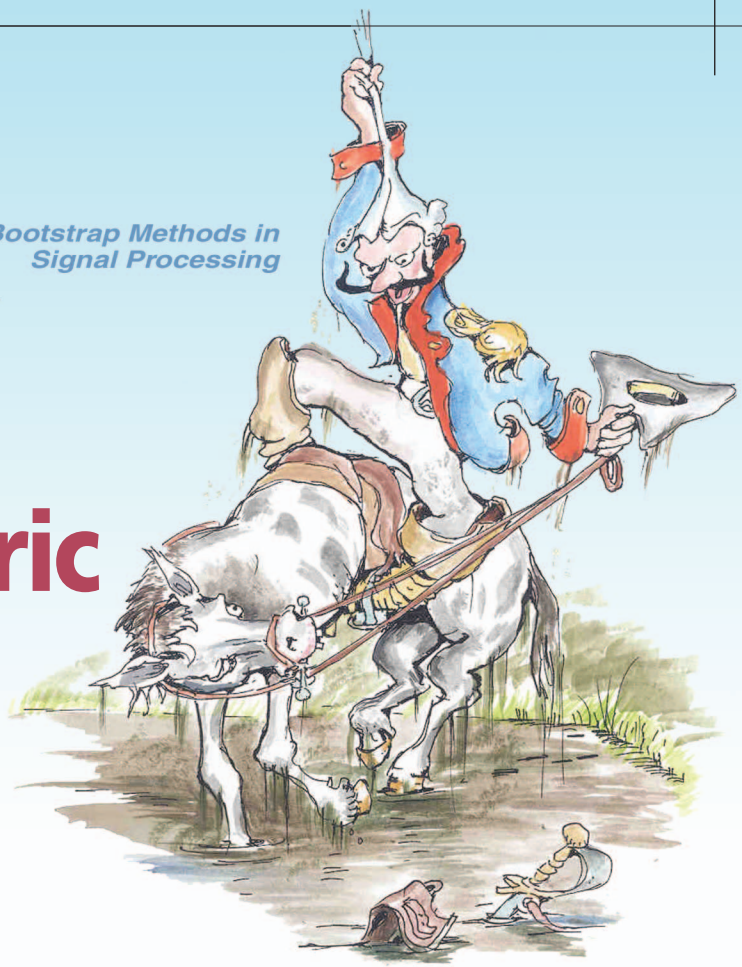
David H. Foster and Kamila Żychaluk

Bootstrap Methods in Signal Processing



PROF. DR. KARL HEINRICH HOFMANN

Nonparametric Estimates of Biological Transducer Functions

Using bootstrap local fitting to overcome parametric regression problems

A common task in applying signal-processing methods to biological systems is estimating a transducer function. The particular system being analyzed may range from the very small, such as a retinal photoreceptor producing a voltage response on being stimulated with a flash of light, to the large and complex, such as a human patient pressing a switch on hearing a test tone through headphones. Achieving a good estimate of the transducer function from a set of data may be an important first step in understanding the underlying biological processes as well as in helping to describe the system more generally in terms of its critical components.

In some applications, the form of the transducer function is already known, and estimating it may involve the optimization of just a few parameters to achieve a fit of a model curve to the experimental data. In many other applications, however, there is no standard model. This may be because the underlying process is poorly understood or the function itself represents several simpler processes interacting with each other in a complicated way.

The problem of estimating a transducer function when its form is unknown can be addressed in several ways. To help set the context of the bootstrap nonparametric approach to

1053-5888/07/$25.00©2007IEEE

this problem, it is useful to review two classical parametric approaches, one based on linear regression [9] and the other on a certain class of nonlinear functions [14].

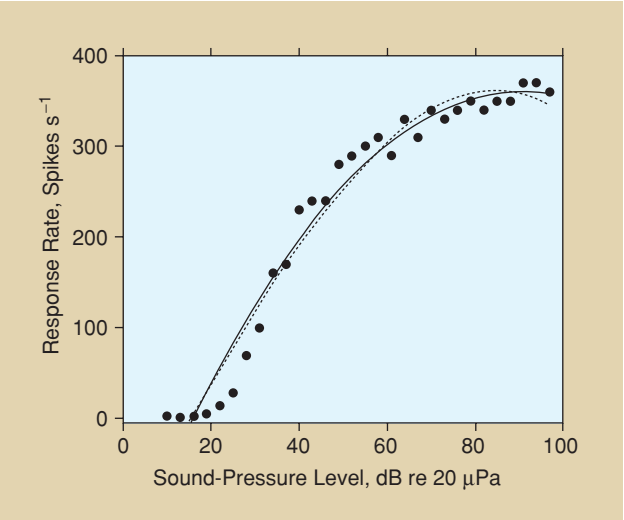
Figure 1 shows the results of recordings from an auditory nerve fiber of the guinea pig. The data were provided by N.P. Cooper from unpublished measurements summarized in [15]. The response rate y in spikes per second evoked by a sound stimulus is plotted as a function of the relative sound-pressure level x in decibels. The symbols are the experimental data, and the continuous and dotted lines are quadratic $y = \beta_0 + \beta_1 x + \beta_2 x^2$ and cubic $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ polynomial regression fits, respectively. Both fits have evident biases, i.e., regions where the data points are all systematically below or above the curve.

In general, the estimated transducer function is assumed to be monotonic, so the response increases as the level of the stimulus increases, and bounded, so the response does not go below nor above a certain level (because of limitations on the range of the stimuli, these bounds may not always be apparent). It is evident in Figure 1 that, at small x, both regression curves take on negative values, although the data are always positive, and, at large x, the cubic fit actually turns down. Choosing a higher degree of polynomial would improve the fit, but it is not obvious how high the degree should be.

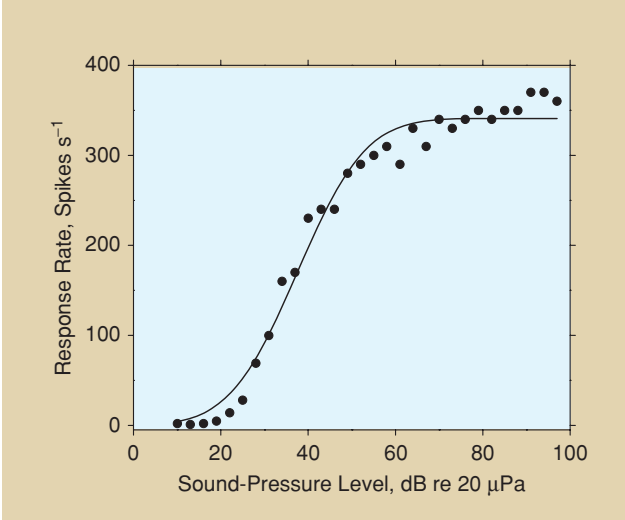
These properties of monotonicity and boundedness are captured by functions such as the Gaussian cumulative distribution function $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-u^2/2) du$. The continuous curve in Figure 2 shows the function $y = y_0\, \Phi(\beta_0 + \beta_1 x)$, where the coefficients y_0, β_0, and β_1 have been adjusted for best fit. Although the fitted curve is now positive at small x, the fit remains biased here and at medium to large x. Other sigmoidal functions might be tested to see whether they improve the fit, but again it is not obvious how these choices should be made.

Local nonparametric fitting provides a powerful alternative approach in which fitting is performed locally over small neighborhoods defined along the stimulus range [4], [12]. But the size of these neighborhoods, the bandwidth, is critical. A bandwidth that is too large gives a very smooth estimate of the function (similar to a parametric estimate), which suffers from large bias. Conversely, a bandwidth that is too small gives an estimate that follows the data very closely but has large variance, as the fit includes noise or other random variation in the data. The aim here is to show how the bootstrap resampling method [10] can provide an automatic and accurate method of estimating an optimal bandwidth for local fitting. Some other relevant applications of the bootstrap have been described elsewhere [30].

The organization of this article is as follows. First, the estimation problem is formalized for continuous response variables, and then the local least-squares approach to fitting is developed. Next, the central role of the bandwidth in balancing bias and variance is made explicit, and the bootstrap method for selecting this bandwidth is introduced. Two particular bootstrap methods are then formulated: one where the distribution of responses at each stimulus level is unknown and the other where it is known to be a particular member of the exponential family of distributions. This family includes the Gaussian distribution, the Poisson distribution (describing, e.g., the number or frequency of events at each level), and the binomial distribution (describing, e.g., the number of successes in a given number of trials at each level). For this family, fitting based on local least squares is replaced by fitting based on local log likelihood. Methods for calculating the goodness of fit, such as the deviance, and for estimating confi-



[FIG1] **Response rate of an auditory nerve fiber of the guinea pig as a function of sound-pressure level. The symbols are from unpublished source data summarized in [5]. The continuous and dotted lines show quadratic and cubic polynomial fits, respectively. Notice the large bias at low sound-pressure levels.**



[FIG2] **Response rate of an auditory nerve fiber of the guinea pig as a function of sound-pressure level. The symbols are from unpublished source data summarized in [5]. The continuous line shows a best-fitting Gaussian cumulative distribution function. Notice the bias at low and medium to high sound-pressure levels.**

dence intervals for the transducer function are briefly considered, along with approaches to securing the monotonicity of the function.

Bootstrap local fitting is then applied to a range of biological data, including the photoreceptor voltage response in the turtle retina, feedback current response in the goldfish retina, auditory-nerve spike rate in the guinea pig, frequency-of-seeing performance in a human patient with glaucoma, and a psychometric function for a normal human observer performing an image-discrimination task. In all these applications, the bootstrap method delivers an optimal or close-to-optimal estimate of the bandwidth. Finally, a potential weakness of the method and some alternative ways of selecting the bandwidth are considered in the conclusion to this article.

IN SOME APPLICATIONS, THE FORM OF THE TRANSDUCER FUNCTION IS ALREADY KNOWN. IN MANY OTHER APPLICATIONS, HOWEVER, THERE IS NO STANDARD MODEL.

THE FITTING PROBLEM

The fitting problem may be formalized in the following way. Suppose that the data consist of N stimulus levels $x_1, \ldots, x_N$ at which responses $y_1, \ldots, y_N$ are recorded. These responses are assumed, for the moment, to have continuous rather than discrete values but contaminated by additive noise, so their dependence on the stimulus level may be represented by

$$y_i = T(x_i) + \varepsilon_i, \quad 1 \leqslant i \leqslant N, \tag{1}$$

where T is the unknown transducer function and the ε_i represent noise, i.e., random errors in the observations or variability from sources not included in the x_i. These errors are assumed to be independent with zero mean and the same finite variance. They are often taken to be normally distributed, but this assumption is unnecessary, and no specific distribution is assumed here. No constraint is placed on the form of T at this stage, except that is should be smooth (monotonicity and boundedness are considered later).

In the approach to fitting based on polynomial regression [9], the function T is estimated by a polynomial $\hat{y}$ of degree k, say, in the x_i,

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k, \quad 1 \leqslant i \leqslant N,$$

where $k < N - 1$ and, in general, $k \ll N$. The coefficients β_j may be estimated by minimizing the sum of squares

$$\sum_{i=1}^{N} \left(y_i - \left(\beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k \right) \right)^2. \tag{2}$$

As already noted, the problem with polynomial regression is that it can introduce large biases in the fit. In addition, individual observations may have a strong influence on remote parts of the curve, the degree of the polynomial cannot be controlled continuously, and the high degree necessary for a good fit may lead to large variability in the estimated coefficients [12]. Local polynomial regression avoids most of these problems.

LOCAL POLYNOMIAL REGRESSION

The principle of local polynomial regression has already been outlined. For any selected stimulus level x_0, the regression is applied to a fraction of the data around x_0, the fitted value at x_0 is then calculated, the procedure is repeated at the next selected level, and so on. In more detail, assume that the transducer function T can be adequately approximated locally by a Taylor expansion, e.g., by a quadratic function of x; i.e., for each selected x_0, values of T at points x near x_0 are given by

$$T(x) \approx \alpha_0 + \alpha_1 (x - x_0) + \frac{1}{2} \alpha_2 (x - x_0)^2,$$

where the coefficients α_j are related to the derivatives of the function T and need to be estimated, and $|x - x_0|$ is less than some typical distance, the bandwidth $h(x_0)$, which generally depends on x_0. Notice that $\alpha_0 = T(x_0)$ here.

For each x_0, estimates $\hat{\alpha}_j$ of the coefficients α_j can be obtained by minimizing the locally weighted sum of squares, where the weight function w_h at x_0 reflects the bandwidth $h(x_0)$, i.e.,

$$\sum_{i=1}^{N} w_{h(x_0)}(x_i - x_0) \Big(y_i - \Big(\beta_0 + \beta_1 (x_i - x_0) + \beta_2 (x_i - x_0)^2 \Big) \Big)^2, \tag{3}$$

where, as before, N is the number of stimulus levels (or, more generally, number of observations, if the experiment is repeated at one or more levels). The solution $\hat{\beta}_j = \hat{\alpha}_j / j!$ and the local regression estimate $\hat{T}(x_0)$ of $T(x_0)$ is then simply the estimate $\hat{\beta}_0$ at the selected level x_0.

The weight function is usually expressed in terms of a kernel K, so $w_{h(x_0)}(x_i - x_0) = K((x_i - x_0)/h(x_0))$. For data where the levels are widely spaced, it is useful if the kernel has unbounded support, as with, e.g., a Gaussian function, $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$. A Gaussian kernel is used in all the examples of this article.

The foregoing leaves several issues unresolved. Fortunately, there is general agreement [28] that the precise shape of the weight function is not important and that the degree of polynomial usually need not be more than two. In fact, with transducer functions, there is an advantage in fixing the degree at one for the control of monotonicity. The degree of the polynomial is one in all the examples of this article. As emphasized earlier, however, the choice of bandwidth h is central. There are many ways of

estimating a suitable h, but the bootstrap offers a particularly appealing method [13], [19].

BOOTSTRAP BANDWIDTH SELECTION

The goal is to obtain an optimal value of the bandwidth h that minimizes the difference between the local regression estimate $\hat{T}_h(x)$ based on the data set $(x_1, y_1), \ldots, (x_N, y_N)$ and the true value $T(x)$, evaluated over the domain of definition of x. To indicate its importance, the dependence of the estimate $\hat{T}_h(x)$ on h is made explicit, and it is sufficient here to assume that h is constant with x. The difference between $\hat{T}_h$ and T may be quantified by various measures, in particular, by the mean integrated squared error (MISE), as a function of h, i.e.,

$$\mathrm{MISE}(h) = \mathrm{E} \int \left(\hat{T}_h(x) - T(x) \right)^2 dx, \tag{4}$$

where E is the expectation (the mean value that would have been obtained if the experiment had been repeated infinitely many times). The right-hand side of (4) may be reexpressed as the integral of the sum of a variance term and a bias-squared term, $\mathrm{E}((\hat{T}_h(x) - \mathrm{E}\hat{T}_h(x))^2) + (\mathrm{E}\hat{T}_h(x) - T(x))^2$. The other terms in the integrand vanish. Minimizing the difference in (4) is therefore intuitive, as this results in an estimate that is neither too noisy nor too biased.

The true value $T(x)$ in (4) is, of course, unknown. A simple application of the bootstrap method would entail replacing the values of the data that would have been obtained by repeating the experiment many times by values of the data obtained by resampling from the given data set [10]. More specifically, the procedure would be as follows. Sample with replacement from the data set $(x_1, y_1), \ldots, (x_N, y_N)$ a large number of times, say B times in all, and, for each bootstrap sample, indexed by $b = 1, \ldots, B$, construct a bootstrap estimate $\hat{T}_h^{*b}(x)$ of $T(x)$ at bandwidth h (the asterisk signifies that the estimate is a bootstrap estimate). Next, apply the bootstrap principle [10], and replace the unknown function $T(x)$ in (4) by the estimate $\hat{T}_h(x)$ and replace $\hat{T}_h(x)$ in (4) by its bootstrap version $\hat{T}_h^{*b}(x)$. The different ways in which these bootstrap samples may be generated are considered later. The bootstrap estimate of the MISE is then

$$\widehat{\mathrm{MISE}}(h) = B^{-1} \sum_{b=1}^{B} \int \left(\hat{T}_h^{*b}(x) - \hat{T}_h(x) \right)^2 dx, \tag{5}$$

where averaging over the B bootstrap replications has taken the place of the expectation in (4). The bootstrap bandwidth $\hat{h}$ is then defined as the minimizer of this estimate of the mean integrated square error in (5).

Unfortunately, there is a problem with the formulation of the right-hand side of (5). When, as before, it is reexpressed as the integral of the sum of a variance term and a bias-squared term, the bias-squared term $(\mathrm{E}^* \hat{T}_h^{*b}(x) - \hat{T}_h(x))^2$ also vanishes. Notice that the expectation E^* is taken with respect to the empirical distribution, which assigns a probability $1/N$ to each data point (x_i, y_i), where N is the number of points. This simple application of the bootstrap would therefore fail as the bias certainly increases with sufficiently large h.

To avoid this problem, the sample estimate $\hat{T}_h(x)$ in (5) is replaced by an estimate $\hat{T}_{h_0}(x)$ based on a pilot bandwidth h_0. This pilot bandwidth needs to be larger than the optimal bandwidth for estimating the transducer function in order to obtain a good approximation of the bias contribution to the MISE [19]. The bootstrap bandwidth $\hat{h}$ is then defined [13] as the minimizer of the revised estimate of the mean integrated square error; i.e.,

$$\widehat{\mathrm{MISE}}(h) = B^{-1} \sum_{b=1}^{B} \int \left(\hat{T}_h^{*b}(x) - \hat{T}_{h_0}(x) \right)^2 dx. \tag{6}$$

This still leaves the problem of estimating the pilot bandwidth h_0, but there are several methods available [12], [18], and the hope is that the bootstrap bandwidth $\hat{h}$ will not depend too strongly on h_0 [10]. The plug-in method [11], which uses estimates of unknown quantities in a formula for the asymptotically optimal bandwidth, provides an effective pilot bandwidth after multiplying by a factor $1.5\,N^{0.1}$, as suggested in [19]. This is the method used in all the examples of this article, and it resulted in good fits to the data.

The estimate $\widehat{\mathrm{MISE}}(h)$ is evaluated for bandwidths h drawn from an interval assumed to contain the optimal bandwidth. A plausible lower limit for this interval is the smallest distance between the stimulus levels, and a plausible upper limit is a multiple of the distance between the lowest and highest stimulus levels.

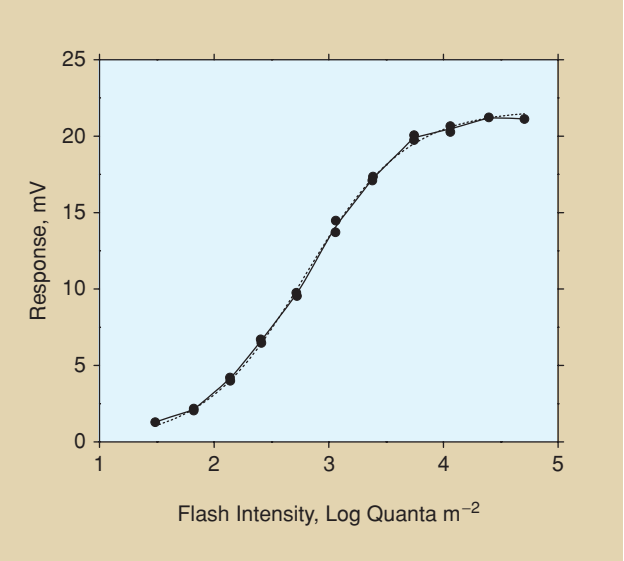
For the bootstrap local-fitting method to be put into practice, a decision has to be made about the mechanism of resampling. This decision depends on how much information about the distribution of responses at each stimulus level is available.

THE WILD BOOTSTRAP

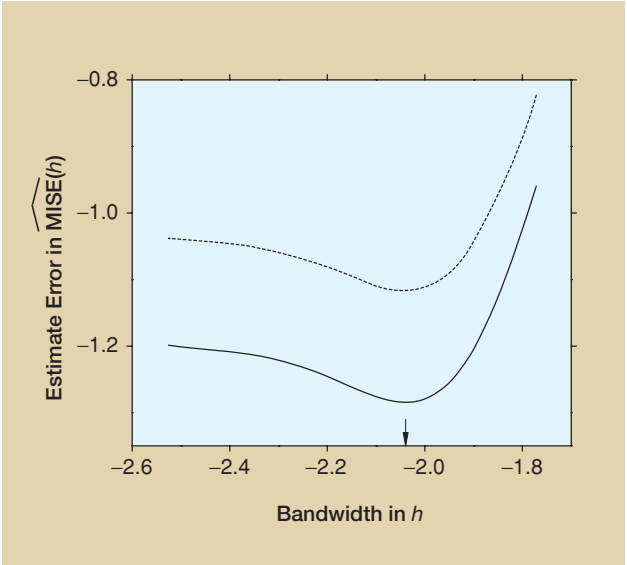
In the initial description of the stimulus-response function in (1), it was assumed that, although their distribution was unknown, the errors ε_i had the same variance. But, in many applications, the assumption of constant variance does not hold. Even so, providing that the variance does not change too rapidly with stimulus level, it can be treated as being approximately constant over each neighborhood with little effect on local regression.

Constant variance is, however, important if the optimal bandwidth is estimated by a bootstrap based on resampling the estimated errors or residuals over all stimulus levels, i.e., resampling $\hat{\varepsilon}_i = y_i - \hat{T}(x_i)$ with $i = 1, \ldots, N$.

One way to preserve this resampling approach with nonconstant error variance is to estimate the variance locally (see, e.g., [16]) and use the result to obtain normalized residuals. Another way is to use the wild bootstrap [15], [20]. In this method, the bootstrap samples are generated from a minimalist discrete two-point distribution that attempts to reconstruct the unknown distribution of each residual (although any distribution satisfying the moment assumptions in the following Step 4 will do). It



[FIG3] Voltage response of a red-sensitive cone photoreceptor of the turtle retina as a function of the logarithm of the stimulus light intensity. The symbols are replotted from [1, Figure 8]. The continuous curve is a local polynomial regression with wild bootstrap bandwidth $h_{\text{wild}} = 0.131$ and is identical to a local fit by a Gaussian general linear model with identity link function and the same bootstrap bandwidth $h_{\text{GLM}} = 0.131$. The dotted curve is a global fit by a model hyperbolic tangent function [26]. The locally and globally fitted curves are almost identical.



[FIG4] Estimate of the MISE as function of the bandwidth h for the data in Figure 3. The continuous curve is for the wild bootstrap, and the dotted curve is for the bootstrap with a Gaussian general linear model. The common location of the minima is indicated by an arrow.

therefore varies as the variance does with stimulus level.

In more detail, the steps involved are as follows:

Step 1) Start with a deliberately large pilot bandwidth h_0.

Step 2) Obtain an estimate $\hat{T}_{h_0}$ of the transducer function from the data at this bandwidth h_0.

Step 3) Obtain the residuals $\hat{\varepsilon}_i = y_i - \hat{T}_{h_0}(x_i)$ at $i = 1, \ldots, N$.

Step 4) Generate B bootstrap residuals ε_i^{*b}, $b = 1, \ldots, B$, from a two-point distribution with mean 0, variance $\hat{\varepsilon}_i^2$, and third moment $\hat{\varepsilon}_i^3$, at each $i = 1, \ldots, N$.

Step 5) Add the bootstrap residuals ε_i^{*b}, $b = 1, \ldots, B$ to the pilot estimate $\hat{T}_{h_0}(x_i)$ to form the corresponding bootstrap observations y_i^{*b}, at each $i = 1, \ldots, N$.

Step 6) For each bootstrap sample $(y_1^{*b}, \ldots, y_N^{*b})$, $b = 1, \ldots, B$, find estimates $\hat{T}_h^{*b}$ of the function over an interval of bandwidths h assumed to contain the optimal value.

For details of the algorithm for generating appropriate samples in Step 4, see [20]. The method described previously is now applied to estimate the MISE defined by (6), the minimizer of which is the wild bootstrap bandwidth $\hat{h}_{\text{wild}}$.

Figure 3 shows an application of this method to microelectrode recordings from a single cone photoreceptor in the turtle retina [1]. The mean peak voltage response evoked by a stimulus light flash of wavelength 644 nm is plotted as a function of the logarithm of the flash intensity. The symbols are the experimental data (replotted from [1, Figure 8]). This example provides a useful test, as there is a known parametric model of the transducer function based on photoreceptor responses in the fish [26]. The continuous curve is a local polynomial regression with bandwidth $\hat{h}_{\text{wild}} = 0.131$ selected by the wild bootstrap method. The dotted curve shows the model function, a hyperbolic tangent function $y = y_0(1 + \tanh(\beta_1(x - x_0)))$, where the coefficients x_0, y_0, and β_1 have been adjusted for best fit. The two curves are almost identical.

The continuous curve in Figure 4 shows how the wild bootstrap bandwidth $\hat{h}_{\text{wild}}$ was obtained from the estimated MISE. The logarithm of the estimated MISE is plotted against the logarithm of the bandwidth h. There is a minimum at $\ln h = -2.03$; i.e., $h = 0.131$. The dotted curve is for a different bootstrap method, the GLM bootstrap, described later. The minimum is at the same location $\ln h = -2.03$; i.e., $h = 0.131$. This example is considered further in "Turtle Photoreceptor Voltage Response."

As a second example but without a model function, Figure 5 shows an application of bootstrap local fitting to signals recorded from the goldfish retina. The symbols are the feedback current from a horizontal cell as a function of the intensity of a light stimulus of wavelength 550 nm [23]. The continuous curve is a local polynomial regression with bandwidth $\hat{h}_{\text{wild}} = 0.190$ selected by the wild bootstrap method. The dotted curve is the local fit obtained with the GLM bootstrap described later. This example is considered further in "Goldfish Horizontal-Cell Feedback Current."

In some applications, the form of the distribution of the responses at each stimulus level is known in principle. Although the wild bootstrap may still be applied, the nonparametric fitting approach and the bootstrap estimate of the optimal bandwidth can both be improved by taking advantage of the additional information about the distribution. The following section provides a general framework.

GENERALIZED LINEAR MODELS

Generalized linear models (GLMs) are an extension of classical linear models in which the component describing the distribution of observations is a member of the exponential family. This family, as mentioned in the beginning of this article, includes the Gaussian, Poisson, and binomial distributions. For any member of this family, the conditional density $f(y|x)$ of y given x is defined [25] by

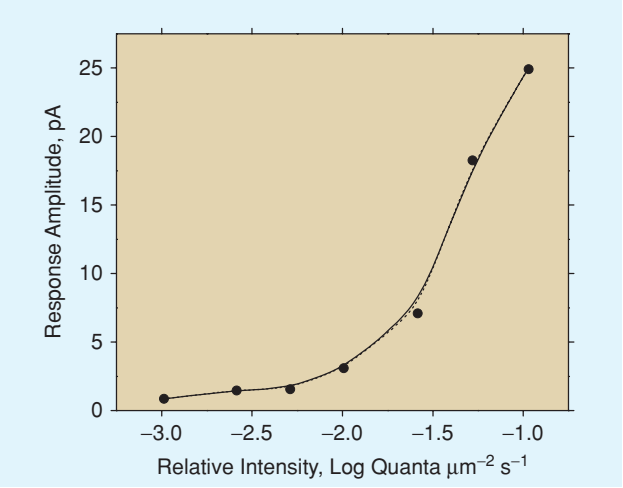
$$f(y|x) = \exp\left[\frac{\theta(x)y - b(\theta(x))}{a(\phi)} + c(y,\phi)\right]$$

for functions a, b, and c, and canonical parameter θ and dispersion parameter ϕ. The expectation $\mathrm{E}(y|x) = b'(\theta(x))$ and the variance $\mathrm{Var}(y|x) = b''(\theta(x))a(\phi)$, where the primes denote differentiation with respect to θ [25]. The variance is the product of two functions: $b''(\theta(x))$, which depends on the canonical parameter $\theta(x)$ and therefore only on the mean $m(x)$, and $a(\phi)$, which is independent of $\theta(x)$ and depends only on ϕ (in this context, it is convenient to use the notation $m(x)$ for $\mathrm{E}(y|x)$). By assumption, $m(x)$ coincides with the value $T(x)$ of the transducer function at x.

A property of GLMs is that the mean $m(x)$, after transformation by a link function g, is modeled linearly. For each distribution in this exponential family, the link function for which $g(m(x)) = \theta(x)$ is the canonical link function. This function has some interesting statistical properties (see [25] for details), but there is no guarantee that $\theta(x)$ is a simple function of x, such as a polynomial, although the variance is usually more stable over x after such a transformation.

For a Gaussian distribution with constant variance,

$$f(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - m(x))^2}{2\sigma^2}\right).$$



[FIG5] Feedback current from a horizontal cell of the goldfish retina as a function of light intensity. The symbols are replotted from [23, Figure 8]. The continuous curve is a local polynomial regression with wild bootstrap bandwidth $\hat{h}_{\mathrm{wild}} = 0.190$. The dotted curve is a local fit by a Gaussian general linear model with identity link function and corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 0.177$. The two curves are almost identical.

The dispersion parameter $\phi = \sigma^2$ and the canonical link function is the identity, $g(y) = y$. For a Poisson distribution,

$$f(y|x) = e^{-\lambda(x)}\frac{\lambda(x)^y}{y!},$$

where $y = 0, 1, 2, \ldots$, and $\lambda(x) > 0$. The canonical link function $g(y) = \ln(y)$ and the canonical parameter $\theta(x) = \ln\lambda(x)$. Finally, for a binomial distribution,

$$f(y|x) = \frac{n!}{y!(n-y)!}p(x)^y(1-p(x))^{n-y},$$

where $y = 0, 1, \ldots, n$, with n depending, in general, on x, and $0 \leqslant p(x) \leqslant 1$. The canonical link function $g(y) = \ln(y/(n-y))$, which is known as the logistic transformation. The canonical parameter $\theta(x) = \ln(p(x)/(1-p(x)))$. Examples are given later.

For GLMs, maximum likelihood rather than least squares is used in the fitting procedure. In fact, it is easier to work with the logarithm of the likelihood, which then takes the form

$$\sum_{i=1}^{N}\left[\frac{\theta(x_i)y_i - b(\theta(x_i))}{a(\phi)} + c(y_i,\phi)\right], \tag{7}$$

where, as noted earlier, $\theta(x)$ and $m(x)$ are related by $\theta(x) = (b')^{-1}(m(x))$. As with least squares, the transformed mean $g(m(x))$ may be estimated by a polynomial of degree k, say, in the x_i,

$$g(m(x_i)) = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k, \quad 1 \leqslant i \leqslant N,$$

where $k < N - 1$. The coefficients β_j can then be estimated by maximizing the log likelihood. For the Gaussian family, maximizing likelihood is equivalent to minimizing least squares [25].

The problems of polynomial regression with maximum likelihood are similar to those of polynomial regression with least squares. The fit may suffer from large biases, it can be strongly influenced by individual observations, the degree of the polynomial cannot be controlled continuously, and there may be large variability in the estimated coefficients. Additionally, the choice of what constitutes an appropriate link function so the mean becomes a polynomial in the stimulus levels is not always obvious. A wrongly chosen link function may result in a poor fit and misleading inferences [6].

In the local approach to fitting by least squares, the sum of squares (2) was replaced by its locally weighted version (3). Analogously, in the local approach to fitting by maximum likelihood, the log likelihood (7) is replaced by its locally weighted version, which, at each x_0, takes the form

$$\sum_{i=1}^{N} w_{h(x_0)}(x_i - x_0)\left[\frac{\theta(x_i)y_i - b(\theta(x_i))}{a(\phi)} + c(y_i,\phi)\right],$$

where $\theta(x) = (b')^{-1}(m(x))$. The transformed mean $g(m(x))$ is assumed to be adequately approximated locally by a Taylor

expansion, e.g., a quadratic function of x; i.e., for each x_0, values of $g(m(x))$ at points x near x_0 are given by

$$g(m(x)) \approx \alpha_0 + \alpha_1(x - x_0) + \frac{1}{2}\alpha_2(x - x_0)^2 .$$

The estimate of $m(x_0)$ is then $g^{-1}(\hat{\beta}_0)$, where $\beta_j = \alpha_j/j!$ for each j, as before.

Local maximum-likelihood fitting resolves most of the problems with global fitting. Most importantly, the choice of link function has a very small effect on the fit. But issues similar to those identified with local least-squares fitting arise in local maximum-likelihood fitting, and the choice of the bandwidth remains critical. The bootstrap again provides an attractive way to find a suitable bandwidth.

BOOTSTRAP BANDWIDTH SELECTION IN LOCAL MAXIMUM-LIKELIHOOD FITTING

The goal, as with local least-squares fitting, is to find an optimal value of the bandwidth h that minimizes the difference between the estimate $\hat{m}_h(x) = \hat{T}_h(x)$ and the true value $m(x) = T(x)$, evaluated over the domain of definition of x. This difference may again be quantified by the MISE but evaluated here for the transformed values, i.e., for $g(m(x))$ and $\hat{\beta}_0$ at each x. As indicated earlier, by fitting the transformed mean $g(m(x))$, the variance is more likely to be stable over x.

LOCAL FITTING WITH BOOTSTRAP BANDWIDTH SELECTION PROVIDES A POWERFUL METHOD FOR ESTIMATING A TRANSDUCER FUNCTION FROM A SET OF BIOLOGICAL DATA.

If the density $f(y|x)$ were known exactly, samples could be generated and used to estimate the MISE as in (4). But, in practice, the parameters of the distribution, in particular, the mean $m(x)$, need to be estimated. As before, the sample estimate $\hat{m}_{h_0}$ is based on a pilot bandwidth h_0. Resampling with GLMs is parallel to resampling with the discrete two-point distribution used in the wild bootstrap, except that the samples are drawn from $f(y|x)$.

In more detail, the steps are as follows:

Step 1) Start with a deliberately large pilot bandwidth h_0.

Step 2) Obtain an estimate $\hat{m}_{h_0}$ of the mean from the data at this bandwidth h_0.

Step 3) Generate B bootstrap observations y_i^{*b}, $b = 1, \ldots, B$, from the distribution with density $f(y|x_i)$ whose mean $\hat{m}_{h_0}(x_i)$ was estimated in Step 2.

Step 4) For each bootstrap sample $(y_1^{*b}, \ldots, y_N^{*b})$, $b = 1, \ldots, B$, find estimates $\hat{m}_h^{*b}$ of the mean over an interval of bandwidths h assumed to contain the optimal value.

The method described previously is now applied again to estimate the MISE defined by (6), the minimizer of which is the GLM bootstrap bandwidth $\hat{h}_{\mathrm{GLM}}$.

Applications of this method are described in "Examples."

GOODNESS OF FIT

To assess the adequacy of the estimated transducer function in accounting for the data, it is helpful to have a measure of goodness of fit. In global least-squares fitting, a commonly used measure is the residual sum of squares, which, for Gaussian errors, has a χ^2 distribution. The equivalent measure for GLMs is the deviance D (see [25] for details). Thus, for a Gaussian distribution,

$$D = \sum_{i=1}^{N} (y_i - \hat{m}(x_i))^2 .$$

For a Poisson distribution,

$$D = 2\sum_{i=1}^{N} y_i \ln\left(\frac{y_i}{\hat{m}(x_i)}\right) - (y_i - \hat{m}(x_i)) ,$$

where $y_i = 0, 1, 2, \ldots$. Finally, for a binomial distribution,

$$D = 2\sum_{i=1}^{N} y_i \ln\left(\frac{y_i}{\hat{m}(x_i)}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{m}(x_i)}\right) ,$$

where $y_i = 0, 1, \ldots, n_i$ [25]. All three are used in "Examples."

If the model is correct, then the deviance divided by the dispersion parameter is distributed approximately as χ^2 with residual degrees of freedom (DoF) given by the number of data points N minus the number of estimated coefficients; i.e., DoF $= N - k - 1$ [25]. For the normal distribution with unknown dispersion, the variance has to be estimated.

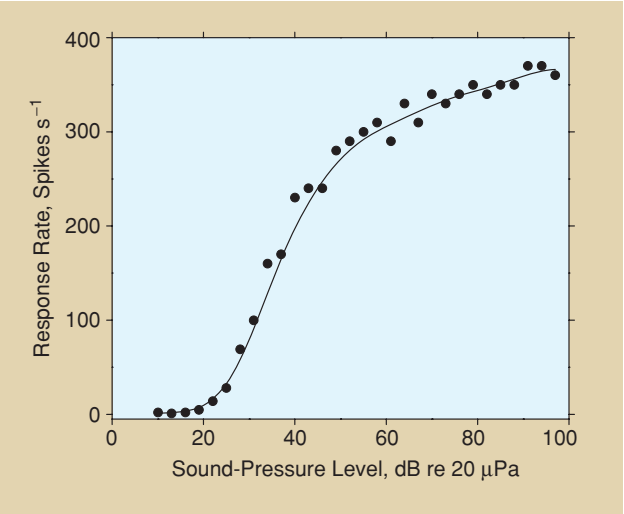
In local fitting, the deviance may be also used as a measure of goodness of fit. But the concept of DoF is more complex and there are several definitions based on the so-called hat matrix $\mathbf{H}$, which relates the observations to the estimates, i.e., $\mathbf{H}y_i = \hat{y}_i$ [22], [24]. One simple definition, which is an extension of the definition of DoF for parametric models, is the trace of the hat matrix $\mathrm{tr}(\mathbf{H})$. For a global polynomial regression of degree k, the trace $\mathrm{tr}(\mathbf{H}) = k + 1$. For a local polynomial regression, the trace will generally be noninteger.

The approximate distribution of the deviance for a non-Gaussian distribution has to be used with caution. The result is only asymptotic [25], and, with finite samples, it is unclear how good the approximation is. For discussion of measures of goodness of fit and definitions of DoF, see [21], [22], and [24].

CONFIDENCE INTERVALS

One of the advantages of the bootstrap method is the ready estimation of simultaneous confidence intervals. For suitably chosen statistics, intervals obtained by this method may converge more quickly than intervals obtained from asymptotic theory [8].

Resampling schemes similar to those described for the wild bootstrap and the GLM bootstrap can be used to obtain confidence intervals for the estimated transducer function. Steps 1–5 (for the wild bootstrap) and Steps 1–3 (for the GLM bootstrap)



[FIG6] Response rate of an auditory nerve fiber of the guinea pig as a function of sound-pressure level. The symbols are from unpublished source data summarized in [5] and the continuous curve is a local fit by a Poisson GLM with canonical log link function and corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 6.02$. This fit is much better than the global parametric fits of Figures 1 and 2.

are executed, but now the bootstrap bandwidth $\hat{h}$ is used in place of the pilot bandwidth h_0.

The simplest estimates of confidence intervals may then be derived from the empirical centiles of the bootstrap samples. More advanced methods may be applied to obtain better approximations and faster rates of convergence. For discussion of different bootstrap confidence intervals, see, e.g., [8].

MONOTONICITY AND BOUNDEDNESS

As observed elsewhere [2], there are two approaches to the requirement of monotonicity in the fitted function. One approach is to assume that any good nonparametric estimator will reflect the monotonicity of the data, so monotonicity constraints need not be introduced. An alternative approach is to incorporate a monotonicity constraint directly into the fitting procedure. This may be done in several ways (see [7] and references therein).

The approach adopted in this article is the former: to allow the fit to indicate the monotonicity. With local fitting based on a polynomial of degree one, it is possible to impose monotonicity on a nonmonotonic estimate by progressively increasing the bandwidth from its nominal optimal value. An independent test of the monotonicity of the data is described in [2] and [17].

A similar argument might be applied to assumptions about the boundedness of the fitted function, but here the risk of violations in an unconstrained fit are more serious. With binomial and Poisson data, the response variables can never fall below zero, and with binomial data, the number of successes can never exceed the number of trials. Using the corresponding link functions in local fitting avoids excursions into physically unrealizable parts of the response range.

EXAMPLES

TURTLE PHOTORECEPTOR VOLTAGE RESPONSE

As noted earlier, the experimental data in Figure 3, showing the photoreceptor voltage response to a light stimulus in the turtle retina, are well fitted by a parametric model (dotted curve), a hyperbolic tangent function for which the deviance $D = 1.34$ on 16 residual DoF. A nonparametric estimate [16] of the dispersion parameter ϕ gives an estimated scaled deviance $D/\phi = 5.29$, which is still small compared with the residual DoF. Local polynomial regression, with a wild bootstrap bandwidth $\hat{h}_{\mathrm{wild}} = 0.131$, produced an almost identical result (continuous curve). Remember that here and in the subsequent examples the polynomial is of degree one. For comparison, local fitting was also performed with a Gaussian GLM and canonical identity link function. The corresponding bootstrap bandwidth was the same as with the wild bootstrap, i.e., $\hat{h}_{\mathrm{GLM}} = 0.131$. For each local fit, the deviance $D = 0.541$ on 8.78 residual DoF. The estimated scaled deviance $D/\phi = 2.13$, which is still small.

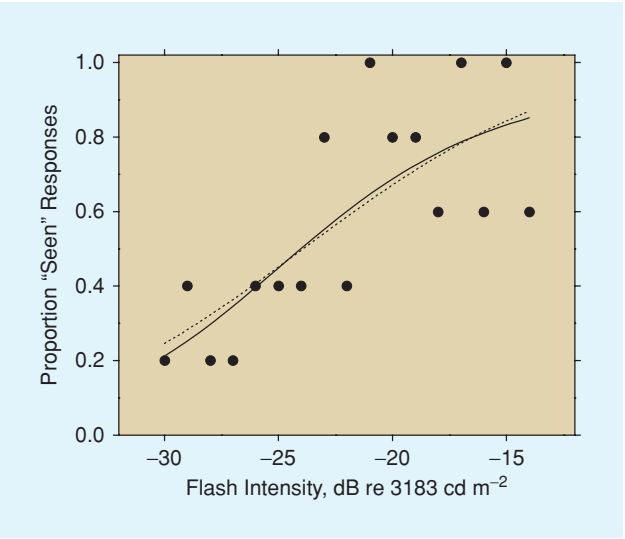
GOLDFISH HORIZONTAL-CELL FEEDBACK CURRENT

As also noted earlier, the experimental data in Figure 5, showing the feedback current response to a light stimulus in the goldfish retina, are well fitted by local polynomial regression (continuous curve). The wild bootstrap bandwidth $\hat{h}_{\mathrm{wild}} = 0.190$, for which the deviance $D = 2.21$ on 1.51 residual DoF. The estimated scaled deviance $D/\phi = 1.13$, which is small. For comparison, the dotted curve was obtained by local fitting with a Gaussian GLM and identity link function. The corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 0.177$, for which the deviance $D = 1.45$ on 1.25 residual DoF. The estimated scaled deviance $D/\phi = 0.743$, which is also small. The two curves are almost identical.

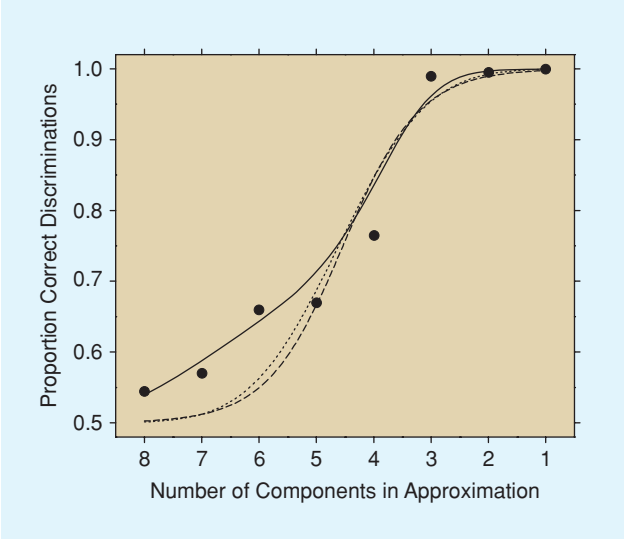
GUINEA PIG AUDITORY SPIKE FREQUENCY

Figure 6 shows the response rate in spikes per second from an auditory nerve fiber of the guinea pig taken from unpublished source data summarized in [5]. The data were used earlier to illustrate global parametric fitting. The distribution of responses at each level is known to be approximately Poisson, although there are also fractal components. The quadratic and cubic regressions gave a poor fit (Figure 1), with deviance $D = 206$ on 26 residual DoF for the cubic regression. The estimated scaled deviance $D/\phi = 257$, which is large. Although the Gaussian cumulative distribution function (Figure 2) gave a somewhat improved fit with deviance $D = 105$ on 27 residual DoF, the estimated scaled deviance $D/\phi = 131$, which is still large.

The continuous curve in Figure 6 was obtained by local fitting with a Poisson GLM and canonical log link function. The corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 6.02$, for which the deviance $D = 24.8$ on 23.1 residual DoF. The estimated scaled deviance $D/\phi = 31.1$, which is not too large. The curve fits the data closely. In the original experimental study [5], freehand curves were drawn through similar data.



[FIG7] **Frequency of seeing by a glaucoma patient as function of stimulus light intensity. The symbols are replotted from [3, Figure 5, bottom]. The continuous curve is a local fit by a binomial GLM with canonical logistic link function and corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 9.02$. The dotted curve is a global fit by a binomial GLM with probit link function.**



[FIG8] **Human performance in discriminating between images and their approximations as a function of the number of components in the approximation. The symbols are replotted from unpublished source data summarized in [27]. The continuous curve is a local fit by a binomial GLM with canonical logistic link function and corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 1.07$. The dashed curve is a global fit by a binomial GLM with logistic link function, and the dotted curve is the same but with the closely related probit link function, each with chance response level 0.5. Both global fits are poorer than the local fit.**

FREQUENCY-OF-SEEING PERFORMANCE IN GLAUCOMA

Figure 7 shows frequency-of-seeing data for a human patient with glaucoma [3]. A flash of light was presented a fixed number of times at a particular location on the retina, and the patient reported whether the flash was seen as its intensity was varied. The symbols are the proportion of positive responses in five trials at each stimulus level (replotted from [3, Figure 5, bottom]). Frequency-of-seeing data are customarily fitted globally by a parametric model [14], a binomial GLM with a probit link function, i.e., the inverse Φ^{-1} of the Gaussian cumulative distribution function. This model provides a useful reference, and the fit here, shown by the dotted curve, is good, with deviance $D = 17.1$ on 15 residual DoF (the dispersion parameter for the binomial distribution $\phi = 1$). A local fit by a binomial GLM with canonical logistic link function is shown by the continuous curve. The corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 9.02$, for which the deviance $D = 16.4$ on 14.7 residual DoF. The fit is also good, but it is a little less symmetric. Notice that the DoF for the local fit is 2.26, slightly more than the two DoF of the global parametric fit. The link functions for the global and local models are different, but this has little effect here.

HUMAN TWO-ALTERNATIVE FORCED-CHOICE PATTERN DISCRIMINATION

Figure 8 shows performance by a normal human observer discriminating between images of natural colored scenes and approximations to those images synthesized with varying levels of fidelity [27]. The symbols are the proportion of correct responses out of 200 trials as a function of the number of components in the approximation in reverse order (replotted from unpublished source data summarized in [27]). Because each trial was a two-alternative forced-choice task, chance performance level is 50%. Global fitting by a binomial GLM with the commonly used link functions is poor. Thus, for a logistic link function, shown by the dashed curve, the deviance $D = 33.1$ on six residual DoF, and for a probit link function, shown by the dotted curve, the deviance $D = 30.4$ also on six residual DoF. Local fitting by a binomial GLM with canonical logistic link function, shown by the continuous curve, is better. The corresponding bootstrap bandwidth $\hat{h}_{\mathrm{GLM}} = 1.07$. The fitted curve follows the data more closely at $x = 6, 7, 8$, and has slightly more of an inflexion at $x = 3$. Even so, the fit is not completely satisfactory. The deviance $D = 15.2$ on 4.31 residual DoF.

CONCLUSIONS

Local fitting with bootstrap bandwidth selection provides an effective method for estimating a transducer function from a set of biological data, overcoming many of the problems with parametric regression. As the examples showed, where a valid parametric model was already known, as with the turtle photoreceptor response and patient frequency-of-seeing performance, local fitting gave an almost identical result. Where a valid parametric model was unknown, local fitting generally provided an excellent description of the data. Even when the response was a complicated function of stimulus level, as with human discrimination of image approximations, local fitting produced a good description that preserved monotonicity. Moreover, where local fits using the wild bootstrap and the GLM bootstrap were both obtained, the results were both identical (Figure 3) or almost identical (Figure 5).

The goodness of the bootstrap estimate of the optimal bandwidth used in these local fits depends necessarily on the pilot bandwidth estimate but, in principle, not strongly. A severely over-smoothed pilot estimate will probably result in a larger-than-optimal bootstrap bandwidth and, conversely, for an under-smoothed pilot estimate.

The bootstrap is not the only method available for estimating bandwidths. Another possibility is cross-validation, which estimates the optimal bandwidth by minimizing the difference between the data and the local fits obtained with successive data points omitted [22], [24]. Although bandwidth selection by bootstrap compares favorably with that by cross-validation, under some conditions, the bootstrap may produce estimates that are slightly too large, but this may be preferable to producing estimates that are too small, as cross-validation sometimes does [29]. The plug-in method, used here to obtain the pilot bandwidth, can also be used directly to estimate the optimal bandwidth, but its calculation can be complicated and it does rely on asymptotic results, which might not hold for small data sets. As with any automatic method, it is important to assess the results of the fit. Graphical methods are a useful aid to this end, at least in low dimensions [22].

In summary, given the often limited knowledge of the biological processes underlying a transducer function, bootstrap local fitting of stimulus-response data seems remarkably effective. It can provide an accurate estimate of the function, preserving the expected critical features, its monotonicity, and, where known, its asymptotic behavior at small and large stimulus levels. Confidence intervals for the fitted function can also be readily estimated. Implementation of the local-fitting and bootstrap routines is straightforward, and computational overheads need not limit its practical application.

ACKNOWLEDGMENTS

The authors thank A.W. Bowman and P.H. Artes for advice; N.P. Cooper, S.M.C. Nascimento, and K. Amano for making available unpublished data; and K. Amano and I. Marín Franch for critically reading the manuscript. This work was supported by the EPSRC (Grant EP/C003470/1).

AUTHORS

David H. Foster (d.h.foster@manchester.ac.uk) received a B.Sc. in physics in 1966 and a Ph.D. in 1970, both from Imperial College London. He received a D.Sc. in 1982 from London University. He was appointed lecturer at Imperial College in 1970 and subsequently has held professorships at Keele University, Aston University, UMIST, and Manchester University. He is a fellow of the Institute of Physics, the Institute of Mathematics and its Applications, and the Optical Society of America. He is on the editorial boards of *Spatial Vision*, *Computers in Biology and Medicine*, and *Vision Research*. His research has concentrated on vision and mathematical modeling of visual processing.

Kamila Żychaluk (kamila.zychaluk@manchester.ac.uk) received an M.Sc. from the Wroc´law University of Technology in 2000 and a Ph.D. from the University of Birmingham in 2004, both in statistics. She was a Research Associate at Sheffield University from 2003 to 2005 and then at Manchester University from 2005. She is a member of the Royal Statistical Society. Her research interests are in nonparametric and Bayesian statistics with applications in biology and vision.

REFERENCES

[1] D.A. Baylor and A.L. Hodgkin, "Detection and resolution of visual stimuli by turtle photoreceptors," *J. Phys.*, vol. 234, no. 1, pp. 163–198, 1973.

[2] A.W. Bowman, M.C. Jones, and I. Gijbels, "Testing monotonicity of regression," *J. Computat. Graphical Statist.*, vol. 7, no. 4, pp. 489–500, 1998.

[3] B.C. Chauhan, J.D. Tompkins, R.P. LeBlanc, and T.A. McCormick, "Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma," *Investigative Ophthalmology Visual Sci.*, vol. 34, no. 13, pp. 3534–3540, 1993.

[4] W.S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Amer. Stat. Assoc.*, vol. 74, no. 368, pp. 829–836, 1979.

[5] N.P. Cooper and G.K. Yates, "Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve-fibres: variations with characteristic frequency," *Hearing Res.*, vol. 78, no. 2, pp. 221–234, 1994.

[6] C. Czado and T.J. Santner, "The effect of link misspecification on binary regression inference," *J. Stat. Planning Inference*, vol. 33, no. 2, pp. 213–231, 1992.

[7] H. Dette and K.F. Pilz, "A comparative study of monotone nonparametric kernel estimates," *J. Stat. Comput. Simul.*, vol. 76, no. 1, pp. 41–56, 2006.

[8] T.J. DiCiccio and B. Efron, "Bootstrap confidence intervals. With comments and a rejoinder by the authors," *Stat. Sci.*, vol. 11, no. 3, pp. 189–228, 1996.

[9] N.R. Draper and H. Smith, *Applied Regression Analysis*. New York: Wiley, 1998.

[10] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1993.

[11] J. Fan, N.E. Heckman, and M.P. Wand, "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," *J. Amer. Stat. Assoc.*, vol. 90, no. 429, pp. 141–150, 1995.

[12] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall, 1996.

[13] J.J. Faraway and M. Jhun, "Bootstrap choice of bandwidth for density estimation," *J. Amer. Stat. Assoc.*, vol. 85, no. 412, pp. 1119–1122, 1990.

[14] D.J. Finney, *Probit Analysis*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1971.

[15] E. Flachaire, "Bootstrapping heteroskedastic regression models: Wild bootstrap vs. pairs bootstrap," *Computat. Stat. Data Anal.*, vol. 49, no. 2, pp. 361–376, 2005.

[16] T. Gasser, L. Sroka, and C. Jennen-Steinmetz, "Residual variance and residual pattern in nonlinear regression," *Biometrika*, vol. 73, no. 3, pp. 625–633, 1986.

[17] P. Hall and N.E. Heckman, "Testing for monotonicity of a regression mean by calibrating for linear functions," *Ann. Statist.*, vol. 28, no. 1, pp. 20–39, 2000.

[18] P. Hall and I. Johnstone, "Empirical functionals and efficient smoothing parameter selection," *J. Roy. Stat. Soc., Series B*, vol. 54, no. 2, pp. 475–530, 1992.

[19] W. Härdle and A.W. Bowman, "Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands," *J. Amer. Stat. Assoc.*, vol. 83, no. 401, pp. 102–110, 1988.

[20] W. Härdle and J.S. Marron, "Bootstrap simultaneous error bars for nonparametric regression," *Ann. Statist.*, vol. 19, no. 2, pp. 778–796, 1991.

[21] J.D. Hart, *Nonparametric Smoothing and Lack-of-Fit Tests*. New York: Springer-Verlag, 1997.

[22] T.J. Hastie and R.J. Tibshirani, *Generalized Additive Models*. London, U.K.: Chapman & Hall, 1990.

[23] D.A. Kraaij, M. Kamermans, and H. Spekreijse, "Spectral sensitivity of the feedback signal from horizontal cells to cones in goldfish retina," *Vis. Neuroscience*, vol. 15, no. 5, pp. 799–808, 1998.

[24] C. Loader, *Local Regression and Likelihood*. New York: Springer-Verlag, 1999.

[25] P. McCullagh and J.A. Nelder, *Generalized Linear Models*. London, U.K.: Chapman & Hall, 1989.

[26] K.I. Naka and W.A.H. Rushton, "S-potentials from colour units in the retina of fish (Cyprinidae)," *J. Phys.*, vol. 185, no. 3, pp. 536–555, 1966.

[27] S.M.C. Nascimento, D.H. Foster, and K. Amano, "Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 22, no. 6, pp. 1017–1022, 2005.

[28] J.S. Simonoff, *Smoothing Methods in Statistics*. New York: Springer-Verlag, 1996.

[29] C.C. Taylor, "Bootstrap choice of the smoothing parameter in kernel density estimation," *Biometrika*, vol. 76, no. 4, pp. 705–712, 1989.

[30] A.M. Zoubir and D.R. Iskander, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, May 2004. SP