Linear Prediction as Applied to Interpolation over CCD Defects

Robert Lupton

October 18, 2002

The use of linear prediction in this context was suggested by Bill Press; a useful discussion is in Press et al. (1993); thanks also to Jeremy Goodman for helpful discussions.

1 General Theory

If we have a stationary process,

$$y_i \equiv s_i + n_i$$

where s_i is the signal, and n_i the noise, let us try to predict the value of the signal s^* at some point $*$ as

$$\tilde{s}^* = \sum_i d_i^* y_i. \tag{1}$$

Taking the expectation value of this equation, we see that

$$\langle \tilde{s}^* \rangle = \langle \sum_i d_i^* y_i \rangle$$

As the process is stationary (and as $\langle n \rangle = 0$ is practically the definition of noise),

$$\langle s \rangle = \sum_i d_i^* \langle s \rangle,$$

i.e.

$$\sum_i d_i^* = 1 \tag{2}$$

is the condition that the estimator be unbiased.

The expectation value of the residual is

$$\begin{aligned} R &\equiv \left\langle (s^* - \tilde{s}^*)^2 \right\rangle \\ &= \left\langle \left(s^* - \sum d_i^* (s_i + n_i) \right)^2 \right\rangle ; \end{aligned}$$

let us choose the d_i^* to minimise R, subject to the constraint of equation 2; that is we wish to minimise

$$R - 2\lambda \sum d_i^* .$$

Differentiating with respect to d_i^*,

$$\langle \tilde{s}^* (s_i + n_i) \rangle - \sum_j d_j^* \langle (s_j + n_j)(s_i + n_i) \rangle - \lambda = 0.$$

Noting that s_i and n_j are independent, and thus that $\langle sn \rangle = 0$, this becomes

$$(S + N)\mathbf{d}^* = \mathbf{S}^* - \lambda \mathbf{E},$$

i.e.

$$\mathbf{d}^* = (S + N)^{-1} (\mathbf{S}^* - \lambda \mathbf{E}) \tag{3}$$

where we have written $\langle s_i s_j \rangle$ and $\langle n_i n_j \rangle$ as the matrices S and N, $\langle s^* s_i \rangle$ and d_i as the vectors $\mathbf{S}^*$ and $\mathbf{d}^*$, and defined $\mathbf{E}$ as the vector with all components equal to one. The Lagrange multiplier λ is determined by the condition $\sum d_i^* = 1$, i.e.

$$\mathbf{E}^T \mathbf{d}^* = 1 = \mathbf{E}^T (S + N)^{-1} \mathbf{S}^* - \mathbf{E}^T (S + N)^{-1} \mathbf{E} \lambda$$

so

$$\lambda = \frac{\mathbf{E}^T (S + N)^{-1} \mathbf{S}^* - 1}{\mathbf{E}^T (S + N)^{-1} \mathbf{E}} \tag{4}$$

Instead of the auto-correlations S and N, it may be more convenient to calculate the structure function, defined as

$$\begin{aligned} V_{ij} &= \frac{1}{2} \left\langle (y_i - y_j)^2 \right\rangle \\ &= \left\langle y^2 \right\rangle + \langle y_i y_j \rangle . \end{aligned}$$

Note that the estimator of equation 3 is independent of any constant added to all of the y_i (a result that depends on our having insisted on an unbiased estimator for $\mathbf{d}^*$). Although we shall not make use of this, it also implies that we may use structure functions in place of the auto-correlations in equation 3 — and this remains true even if the process is low-frequency divergent and thus the true $\langle y^2 \rangle$ is unknown, or even non-existent.

2 Application to Seeing-convolved Data

2.1 Infinite Signal/Noise Ratio

Let us apply this theory to interpolation over defects in images. We shall assume that the images consist purely of point sources, so that the signal s is given by

$$s(\mathbf{x}) = P \otimes \sum (B + \delta(\mathbf{x} - \mathbf{x}_i))$$

where B is the background, P is the point-spread function, $\otimes$ signifies a convolution, and the $\mathbf{x}_i$ are the positions of the sources. Setting $B = 0$ (which, as noted at the end of the previous section, will not affect our $\mathbf{d}^*$), and assuming that the sources are scattered randomly,

$$\begin{aligned} S &= \langle s_i s_j \rangle = \langle P_i P_j \rangle = \langle P(\mathbf{x}_i) P(\mathbf{x}_j) \rangle = \langle P(\mathbf{0}) P(\mathbf{x}_i - \mathbf{x}_j) \rangle \\ \mathbf{S}^* &= \langle P_i P^* \rangle = \langle P(\mathbf{0}) P(\mathbf{x}_i - \mathbf{x}_*) \rangle. \end{aligned}$$

We may take the noise to be uncorrelated, so that N is diagonal.

In order to apply this to interpolation, let us set the noise to be infinite in the bad columns. Consider the noiseless case; equation 3 shows, unsurprisingly, that if there are no bad columns the best estimate of y_j is y_j itself. Let us now take the PSF P to be a Gaussian of the form

$$P = \exp(-|\mathbf{x}|^2/(2\alpha^2)),$$

so

$$S(\mathbf{x}_i - \mathbf{x}_j) = \exp(-|\mathbf{x}_i - \mathbf{x}_j)|^2/(4\alpha^2)),$$

(a result most easily seen by noting that the autocorrelation function is the Fourier transform of the power spectrum, and that the power spectrum of a set of uncorrelated PSFs is the sum of the individual power spectra), and

M	$\sum d_i^2$					$\mathbf{d}^*$				
1	0.500				0.500	0	0.500			
2	1.347			-0.274	0.774	0	0.774	-0.274		
3	2.004		0.166	-0.510	0.844	0	0.844	-0.510	0.166	
4	2.717	-0.108	0.351	-0.643	0.900	0	0.900	-0.643	0.351	-0.108

Table 1: Interplation coefficients for a single bad column, and various values of M Also listed are the values of $\sum d_i^2$, which give the variance in the interpolated pixels (in units of the variance of the good pixels).

restrict ourselves to only $2 * M + 1$ terms (centred on the bad column) in estimating the d_i. Table 1 shows the results for a variety of values of M, for $\alpha = 1$ and considering only one-dimensional interpolation.

It is clear that we may calculate such coefficients for any pattern of bad columns. If there are N_{bad} bad columns, and $M = 2$, the resulting coefficients are given in table 2; the corresponding values for one-sided interpolation are given in tables 4 and 5; and the coefficients for interpolating over long sections of bad data are given in table 3 and 6. Finally, the coefficients for $M = 2$ but with one of the 'good' columns actually bad (so $M_{\rm left} = 1$ and $M_{\rm right} = 2$) are in table 7.

In addition to the coefficients, the tables give the values of $\sum d_i^2$, the variance in the interpolated value in units of the variance of the good pixels. By choosing M suitably it is possible to tune the variance of the interpolated pixels, although whether this is worthwhile is not clear.

A simple image (the sum of a wide Gaussian and two delta functions convolved with the PSF, as illustrated in the bottom left panel) is shown along with the interpolated values in the accompanying figures.

2.2 The Effects of a Finite Signal/Noise Ratio

If we take the signal-noise ratio of the image to be finite, the interpolation coefficients change.

We have seen that when large amounts of data are absent, our estimate becomes equal to the mean of the end points, with transients of width consistent with the PSF.

In the other limit, that of zero S/N, the interpolated values are all equal to the mean of the end points (as there is no information in the signal, the best

N_{bad}	$\sum d_i^2$	$\mathbf{d}^*$									
1	1.347	-0.274	0.774	$\underline{0}$	0.774	-0.274					
2	1.972	-0.479	1.190	$\underline{0}$	0	0.521	-0.232				
2	1.972	-0.232	0.521	0	$\underline{0}$	1.190	-0.479				
3	2.073	-0.523	1.316	$\underline{0}$	0	0	0.254	-0.047			
3	1.524	-0.314	0.814	0	$\underline{0}$	0	0.814	-0.314			
3	2.073	-0.047	0.254	0	0	$\underline{0}$	1.316	-0.523			
4	2.076	-0.520	1.337	$\underline{0}$	0	0	0	0.123	0.060		
4	1.035	-0.292	0.891	0	$\underline{0}$	0	0	0.394	0.007		
4	1.035	0.007	0.394	0	0	$\underline{0}$	0	0.981	-0.292		
4	2.076	0.060	0.123	0	0	0	$\underline{0}$	1.337	-0.520		
5	2.079	-0.518	1.340	$\underline{0}$	0	0	$\cdots$	0.093	0.085		
5	0.973	-0.280	0.907	0	$\underline{0}$	0	$\cdots$	0.223	0.149		
5	0.405	0.053	0.447	0	0	$\underline{0}$	$\cdots$	0.447	0.053		
6	2.079	-0.518	1.340	$\underline{0}$	0	0	$\cdots$	0.089	0.089		
6	0.972	-0.277	0.909	0	$\underline{0}$	0	$\cdots$	0.188	0.180		
6	0.331	0.068	0.461	0	0	$\underline{0}$	$\cdots$	0.272	0.198		
7	2.079	-0.518	1.340	$\underline{0}$	0	0	0	$\cdots$	0.089	0.089	
7	0.972	-0.277	0.910	0	$\underline{0}$	0	0	$\cdots$	0.184	0.183	
7	0.329	0.070	0.464	0	0	$\underline{0}$	0	$\cdots$	0.237	0.229	
7	0.255	0.213	0.287	0	0	0	$\underline{0}$	$\cdots$	0.287	0.213	
8	2.079	-0.518	1.340	$\underline{0}$	0	0	0	$\cdots$	0.089	0.089	
8	0.972	-0.277	0.910	0	$\underline{0}$	0	0	$\cdots$	0.184	0.183	
8	0.329	0.070	0.464	0	0	$\underline{0}$	0	$\cdots$	0.234	0.232	
8	0.253	0.215	0.290	0	0	0	$\underline{0}$	$\cdots$	0.251	0.243	
9	2.079	-0.518	1.340	$\underline{0}$	0	0	0	0	$\cdots$	0.089	0.089
9	0.972	-0.277	0.910	0	$\underline{0}$	0	0	0	$\cdots$	0.184	0.183
9	0.329	0.070	0.464	0	0	$\underline{0}$	0	0	$\cdots$	0.233	0.233
9	0.253	0.216	0.290	0	0	0	$\underline{0}$	0	$\cdots$	0.247	0.247
9	0.250	0.246	0.254	0	0	0	0	$\underline{0}$	$\cdots$	0.254	0.246
10	2.079	-0.518	1.340	$\underline{0}$	0	0	0	0	$\cdots$	0.089	0.089
10	0.972	-0.277	0.910	0	$\underline{0}$	0	0	0	$\cdots$	0.184	0.183
10	0.329	0.070	0.464	0	0	$\underline{0}$	0	0	$\cdots$	0.233	0.233
10	0.253	0.216	0.290	0	0	0	$\underline{0}$	0	$\cdots$	0.247	0.247
10	0.250	0.246	0.254	0	0	0	0	$\underline{0}$	$\cdots$	0.250	0.249

Table 2: Interpolation coefficients for $\alpha = 1$, $M = 2$, $S/N = \infty$. The position of the interpolated value is underlined. Note that for $N_{bad} \geq 5$, only the first half of the coefficients are listed (the remaining ones are given by symmetry). For larger values of N_{bad}, the coefficients for $N_{bad} = 10$ may be used, as the central interpolated values are simply the mean of the end points.

$\sum d_i^2$							$\mathbf{d}^*$							
2.079	-0.518	1.340	$\underline{0}$	0	0	0	0	0	0	$\cdots$	0	0	0.089	0.089
0.972	-0.277	0.910	0	$\underline{0}$	0	0	0	0	0	$\cdots$	0	0	0.183	0.183
0.329	0.071	0.464	0	0	$\underline{0}$	0	0	0	0	$\cdots$	0	0	0.233	0.233
0.253	0.216	0.290	0	0	0	$\underline{0}$	0	0	0	$\cdots$	0	0	0.247	0.247
0.250	0.246	0.254	0	0	0	0	$\underline{0}$	0	0	$\cdots$	0	0	0.250	0.250
0.250	0.250	0.250	0	0	0	0	0	$\underline{0}$	0	$\cdots$	0	0	0.250	0.250

Table 3: Interpolation coefficients for $\alpha = 1$, $M = 2$, $S/N = \infty$ for wide stretches of bad data. The position of the interpolated value is underlined. Note that this table is equivalent to the last entry in table 2

$\sum d_i^2$				$\mathbf{d}^*$				
2.225	-0.429	1.429	$\underline{0}$	0	0	0	0	0
1.204	-0.093	1.093	0	$\underline{0}$	0	0	0	0
0.577	0.303	0.697	0	0	$\underline{0}$	0	0	0
0.503	0.463	0.537	0	0	0	$\underline{0}$	0	0
0.500	0.496	0.504	0	0	0	0	$\underline{0}$	0
0.500	0.500	0.500	0	0	0	0	0	$\underline{0}$

Table 4: Interpolation coefficients for $\alpha = 1$, $M = 2$, $S/N = \infty$, interpolating from one side only. The position of the interpolated value is underlined.

$\sum d_i^2$					$\mathbf{d}^*$				
5.092	0.637	-1.338	1.702	$\underline{0}$	0	0	0	0	0
5.230	0.946	-1.445	1.499	0	$\underline{0}$	0	0	0	0
3.137	0.933	-1.031	1.097	0	0	$\underline{0}$	0	0	0
2.290	0.890	-0.809	0.919	0	0	0	$\underline{0}$	0	0
2.118	0.877	-0.758	0.880	0	0	0	0	$\underline{0}$	0
2.098	0.876	-0.751	0.876	0	0	0	0	0	$\underline{0}$

Table 5: Interpolation coefficients for $\alpha = 1$, $M = 3$, $S/N = \infty$ interpolating from one side only. The position of the interpolated value is underlined.

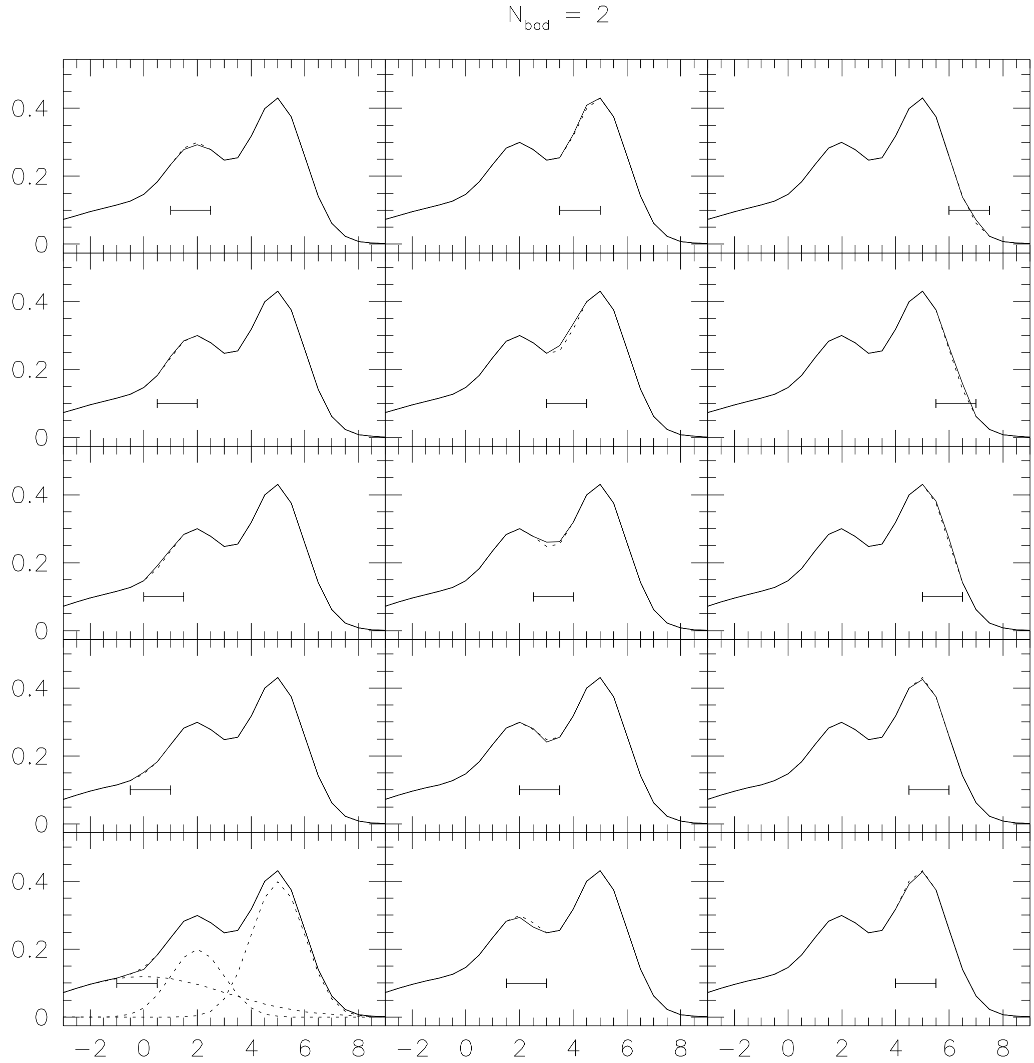

Figure 1: The effect of interpolating over two bad columns. The bottom left panel shows (as dotted lines) the model being interpolated over; the horizontal 'error bar' shows the range of missing data). The correct data is shown dotted, the reconstruction is shown as a solid line.

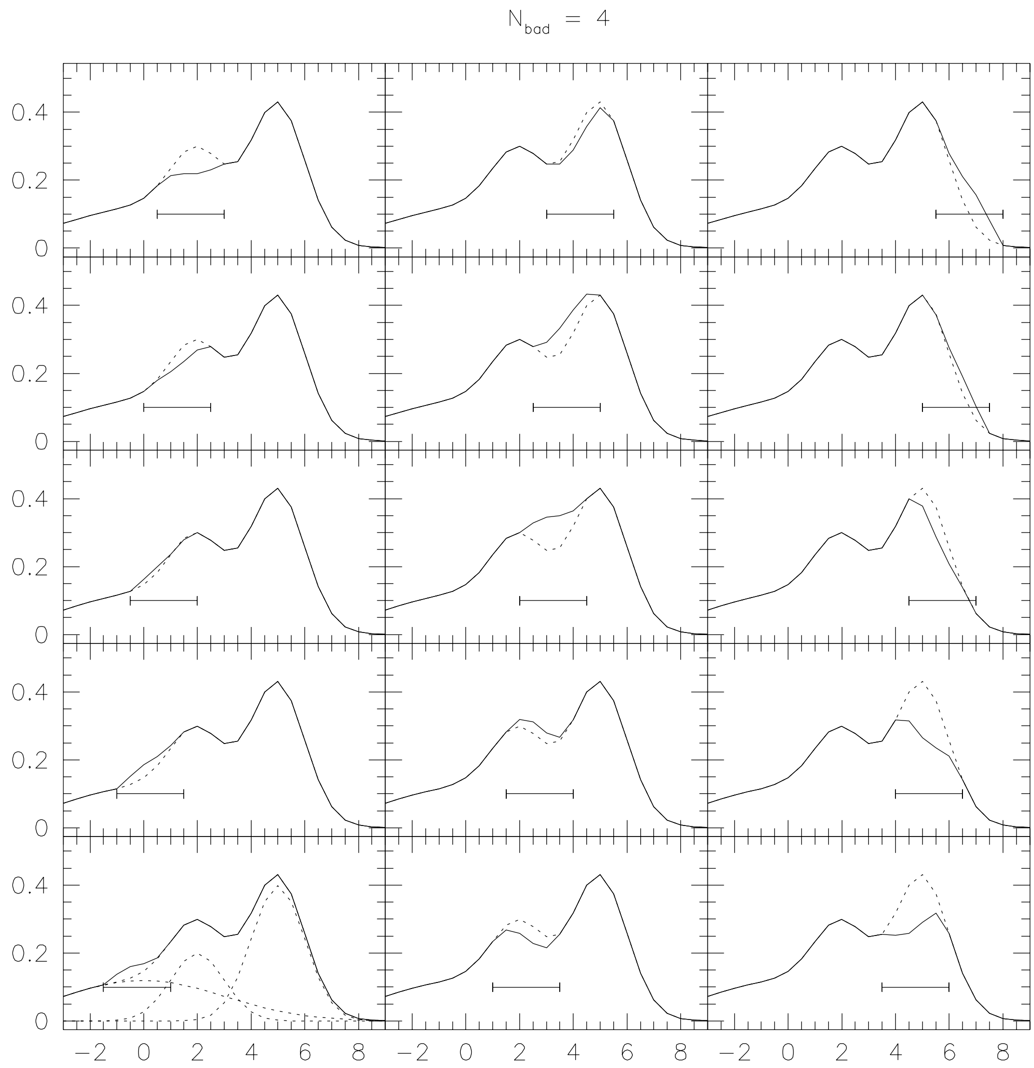

Figure 2: The effect of interpolating over four bad columns; see figure 1 for a description of the symbols.

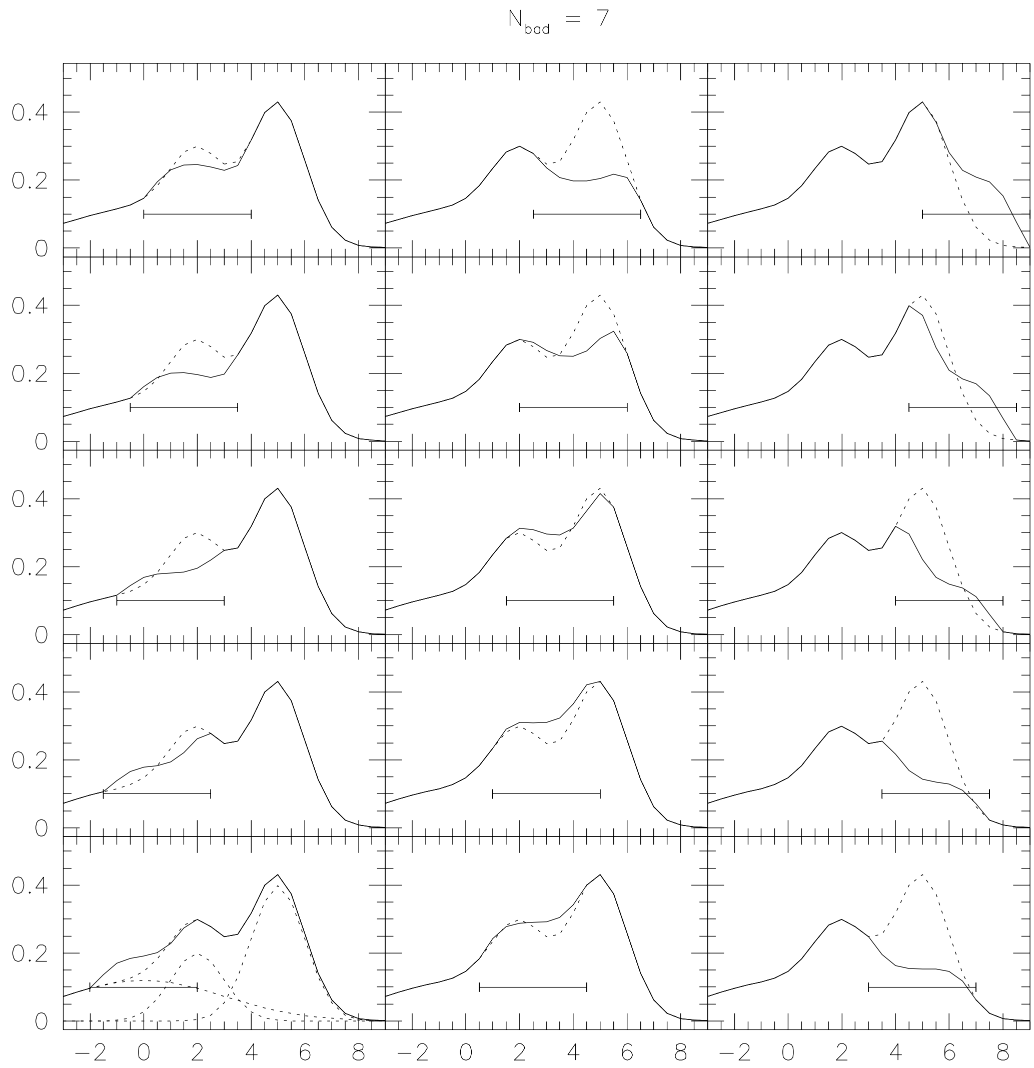

Figure 3: The effect of interpolating over seven bad columns; see figure 1 for a description of the symbols.

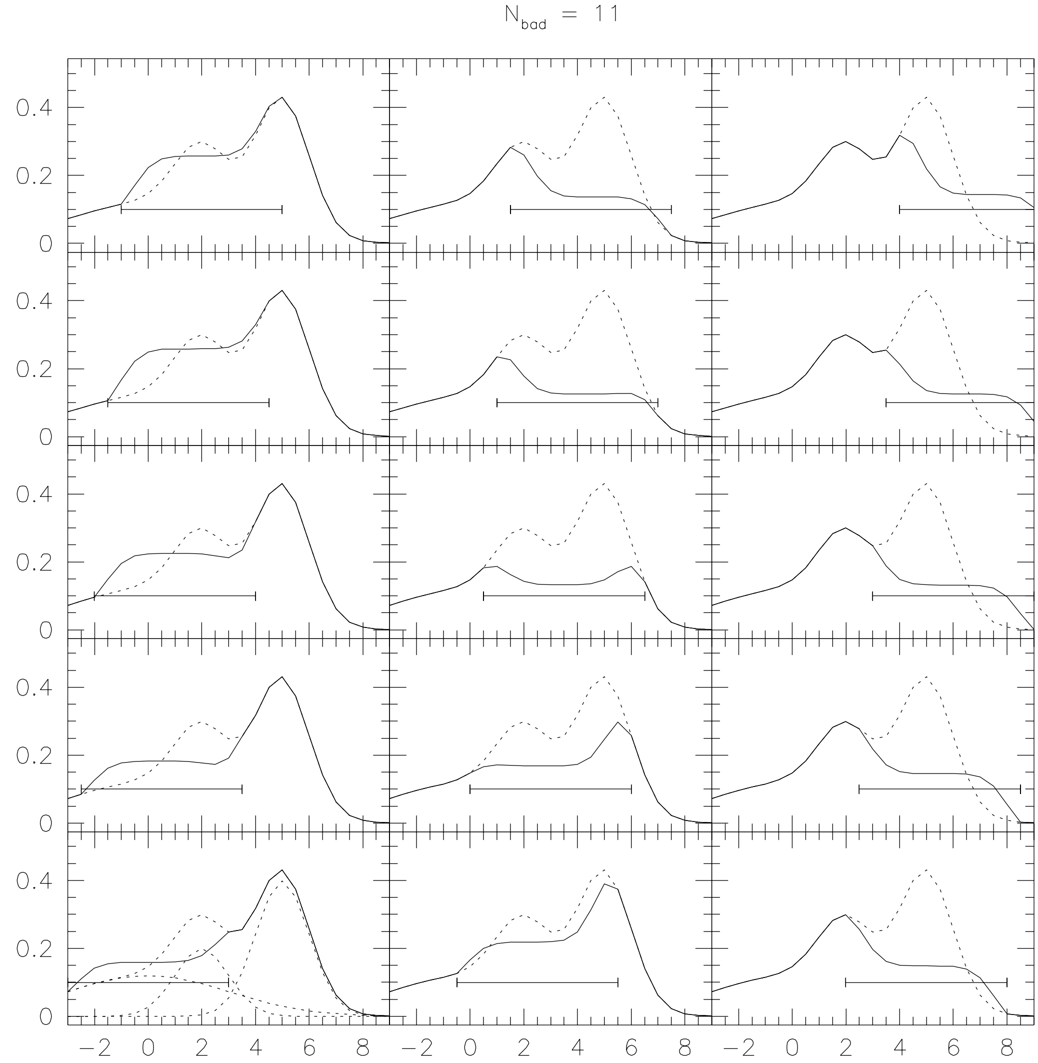

Figure 4: The effect of interpolating over eleven bad columns; see figure 1 for a description of the symbols. Note that the interpolated values settle down to the means of the two end values as fast as is permitted by the PSF.

$\sum d_i^2$	$\mathbf{d}^*$														
4.557	0.554	-1.268	1.620	$\underline{0}$	0	0	0	0	0	0	$\cdots$	0	0.082	-0.070	0.082
3.717	0.693	-0.228	1.246	0	$\underline{0}$	0	0	0	0	0	$\cdots$	0	0.253	-0.217	0.253
1.712	0.553	-0.704	0.717	0	0	$\underline{0}$	0	0	0	0	$\cdots$	0	0.380	-0.326	0.380
1.150	0.463	-0.442	0.491	0	0	0	$\underline{0}$	0	0	0	$\cdots$	0	0.427	-0.366	0.427
1.059	0.441	-0.383	0.444	0	0	0	0	$\underline{0}$	0	0	$\cdots$	0	0.437	-0.374	0.437
1.049	0.438	-0.376	0.438	0	0	0	0	0	$\underline{0}$	0	$\cdots$	0	0.438	-0.375	0.438
1.048	0.438	-0.376	0.438	0	0	0	0	0	$\underline{0}$	0	$\cdots$	0	0.438	-0.375	0.438

Table 6: Interpolation coefficients for $\alpha = 1$, $M = 3$, $S/N = \infty$ for wide stretches of bad data. The position of the interpolated value is underlined.

estimate of s^* is simply the background, and the best estimate of *that* is the mean of the available data). Table 8 gives the coefficients for interpolating over one bad column as a function of the S/N ratio.

3 References

Large-Scale Linear Methods for Interpolation, Realization, and Reconstruction of Noisy Irregularily Sampled Data William H. Press and George B. Rybicki in *Time Series Prediction: Forcasting the Future and Understanding the past* Eds. A.S. Weigend and N.A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley, 1993 (pp 493-512).

N_{bad}	$\sum d_i^2$	$\mathbf{d}^*$								
1	1.187	$*$	0.487	$\underline{0}$	0.896	-0.383				
2	0.994	$*$	0.754	$\underline{0}$	0	0.568	-0.322			
2	1.841	$*$	0.310	0	$\underline{0}$	1.213	-0.523			
3	0.793	$*$	0.853	$\underline{0}$	0	0	0.239	-0.092		
3	1.052	$*$	0.536	0	$\underline{0}$	0	0.806	-0.341		
3	2.052	$*$	0.212	0	0	$\underline{0}$	1.315	-0.527		
4	0.781	$*$	0.878	$\underline{0}$	0	0	0	0.094	0.028	
4	0.543	$*$	0.633	0	$\underline{0}$	0	0	0.378	-0.011	
4	1.040	$*$	0.394	0	0	$\underline{0}$	0	0.981	-0.292	
4	2.093	$*$	0.176	0	0	0	$\underline{0}$	1.340	-0.516	
5	0.786	$*$	0.882	$\underline{0}$	0	0	0	0	0.063	0.055
5	0.496	$*$	0.660	0	$\underline{0}$	0	0	0	0.207	0.133
5	0.405	$*$	0.447	0	0	$\underline{0}$	0	0	0.447	0.053
5	1.038	$*$	0.223	0	0	0	$\underline{0}$	0	0.907	-0.280
5	2.100	$*$	0.168	0	0	0	0	$\underline{0}$	0.135	-0.513
6	0.786	$*$	0.883	$\underline{0}$	0	0	0	$\cdots$	0.059	0.058
6	0.498	$*$	0.665	0	$\underline{0}$	0	0	$\cdots$	0.172	0.163
6	0.389	$*$	0.521	0	0	$\underline{0}$	0	$\cdots$	0.276	0.202
6	0.430	$*$	0.448	0	0	0	$\underline{0}$	$\cdots$	0.473	0.079
6	1.038	$*$	0.347	0	0	0	0	$\cdots$	0.920	-0.267
6	2.100	$*$	0.167	0	0	0	0	$\cdots$	1.345	-0.513
7	0.786	$*$	0.883	$\underline{0}$	0	0	0	$\cdots$	0.059	0.059
7	0.499	$*$	0.665	0	$\underline{0}$	0	0	$\cdots$	0.168	0.167
7	0.389	$*$	0.526	0	0	$\underline{0}$	0	$\cdots$	0.241	0.233
7	0.366	$*$	0.475	0	0	0	$\underline{0}$	$\cdots$	0.300	0.225
7	0.428	$*$	0.439	0	0	0	0	$\cdots$	0.477	0.084
7	1.038	$*$	0.346	0	0	0	0	$\cdots$	0.921	-0.266
7	2.100	$*$	0.167	0	0	0	0	$\cdots$	1.345	-0.512
8	0.786	$*$	0.883	$\underline{0}$	0	0	0	$\cdots$	0.059	0.059
8	0.499	$*$	0.665	0	$\underline{0}$	0	0	$\cdots$	0.167	0.167
8	0.389	$*$	0.526	0	0	$\underline{0}$	0	$\cdots$	0.237	0.236
8	0.366	$*$	0.480	0	0	0	$\underline{0}$	$\cdots$	0.264	0.256
8	0.363	$*$	0.466	0	0	0	0	$\cdots$	0.304	0.230
8	0.427	$*$	0.438	0	0	0	0	$\cdots$	0.478	0.084
8	$\cdots$									
9	0.786	$*$	0.883	$\underline{0}$	0	0	0	$\cdots$	0.059	0.059
9	0.499	$*$	0.665	0	$\underline{0}$	0	0	$\cdots$	0.167	0.167
9	0.389	$*$	0.527	0	0	$\underline{0}$	0	$\cdots$	0.237	0.237
9	0.366	$*$	0.480	0	0	0	$\underline{0}$	$\cdots$	0.260	0.260
9	0.362	$*$	0.471	0	0	0	0	$\cdots$	0.268	0.260
9	0.362	$*$	0.465	0	0	0	0	$\cdots$	0.304	0.230
9	$\cdots$									

Table 7: Interpolation coefficients for $\alpha = 1$, $M = 2$, $S/N = \infty$. One of the two 'good' columns to the left of the region to be interpolated over is taken to be, in fact, bad. The position of the interpolated value is underlined. For larger values of N_{bad}, the coefficients for $N_{bad} = 9$ may be used, note that the central coefficients are asymptotically $*$, 0.471, 0, $\cdots$, 0, 0.265, 0.265. end points.

S/N	$\sum d_i^2$			$\mathbf{d}^*$		
∞	1.347	-0.274	0.774	$\underline{0}$	0.774	-0.274
500	1.333	-0.270	0.770	$\underline{0}$	0.770	-0.270
100	1.279	-0.257	0.757	$\underline{0}$	0.757	-0.257
50	1.218	-0.242	0.742	$\underline{0}$	0.742	-0.242
10	0.876	-0.146	0.646	$\underline{0}$	0.646	-0.146
5	0.654	-0.068	0.568	$\underline{0}$	0.568	-0.068
2	0.410	0.050	0.450	$\underline{0}$	0.450	0.050
1	0.311	0.126	0.374	$\underline{0}$	0.374	0.126
0	0.250	0.250	0.250	$\underline{0}$	0.250	0.250

Table 8: The interpolation coefficients for a single column defect as a function of S/N.