Automatic repeated-loess decomposition of data consisting of sums of oscillatory curves

DAVID H. FOSTER

Computational Neuroscience Group, Department of Optometry and Neuroscience,
University of Manchester Institute of Science and Technology, Manchester M60 1QD, UK
d.h.foster@umist.ac.uk

Repeated loess is a nonparametric procedure that uses progressive smoothing and differencing to decompose data consisting of sums of curves. Smoothing is by locally weighted polynomial regression. Here the procedure was developed so that the decomposition into components was controlled automatically by the number of maxima in each component. The level of smoothing of each component was chosen to maximize the estimated probability of the observed number of maxima. No assumptions were made about the periodicity of components and only very weak assumptions about their shapes. The automatic procedure was applied to simulated data and to experimental data on human visual sensitivity to line orientation.

Keywords: smoothing, locally weighted polynomial regression, repeated loess, bandwidth selection, periodicity, orientation sensitivity, circular data

1. Introduction

Many biological processes give rise to sequential data that may be treated as sums of oscillatory curves, where the independent variable might be time, distance, orientation, or some other ordered quantity. Although a variety of methods of analysis are potentially available from the time-series and spectral-analysis literature, there is a particular problem in dealing with data in which the component curves have an unknown shape and periodicity, and which may vary systematically from sample to sample. An example comes from data on human and animal spatial vision, the early stages of which depend on the eye detecting lines and edges of different orientation in the stimulus scene. From psychophysical measurements, it is known that detection performance varies with line orientation in a complex way (Regan and Price 1986, Wolfe *et al.* 1992). These variations almost certainly reflect the activity in the brain of neurons or groups of neurons with different orientation sensitivities (De Valois *et al.* 1982, Celebrini *et al.* 1993), but how the responses of such neurons are combined to form an overall behavioral response is not well understood. It would be useful to be able to decompose a behavioral response into components at different orientation scales, which might eventually be related to the

orientation sensitivity of individual neurons or groups of neurons. This decomposition problem was the subject of the present work.

An analysis of sequential data of the kind just described could be made in terms of global (e.g. Fourier) or local (e.g. wavelet) basis functions, but such representations do not always allow one to readily distinguish the components underlying the process (e.g. Diggle 1990, Sections 4.10 and 4.11). Critically, each representation depends on the choice of basis function (Macaulay 1931, Farge 1992). If the form and periodicity of the underlying components were known in advance, then an appropriately matched basis could be contrived, but without this information a more general, nonparametric approach is required.

A method of analyzing sequential data that does not involve assumptions about the underlying components was proposed by Cleveland (1993). This method, called "repeated loess fitting", is a nonparametric statistical filtering procedure that decomposes a set of data into several components by a process of progressive smoothing and differencing. Smoothing is achieved by locally weighted polynomial regression (i.e. loess, Cleveland 1979, Cleveland and Devlin 1988), in which the polynomial is typically linear or quadratic and the neighborhood over which each local fit takes place is characterized by a bandwidth h. In

the repeated-loess procedure, the slowest-varying component is initially fitted with bandwidth h_1 say. This component is then subtracted from the data to form residuals. The slowest-varying component of the residuals is then fitted, with bandwidth h_2 say, where $h_2 < h_1$. This new component is then subtracted from the residuals to form new residuals, and so on. Further passes through the data may be performed in which each component is fitted again, but in a different order to assess whether the allocation of variance over the different components has been stabilized.

When the periodicities are known in advance, an elaborated version of repeated loess ("seasonal loess", Cleveland *et al.* 1990) is available in which loess fitting is applied to each cycle subseries. From knowledge of the series and with the aid of diagnostic methods, one can make estimates of the most suitable bandwidths (Cleveland *et al.* 1990, Sections 4 and 5). When the periodicities are not known in advance, it is less obvious how to proceed: in Cleveland (1993), the bandwidths for the components were obtained by experimenting with different values. Nevertheless, it might be useful to have a method that did not require user intervention. In the version of repeated loess proposed here, the decomposition is controlled automatically by a characteristic property of the components, namely, the number of maxima that each contains. The bandwidth for each component is chosen to maximize the estimated probability of the observed number of maxima.

This nonparametric procedure should be distinguished from related but more general approaches to discovering features in sequential data by automatic smoothing techniques. For example, Silverman's (1986) "bump-hunting" procedure provides a test for a unimodal or multimodal density by critically smoothing the data and using centiles of the bootstrap distribution from this critical smooth. The graphical device, SiZer, due to Chaudhuri and Marron (1999), displays the significance of features in a set of sequential data with respect to both their location and scale defined by the smoother bandwidth. Without denying the usefulness of these approaches, neither was intended to provide a solution to the decomposition problem.

The objective, then, of the present work was to formulate an automatic version of repeated loess and to illustrate its application to the analysis of data on visual sensitivity to line orientation. The description of the application is limited to this illustrative role; consequently, no attempt is made to construct an explicit statistical model of subjects' performance (this would properly involve detailed assumptions about the underlying neurophysiological processes, which would be out of place here).

The organization of this article is as follows. First, some general background to the decomposition problem is reviewed in Section 2. The automatic version of repeated loess is then set out in Section 3. The results of testing it on some simulated data are described in Section 4, together with the results of Fourier analyses for comparison. The results of applying it to experimental orientation-sensitivity data are then presented: the decompositions are described in Section 5 and some derived density estimates are described in Section 6, together with the

results of a principal component analysis and an independent component analysis for comparison. Finally, some advantages and disadvantages of automatic repeated loess are considered in Section 7.

2. The decomposition problem

Suppose that the data consist of a sequence of I observations $Y_1, \ldots, Y_I$ at fixed design points $x_1, \ldots, x_I$, where $x_1 < x_2 < \cdots < x_I$, and suppose that the data can be modeled by a function f defined on the closed interval $[x_1, x_I]$, thus

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, I, \tag{1}$$

where the error terms ϵ_i are independent random variables representing noise, with $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \sigma_i^2$. (In the present application, the assumption of independence was appropriate, and the σ_i were, in general, nonconstant). Assume that the function f consists of a sum of J component curves c_j defined over a neighborhood of the interval $[x_1, x_I]$, so that

$$f(x) = \sum_{j=1}^{J} c_j(x), \quad x_1 \leq x \leq x_I, \tag{2}$$

where each c_j is oscillatory in having at least one maximum in the open interval (x_1, x_I); each c_j is locally simple; and each c_j is coarser (has fewer oscillations) than c_k for all $k > j$. More precisely, assume that the c_j satisfy the following conditions, the rationale for which is set out afterwards.

(a) Each component c_j, where $1 \leq j \leq J$, has continuous derivatives up to at least the third order over the interval $[x_1, x_I]$ and has n_j maxima ($n_j > 0$) in the interior (x_1, x_I). (Any monotonic trend component, providing it is not too large, can be merged into the coarsest component c_1; see Section 7.1.)

(b) If at a point x_0 in the interval $[x_1, x_I]$, the first derivative $c'_j(x_0)$ of component c_j is zero, then its second derivative $c''_j(x_0)$ is non-zero. The c_j have therefore only nondegenerate critical points.

(c) At each point x_0 in the interval $[x_1, x_I]$, each component c_j may be adequately approximated by its 2nd-order Taylor polynomial within a neighborhood of x_0; that is, at each x_0 there is a number $d(x_0) > 0$ such that for all x in $[x_1, x_I]$ with $|x - x_0| < d(x_0)/2$,

$$c_j(x) \approx c_j(x_0) + c'_j(x_0)(x - x_0) + \frac{c''_j(x_0)}{2}(x - x_0)^2. \tag{3}$$

The set of all such open intervals $(x_0 - d(x_0)/2, x_0 + d(x_0)/2)$ necessarily contains a finite subcover of the closed interval $[x_1, x_I]$. Let $d_j^{\min}$ be the width of the smallest interval in this subcover, and let d_j be the least upper bound on the $d_j^{\min}$ over all such subcovers for component c_j. Over all c_j, the d_j are ordered so that $d_1 > d_2 > \cdots > d_J$.

(d) The numbers n_j of maxima in the components c_j are ordered so that $n_1 < n_2 < \cdots < n_J$.

Condition (a) is basic. Condition (b) ensures that maxima (and minima) in c_j are strict; that is, if c_j has a maximum (resp. minimum) at $x = x_0$, then $c_j(x) < c_j(x_0)$ (resp. $c_j(x) > c_j(x_0)$) for all $x \neq x_0$ in some neighborhood of x_0. Between neighboring maxima and minima, components are strictly monotonic increasing or strictly monotonic decreasing. As a consequence, arbitrarily small perturbations in c_j will not generate new maxima and minima. Condition (c) limits the sharpness of the maxima and minima in each c_j (by defining the minimum distance between neighboring points of inflexion) and therefore the maximum number of maxima; that is, n_j is less than or approximately equal to $(x_I - x_1)/(2d_j)$. But (c) does not limit the minimum number of maxima in c_j, and condition (d) sets the weakest constraint on that number. The last two conditions are important, for they help characterize components in terms of the different numbers of maxima in each. No other assumption was made about the periodicity or shape of components.

Smoothing was effected by locally weighted quadratic fitting, which produces less bias in regions of high curvature (Cleveland *et al.* 1990).

How is one particular component, c_k say, estimated from the Y_i? Suppose that all other components c_j, with $j \neq k$, are known. Let C_{-k} be the sum of the c_j with c_k omitted; that is, $C_{-k}(x) = \sum_{j=1, j \neq k}^{J} c_j(x)$, and let e_i denote the residuals $Y_i - C_{-k}(x_i)$, for $i = 1, \ldots, I$. The aim is to smooth the e_i to obtain an estimate s_k of c_k. As in Fan and Gijbels (1996), suppose that the kernel function K_h is defined by a symmetric unimodal density function with bandwidth $h = h_k$ (so $K_{ah}(x) = K_h(x/a)/a$ for $a > 0$). For each point x_0 in the interval $[x_1, x_I]$, find values of coefficients $\gamma_0, \gamma_1, \gamma_2$ that minimize

$$\sum_{i=1}^{I} (e_i - \gamma_0 - \gamma_1(x_i - x_0) - \gamma_2(x_i - x_0)^2)^2 K_h(x_i - x_0). \quad (4)$$

If $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2$ is the solution to this weighted least-squares problem, then comparison of (4) with (3) suggests that, for $\nu = 0, 1, 2$, the product $\nu! \, \hat{\gamma}_\nu$ is an estimate of the νth derivative $c_k^{(\nu)}(x_0)$. Set $s_k(x_0) = \hat{\gamma}_0$. The entire smooth s_k is obtained by solving the least-squares problem at all points x_0. (In practice, s might be calculated only at the points $x_1, \ldots, x_I$, and interpolation used elsewhere.)

It is clear that there is a close relationship between the bandwidth $h = h_k$ in (4) and the least upper bound d_j on the intervals defined in condition (c) for $j = k$. The problem, as noted by Cleveland (1993) and others, is that h_k is difficult to determine independently for each estimate s_k: if h_k is too large, then s_k will be biased and some of the variation in s_k will leak into s_{k+1}; and if h_k is too small, then s_k will be noisy and will absorb some of the variation in s_{k+1}.

As indicated earlier, in the automatic version of repeated loess described here, this problem was addressed by finding the value of the bandwidth h_k that maximized the estimated probability of the observed number n_k of maxima in s_k. This probability was estimated with a bootstrap. With the use of a criterion based on the numbers of maxima in a curve and a bootstrap to determine an optimum bandwidth, there are parallels with Silverman's (1986) bump-hunting procedure mentioned in the Introduction.

Whatever the method of decomposition, the closeness to the data Y_i of the sum of the smooths, $S = \sum_{j=1}^{J} s_j$, needs to be quantified, in order, as shown later, to decide when the decomposition procedure should be terminated. Goodness of fit can be measured in several standard ways (e.g. Hart 1997). A measure such as the residual sum of squares $\sum_{i=1}^{I} (Y_i - S(x_i))^2$ is insensitive to the nonhomogeneity of the error variance (1), which may be large (see Section 5). The residuals could be standardized to form the quantity

$$X^2 = \sum_{i=1}^{I} \frac{(Y_i - S(x_i))^2}{\hat{\sigma}_i^2}, \quad (5)$$

where $\hat{\sigma}_i^2$ is some estimate of σ_i^2, possibly derived from replications of the original measurements. If replications are not available, then pseudo-residuals (Gasser *et al.* 1986) such as $Y_i - Y_{i-1}$ might be used to estimate a common variance: $\hat{\sigma}^2 = \sum_{i=2}^{I} (Y_i - Y_{i-1})^2/(2(I-1))$. Under the assumption that the ϵ_i are normally distributed, values of X^2 in (5) may be compared with a χ^2 distribution on $I - p$ d.f., where p is the d.f. associated with the hat matrix yielding S. Caution should be exercised in assessing the nominally significant values of X^2, as the upper tail of the distribution of X^2 will tend to be inflated. The emphasis here, however, is not on accepting or rejecting the overall fit, which, in the present context, might be better decided with resampling methods, but simply in establishing a numerical stopping criterion.

3. Automatic repeated loess

There are two stages to the automatic repeated-loess procedure. In the first stage, the numbers of maxima in a set of candidate components are estimated. In the second stage, an optimum set of components with these maxima are estimated and fitted to the experimental data.

Assume for the present that the number J of components to be estimated is known or can plausibly be estimated. (The procedure may, in some applications, be applied without knowing J; see Sections 4 and 7.2.) If the numbers $n_1, \ldots, n_J$ of maxima in the J components are also known, then Stage 1 may be omitted.

3.1. *Stage 1. Estimate numbers of maxima in candidate components*

In this stage, the data $Y_1, \ldots, Y_I$ are decomposed into L candidate components with maxima $m_1, \ldots, m_L$. This set of maxima is then reduced to J values $n_1, \ldots, n_J$, which are used as inputs to Stage 2 (assume for the moment that $J \leq L$). For brevity, let #max(s) denote the number of maxima in a smooth s.

1. Oversmooth the data $Y_i, i = 1, \ldots, I$, by setting the smoother bandwidth h to a value h_0 say for which the corresponding smooth s_0 has no maxima over the interval (x_1, x_I).

2. Progressively reduce the amount of smoothing by decrementing h until a critical value, h_1^{crit} say, is reached when the number of maxima in the corresponding smooth s_1^{crit} first exceeds zero (so $h_1^{\mathrm{crit}} = \max\{h \mid \#\mathrm{max}(s) > 0\}$). Set $m_1 = \#\mathrm{max}(s_1^{\mathrm{crit}})$.

3. Subtract the smooth s_1^{crit} from the data Y_i to obtain residuals e_i.

4. Oversmooth the residuals e_i and then progressively reduce the amount of smoothing by decrementing h until a new critical value, h_2^{crit} say, is reached when the number of maxima in the corresponding smooth s_2^{crit} first exceeds m_1 (so $h_2^{\mathrm{crit}} = \max\{h \mid \#\mathrm{max}(s) > m_1\}$).

   In this way, a family of smooths is obtained over the bandwidth interval $(h_2^{\mathrm{crit}}, h_1^{\mathrm{crit}}]$ such that for each s in this family the number $\#\mathrm{max}(s)$ of maxima equals m_1 (possible lapses in this equality due to the nonmonotonicity of $\#\mathrm{max}(s)$ with h are discussed in Appendix 1). The next step is to find an optimum value h_1^{opt} of h in $(h_2^{\mathrm{crit}}, h_1^{\mathrm{crit}}]$ which, as explained earlier, is neither so large that the variation in the corresponding smooth s_1^{opt} leaks into the estimate of the next component nor so small that s_1^{opt} absorbs variation from the estimate of the next component. In the present context, where components are characterized by numbers of maxima, the loss or gain of variation may be interpreted as a tendency to lose or gain maxima: if an estimate $\hat{h}_1^{\mathrm{opt}}$ of h_1^{opt} is too large, then the corresponding smooth, $\hat{s}_1^{\mathrm{opt}}$ say, is more likely to lose maxima; if $\hat{h}_1^{\mathrm{opt}}$ is too small, then $\hat{s}_1^{\mathrm{opt}}$ is more likely to gain maxima. Accordingly, h_1^{opt} is assumed to be that value of $\hat{h}_1^{\mathrm{opt}}$ which maximizes the estimated probability of the number of maxima in $\hat{s}_1^{\mathrm{opt}}$ coinciding with the observed number m_1. This optimum value h_1^{opt} is calculated as follows.

5. Generate a bootstrap sample e_i^* from the residuals e_i, for $i = 1, \ldots, I$. Specifically, if the errors ϵ_i in (1) are each normally distributed, then sample from the normal distributions $N(e_i, \hat{\sigma}_i^2)$, where the $\hat{\sigma}_i^2$ are assumed to be derived from replications of the original measurements. Alternatively, the bootstrap sample may be generated directly from the empirical distribution of the replications, or, if replications are not available, then from pseudo-residuals. Smooth the bootstrap residuals e_i^* at the estimated bandwidth $\hat{h}_1^{\mathrm{opt}}$ to obtain a bootstrap smooth s_1^* say. Repeat the sampling and smoothing a further $B - 1$ times to obtain B bootstrap smooths $s_1^*(b)$, with $b = 1, \ldots, B$. Each $s_1^*(b)$ has say $m_1^*(b)$ maxima, which may or may not coincide with the observed value m_1. Let $P(\hat{h}_1^{\mathrm{opt}})$ be the proportion of the $s_1^*(b)$ that do have m_1 maxima (see Appendix 1); that is,

$$P\big(\hat{h}_1^{\mathrm{opt}}\big) = \#\{s_1^*(b) \mid \#\mathrm{max}(s_1^*(b)) = m_1\}/B. \quad (6)$$

Set h_1^{opt} to the smallest value of $\hat{h}_1^{\mathrm{opt}}$ in $(h_2^{\mathrm{crit}}, h_1^{\mathrm{crit}}]$ at which $P(\hat{h}_1^{\mathrm{opt}})$ is maximum, and let s_1^{opt} be the corresponding smooth.

6. Subtract this optimum smooth s_1^{opt} from the residuals e_i to obtain new residuals $e_i^{(1)}$ say.

7. Starting now with the residuals $e_i^{(1)}$, apply steps 4–6 to obtain a second optimum smooth s_2^{opt} with m_2 maxima. Repeat this process a further $L - 2$ times to obtain, in all, L optimum smooths $s_1^{\mathrm{opt}}, \ldots, s_L^{\mathrm{opt}}$ with, respectively, $m_1, \ldots, m_L$ maxima, the process terminating at a limiting bandwidth h_{L+1}^{crit}, the smallest compatible with the sampling interval defined by the design points $x_1, \ldots, x_I$. If, for any l' with $1 \leq l' < L$, the residuals are random, that is, the smooth $s_{l'}^{\mathrm{opt}}$ fits the residuals $e_i^{(l'-1)}$ according to (5) (or an equivalent measure), then set s_l^{opt} to zero for all l such that $l' < l \leq L$.

Rank the smooths $s_1^{\mathrm{opt}}, \ldots, s_L^{\mathrm{opt}}$ in order of decreasing amplitude (or probability) and take the top J smooths, that is, $s_{\beta(1)}^{\mathrm{opt}}, \ldots, s_{\beta(J)}^{\mathrm{opt}}$, where β is the corresponding permutation (recall that the number J of components to be estimated is assumed known at this point; if $L < J$, set $J = L$). The corresponding numbers $m_{\beta(1)}, \ldots, m_{\beta(J)}$ of maxima, reranked in ascending order, $n_1, \ldots, n_J$ say, are used as inputs to the next stage.

3.2. *Stage 2. Estimate components*

In this stage, the data $Y_1, \ldots, Y_I$ are decomposed into J new optimum smooths $s_1^{\mathrm{opt}}, \ldots, s_J^{\mathrm{opt}}$ with, respectively, the given numbers $n_1, \ldots, n_J$ of maxima.

1. Starting with the original data set Y_i, repeat the smoothing, differencing, and optimizing operations of Stage 1, but extract new optimum smooths $s_1^{\mathrm{opt}}, \ldots, s_J^{\mathrm{opt}}$ only where the corresponding numbers $m_1, \ldots, m_J$ of maxima coincide with the given numbers $n_1, \ldots, n_J$.

2. If the sum S of the optimum smooths $s_1^{\mathrm{opt}}, \ldots, s_J^{\mathrm{opt}}$ does not fit the data Y_i according to (5) or an equivalent measure, then reduce the smoothing associated with s_J^{opt} until it does (the larger the variance σ_i^2 of the ϵ_i, the fewer the components estimated).

3. If, for any J' with $1 \leq J' < J$, the partial sum $\sum_{j=1}^{J'} s_j^{\mathrm{opt}}$ fits the Y_i, then set the remaining curves s_j^{opt} to zero for all j such that $J' < j \leq J$.

The whole procedure, comprising Stages 1 and 2, was fully automated. It was implemented in S-Plus (MathSoft, Inc., Seattle, U.S.A), and, in a preliminary version, partially in GLIM (Numerical Algorithms Group, Oxford, U.K.).

4. Testing with simulated data

To establish that the automatic repeated-loess procedure operated correctly with well-behaved functions, it was tested on trigonometric and smoothed sawtooth functions with added noise of varying amplitude. Robust fitting or outlier suppression, part of the original loess procedure, was not incorporated. The kernel function K_h was the usual tricube.

Results of an automatic repeated-loess analysis of a set of trigonometric data are shown in the left panels of Fig. 1. The test data Y_i are shown by the circles in (a), with small vertical
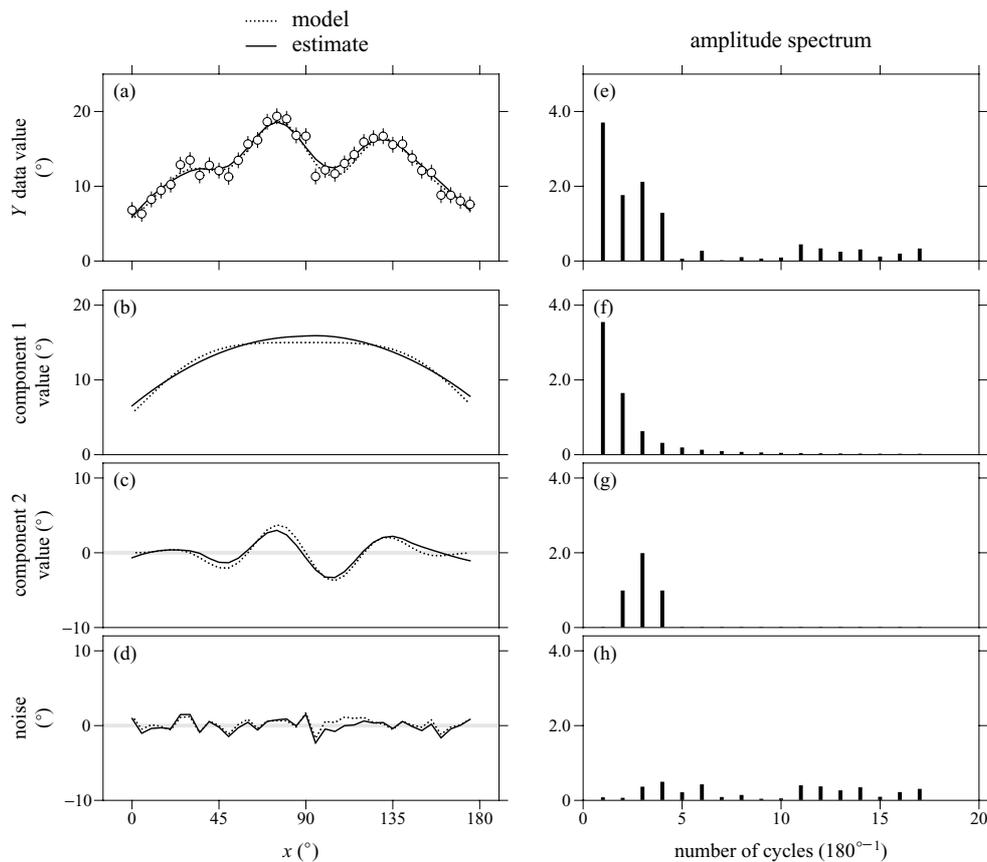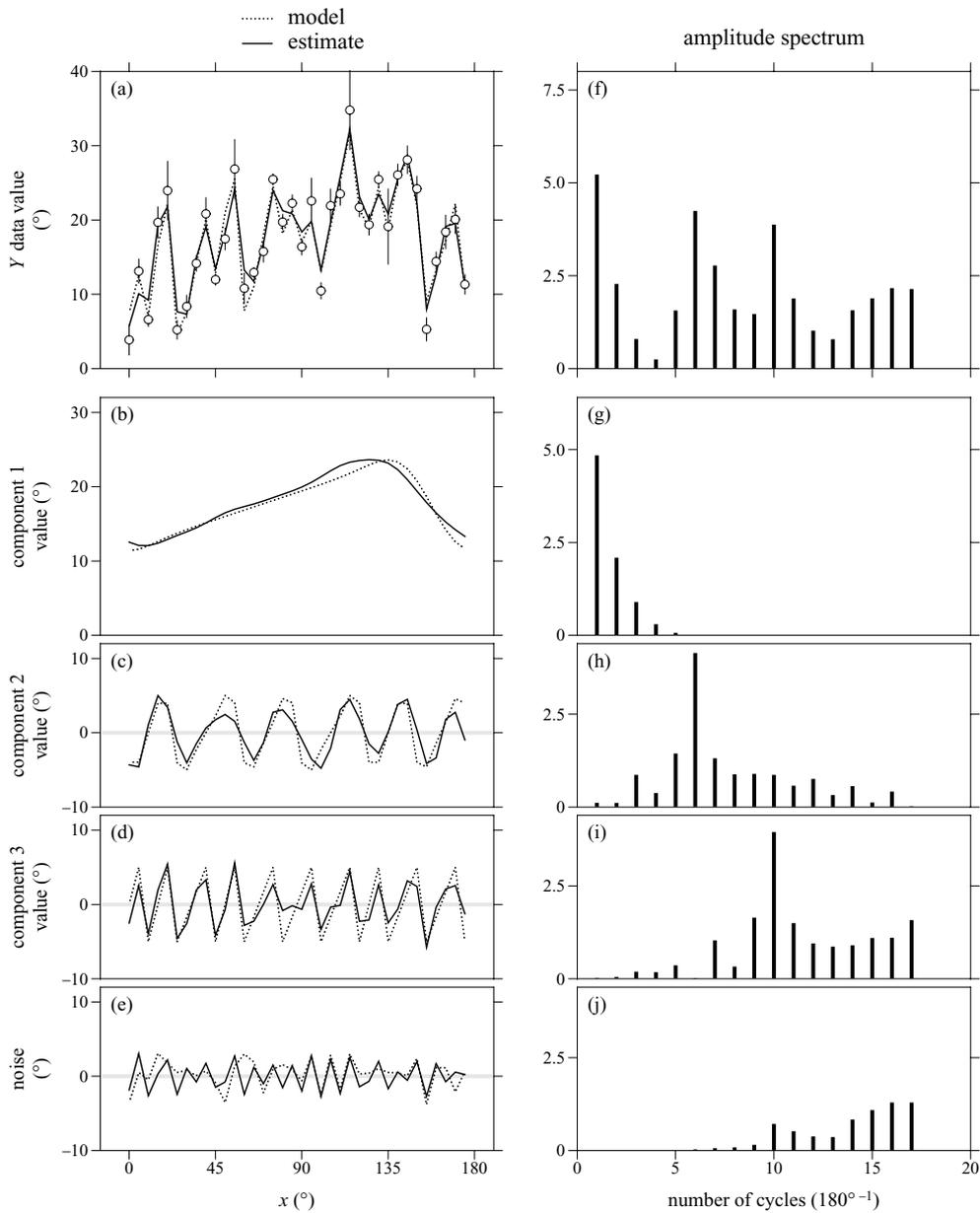
Fig. 1. *Trigonometric test data, their decomposition into components, and corresponding Fourier spectra. In the left panels, the model function, its components, and added noise are shown by dotted curves and their estimates obtained by automatic repeated loess are shown by continuous curves. The model function in (a) is the sum of an exponential function (b) and a sinecurve with a sinusoidal envelope (c). This sum, along with the model noise in (d), formed the test data (open symbols in (a)), with the vertical bars indicating ±1 estimated SE. In the right panels, the vertical bars in (e)–(h) show, respectively, the Fourier amplitude spectra of the test data in (a), the model components in (b) and (c), and the residuals in (d). In each spectrum, the amplitude at 0 cycles was set to zero*

bars indicating ±1 estimated SE, in general calculated from the data but here assumed known and given identically by a model value σ, defined shortly. The model function and its constituent curves (2) in the panels below are shown by the dotted lines. The model function (a) consisted of the sum of an exponential function (b) and a sinecurve with a sinusoidal envelope (c). The test data were formed from the model function (see (1)) by adding model noise (d) drawn from the normal distributions $N(0, \sigma_i^2)$ with $\sigma_i = \sigma = 1.0$. For this first demonstration, although the x_i were drawn from the range $0°, 5°, \ldots, 175°$, it was assumed that the design was not axial; that is, $x_{36} = 175°$ was not adjacent to $x_1 = 0°$ (this decision only affects smoothing over 2–3 points at the ends of the range). The automatic repeated-loess estimates of the components (b) and (c) are shown by the continuous curves and were derived with the number J of components assumed to be 2 (without this constraint, a spurious component appeared with 2 maxima and amplitude 3% of the next largest component (c)). The continuous curve in (a) is the sum of the estimated components in (b) and (c), and the continuous curve in (d) shows the residuals; that is, the differ-

ence between the sum of the estimated components in (a) and the test data. Overall, the components were well estimated by the procedure, although there was a slight oversmoothing of the data.

Results of a Fourier analysis of both this set of trigonometric data and its decomposition are shown in the right panels of Fig. 1. Phase spectra have been omitted. The vertical bars in (e)–(h) indicate, respectively, the amplitude spectra of the test data in (a) and of the model components in (b) and (c) (dotted curves), and of the residuals of the estimates in (d) (continuous curve). The spectrum of the model noise in (d) (dotted curve) is, on average, constant, and is not shown. A priori, it is not obvious how the data spectrum in (e) would be partitioned to obtain the component spectra in (f) and (g), which greatly overlap.

Results of an automatic repeated-loess analysis of a set of smoothed sawtooth data are shown in the left panels of Fig. 2. As in Fig. 1, the test data Y_i are shown by the circles in (a), with small vertical bars indicating ±1 estimated SE given by model values σ_i. The model function and its constituent curves in the panels below are shown by the dotted lines. The model function

Fig. 2. *Sawtooth test data, their decomposition into components, and corresponding Fourier spectra. In the left panels, the model function, its components, and added noise are shown by dotted curves and their estimates obtained by automatic repeated loess are shown by continuous curves. The model function in (a) is the sum of a smoothed ramp function (b) and two smoothed sawtooth functions with varying periods (c) and (d). This sum, along with the model noise in (e), formed the test data (open symbols in (a)), with the vertical bars indicating ± 1 estimated SE. In the right panels, the vertical bars in (f)–(j) show, respectively, the Fourier amplitude spectra of the test data in (a), the model components in (b)–(d), and the residuals in (e). In each spectrum, the amplitude at 0 cycles was set to zero*

(a) consisted of the sum of a smoothed ramp function (b) and two smoothed sawtooth functions with varying periods (c) and (d) (because of the size of the sampling interval, the smoothing in the last is not apparent). The test data were formed from the model function by adding model noise (e) drawn from the normal distributions $N(0, \sigma_i^2)$ in which, to reflect their nonconstancy, the σ_i^2 were sampled from one of the sets of estimates reported in Section 5. For this demonstration, it was assumed that the design was axial; the data to be smoothed were extended by the

periodicity relation $Y_{i+36q} = Y_i$, where q is integer; smoothing was then applied to the extended set; and the smooth was then restricted to the original domain. Smoothing was not weighted locally by the σ_i^2 (see Section 5). The automatic repeated-loess estimates of the components (b)–(d) are shown by the continuous curves and were derived with the number J of components assumed to be 3 (without this constraint, a spurious component appeared with 5 maxima and amplitude 25% of the next largest component (c)). The continuous curve in (a) is the sum of the

estimated components in (b)–(d), and the continuous curve in (e) shows the residuals.

Given the increased level of noise in this data set and the closeness of the mean periodicity of the third component (d) to the sampling interval defined by the design points x_i, the components were reasonably well estimated by the procedure. The estimates might have been improved further if the kernel function, the tricube, did not have bounded support (so that the fit did not become degenerate at small bandwidths).

Results of a Fourier analysis of both this set of smoothed sawtooth data and its decomposition are shown in the right panels of Fig. 2. The vertical bars in (f)–(j) indicate, respectively, the amplitude spectra of the test data in (a) and of the model components in (b)–(d) (dotted curves), and of the residuals of the estimates in (e) (continuous curve). Notice that, unlike the residuals spectrum in Fig. 1(h), the spectrum in (j) is attenuated at low-to-medium frequencies. The component spectra in (g), (h), and (i) overlap moderately.

5. Application to psychophysical data

An automatic repeated-loess analysis was applied to data from a psychophysical experiment (Foster and Westland 1998) on human visual sensitivity to differences in line orientation. Each subject was presented briefly with a stimulus consisting of an array of identical lines, all with the same orientation except for one, a "target" line, which appeared within the array with probability 0.5. In each such trial, the subject had to indicate whether the target was present. The orientations of the target and the other "background" lines in the array varied randomly from trial to trial. Ten subjects participated, each performing, on average, 13,000 trials over a period of several months. Experimental details are given in Foster and Westland (1998). From these data, a curve was derived for each subject showing the smallest detectable difference in orientation of target and background lines as a function of the orientation of the background lines, sometimes referred to as an orientation increment-threshold function. The derivation is summarized in Appendix 2. As suggested in the Introduction, these increment-threshold functions appear to represent the sum of several oscillatory components of unknown form and periodicity. The task was to decompose the functions obtained from individual subjects into orientation components that might, in turn, be related to the activity of orientation-sensitive mechanisms within the human visual system.

The curves in Fig. 3 show, superposed, the orientation increment-threshold functions for the 10 subjects. There is a rapid oscillation of threshold orientation with background orientation, which varies markedly from subject to subject, along with a more general underlying slow variation with a period of about 90°.

Results of an automatic repeated-loess analysis of the data from an individual subject are shown in the left panels of Fig. 4. The increment-threshold function is shown by the circles in (a), with small vertical bars indicating ±1 estimated SE, obtained
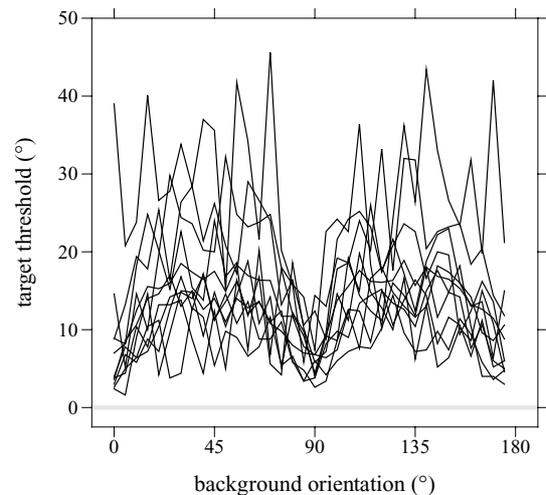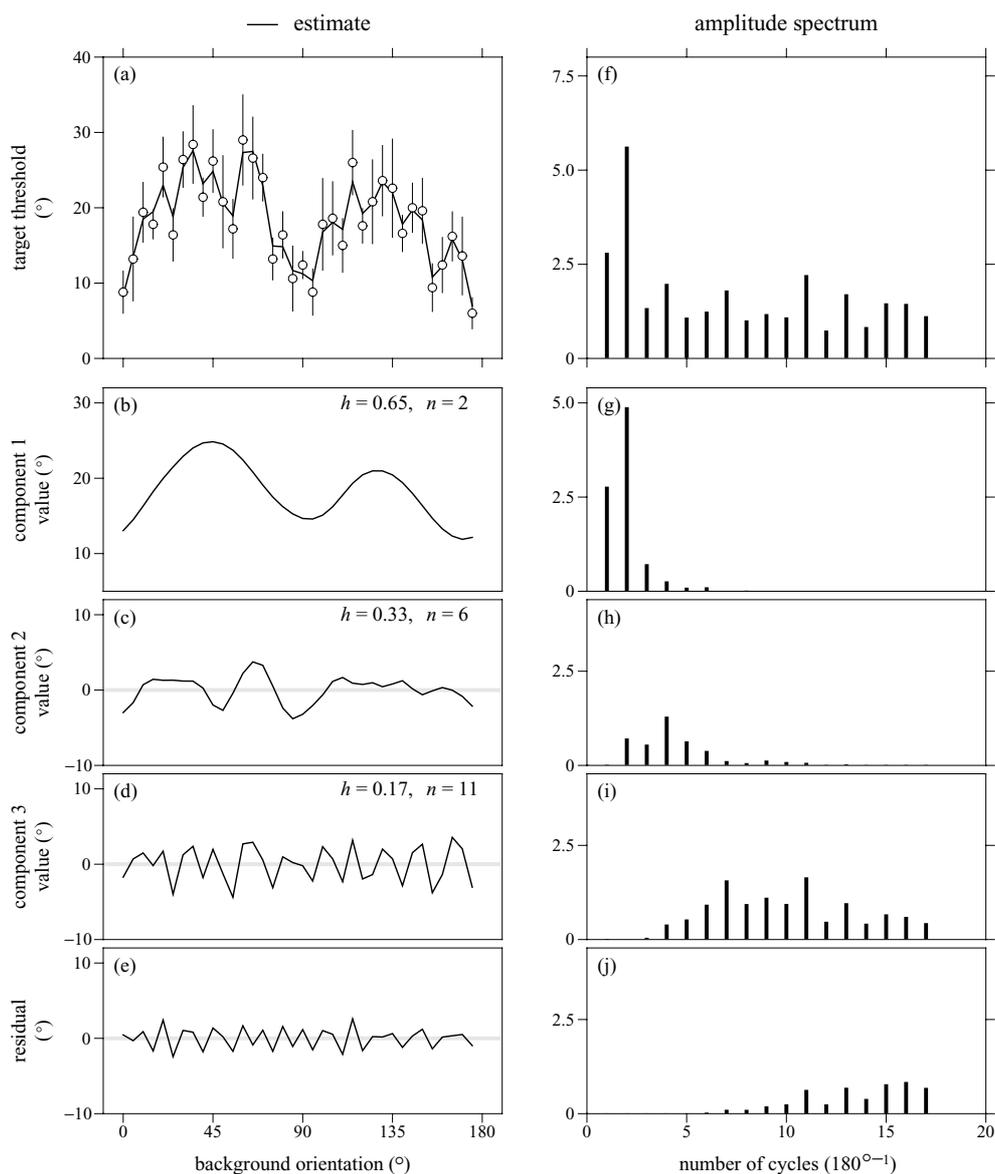

Fig. 3. *Orientation increment-threshold functions for 10 subjects. The continuous curves show the smallest detectable difference in orientation of target and background lines as a function of the orientation of the background lines. (An orientation of 0° corresponds to the vertical in the frontoparallel plane)*

from the experimental data by a bootstrap. The variation in the estimated SEs is evident (and more so in a second example to follow). They were mildly correlated with the threshold values: averaged over all subjects, the sample correlation coefficient was about 0.4. The repeated-loess procedure was applied as in the second example in Section 4, for axial data. Although the estimated SEs entered the goodness-of-fit measure (5), they were not used to weight the smoothing locally, as this would then have biased the fits downwards. (Apart from the local scaling effect of the threshold value, the most important other cause of the nonhomogeneous variance was probably variation in the shape of the local curve relating target-detection performance to orientation difference between target and background lines at each background-line orientation; see Appendix 2.) The number J of components to be estimated was fixed so that their statistics could more easily be summarized over subjects. On the basis of exploratory analyses, J was set to 3; the effect of allowing J to vary is considered in Sections 6.1 and 7.2.

The components estimated from the data in Fig. 4(a) are shown by the continuous curves in (b)–(d). The continuous curve in (a) is the sum of these orientation components. The residuals are shown in (e).

In the light of the results with the test data in Section 4, a Fourier analysis was not expected to be especially revealing. Nevertheless, for completeness, results are shown in the right panels of Fig. 4. The vertical bars in (f)–(i) indicate, respectively, the amplitude spectra obtained from the (smoothed) increment-threshold function in (a) (continuous curve) and from the estimated components in (b)–(d). There is no obvious partition of the spectrum in (f) that would yield the spectra in (g)–(i), which, as with the test data, overlap considerably. The spectrum in (j) of the residuals contained high-frequency components only.
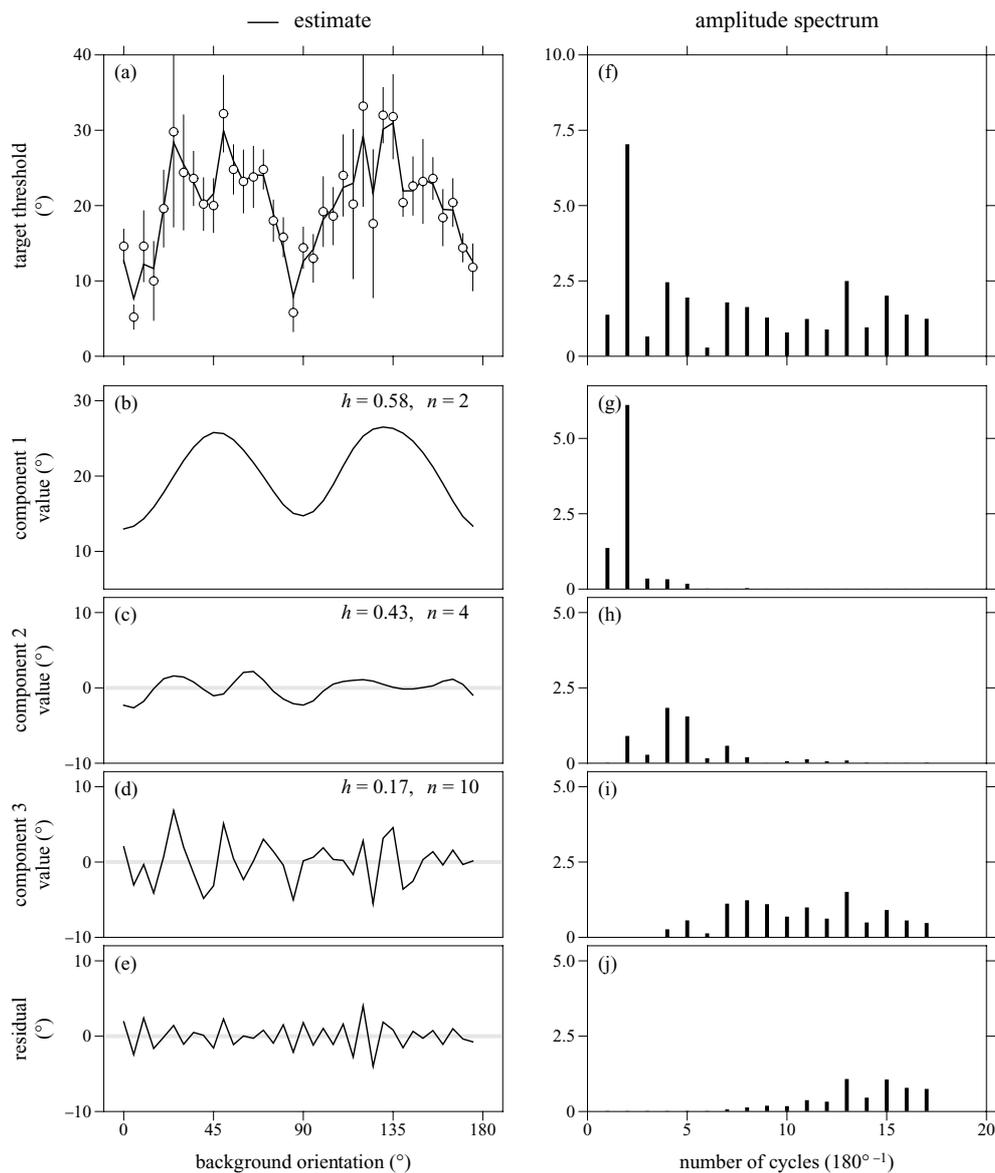
Fig. 4. *Orientation increment-threshold function for subject S.W., its decomposition into orientation components, and corresponding Fourier spectra. In the left panels, the increment-threshold function is shown by the circles in (a), with small vertical bars indicating ±1 estimated SE, and the first, second, and third components estimated by automatic repeated loess are shown by the continuous curves in (b)–(d). The residuals are shown in (e). The continuous curve in (a) is the sum of the orientation components in (b)–(d). Values of the bandwidth h and number n of maxima associated with each component are indicated. In the right panels, the vertical bars in (f)–(j) show, respectively, the Fourier amplitude spectra of the test data in (a), the estimated components in (b)–(d), and the residuals in (e). In each spectrum, the amplitude at 0 cycles was set to zero*

Results of an automatic repeated-loess analysis and corresponding Fourier analysis of data from a second subject are shown, respectively, in the left and right panels of Fig. 5. Although a partition of the amplitude spectrum in (f) might be based on the minima at frequencies of 3 and 6 cycles per 180°, the overlap of the spectra in (g) and (h) and in (h) and (i) would be lost.

In both Figs. 4 and 5, the mean periodicity of the third estimated component in (d) is close to the sampling interval, which may contribute to its appearance of randomness (as with the estimated component in (d) for the smoothed sawtooth data, Fig. 2), but its spectrum in (i) is different from the residuals spectrum in (j). Further discussion of the non-randomness and reproducibility of these fine components may be found in Foster and Westland (1998, Appendix B, available at http://www.pubs.royalsoc.ac.uk).

There are clear similarities both in the first and in the second estimated components for the two subjects, and the results of automatic repeated-loess analyses of data from the remaining eight subjects were also similar. Averaged over all ten subjects, the mean periodicity of the first estimated component was 89°, of the second one 41°, and of the third 19°. In the next section,

Fig. 5. *Orientation increment-threshold function for subject A.C., its decomposition into orientation components, and corresponding Fourier spectra. Other details as for Fig. 4*

some distributional properties of the components are considered in relation to possible underlying mechanisms.
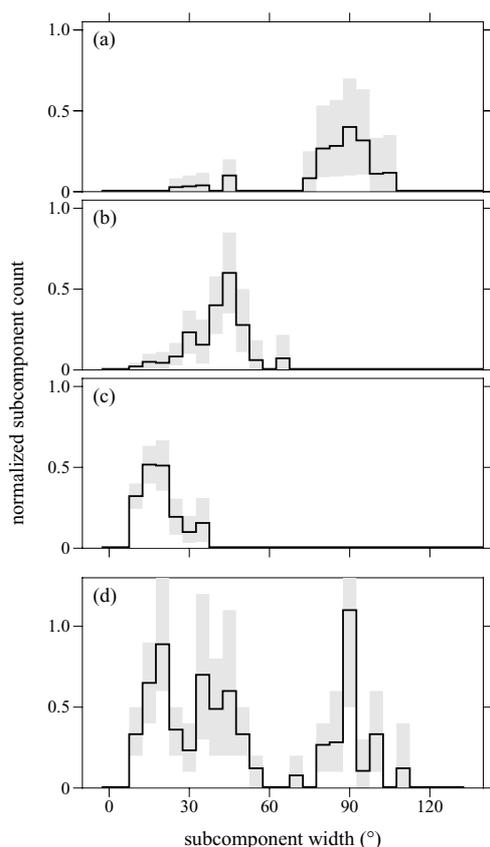
6. Statistics of components

6.1. *Subcomponent analysis*

This application of automatic repeated loess may be taken a little further while maintaining the generality of a nonparametric approach. For descriptive purposes, then, the results of the automatic repeated-loess analyses for the 10 subjects were summarized in terms of the estimated densities of the subcomponents making up each component: subcomponents were defined by segmenting a component at its minima, a

process that yielded a set of individual bump-shaped curves. Each of these subcomponents may be thought of as the result of an interaction between signals from individual orientation-sensitive visual mechanisms with adjacent preferred orientations (Foster and Ward 1991). The angular width of each subcomponent would then correspond to the angle between the preferred orientations of these mechanisms, that is, their orientation spacing. For each component (first, second, third) a histogram was constructed by counting at each angular width the number of subcomponents with that width, over all subjects.

For a uniform distribution of subcomponent widths, the number of counts for a small subcomponent width is greater than for a large subcomponent width; so, to make departures from

each subject, and the same three modes emerged in a histogram of subcomponent widths derived from the set of all components, replotted here in Fig. 6(d). The total number of counts in (d) is a little higher than the sum of the totals in (a)–(c), by the mode at 90° in (d) is sharper than in (a).

6.2. *Principal component analysis and independent component analysis*

An examination (Foster and Westland 1998) of the relative phases of the subcomponents suggested that coarse and intermediate components were relatively stable over subjects, whereas fine subcomponents varied from subject to subject. A principal component analysis (PCA) of the increment-threshold functions over subjects should therefore identify only the fine orientation components of the automatic repeated-loess analysis. As in other applications (Grambsch *et al.* 1995, see also Ramsay and Dalzell 1991, Rice and Silverman 1991), the PCA was applied to the smoothed increment-threshold functions from each of the 10 subjects (for the two subjects in Figs. 4 and 5, the continuous curves in (a)). The four leading principal components, accounting for 94% of the variance in the data, were each found to oscillate rapidly, with periods 14°–23°. As anticipated, this range contained the main mode at 15°–20° characterizing the fine orientation subcomponents of the automatic repeated-loess analysis (Fig. 6(c)). PCA was not informative about the coarse and intermediate components of the automatic repeated-loess analysis.

An independent component analysis (ICA) was also undertaken (Bell and Sejnowski 1995, Comon 1994). ICA is similar to PCA, but is designed to obtain components that are maximally independent, not merely decorrelated. It was, however, no more informative than PCA.

Fig. 6. *Histograms of subcomponent widths. Mean normalized subcomponent counts at each subcomponent width for (a) the first, (b) the second, and (c) the third components estimated in automatic repeated-loess decompositions of all ten subjects' data (illustrated for two subjects in Figs. 4 and 5). The histogram in (d) is replotted from an approximate analysis (Foster and Westland 1998, Fig. 4a). The shaded regions mark 5–95% pointwise confidence intervals. The data were normalized to reveal departures from uniformity over subcomponent widths*

6.3. *Comparison with other experimental data*

At present, only broad comparisons can be made between the properties of the estimated subcomponent densities of Fig. 6 and data from the primate neurophysiological literature, primarily because of the different ways in which responses are measured. Thus, electrical recordings are normally made from single cells rather than from groups of cells; the recorded orientation characteristics of individual cells depend on which layers of the cortex are sampled (e.g. De Valois *et al.* 1982, Henry *et al.* 1994) and on the size of the stimuli (Schiller *et al.* 1976), and these characteristics may vary with time after the onset of the stimulus (Ringach *et al.* 1997). Even so, population data from some single-cell recordings show a 90° anisotropy, with more cells tuned to vertical and horizontal than to oblique directions (De Valois *et al.* 1982). The orientation-tuning bandwidths of cells, which might be related to subcomponent widths, is large: full-width half-height values range from about 8° to more than 100° (De Valois *et al.* 1982, see also Celebrini *et al.* 1993). And, albeit less reliably, the estimated

uniformity clearer, the recorded number of counts at each spacing, T say, was multiplied by $T/90°$.

Figure 6 shows this normalized subcomponent count, averaged over the 10 subjects, plotted against subcomponent width for (a) the first, (b) the second, and (c) the third component. The shaded regions mark 5–95% pointwise confidence intervals, estimated by the bootstrap percentile method.

Despite the variations present in the individual decompositions in Figs. 4 and 5, the histograms of the subcomponent widths showed three well-defined modes. For the first component, the mode is at 90° (Fig. 6(a)); for the second component, it is at 45° (Fig. 6(b)) and for the third component, it is at 15°–20° (Fig. 6(c)).

If the number J of components to be estimated had not been fixed, it would have been difficult to pool the partitioned subcomponent counts over different subjects. But it seems unlikely that these three estimated densities are an artifact of setting J to 3. In an earlier approximate analysis (Foster and Westland 1998; see Section 7.2 here), J was given a maximum value of 5 for

densities of orientation bandwidths have modes variously at about 20°, 30°–40°, and 90°–180°, depending on retinal-input location and cell type (De Valois, Yund and Hepler 1982, see also Vogels and Orban 1990).

Comparisons with other psychophysical data are more straightforward, and have been reviewed elsewhere (Foster and Ward 1991, Baddeley and Hancock 1991, Foster and Westland 1998). In brief, evidence is available for 90° periodicities (Marendaz 1998, Regan and Price 1986, see also the general literature on the "oblique effect", Appelle 1972, Gentaz and Ballaz 2000), for 45° periodicities (Regan and Price 1986, Stivalet *et al.* 1995), and for the possibility of smaller periodicities still (Wolfe *et al.* 1992, Regan and Price 1986).

7. Discussion

7.1. *Advantages and disadvantages*

Automatic repeated loess has several advantages as a procedure for decomposing data consisting of sums of oscillatory curves.

1. Based on repeated loess, it is nonparametric. No assumptions need be made about the periodicity of components and only very weak assumptions about their shapes. It can incorporate robust fitting; deal routinely with missing values; and provide an immediate graphical interpretation.
2. It retains the particular advantages of locally weighted quadratic fitting in minimizing bias.
3. It is automatic, and does not require user intervention to determine bandwidths.

The success of the procedure depends on what is meant by a component, but the interpretation adopted here, based on the number of maxima in a curve and the range over which the curve can be approximated adequately by a 2nd-order Taylor polynomial, seems sufficiently general to capture many biological (and physical) processes.

The disadvantages of the procedure stem mainly from its nonparametric foundations.

1. It may fail to decompose satisfactorily sums of curves with large-amplitude noise that produces significant additional maxima or hides existing maxima. It may also fail to detect small-amplitude components in the presence of a large-amplitude trend. For example, if the model function f in (2) is defined by the sum $f(x) = a_1 x + a_2 \sin \omega x$, with $x_1 \leq x \leq x_I$, and the constants a_1, a_2, ω are such that $a_1 > a_2 \omega$, then f has no maxima in (x_1, x_I).
2. Subcomponents may be split over adjacent components, for example, where maxima coincide. As a test for splitting, a histogram of the number of subcomponents at each subcomponent width may be compared with a plot of the summed subcomponent amplitudes at each subcomponent width. (By this measure, there was little subcomponent splitting in the psychophysical data; see Foster and Westland 1998.)

3. Occasional spurious interaction components may be generated by closely similar components. With the test data, it was found that the amplitudes of the interaction components were smaller than those of the true components (see Section 4).

Unlike some other decomposition procedures such as independent component analysis, there is no constraint that components estimated by automatic repeated loess should be independent. This requirement could, if desired, be introduced during the optimization of the smoother bandwidths in Stage 1 and Stage 2 of the procedure (Section 3).

7.2. *Efficiency considerations*

A large part of the computational burden of the procedure came from the optimizing smoother bandwidths. In the optimizations, the value of the smoother bandwidth h producing a constant number of maxima in an estimated component was decremented in multiplicative steps of 0.95, and, at each of these values, B bootstrap replications were performed. A value of $B = 50$ was found to be just acceptable. To reduce the time spent on optimization, the size of the decrement in h might have been increased, but it would then have been necessary to make B larger. Alternatively, the accuracy of the estimated optimum bandwidth could be improved by smoothing the estimated probability $P(h)$ in (6) with respect to h.

For any particular application, it is possible to reduce computation time by using model data to guide the choice of bandwidths. This approach was used in an approximate analysis referred to earlier (Foster and Westland 1998). Thus, for a given bandwidth interval $(h_{l+1}^{\mathrm{crit}}, h_l^{\mathrm{crit}}]$ producing a constant number of maxima in an estimated component, an approximately optimum bandwidth h_l was selected by multiplying h_{l+1}^{crit} by a factor κ, where κ was the largest value between 1.0 and 1.5 for which $h_l \leq h_l^{\mathrm{crit}}$. The limit of 1.5 was found by experimenting with several sets of model data thought likely to be similar in structure to the experimental data. The number J of components to be estimated was not fixed, but given a maximum value of 5. This reduced procedure produced decompositions broadly similar to those obtained here, as already noted in the analysis of the subcomponent distributions (Fig. 6). Over the 10 subjects, the number of components estimated varied from 2 to 4. Although in its application this procedure was still controlled by the number of maxima obtained in each component, it was not fully optimal, and, in requiring preliminary user intervention, it was closer in spirit to Cleveland's original repeated-loess procedure.

Appendix 1: Monotonicity of numbers of maxima

Silverman's (1981) proof of the monotonic decrease in the number of modes with increasing bandwidth of the smoother depends on the use of a normal density function as a kernel smoother (see Chaudhuri and Marron 2000, for discussion). With the

locally weighted quadratic smoother used here, this monotonicity property held only approximately. By construction of the bandwidth intervals $(h_{l+1}^{\mathrm{crit}}, h_l^{\mathrm{crit}}]$, it follows that for each s in the corresponding family of smooths $\{s \mid h_{l+1}^{\mathrm{crit}} < h \leq h_l^{\mathrm{crit}}\}$ the number of maxima #max(s) does not exceed #max(s_l^{crit}), where s_l^{crit} is the smooth corresponding to h_l^{crit}. For the present data, equality was found to hold over 93–94% of the bandwidth range, with #max(s) equal to #max$(s_l^{\mathrm{crit}}) - 1$ over the remainder. As the estimated probability $P(h)$ at h in (6) was calculated only for smooths in which the number of maxima was equal to #max(s_l^{crit}), lapses were automatically excluded.

Appendix 2: Derivation of increment-threshold functions

In the psychophysical experiment, the orientation θ of the background lines varied over the range $0°, 5°, \ldots, 175°$, and the difference $\delta\theta$ in orientation between target and background lines varied over the range $5°, 10°, \ldots, 40°$. Target detection at each combination of θ and $\delta\theta$ was summarized by the discrimination index, denoted by d', from signal-detection theory (Green and Swets 1966). Thus, if HR is the detection hit rate, FAR the false-alarm rate, and Φ^{-1} the inverse of the (cumulative) unit normal distribution, then $d' = \Phi^{-1}(\mathrm{HR}) - \Phi^{-1}(\mathrm{FAR})$. Hence, d' linearizes and combines responses to trials with and without targets.

For each θ, a graph of d' against $\delta\theta$ was obtained, from which a threshold value $\Delta\theta$ of $\delta\theta$ was derived, in the following way. As some graphs had both concave and convex sections, a cubic curve g was fitted, which accounted adequately for the variance in the data (maximum $\chi^2 = 196$, d.f. $= 180$). Fitting was by weighted least-squares, constrained so that $g(0) = 0$. For a selected criterion level d_0' of d' (e.g. $d_0' = 0.5$), the corresponding threshold value $\Delta\theta$ of $\delta\theta$ was calculated as the (positive) value of $\delta\theta$ nearest zero such that $g(\delta\theta) = d_0'$ (such a value could always be found). The curve of $\Delta\theta$ against θ defined the increment-threshold function. At each value of θ the standard deviation of $\Delta\theta$ was estimated with a bootstrap. The choice of criterion level d_0' was constrained by two requirements: that the $\Delta\theta$ should be stable and that the shape of the curve of $\Delta\theta$ against θ should be invariant under modest changes in d_0'. For values of d_0' greater than about 0.8 one or other of these requirements was not satisfied (Foster and Ward 1991). Further details of methodology are given in Foster and Westland (1998).

Acknowledgments

References

Appelle S. 1972. Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. Psychological Bulletin 78: 266–278.

Baddeley R.J. and Hancock P.J.B. 1991. A statistical analysis of natural images matches psychophysically derived orientation tuning curves. Proceedings of the Royal Society of London, Series B 246: 219–223.

Bell A.J. and Sejnowski T.J. 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7: 1129–1159.

Celebrini S., Thorpe S., Trotter Y., and Imbert M. 1993. Dynamics of orientation coding in area V1 of the awake primate. Visual Neuroscience 10: 811–825.

Chaudhuri P. and Marron J.S. 1999. SiZer for exploration of structures in curves. Journal of the American Statistical Association 94: 807–823.

Chaudhuri P. and Marron J.S. 2000. Scale space view of curve estimation. Annals of Statistics 28: 408–428.

Cleveland R.B., Cleveland W.S., McRae J.E., and Terpenning I. 1990. STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6: 3–73.

Cleveland W.S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74: 829–836.

Cleveland W.S. 1993. Visualizing Data. Hobart Press, Summit, N.J.

Cleveland W.S. and Devlin S.J. 1988. Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association 83: 596–610.

Comon P. 1994. Independent component analysis, A new concept? Signal Processing 36: 287–314.

De Valois R.L., Yund E.W., and Hepler N. 1982. The orientation and direction selectivity of cells in macaque visual cortex. Vision Research 22: 531–544.

Diggle P.J. 1990. Time Series. A Biostatistical Introduction. Clarendon Press, Oxford.

Fan J. and Gijbels I. 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

Farge M. 1992. Wavelet transforms and their applications to turbulence. Annual Review of Fluid Mechanics 24: 395–457.

Foster D.H. and Ward P.A. 1991. Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. Proceedings of the Royal Society of London, Series B 243: 75–81.

Foster D.H. and Westland S. 1998. Multiple groups of orientation-selective visual mechanisms underlying rapid orientated-line detection. Proceedings of the Royal Society of London, Series B 265: 1605–1613.

Gasser T., Sroka L., and Jennen-Steinmetz C. 1986. Residual variance and residual pattern in nonlinear regression. Biometrika 73: 625–633.

Gentaz É. and Ballaz C. 2000. La perception visuelle des orientations et "l'effet de l'oblique." Année Psychologique 100: 715–744.

Grambsch P.M., Randall B.L., Bostick R.M., Potter J.D., and Louis T.A. 1995. Modeling the labeling index distribution: An application

of functional data analysis. Journal of the American Statistical Association 90: 813–821.

Green D.M. and Swets J.A. 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Hart J.D. 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer, New York.

Henry G.H., Michalski A., Wimborne B.M., and McCart R.J. 1994. The nature and origin of orientation specificity in neurons of the visual pathways. Progress in Neurobiology 43: 381–437.

Macaulay F.R. 1931. The Smoothing of Time Series. National Bureau of Economic Research Inc, New York.

Marendaz C. 1998. Nature and dynamics of reference frames in visual search for orientation: Implications for early visual processing. Psychological Science 9: 27–32.

Ramsay J.O. and Dalzell C.J. 1991. Some tools for functional data analysis. Journal of the Royal Statistical Society, B 53: 539–572.

Regan D. and Price P. 1986. Periodicity in orientation discrimination and the unconfounding of visual information. Vision Research 26: 1299–1302.

Rice J.A. and Silverman B.W. 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society, B 53: 233–243.

Ringach D.L., Hawken M.J., and Shapley R. 1997. Dynamics of orientation tuning in macaque primary visual cortex. Nature 387: 281–284.

Schiller P.H., Finlay B.L., and Volman S.F. 1976. Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. Journal of Neurophysiology 39: 1320–1333.

Silverman B.W. 1981. Using kernel density estimates to investigate multimodality. Journal of the Royal Statistical Society Series B—Statistical Methodology 43: 97–99.

Silverman B.W. 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London

Stivalet P., Marendaz C., Barraud P.-A., and Raphel C. 1995. Is the 45° oblique a third reference axis for the coding of orientation in preattentive vision? Perception 24 (supplement): 140.

Vogels R. and Orban G.A. 1990. How well do response changes of striate neurons signal differences in orientation: A study in the discriminating monkey. Journal of Neuroscience 10: 3543–3558.

Wolfe J.M., Friedman-Hill S.R., Stewart M.I., and O'Connell K.M. 1992. The role of categorization in visual search for orientation. Journal of Experimental Psychology—Human Perception and Performance 18: 34–49.