}_o - \bar{\mathbf{f}}_o)$. The iteration cycle is initialized by setting $\boldsymbol{\beta}^0 = \mathbf{0}$ and $\mathbf{r}^0 = \mathbf{f}'_o$, and the interpolation, onto analysis sites following k iterations subsequently obtained by substituting $\boldsymbol{\beta}^k$ into (7). According to Franke (1988), the rate of convergence (in our case toward $\boldsymbol{\beta}^\infty$) can be increased by prescribing a value of α greater than unity, but values close to the upper "optimal" limit of 2 can increase analysis error for small values of k. For the type of field described in section 3 we found that the number of iterations required to reduce rms residual below a prescribed value could be nearly halved by using a value of $\alpha = 2$, and this had little effect on analysis accuracy.

A useful feature of the aforementioned algorithm is that following k iterations the vector $\mathbf{r}^k$ is related to the residual increments on observing sites $\boldsymbol{\epsilon}^k = \mathbf{f}_o - \mathbf{f}_o^k$ by

$$\boldsymbol{\epsilon}^k = \mathbf{r}^k + \gamma_o\boldsymbol{\beta}^k, \tag{A3}$$

where γ_o is the smoothing parameter that models the noise-to-signal variance ratio of observations, as defined in section 2c of the main text. (Obviously, the residual vector $\boldsymbol{\epsilon}^k$ equals $\mathbf{r}^k$ if no smoothing parameter is included in the specification of the weights in $\mathbf{\Omega}_o$, which is equivalent to setting $\gamma_o = 0$.) This provides useful diagnostic information about the rate of convergence of the analysis without the need to interpolate onto analysis locations. An obvious measure of convergence is the rate of change of mean-square residual $(\bar{\epsilon}^k)^2 = [\Sigma(\epsilon_j^k)^2]/n$, considered as a function of k.

As is apparent from (A2), the update algorithm formally involves storing at least the upper triangle of the $n \times n$ symmetric influence function matrix $\mathbf{\Omega}_o$, unless its elements are recalculated at each iteration (which could be numerically expensive). However, storage requirements can be reduced considerably by addressing only those elements of $\mathbf{\Omega}_o$ that are considered to be significantly greater than zero relative to its diagonal entries (which equal 1 if $\gamma_o = 0$). In the case of the normal-error influence function given by (3), we have found that ignoring all weights corresponding to spatial separations greater than the order of $3L$ has no significant impact on the successive correction analysis when L is approximately equal to the mean distance between observing sites. For the types of station distribution considered in section 2 and L taken as equal to the average distance between observing sites, this approximation reduces the storage requirement from $n(n + 1)/2$ (for the complete upper triangle of $\mathbf{\Omega}_o$) to the order of $20n$, which represents a considerable saving in computational resources for $n > 100$. Furthermore, the time required to perform each iteration of (A1) and (A2) then tends to be very much less than the time required to solve (7) (as the "final step" of the analysis), provided that the number of analysis locations is at least comparable with the number of observing sites. We have found that for as many as 50 iterations this successive correction analysis system is only marginally more expensive to operate than a Barnes two-pass algorithm, which, though based on only two iterations, involves solving (1) explicitly and recalculating the weight matrices $\mathbf{W}_o$ and $\mathbf{W}_g$ at each iteration.

APPENDIX B

Gaussian Influence and Filter Scales for Multidimensional Sampling

In d-dimensional space, the separability property of an isotropic Gaussian influence function allows us to write (3) as

$$\Omega(x) = \prod_{l=1}^{d} \exp\left(-\frac{1}{2}\frac{x_l'^2}{L^2}\right), \tag{B1}$$

where x_l' is the distance between two points measured along the lth coordinate direction. Thus, L acts as both the component and multidimensional length scale of an equivalent spatial filter. In considering the choice of L to satisfy a prescribed wavenumber response, it is therefore sufficient to regard the multidimensional response as the product of d one-dimensional response functions and to use Fig. 1 as a basis for calculating L as a multiple of a component separation scale. If spatial sampling is fairly regular and isotropic, then a reasonable measure of component separation scale might simply be the average distance between nearest neighboring observing sites.

The aforementioned approach is obviously not appropriate if sampling density varies substantially along different coordinate directions, implying a significant

directional dependence in the spatial resolution offered by the observing system. However, if sampling density along orthogonal coordinate directions is fairly well defined, as is often the case with meteorological and oceanographic observing systems, then a simple solution is to work in terms of suitably scaled separation variables. Thus, we can replace x'_l in (B1) by $x'_l/\bar{x}_l$, where $\bar{x}_l$ is the sample separation scale along a given coordinate direction. Then L represents a single filter scale measured in units of separation scale along any of the d coordinate directions. In the context of SI analysis based on a Gaussian covariance model for observed increments, this specification is equivalent to assuming an elliptical covariance function of spatial lag whose principal axes coincide with the given coordinate directions.

The theory given in section 2d of the main text may similarly be modified to allow for different degrees of spatial filtering along different coordinate directions. It is again convenient to work in terms of scaled coordinate variables, so that x_{jl} here represents the lth component of location $\mathbf{x}_j$ measured in units of sample separation scale along the lth coordinate direction. The values of a filtered influence function are then obtained by replacing (18) in the main text by

$$\tilde{\Omega}^K(\mathbf{x}_g - \mathbf{x}_j) = \sum_{t=1}^{K} a_t \prod_{l=1}^{d} \left(\frac{L}{L_{tl}}\right) \exp\left[-\frac{1}{2}\frac{(x_{gl} - x_{jl})^2}{L_{tl}^2}\right], \tag{B2}$$

where

$$L_{tl} = (tL_{gl}^2 + L^2)^{1/2}.$$

Here L_{gl} is a component filter scale measured in units of separation scale along the lth coordinate direction.

REFERENCES

Achtemeier, G. L., 1989: Modification of a successive corrections objective analysis for improved derivative calculations. *Mon. Wea. Rev.*, **117**, 78–86.

Barnes, S., 1964: A technique for maximizing details in numerical map analysis. *J. Appl. Meteor.*, **3**, 395–409.

Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus*, **38**(A), 439–447.

Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.*, **23**, 559–582.

Buzzi, A., D. Gomis, M. A. Pedder, and S. Alonso, 1991: A method to reduce the adverse impact that inhomogeneous station distributions have on spatial interpolation. *Mon. Wea. Rev.*, **119**, 2465–2491.

Cressman, G., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **88**, 327–342.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Errico, R. M., 1985: Spectra computed from a limited area grid. *Mon. Wea. Rev.*, **113**, 1554–1562.

Franke, R., 1988: Statistical interpolation by iteration. *Mon. Wea. Rev.*, **116**, 961–963.

Gomis, D., A. Buzzi, and S. Alonso, 1990: Diagnosis of mesoscale structures in cases of lee cyclogenesis during ALPEX. *Meteor. Atmos. Phys.*, **43**, 49–57.

Grenander, U., and M. Rosenblatt, 1957: *Statistical Analysis of Stationary Time Series.* Wiley, 300 pp.

Le Traon, P. Y., 1990: A method for optimal analysis of fields with spatially variable mean. *J. Geophys. Res.*, **95**(C8), 13 543–13 547.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **99**, 303–319.

——, 1992: Iterative analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.*, **118**, 569–591.

Maddox, R. A., 1980: An objective technique for separating macroscale and mesoscale features in meteorological data. *Mon. Wea. Rev.*, **108**, 1108–1121.

Ooyama, K. V., 1987: Scale-controlled objective analysis. *Mon. Wea. Rev.*, **115**, 2479–2506.

Pauley, P., and X. Wu, 1990: The theoretical, discrete, and actual response of the Barnes objective analysis scheme for one- and two-dimensional fields. *Mon. Wea. Rev.*, **118**, 1145–1163.

Seaman, R. S., 1988: Some real data tests of the interpolation accuracy of Bratseth's successive correction method. *Tellus*, **40**(A), 173–176.

Thiebaux, H. J., 1976: Anisotropic correlation functions for objective analysis. *Mon. Wea. Rev.*, **104**, 994–1002.

——, and M. A. Pedder, 1987: *Spatial Objective Analysis with Applications in Atmospheric Science.* Academic Press, 295 pp.

——, H. L. Mitchell, and D. W. Shantz, 1986: Horizontal structure of hemispheric forecast error correlations of geopotential and temperature. *Mon. Wea. Rev.*, **114**, 1048–1066.

Wahba, G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.*, **108**, 1122–1143.