Filter Transfer Functions for the Method of Successive Corrections

MICHAEL G. SCHLAX AND DUDLEY B. CHELTON

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

(Manuscript received 13 March 2001, in final form 12 June 2001)

ABSTRACT

The frequency-domain characteristics of the successive corrections method in one dimension are investigated through the calculation of smoother weights and filter transfer functions. The successive corrections algorithm acts as a low-pass filter that behaves similarly to noniterative smoothers. The spectral content of fixed-span successive corrections estimates depends upon the number of iterations, the selected weighting function and the grid to which the dataset is interpolated. For a given weighting function and grid, increasing the number of iterations for the fixed-span case results in filter transfer functions with increased cutoff frequency and rolloff. Within data gaps, the use of more than one iteration leads to estimates that are more likely to be contaminated by high-frequency variability in the data. It is shown that variable-span successive corrections estimates are nearly independent of the choice of weights for the initial iterations and are almost equivalent to estimates obtained using a single iteration. The greater computational requirements of multiple-iteration successive corrections is a disadvantage for general applications.

1. Introduction

The method of successive corrections is an iterative procedure used to interpolate meteorological and oceanographic data onto uniform space–time grids (cf. chapter 3 of Daley 1991, and references therein). The successive corrections algorithm is based on a sequence of differencing and smoothing operations. At each iteration cycle, the current grid of estimates are interpolated to the data locations. The differences between these interpolants and the data are formed and smoothed back to the grid by forming weighted averages defined by a pre-specified weighting function parameterized by a “radius of influence,” or span. The smoothed differences are then added to the current estimates, and the next iteration is performed. The span of the weighting function may be fixed or varied from one iteration to the next. In its original form (Berghorsson and Doos 1955; Cressman 1959), the method was initialized by a grid of initial or “background” values typically obtained from forecasts or climatology. Subsequent modifications replaced the background fields with smoothed values obtained from the data (e.g., Barnes 1964, 1967; Koch et al. 1983), or provided approximants to optimal estimates (Bratseth 1986).

While supplanted to a degree by other methods of statistical interpolation and assimilation schemes (e.g., 3DVAR and 4DVAR, as described by Rabier et al. 1998; Courtier et al. 1998; Andersson et al. 1998; Rabier et al. 2000; Mahfouf and Rabier 2000; Klinker et al. 2000), the method of successive corrections continues to be applied to various meteorological and oceanographic datasets (e.g., Levitus 1982; Vasquez et al. 1990; Perigaud 1990; Levitus and Boyer 1994; daSilva et al. 1994; Josey et al. 1998, 1999; Liu et al. 1998; Kent et al. 2000). Even if the use of successive corrections is declining, its widespread historical use makes a clear understanding of the properties of the method useful.

Successive corrections is linear in the data, and acts as a low-pass filter, or smoother, whose filter transfer functions have been investigated for certain cases. Barnes (1964) analyzed the spectral content of fixed-span successive corrections for the continuous case with a fixed span and concluded that as the number of iterations increased, shorter scales of variability in the data were included in the gridded estimates. In his terms, the gridded estimates “converged” toward the “true” field values. The number of iterations could be varied depending on the data distribution, measurement errors, and the scales of variability desired in the final grid. Thus, the degree of smoothing, or the effective filter cutoff, could be controlled through setting the span and number of iterations. Barnes (1967) [see Koch et al. (1983) for a more readily accessible review] realized that, by varying the span of the weighting function, the desired degree of filtering could be reached in fewer iterations. By selecting the decrease in span appropriately, the desired filter cutoff could be reached in only
two iteration cycles at the expense of a somewhat less sharp filter (i.e., a slower rolloff of the filter transfer function at the cutoff wavelength). Achnetemier (1987) performed a similar analysis and concluded that superior filtering properties (more rapid rolloff) were obtained by fixed-span iterations. Daley (1991) calculated filter transfer functions for two-iteration, fixed-span successive corrections for the continuous case and for regular, discrete data distributions. Using the elegant formulation of Stephens and Stitt (1970), Stephens and Polan (1971) derived transfer functions for one and two-iteration, fixed-span successive corrections estimates for randomly distributed data.

While use of the Barnes (1967) two-iteration implementation of the successive corrections method is widespread, and supported by his theoretical analyses, the application of variable-span successive corrections with more than two iterations is also common. For example, Levitus (1982) applied a four-iteration, variable-span implementation and provided a filter transfer function, but it is not clear how this was derived based on the fixed-span formula of Barnes (1964) that is cited. Some applications of this technique are ad hoc: Vasquez et al. (1990) and Liu et al. (1998) apparently gave little consideration to the effect of what seem to be arbitrarily selected numbers of iterations and spans.

In this paper, a matrix formulation of the successive corrections algorithm is presented, from which the smoother weights (the a posteriori weights of Daley 1991) can be readily obtained for general implementations of the successive corrections algorithm, and for arbitrary distributions of data. Using the smoother weights so derived, filter transfer functions for a number of examples of successive corrections in one dimension are presented. The effects of varying the number of iterations and the spans are discussed, as is the influence of the estimation grid spacing. Comparison is made with the loess smoother, a noniterative smoothing algorithm.

As will be shown, fixed-span successive corrections is capable of generating smoother weights that lead to an effective low-pass filter with a rapid rolloff, in accordance with the earlier works of Barnes (1964, 1967) and Achnetemier (1987). The variable-span implementation is shown to depend almost entirely on the span used in the final iteration, obviating the need for the earlier iterations. As with any smoothing algorithm, the filtering properties of successive corrections depends heavily on the distribution of the data around any grid point. The properties of any specific implementation of successive corrections thus need to be evaluated on a case-by-case basis. For this reason no specific set of parameters is recommended.

2. Matrix formulation of successive corrections

Consider a set of N observations \( y_m \) of some variable of interest at the locations (in one, two, or three dimensions) \( x_m, n = 1, \ldots, N \). From these data, we seek estimates \( y_n \) at \( M \) selected locations \( \eta_n, m = 1, \ldots, M \). Ordinarily, the \( \eta_n \) define a regular grid in one or more dimensions. The successive corrections method iterates to obtain a sequence of estimates \( g_m(n) \) as

\[
g_m(n + 1) = g_m(n) + \sum_{n=1}^{N} s_{mn}(n + 1) [y_n - \hat{y}_n(n)],
\]

where \( n \) is the iteration number, \( s_{mn}(n) \) are specified weights (here, referred to as the analytical weights to distinguish them from the smoother weights calculated later), and the \( \hat{y}_n(n) \) are interpolated (using a method assumed here to be linear) values of the variable at the locations \( x_n \) obtained from the \( g_m(n) \):

\[
\hat{y}_n(n) = \sum_{m=1}^{M} I_{mn} g_m(n).
\]

To use matrix notation, define the vectors \( \mathbf{y} = [y_n] \), \( \mathbf{g} = [g_m] \), and \( \mathbf{\hat{y}} = [\hat{y}_n] \), the \( M \times N \) smoother matrices \( \mathbf{S}(n) = [s_{mn}(n)] \), and the \( N \times M \) interpolation matrix \( \mathbf{I} = [I_{mn}] \), so that

\[
\mathbf{g}(n + 1) = \mathbf{g}(n) + \mathbf{S}(n + 1) [\mathbf{y} - \mathbf{\hat{y}}(n)]
\]

\[
= \mathbf{g}(n) + \mathbf{S}(n + 1) [\mathbf{y} - \mathbf{L} \mathbf{g}(n)]
\]

\[
= [\mathbf{I} - \mathbf{S}(n + 1) \mathbf{L}] \mathbf{g}(n) + \mathbf{S}(n + 1) \mathbf{y}.
\]

Thus, given a background field \( \mathbf{g}(0) \),

\[
\mathbf{g}(n) = \begin{bmatrix} [1 - \mathbf{S}(1) \mathbf{L}] \mathbf{g}(0) + \mathbf{S}(1) \mathbf{y}, & \mathbf{g}(2) = [1 - \mathbf{S}(2) \mathbf{L}] \mathbf{g}(1) + \mathbf{S}(2) \mathbf{y} \\
\end{bmatrix}
\]

\[
= [1 - \mathbf{S}(2) \mathbf{L}] [1 - \mathbf{S}(1) \mathbf{L}] \mathbf{g}(0) + \mathbf{S}(1) \mathbf{y} + \mathbf{S}(2) \mathbf{y}
\]

\[
+ [1 - \mathbf{S}(2) \mathbf{L}] \mathbf{S}(1) \mathbf{y} + \mathbf{S}(2) \mathbf{y},
\]

or in general, for \( K \) iterations

\[
\mathbf{g}(K) = \left[ \prod_{k=K}^{1} [1 - \mathbf{S}(k) \mathbf{L}] \right] \mathbf{g}(0)
\]

\[
+ \sum_{k=1}^{K-1} \left[ \prod_{k=K}^{1} [1 - \mathbf{S}(k) \mathbf{L}] \right] \mathbf{S}(\mathbf{K} \mathbf{y})
\]

As equation (1) shows, before applying the successive corrections method, one must select the background field \( \mathbf{g}(0) \), an estimation grid \( \eta_n \) and interpolation algorithm that are both embodied in \( \mathbf{L} \), the number of iterations, \( K \), and the analytical weights for each datum and iteration, \( \mathbf{S}(n) \).

A variety of functions have been used historically to define the analytical weights. In general these weights are defined by \( s_{mn} = s(r_{mn})b_{mn} \), where \( r_{mn} = \| \eta_m - x_n \| \) is the distance between the datum and the grid point,
3. Examples of successive corrections

Examples of different implementations of successive corrections in one dimension (referred to here as time) with regularly spaced data are presented in this section. In each case here and in section 4, the background field is taken to be \( g(0) = 0 \), the interpolation step is accomplished using cubic Lagrange interpolation, and the function defining the analytical weights (with the exception of one case, cf. Fig. 2) is a Gaussian, \( s(r) = \exp(-r^2/\alpha^2) \). With \( g(0) = 0 \), equation (1) may be written as \( g(K) = Ay \), and thus for each grid point, the procedure results in a linear estimate or smoother,

\[
g_m = \sum_n a_{mn} y_n.
\]

For each example, \( A \) is constructed and the smoother weights \( a_{mn} \) for a specific grid point \( \eta_m \) are displayed. The filtering properties of the estimate are examined using the modulus of the equivalent transfer function for that grid point:

\[
\| \hat{P}_m(f) \| = \left\| \sum_{n=1}^{N} a_{mn} \exp(-2\pi j f \eta_n) \right\|.
\]

The equivalent transfer function bounds the spectral content of a specific estimate and is a generalization of the usual transfer function (Schlax and Chelton 1992).\(^1\)

The equivalent transfer function modulus for each example of the successive corrections method can be compared to an ideal low-pass filter with transfer function modulus

\[
\| P(f; f_c) \| = \begin{cases} 
1, & \text{if } f \leq f_c \\
0, & \text{otherwise}
\end{cases}
\]

using the relative error, expressed as

\[
R(f_c; \hat{P}_m) = \frac{\int \| P(f; f_c) \|^2 df - \| \hat{P}_m(f) \|^2 df}{\int \| P(f; f_c) \|^2 df}.
\]

\(^1\) Assuming that the spectral representation (e.g., Priestley 1981) holds, then \( y_n = \int_s \exp(2\pi j f \eta_n) F(s) ds \) where \( F(s) \) is the Fourier transform associated with the stochastic process generating the \( y_n \). Direct substitution into (2) leads to \( g_m = \int_s \hat{P}_m(f) F(s) df \), the form applied by Schlax and Chelton (1992). This form also follows from the power theorem (e.g., Bracewell 1986). The earlier works of Jones (1972) and Yang and Shapiro (1973) (that were drawn to our attention by M. Askelson) preferred to write the equivalent form \( g_m = \int_s \phi(s) F(s) \exp(2\pi i f \eta_n) df \) with \( \phi(s) = \hat{P}_m(f) \exp(-2\pi j f \eta_n) \). Since the description of the filtering properties of the successive corrections algorithm here depends only upon \( \| \hat{P}_m(f) \| \), this difference in notation is not significant for present purposes. The bias calculations of Schlax and Chelton (1992) are consistent with their definition of \( \hat{P}_m(f) \), and this formulation is in accord with their applications in which it is assumed that only \( \| P(f) \| \) is known a priori.
The cutoff frequency for a given successive corrections estimate is defined as the value of \( f_c \) that minimizes the relative error. In the examples analyzed here, the integral in the numerator of (3) was evaluated numerically, and the minimization required to define \( f_c \) was performed using the golden section algorithm (Press et al. 1992).

\[ a. \text{ Coincident data and grid locations} \]

When the data locations and grid points are the same, the interpolation matrix \( L \) is the identity. This ideal case has been studied by previous authors (e.g., Daley 1991) to provide a simplified analysis of the successive corrections method. It is reexamined here in order to isolate the effects of the choices of span and the number of iterations.

For the examples in this section, the data locations \( x_i \) and the estimation grid points \( \eta_y \) are coincident on a grid with spacing 0.05. The specific grid point is selected so that the edges of the dataset will not influence the smoothing provided by \( T \). The grid extends to 5 \( \sigma_m \) on either side of the selected grid point, where \( \sigma_m \) is the largest value of the Gaussian e-folding scale \( \sigma \) applied in a given example.

\[ 1) \text{ FIXED-span iterations} \]

The first case considered is iteration with a fixed span, that is, \( S(v) \) is constant with a Gaussian e-folding scale of \( \sigma = 1 \). The number of iterations, \( K \), was varied from 1 through 20. The results are shown in Fig. 1 for selected values of \( K \). It is apparent from Fig. 1b that the successive corrections method is a low-pass filter. For \( K = 1 \), the successive corrections method is equivalent to a simple, Gaussian-weighted smoother. As the number of iterations is increased, the analytical weights are modified by the algorithm and the filter passes successively higher frequencies and has a steeper rolloff (lower relative error), in agreement with the results of Barnes (1964, 1967) and Auchmeier (1987). The cutoff frequency increases by almost a factor of 2 from one iteration to four iterations. This implementation is convergent when using Gaussian analytical weights (e.g., Daley 1991); both the relative error and the cutoff frequency approach asymptotic values with increasing \( K \).

It is noteworthy that the widely used Cressman weighting function results in a divergent iteration sequence for \( K > 4 \) (Fig. 2). This fundamental shortcoming of the Cressman function as a weighting function for the successive corrections method has previously been pointed out by Daley (1991).

\[ 2) \text{ VARIABLE-span, two iterations} \]

This example is the case presented by Barnes (1967) and Koch et al., (1983). There are two iterations with different Gaussian analytical weights, characterized by the e-folding parameters \( \sigma_i \) for the first iteration and \( \sigma_j \) for the second. For this example, \( \sigma_i \) was varied from 1 to 6, while \( \sigma_j \) was held fixed at \( \sigma_j = 1 \). The results are shown in Fig. 3. The smoother weights and equivalent transfer functions are similar for all \( \sigma_i \) except when \( \sigma_i = 1 \), corresponding to the \( K = 2 \) case in the previous example. For \( \sigma_i > 2 \), the cutoff frequencies and relative errors are nearly insensitive to \( \sigma_i \). Comparison with Fig. 1 shows that \( \sigma_i \geq 2 \) results in smoother weights and transfer functions almost identical to those of the \( K = 1 \) example. Apparently the additional iteration with \( \sigma_j < \sigma_i \) has only a slight effect.

To understand this result, consider the final estimate with \( K = 2 \) [recalling that \( L = I \) and \( g(0) = 0 \):]

\[ g(2) = [I - S(2)]S(1)y + S(2)y. \]

The first term on the right-hand side consists of a low-pass filtering operator \( S(1) \) followed by a high-pass filtering operator \([I - S(2)]\). When \( \sigma_i > \sigma_j \), \( S(1) \) smooths the data more than \( S(2) \), so that the lower frequencies retained after low-pass filtering with \( S(1) \) will be attenuated by the high-pass filtering using \([I - S(2)]\). Depending on the smoothing provided by \( S(2) \) relative to \( S(1) \), this first term on the right-hand side may be negligible, so that \( g(2) = S(2)y \). For \( \sigma_i > 2 \) this is the case, while for \( \sigma_i = 2 \), the filtering characteristics of \( S(1) \)
and $S(2)$ are close enough so that the combined effect of the low-pass filter followed by the high-pass filter is small but noticeable (see the dashed line in Figs. 3a and 3b). This situation is similar to that described by Koch et al. (1983), where multiple iterations with a slowly decreasing span improve the filter response (see their Fig. A2).

3) VARIABLE-SPAN, THREE ITERATIONS

For this example, $K = 3$, and the final iteration is fixed at $\sigma_1 = 1$. The analytical weights for the first two iterations vary with $\sigma_1$ and $\sigma_3$, ranging from 1 through 6, with $\sigma_1 \geq \sigma_3$. Only the cutoff frequencies and relative errors are presented in Fig. 4. For fixed $\sigma_3$, the filtering characteristics are almost independent of $\sigma_1$; and for $\sigma_2 \geq 2$, the filtering characteristics are only weakly dependent on $\sigma_3$. The lowest relative errors are found for the three-iteration, fixed-span case, $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The situation is similar to the two-iteration example considered in section 3a(2): with variable spans, the succession of low- and high-pass filtering operators in the second term of (1) essentially cancel one another, and the successive corrections revert to the equivalent of a single pass, with the final, smallest span for the analytical weights.

The relative errors shown in Figs. 1, 3, and 4 demonstrate that iteration with fixed spans results in successively sharper filter transfer functions, while the use of variable-spans does little to enhance the filtering properties of the successive corrections method unless the spans decrease slowly with each iteration.

b. The effects of grid spacing

An integral part of each successive corrections iteration is interpolation from the grid values of the previous step to the data locations. This interpolation is embodied in the matrix $L$ in equation (1). The exact form of $L$ will depend on both the interpolation algorithm used and the configuration of the grid points $\eta_m$ relative to the data. Probably the most common interpolation algorithm used in successive corrections is polynomial or Lagrange interpolation (Press et al. 1992).

As noted above, when the data and estimation grid locations coincide, $L = I$ and the interpolation step has no effect. In realistic applications, where the data are irregularly spaced or otherwise do not coincide with the estimation grid points, this ideal situation will not exist and the nature of the interpolation must be considered.

Figure 5 shows smoother weights (i.e., the $l_{nm}$) and filter transfer function moduli for the cubic Lagrange interpolation (used herein) applied with two different grid spacings, and to situations where the data locations are located either midway or one-quarter of the way between estimation grid points. For the grid spacing of 0.05, smoother weights for both the midpoint and quarter-point are concentrated about those points, and the resulting transfer functions show that effectively none of the existing signal will be attenuated. For the coarse grid spacing of 1.0, on the other hand, the smoother weights are spread out away from the interpolation points and the Lagrange interpolation results in some signal attenuation that changes in quality from the midpoint (greater attenuation) to the quarter-point (less attenuation). Figure 5 demonstrates the fact that cubic Lagrange interpolation based on a coarse grid is less able to approximate a highly variable function than cubic Lagrange interpolation based on a fine grid. Thus, for a coarse estimation grid, if the data locations do not coincide with the estimation grid points, the interpolation step will introduce some filtering that varies from one data location to another. Depending on the spectral content of the process being smoothed, this may in turn degrade the filtering properties of the successive corrections algorithm, since the frequency content of the interpolated time series will vary with location.

The effect of varying the estimation grid spacing with the uniform dataset of section 3b is examined here for
one case each of fixed- and variable-span successive corrections. Figure 6a shows the smoother weights and transfer function moduli for Gaussian-weighted, fixed-span successive corrections estimates with $K = 3$, data with a uniform spacing of 0.05, and grid spacings $\Delta$ varying from 0.05 to 1.0. The span of the Gaussian function was set to $\sigma = 0.4$ so that the transfer function for the $\Delta = 0.05$ case (i.e., when the estimation grid locations coincide with the data points) has unit cutoff frequency. Figure 6a shows that for $\Delta > 0.05$, the fixed-span successive corrections method filters the data more strongly, retaining less high-frequency content. The transfer functions for $\Delta = 0.75$ and 1.0 cases deviate from the other cases in that they both markedly exceed unit value at lower frequencies. Variability at these frequencies is therefore amplified when the interpolation grid is coarse.

The variable-span case (Fig. 6b) is Gaussian-weighted with $K = 3$, $\sigma_1 = 0.274$, $\sigma_2 = 2\sigma_1$, and $\sigma_3 = 4\sigma_1$, which results in a transfer function with unit cutoff frequency for $\Delta = 0.05$. For the same range of $\Delta$ as shown in Fig. 6a, the filtering characteristics are much less sensitive to the change in grid spacing. This is another manifestation of the cancellation of the concatenated low- and high-pass filter operators discussed in section 3a(2).

4. Comparison with the loess smoother and further examples

It was shown in section 3 that the successive corrections algorithm acts as a low-pass filter. There are many other smoothing and interpolation algorithms that behave similarly. One of these methods, the quadratic loess smoother (Cleveland 1979, modified to use a Gaussian weighting function by Schlax et al. 2001) has been selected here for comparison with three cases of Gaussian-weighted successive corrections: one with $K = 1$, and the $K = 3$, fixed- and variable-span examples considered in section 3b. For the $K = 1$ case, $\sigma$ is set to the same value as the final iteration for the $K = 3$ variable-span case. The estimation grid spacing has been set at $\Delta = 0.05$ for the successive corrections estimates. The parameters for the four methods have been chosen to yield unit cutoff frequencies. The loess smoother requires only a single smoothing parameter, with weights obtained using the method of Schlax et al. (2001).

a. Uniformly spaced data

Figure 7a shows the smoother weights and transfer functions for the four smoothers when both the data and the estimation grid have a uniform spacing of 0.05. The smoother weights for all but the $K = 1$ successive corrections are quite similar. The transfer functions all display the general low-pass form with relatively minor differences. The $K = 3$, fixed-span successive corrections transfer function has the lowest relative error (0.076) of the four cases, and thus best approximates the ideal low-pass filter, followed by loess with a slightly higher relative error of 0.083, the $K = 3$ variable-span (0.113) and the $K = 1$ (0.133) successive corrections smoothers. As in the previous examples, the additional iterations in the variable-span case provide little improvement in relative error over a single iteration. Moreover, the variable-span case has much larger relative error than fixed-span successive corrections.

Figure 7b shows how the smoothers respond to a sparser uniform data distribution with spacing 0.3 (and estimation grid spacing $\Delta = 0.05$). The smoother weights for all but the $K = 1$ successive corrections are quite similar. The transfer functions all display the general low-pass form with relatively minor differences. The $K = 3$, fixed-span successive corrections transfer function has the lowest relative error (0.076) of the four cases, and thus best approximates the ideal low-pass filter, followed by loess with a slightly higher relative error of 0.083, the $K = 3$ variable-span (0.113) and the $K = 1$ (0.133) successive corrections smoothers. As in the previous examples, the additional iterations in the variable-span case provide little improvement in relative error over a single iteration. Moreover, the variable-span case has much larger relative error than fixed-span successive corrections.
the selected estimation grid point. The peak in the transfer functions around a frequency of 3 is due to aliasing in the classical sense. Similar peaks will be found for all cases of uniformly spaced data; were the abscissa for the plot of the transfer function for the 0.05 data spacing in Fig. 7a extended to include frequencies up to the Nyquist frequency of 10, the first alias for that example would be observed. The data spacing directly affects the ability of any smoothing algorithm to filter the data. Signal energy at frequencies near 3 would be attenuated by the smoothers in Fig. 7a with the fine data spacing; the same signal energy would contaminate the estimates with the coarse dataset considered in Fig. 7b, thus degrading the performance of the low-pass filters.

b. Nonuniformly spaced data

All of the previous examples considered the smoother weights and transfer functions with uniformly spaced data and at an estimation grid point far removed from the edge of the dataset. In realistic data analyses, data are typically nonuniformly distributed in space and time,
yielding both regions of high data density and regions with significant data gaps. The performance of interpolation algorithms can change drastically under such conditions (Stephens 1967; Jones 1972; Schlax and Chelton 1992). It is thus instructive to consider the behavior of these smoothers under conditions other than the ideal ones considered thus far.

1) Randomly Spaced Data

Figures 8a and 8b show how the smoothers respond to data that are randomly spaced. For these examples, the estimation grid spacing is $\Delta = 0.05$, and the data locations were distributed according to the uniform random distribution with average spacings of 0.05 (Fig. 8a) and 0.3 (Fig. 8b). Randomly spaced data reduces the ability of all of the smoothers to accurately low-pass filter the data. In both Figs. 8a and 8b, it is seen that the transfer functions have lower cutoff frequencies than their counterparts for the uniformly spaced cases in Figs. 7a and 7b. For both cases, more signal energy between frequencies of 1 and 3 will be incorporated into the smoothed fields. This “aliasing” resulting from non-uniformly spaced data has been examined by Jones (1972), Schlax and Chelton (1992), and Chelton and Schlax (1994) and can significantly degrade the performance of smoothing algorithms.

2) Edges and Gaps

An extreme case of nonuniformly spaced data results when the estimation grid point lies at the edge of the dataset. Figure 9 shows the smoother weights and transfer function moduli for the four smoothers in this setting when the data and estimation grid spacing are both 0.05. The performance of all algorithms is significantly degraded. The loess smoother changes most dramatically, amplifying much more of the high-frequency energy than estimates at locations away from the dataset edge, presumably because of the relatively large smoother weights associated with the data points nearest the edge (see the left panel of Fig. 9). The $K = 3$ successive corrections smoothers respond less to the change of estimation location, while the $K = 1$ case (i.e., simple Gaussian smoothing) is the least affected.

Figures 10 and 11 demonstrate how the smoothers behave at a data gap. Once again, where the data exist, they have a uniform data spacing of 0.05, while $\Delta = 0.05$. In both figures the gap width varies from 0.2 through 1.0. Figure 10 shows the smoother weights and transfer functions for estimates located in the center of the gap, while in Fig. 11 the estimation points lie one quarter of the way across the gap.

For an estimate at the center of the gap (Fig. 10), increasing the gap width results in smoother weights that are dominated by the weights associated with the data very near the gap edges. Comparison of Fig. 10 with Fig. 7 shows that the near-edge data receive much more relative weight than the data near the center of the effective span of the smoothers for the case with uniform data. This is most pronounced for the loess smoother, which, for the span used in this example, is not capable of yielding an estimate for the widest gap since the half-span of the smoother is not wide enough to encompass any observations outside of the data gap.
The effect on the transfer functions is to lower the cutoff frequency for the main lobe, and to increase the number and amplitudes of the sidelobes. Thus, estimates at the center of a data gap may be contaminated by high-frequency signal energy.

It is noteworthy that both the fixed and variable span $K = 3$ estimates behave in a similar fashion when used to interpolate across data gaps: contrary to the conventional wisdom that often motivates the use of variable-span successive corrections, the first two iterations do not compensate for the presence of the data gap. Indeed, the utility of multiple interactions for the fixed-span case is questionable in this setting, since the $K = 1$ estimates have the smallest sidelobes in both Figs. 10 and 11, and are thus the least susceptible to aliasing of unresolved variability.

When the estimation point is closer to one of the gap edges (Fig. 11), the data points at the proximal edge are disproportionately weighted when the gap becomes large. The associated transfer functions do not have the distinct sidelobes present in Fig. 10. The difference in the weights and transfer functions between Figs. 10 and 11 imply that the estimates made at various locations across the gap will differ substantially in terms of their frequency content.

As an example of how the successive corrections method responds to a data gap, four implementations of the method were applied to a simulated dataset (Fig. 12). A 161-point realization of a first-order autoregressive process with parameter 0.7 and spacing 12.5 was generated. Utilizing a grid with the same spacing, successive correction estimates were calculated with $K = 1$ and $\sigma = 25$; $K = 3$ with a fixed span and $\sigma = 40$; $K = 5$ with variable spans and $\sigma_i = 500, 250, 125, 60$, and 25 for $i = 1, \ldots, 5$, respectively; and $K = 3$ with variable spans and $\sigma_1 = 125, \sigma_2 = 60$, and $\sigma_3 = 25$. Figure 12a shows the estimates obtained when a gap width of 300 is formed.
in the dataset. Outside of the gap, the estimates from all four implementations of the successive corrections method are very similar. As previously noted, the two variable-span implementations yield results almost identical to those obtained from the single iteration with $\sigma = 25$. For the chosen span, the fixed-span estimates are nearly indistinguishable from the $K = 1$ and variable-span estimates. Within the data gap (the gray shaded area in Fig. 12a) the estimates follow the trend of the last few data near the gap edge. This behavior is consistent with the observations made above: the near-edge data are given a disproportionate weight (see Fig. 11). The sudden jump in the estimates at the center of the gap occurs when the span bridges
the gap and the weights become symmetric at the gap center, and data on both edges of the gap are included in the estimate (see Fig. 10). The $K = 1$ estimates (dashed line) are the least variable across the gap.

The variability of the estimates in the gap is demonstrated by Fig. 12b, where the gap has been slightly narrowed to a width of 275, so that one additional datum on each side of the gap in Fig. 12a is included. This minor change in the dataset completely changes the estimates within the gap. Again, the high relative weight assigned to the near-edge data produces unstable estimates. The similarity of the two variable-span estimates is of particular interest. The inclusion of two extra iterations with relatively large spans (i.e., $\sigma_1 = 500$ and $\sigma_2 = 250$) does not stabilize the estimates, as is generally presumed in the use of successive corrections.

Fig. 11. Same as Fig. 10 except with estimation points located one-quarter of the gap width from the left edge of the gap.
Because of the erroneous nature and extreme sensitivity of the estimates within the gap, the parameters used to generate the grid of estimates in Fig. 12 are clearly not appropriate for extrapolating across the given gap. Figure 13 shows a much smoother set of estimates obtained using the previous sets of spans increased by a factor of 5. The estimates so obtained are insensitive to the slight change in gap width, but clearly do not approximate the data outside of the gap as well as the estimates in Fig. 12. This demonstrates a fundamental trade-off that must be made when smoothing a dataset. High-resolution estimates (with a small amount of smoothing) can well approximate the data in densely sampled parts of the data record, but will be meaningless in data gaps. Low-resolution estimates (with a large amount of smoothing) more strongly filter the data and retain only relatively low-frequency features, but are less sensitive to data gaps. To retain uniform frequency content and accuracy in an interpolated time series, the choice of smoothing parameters is dictated by the widths of the data gaps. It is clear from Fig. 12 that this limitation is not mitigated by variable-span successive corrections, although this is usually the motivation for applying this method.

5. Discussion

The examples considered here demonstrate the complex dependence of the filtering characteristics of the successive corrections method on the choices of the various parameters. For a fixed span, both the cutoff frequency and the relative error vary considerably with the number of iterations as well as with the selected span. More iterations result in a sharper filter rolloff with a higher cutoff frequency. A poor choice for the smoothing weights (e.g., the Cressman weighting function) can result in a divergent sequence of estimates. The configuration of the estimation grid can also affect how fixed-span successive corrections filter the data.

An important conclusion from the results presented here is that a variable-span successive corrections case results in filtering that is essentially independent of the number of iterations, and almost wholly dependent on the span of the function generating the analytical weights for the final iteration. The inutility of multiple iterations for variable-span successive corrections suggests that this method is computationally inefficient and should be avoided. The advantage of the fixed-span method, namely that more iterations result in a sharper filter, appears to be offset in data gaps by the presence
of large sidelobes in the transfer functions that are less extreme for the \( K = 1 \) case of simple Gaussian smoothing. While all of the successive corrections estimates examined here are less sensitive to dataset edges than the loess smoother, it is not clear that the additional computational burden imposed by multiple-iteration successive corrections is compensated for by superior smoothing in general.

The quality of gridded fields produced by successive corrections will depend upon the correct choice of the parameters for the algorithm and an understanding of what the dataset is capable of resolving. As the examples in section 4 show, the filtering properties of any smoother can depend as much on the data distribution as on the selection of the smoother parameters. The fact that a given smoother and dataset allow the selection of a desired set of smoothing parameters does not imply that the nominal filtering will be achieved. If the parameters are chosen so that the low-pass filtering is insufficient to accommodate the data distribution, the resulting grid of estimates will be contaminated with spurious short-scale signals (cf. Figs. 12 and 13; Greenslade et al. 1997; Chelton and Schlax 1994), regardless of how fine the estimation grid spacing is.

Acknowledgments. The authors wish to thank Michael Freilich, Steve Esbensen, and Mark Askelson for thoughtful reviews and comments. The research presented in this paper was supported by Contract 1206715 from the Jet Propulsion Laboratory for funding of Jason Science Working Team activities and by NASA Grant NAS5-32965 for funding of Ocean Vector Winds Science Team Activities.

REFERENCES


