

Stochastic resonance in a statistical model of a time-integrating detector

Ursula U. Müller*

FB 3 Mathematics, University of Bremen, 28334 Bremen, Germany

Lawrence M. Ward[†]

Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

(Received 19 August 1999; revised manuscript received 12 January 2000)

We study an optimal nonparametric regression model for a threshold detector exposed to a noisy, subthreshold signal. The problem of recovering the signal is similar to that faced by neurons in nervous systems, although our model is intended to be normative rather than realistic. In our approach, the time-integrating activity of the neuron is modeled by kernel regression. Several aspects of the performance of the model are studied, including the existence of an optimal amount of noise (stochastic resonance). We construct a sequential, data-driven procedure for estimating the subthreshold signal. The performance of our model for threshold data is compared with kernel estimation for fully observed data. Finally, we discuss differences between our estimator and the best estimator for a constant signal.

PACS number(s): 87.10.+e, 05.45.-a, 02.50.Ph, 05.40.Ca

I. INTRODUCTION

Stochastic resonance (SR) is a nonlinear cooperative effect in which large-scale stochastic fluctuations (e.g., “noise”) are entrained by an independent, often but not necessarily, periodic, weak fluctuation (or “signal”) with the result that the weaker signal fluctuations are amplified (see [1] for a review). Classical SR in physical systems has been generalized to include noise-enhanced signal detection exhibited by a wide variety of information processing systems, including living ones. For example, it has been demonstrated that model excitable systems with one stable state and a threshold to an unstable excited state, such as model neurons, exhibit this generalized form of SR [2–4]. With suitable tuning of the noise, such a model neuron can be so sensitive that it can detect a weak constant signal which elicits on average only a single additional spike. This provides a mechanism for speedy neural response to such signals [5]. Moreover, the generalized form of SR has also been demonstrated to exist in a variety of living systems, including networks of neurons [6]. Indeed, researchers have speculated that SR in this form is a fundamental and general principle of biological information processing (e.g., [7]).

The study of systems that exhibit SR has been greatly facilitated by the demonstration that a simple threshold detector also exhibits SR (e.g., [8]). Such a detector has no dynamics and fires a pulse each time its input exceeds a threshold. Since SR is an aspect of the detectability of a signal in thresholded data, not of the dynamics that cause it [9], studying simple threshold detectors is sufficient for many purposes. Indeed, the simplest model of a neuron is just such a simple threshold detector [10]. In the present paper we consider the output of a single threshold detector

exposed to an arbitrary subthreshold signal embedded in noise. This could be the situation of a single neuron in a neural network. The noise is composed of rapidly varying, unsynchronized inputs from perhaps 1000 other neurons, and the signal may be longer time scale variations in the input of one specific neuron (or a few synchronized ones) that is by itself insufficient to drive the neuron of interest. Nonetheless, the output of the latter neuron could become synchronized with the variations in the input of the subthreshold signaling neuron(s) through the mechanism of SR, with the noise “amplifying” the signal.

Although a simple threshold detector can exhibit SR, it still lacks one other important property of neurons that is possessed by the next-most-simple model neuron, the integrate-and-fire neuron (e.g., [2]). This model integrates its inputs over a moving time window τ . In real neurons, τ varies from about 100 ms for sensory neurons, to 40–60 ms for inputs to the soma of pyramidal interneurons, and can be as short as 10 ms for inputs to the apical tufts of dendrites of pyramidal neurons in the prefrontal cortex [11]. Larger values of τ make neurons act as integrators while smaller values such as those at the apical tufts, make them act as coincidence detectors [12]. The value of τ is set by the time courses of various processes that affect the electrochemical state of the neuron. At any time there is a complex balance of electrochemical forces influenced by synaptic and internal events with specific decay rates, plus the constant diffusion of ions caused by electrical and concentration gradients and active pumps. The time-varying instantaneous firing rate of the neuron ($1000/\Delta t$, where Δt is the interval between two successive action potentials in ms) reflects the momentary strengths of all of these forces, which are represented in integrate-and-fire models by the exponentially decaying influence of previous inputs.

The statistical technique of kernel regression can be regarded as an approach to formally modeling temporal integration of input. In this approach, a parameter is estimated from a set of measurements using a weighting function defined over the set. For example, the probability of a neuron

*Electronic address: uschi@math.uni-bremen.de; URL: <http://www.math.uni-bremen.de/~uschi/>

[†]Electronic address: lward@cortex.psych.ubc.ca; URL: <http://www.psych.ubc.ca/~lward/personnel/ward/>

firing at a particular moment could be estimated from a weighted regression on previous firings. The weighting function is called the “kernel” and corresponds to a filter, or a smoothing function, in physical applications. Such regressions are done routinely, albeit informally, in physics when a smoothing function, such as a Gaussian distribution, is convolved with a noisy signal to obtain time-averaged behavior. In fact there exists an extensive statistical literature on smoothing and signal processing. Müller [13] adapted statistical smoothing methods from this literature to obtain a general and asymptotically optimal procedure for estimating a signal from thresholded data. The asymptotic mean-squared error criterion adopted there is independent of sample size. This criterion avoids a difficulty of correlation-type criteria commonly used, viz., correlations become uniformly high at large sample sizes (see [13] and Sec. II). Previous studies of numerical smoothing of model excitable system outputs with a box-type kernel or of thresholded data with a Gaussian kernel were done by the authors of [2] and [14], respectively. However, both previous studies computed the correlation between smoothed system outputs and signals rather than applying the mean-squared error criterion to estimates of the signal, as we do here.

In this paper we study the approach of Müller [13], which represents a formal theory of the computational problem faced by neurons and other threshold detector systems attempting to extract information from a subthreshold signal. The formal theory provides boundary conditions on possible neuronal computations, rather than describing in detail the way in which neurons actually compute estimates of subthreshold signals. In Sec. I we briefly describe the theory, emphasizing the normal error distribution with variance σ^2 . In Sec. II we provide a data-driven multistage estimation procedure, which we illustrate for some typical signals. Simulations show how the quality of the signal estimation depends on the variance of the error, or noise, distribution.

In Sec. III we discuss how the noise variance affects the mean-average-squared error of estimation, which is also a function of the signal and its first two derivatives. For some examples we determine the variance σ^2 for which the error of estimation is minimal, i.e., the stochastic resonance point.

Since the optimal noise level depends on the signal and its derivatives, the question arises whether there is a choice of σ^2 that is robust in the sense that the estimator for the signal behaves well for a certain range of first and second derivatives. In Sec. IV we show that this is the case.

In Sec. V we compare our estimator with classical kernel regression, which does not use thresholded data but full observations of signal plus noise.

Greenwood *et al.* [15] constructed efficient estimators for constant signal and given error variance. They calculated the stochastic resonance point, the σ^2 for which the asymptotic variance of the estimator is minimal. They argued that their result remains approximately valid for smooth nonconstant signals. However, the asymptotic mean-squared error of our estimator for nonconstant signals contains an additional bias term that is not always negligible. In Sec. VI we study how the stochastic resonance point is affected by this bias term.

II. MODEL AND ESTIMATION

We describe the typical noisy, subthreshold, signal treated in studies of stochastic resonance by what, in statistical

theory, is called a nonparametric regression model

$$Y(t_i) = s(t_i) + \epsilon(t_i), \quad i = 1, \dots, n,$$

with equally spaced time points $t_i = i/n$ on $[0, 1]$, signal $s(t_i)$, and noise variables (called error variables in statistical theory) $\epsilon(t_i)$ that are assumed to be independent and identically distributed (i.i.d.) with mean zero and continuous distribution function F (the distribution function is the integral of the probability density function regarded as a function of the upper limit). Let a be a threshold that is larger than the maximum of the regression function (signal) s over the unit interval. A detector records the times at which the $Y(t_i)$ exceed a . These exceedances can be coded as Bernoulli variables

$$X(t_i) = \mathbf{1}(Y(t_i) > a), \quad i = 1, \dots, n,$$

with distribution parameters

$$p(t_i) = P(Y(t_i) > a) = 1 - F(a - s(t_i)).$$

Hence the signal can be recovered from the just defined probabilities as follows:

$$s(t_i) = a - F^{-1}(1 - p(t_i)).$$

When the noise distribution F is known, we can reconstruct the regression function (signal) from the threshold exceedances. Inserting a kernel estimator $\hat{p}_h(t)$ for p in the above transformation, we can, in particular, estimate it by

$$\hat{s}_h(t) = a - F^{-1}(1 - \hat{p}_h(t)).$$

The estimator $\hat{s}_h(t)$ is consistent, meaning that the probability of its being within an arbitrary distance of the actual signal approaches 1 as the sample size on which the estimate is based approaches infinity.

In the case of a neuron, the noise distribution has been considered to be either Gaussian or Poisson (see [16] for a discussion). In the case of a physical signal, it is usually possible to make a theoretically based assumption as to the noise distribution, or to approximate it empirically. In many cases the Gaussian distribution function will be a reasonable approximation to F . Another possible technique for the case of unknown F is to employ estimates for several different types of F and look for similarities in the form of the recovered signal across distribution types. An estimate \hat{s}_h based on an incorrectly specified F will still recover the form of the signal up to a scale transformation.

If the variance of the noise distribution is small, the realizations $Y(t)$ will rarely exceed the threshold, and the signal will be hard to estimate. If the noise variance is very large, the exceedance probability $p(t)$ will be estimated with a large variance, and the estimator for the regression function will again be imprecise. This suggests that there is a, not necessarily unique, noise variance for which the estimator of the signal has a minimal asymptotic mean-squared error. We call such a variance a stochastic resonance point.

For our investigations we used the Nadaraya-Watson kernel estimator

$$\hat{p}_h(t) = \frac{\sum_{i=1}^n \frac{1}{h} K\left(\frac{t-t_i}{h}\right) X(t_i)}{\sum_{i=1}^n \frac{1}{h} K\left(\frac{t-t_i}{h}\right)}, \tag{2.1}$$

where K denotes the kernel weighting function and h denotes the set of sample values to which the kernel is applied, called its bandwidth. The Nadaraya-Watson estimator is a common kernel estimator used in statistics for many kinds of data, including binary data such as our exceedances. The bandwidth of the kernel can be considered to represent the window of temporal integration when applying the method to the case of a model neuron. For the estimation of $\hat{s}_h(t)$ at time points $t \in [h, 1-h] \subset [0, 1]$, K is a second-order kernel function, i.e., $\int K(u) du = 1$, $\int u K(u) du = 0$, $\int u^2 K(u) du \neq 0$, with bounded derivative and support $[-1, 1]$. Formally $\hat{p}_h(t)$ has to be set to a constant if it takes values close to 0 or 1. For technical details refer to Müller [13], who introduced this approach and where also a discussion of the assumptions can be found.

We consider two possible bandwidths, a local bandwidth that can vary for each time point, and a global bandwidth that is constant for all time points. Our criteria for the choice of optimal local and global bandwidths are the asymptotic mean-squared and mean-average-squared error, AMSE and AMASE. For the asymptotics $n \rightarrow \infty, h = h_n \rightarrow 0$ and $nh^3 \rightarrow \infty$ they are approximations of $E(\hat{s}_h(t) - s(t))^2$ and $1/n \sum_{t \in T} E(\hat{s}_h(t) - s(t))^2$. Here $T \subset (0, 1)$ denotes some subinterval of $(0, 1)$ to which, because of boundary effects, summation is usually restricted. The formulas given in [13] are

$$\Psi^{\text{AMSE}}(h, t) = \frac{1}{f p [F^{-1}(p(t))]^2} \left(\frac{1}{nh} R(K) p(t) (1 - p(t)) + \frac{h^4}{4} \mu_2(K)^2 p''(t)^2 \right), \tag{2.2}$$

$$\Theta^{\text{AMASE}}(h) = \frac{1}{nh} R(K) \frac{1}{n} \sum_{t \in T} \frac{1}{f [F^{-1}(p(t))]^2} p(t) [1 - p(t)] + \frac{h^4}{4} \mu_2(K)^2 \frac{1}{n} \sum_{t \in T} \frac{1}{f [F^{-1}(p(t))]^2} p''(t)^2 \tag{2.3}$$

with kernel constants $R(K) = \int K^2(u) du$ and $\mu_2(K) = \int u^2 K(u) du$. The asymptotically optimal local and global bandwidths derived from these approximations are

$$h_{opt}(t) = n^{-1/5} \left(\frac{R(K) p(t) (1 - p(t))}{\mu_2(K)^2 p''(t)^2} \right)^{1/5}, \tag{2.4}$$

$$h_{opt} = n^{-1/5} \left(\frac{R(K) \sum_{t \in T} \frac{1}{f [F^{-1}(p(t))]^2} p(t) [1 - p(t)]}{\mu_2(K)^2 \sum_{t \in T} \frac{1}{f [F^{-1}(p(t))]^2} p''(t)^2} \right)^{1/5}. \tag{2.5}$$

In this study we used the Epanechnikov kernel $K(u) = 3/4(1 - u^2)\mathbf{1}_{[-1, 1]}(u)$, which minimizes, in the class of kernels considered, AM(A)SE with the optimal bandwidth inserted. For this kernel we have $R(K) = 3/5$ and $\mu_2(K) = 1/5$. Although this choice of kernel is optimal, other kernels would do nearly as well, for example, the Gaussian kernel. An approach using the box or even asymmetrical kernels is also conceivable, although it remains to be worked out. In particular, an asymmetrical kernel would be more appropriate to model the time-integration behavior of actual neurons, since presumably they do not know the future. We can speculate, however, on the basis of some pilot work, that as long as the kernel contains sufficient time points and weights them according to a reasonable function, considerable information about the behavior of the subthreshold signal can be recovered from the exceedances alone.

The bandwidth formulas (2.4) and (2.5) are both of the form $n^{-1/5}$ times a constant c , where c depends on the probability function and its second derivative, i.e., on the unknown signal. To guarantee consistency it is sufficient that h has the order of magnitude $n^{-1/5}$. For good performance in the finite sample situation the constant c should, however, be chosen suitably. For our demonstrations we used plug-in methods, i.e., we estimated the unknown functions and plugged them into the formulas above. This approach has, in the classical setting with fully observed data, been shown to perform more reliably than standard methods such as cross validation (see, for example, [17]).

In this paper we interpret $[0, 1]$ as a time window that shows a representative part of a signal. The local bandwidth approach was discarded since it is known to be too sensitive, even in the simpler setting with full observations (e.g., [18]), and also proved to be so in our preliminary investigations in the present setting. We used the global bandwidth to estimate the signal at all time points in $[0, 1]$ extending the kernel as far as necessary outside that window. This was done by calculating thresholded signal plus noise observations for the extended time interval $[-0.5, 1.5]$. The additional data were used whenever the kernel overlapped the boundaries 0 and 1.

The global bandwidth formula (2.5) contains unknown quantities that must be estimated in a preliminary way to get started. For these pilot estimates we adapted a method from Ruppert *et al.* [19] who investigated plug-in strategies for the classical setting. The probability function p and its second derivative p'' were estimated with preliminary smoothers with asymptotically optimal bandwidths derived analogously to the actual estimator. The unknown quantities appearing in the new bandwidth formulas were estimated using a modification of a ‘‘blocking method’’ developed by Härdle and Marron [20].

For the preliminary estimation of p we used an estimator simpler than the Nadaraya-Watson estimator, namely,

$$\hat{p}_\lambda(t) = \frac{1}{n\lambda} \sum_{i=1}^n K\left(\frac{t_i - t}{\lambda}\right) X(t_i), \tag{2.6}$$

which is the Priestley-Chao estimator for the special case of equidistant time points. The bandwidth is denoted by λ and K is again the Epanechnikov kernel. Since $\hat{p}_\lambda(t)$ and the final probability estimator $\hat{p}_h(t)$ have the same asymptotic behav-

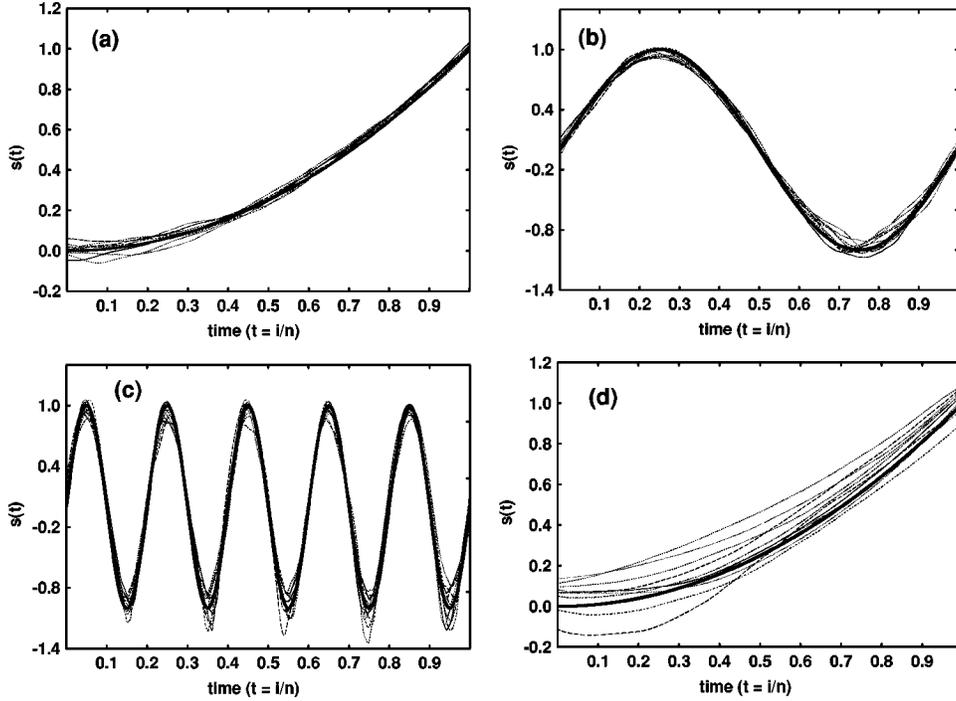


FIG. 1. Ten realizations of the estimation procedure for (a) $s_1(t) = t^2$ with $\sigma = 0.59$, average estimated bandwidth $h = 0.16$, and theoretically optimal bandwidth [Eq. (2.5)] $h_{opt} = 0.16$; (b) $s_2(t) = \sin(2\pi t)$ with $\sigma = 1.08$, average bandwidth $h = 0.09$, and $h_{opt} = 0.08$; (c) $s_3(t) = \sin(10\pi t)$ with $\sigma = 1.08$, average bandwidth $h = 0.03$ and $h_{opt} = 0.02$; and (d) $s_1(t) = t^2$ as in (a) but with $\sigma = 4.0$. The signal and estimated signals are shown by thick continuous and thin broken lines, respectively.

ior, the AMASE expressions of $\hat{p}_\lambda(t)$ and $\hat{s}_h(t) = a - F^{-1}(\hat{p}_h(t))$ coincide up to the density weights coming in through the transformation. In particular, the terms involving the kernel constants are the same in both error formulas, which explains the choice of the Epanechnikov kernel K . The optimal bandwidth for the preliminary smoother $\hat{p}_\lambda(t)$ is

$$\lambda_{opt} = n^{-1/5} \left(\frac{R(K) \sum_{t \in T} p(t)(1-p(t))}{\mu_2(K)^2 \sum_{t \in T} p''(t)^2} \right)^{1/5}.$$

A natural estimator for the second unknown quantity in (2.5), p'' , is

$$\hat{p}_g''(t) = \frac{1}{ng^3} \sum_{i=1}^n \tilde{K} \left(\frac{t_i - t}{g} \right) X(t_i) \quad (2.7)$$

with the kernel \tilde{K} satisfying the moment conditions $\mu_i(\tilde{K}) = \int u^i \tilde{K}(u) du = 0$ for $i = 0, 1, 3$, $\mu_2(\tilde{K}) = 2$ and $\mu_4(\tilde{K}) = 0$, in order to guarantee consistency. We used the kernel

$$\tilde{K}(u) = \left(-\frac{105}{16} + \frac{630}{16}u^2 - \frac{525}{16}u^4 \right) \mathbf{1}_{[-1,1]}(u),$$

which is optimal (with respect to the mean-squared error) in the sense of Müller's theory ([21], p. 52 ff), i.e., it meets the moment condition and has no more sign changes than necessary. The optimal bandwidth derived from the mean-average-squared error of \hat{p}_g'' is

$$g_{opt} = n^{-1/9} \left(\frac{R(\tilde{K}) \sum_{t \in T} p(t)(1-p(t))}{\mu_4(\tilde{K})^2 \sum_{t \in T} p^{(4)}(t)^2} \right)^{1/9}$$

with kernel constants $R(\tilde{K}) = 35$ and $\mu_4(\tilde{K}) = 4/3$.

With (2.6) and (2.7) the necessary formulas for the preliminary smoothing are provided. However, the formulas for the optimal bandwidths λ_{opt} and g_{opt} still depend on p and its derivatives p'' and $p^{(4)}$. These quantities were, in an initial step, determined by a simple procedure. We used a variant of a ‘‘blocking method’’ developed by Härdle and Marron [20] but with the number of blocks chosen by Mallows' C_p criterion [22] as suggested by Ruppert *et al.* [19]. For this method the interval $[0,1]$ is divided into $N \in \{1, \dots, N_{max}\}$ equally sized blocks. For our examples we chose $N_{max} = 10$, which is suitably small to reduce the chance of overfitting and, what is more important here, to ensure that there are enough exceedances in every block so that the estimator exists. In contrast to Ruppert *et al.* who fit quartics, we fitted a logistic curve to each block, i.e., a probability function $p(t) = e^{\alpha + \beta t} / (1 + e^{\alpha + \beta t})$. The combination of the N logistic fits then represents an initial estimate of p . The derivatives are derived from this fit. Since we wanted to estimate probabilities, the logistic regression has the advantage of producing estimates between 0 and 1. Moreover, the fourth derivatives needed for g_{opt} are nonzero. Our rule for choosing an optimal number of blocks, and hence the optimal logistic fit, was the same as in [19]: We took the $\hat{N} \in \{1, \dots, N_{max}\}$ that minimized Mallows' C_p criterion,

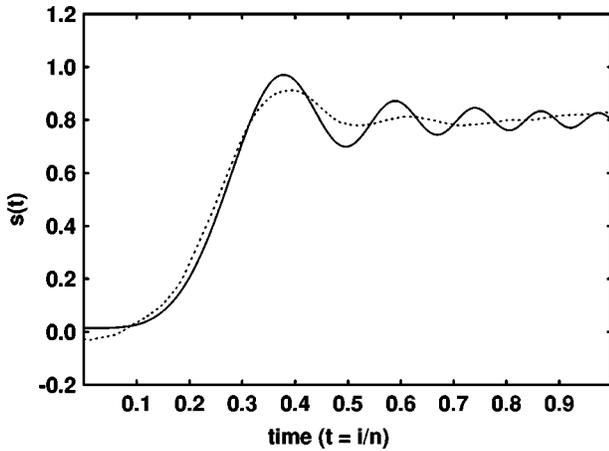


FIG. 2. Signal estimation of $s_4(t) = 0.8 - 0.025 \sin(10\pi t^2)/t^2$ with $\sigma = 0.76$ and estimated bandwidth $h = 0.10$. Signal and estimated signals are shown by continuous and broken lines, respectively.

$$C_p = \frac{R(N)}{R(N_{max})/(n - 5N_{max})} - (n - 10N).$$

Here $R(N)$ stands for the residual sum of squares based on logistic fits over N blocks. The initial estimate of p needed for λ_{opt} and g_{opt} is the logistic fit over \hat{N} blocks.

Although the procedure is applicable to all signals of arbitrary shape, we demonstrate it here for three characteristic examples $s_1(t) = t^2$, $s_2(t) = \sin(2\pi t)$, and $s_3(t) = \sin(10\pi t)$ [see Figs. 1(a)–1(c)] for optimal noise levels from a normal error distribution (see Sec. III) and $n = 10\,000$ time points. In the examples we used threshold $a = 1$ throughout.

These examples illustrate several properties of the estimation technique. One is that estimation is noticeably worse for parts of the signals that are far from threshold than for parts near threshold. This is because the bandwidth used was globally optimum, and could not give equally good performance for all parts of a signal whose distance from the threshold varies widely. Another property of the technique is that estimation is worse for parts of signals where the second derivative is large. Finally, and not shown in Fig. 1, estimation is better the less high frequency content in the signal. This is reflected in the estimated root-mean-squared errors of estimation of the ten realizations of each signal type. The average values are 0.02, 0.05, and 0.10 for signals $s_1(t)$, $s_2(t)$, and $s_3(t)$, respectively. Finally, as expected for large n , a normalized correlation measure, the Pearson product-moment correlation coefficient between signal and estimated signal, is uniformly high for each type of signal: it was 0.99 for every realization of each of the three signal types (see [13] for a discussion).

We also considered a signal $s_4(t) = 0.8 - 0.025 \sin(10\pi t^2)/t^2$. A single realization of this signal is shown in Fig. 2. The distance of the signal from the threshold varies widely, and the signal also contains considerable high-frequency information, albeit only in the part nearest the threshold. In a sense this is a mixture of several types of signals, although it does resemble some that might be of biological importance, for example, that of an approaching predator who ‘‘holds’’ at an attack launch point. The estima-

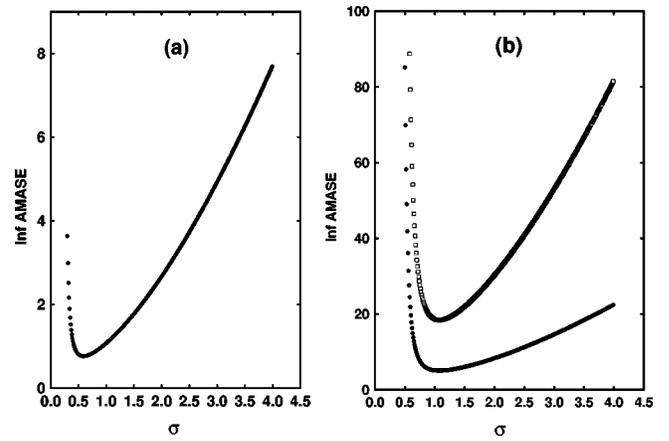


FIG. 3. Plot of $n^{4/5} \inf_h \Theta^{AMASE}(h)$ vs σ for (a) $s_1(t) = t^2$, (b) $s_2(t) = \sin(2\pi t)$ (dots), and $s_3(t) = \sin(10\pi t)$ (open squares).

tion technique does a good job of capturing the low-frequency content, but fails to accurately estimate the signal at the high-frequency end. Again, this is because we used a global optimum bandwidth 0.10, which is near that of $s_2(t)$ and does well for the components near 1 Hz. The bandwidth needs to be narrower to accurately capture the higher frequency variation, although, of course, a higher sample rate would also be needed to provide equal accuracy with the narrower bandwidth. It might be possible to use the local bandwidth in such cases, providing an upper limit, such as the global bandwidth, is used to avoid degeneracy when $p''(t)$ becomes small. This result might point to some limitations of neurons attempting to extract information from naturally occurring subthreshold signals. If a globally optimal bandwidth is used (optimized by evolution), then only certain critical aspects of those signals, such as increases or decreases in amplitude indicating approach or avoidance, can be accurately estimated. Other aspects, such as higher frequency signatures of different causative agents, will be unavailable with any precision, leading to false alarm reactions.

III. STOCHASTIC RESONANCE

In Sec. II we described technical details of an estimation procedure based on kernel methods without mentioning stochastic resonance. Good performance of the procedure is connected to a certain optimality of the noise levels chosen—the stochastic resonance effect. This section discusses the stochastic resonance effect in the context of the approach described in Sec. II.

Stochastic resonance as shown in the literature would entail that our error criteria, i.e., $\Psi^{AMSE}(h, t)$ and $\Theta^{AMASE}(h)$, regarded as functions of the noise, each have a minimum. In the following we will consider only the global approximation $\Theta^{AMASE}(h)$ since, as mentioned before, the local approach based on $\Psi^{AMSE}(h, t)$ [Eq. (2.2)] is known to perform badly. It fails completely if $p''(t) = 0$ and does not work well in practice if $p''(t)$ is close to zero [see formula (2.4) and Sec. VI for more details]. Hence, a closer investigation does not seem to be warranted. Nevertheless, provided that additional information is available, AMSE can be used to determine an optimal σ in a certain minimax sense, which will be demonstrated in Sec. IV.

Consider the global criterion $\Theta^{\text{AMASE}}(h)$ [Eq. (2.3)]. With the optimal bandwidth h_{opt} [Eq. (2.5)] inserted it can be rewritten in terms of the signal function s :

$$\inf_{h>0} \Theta^{\text{AMASE}}(h) = \frac{5}{4} n^{-4/5} \left[\mu_2(K)^2 \frac{1}{n} \sum_{t \in T} \left(\frac{-f'(a-s(t))s'(t)^2 + f(a-s(t))s''(t)}{f(a-s(t))} \right)^2 \right]^{1/5} \\ \times \left(R(K) \frac{1}{n} \sum_{t \in T} \frac{F(a-s(t))[1-F(a-s(t))]}{f^2(a-s(t))} \right)^{4/5}. \quad (3.1)$$

This expression is not helpful for detecting stochastic resonance, particularly not since a restriction to error distributions F with convenient properties such as unimodality seems to be necessary. Restricting attention to the normal $N(0, \sigma^2)$ error distribution and using the Epanechnikov kernel K from Sec. II, we obtain

$$\inf_{h>0} \Theta^{\text{AMASE}}(h) = \frac{5}{4} n^{-4/5} \left[\frac{1}{25n} \sum_{t \in T} \left(\frac{a-s(t)}{\sigma^2} s'(t)^2 + s''(t) \right)^2 \right]^{1/5} \left(\frac{3}{5n} \sum_{t \in T} \frac{\sigma^2 \Phi\left(\frac{s(t)-a}{\sigma}\right) \Phi\left(\frac{a-s(t)}{\sigma}\right)}{\phi^2\left(\frac{s(t)-a}{\sigma}\right)} \right)^{4/5} \quad (3.2)$$

with $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and $\Phi(x) = \int_{-\infty}^x \phi(y) dy$ denoting the density and distribution functions of the standard normal $N(0,1)$ distribution. This formula is more informative: The first term of the product varies like $\sigma^{-4/5}$, which is monotonic in σ . Since stochastic resonance curves are not monotonic, if stochastic resonance emerges, then it is because of the second term (see also Sec. VI).

A noise level σ that is optimal with respect to AMASE is theoretical since it depends on the signal function s . An explicit general formula can obviously not be given. Instead, we compute as examples the optimal σ 's for the signals considered in Sec. II by minimizing AMASE numerically. In Sec. IV we present an approach to determining a robust stochastic resonance point without using the signal function explicitly.

In Fig. 3(a) we plotted AMASE as given above, multiplied by the convergence rate $n^{4/5}$ for $s_1(t) = t^2$. For our

calculations we chose a threshold $a = 1$ and $n = 10\,000$. Since the sums in the formula approximate integrals, the curve is the same for all n sufficiently large. In particular, it has a sharp minimum at the stochastic resonance point $\sigma = 0.59$. Analogously, we obtained the optimal noise level $\sigma = 1.08$ for both signals s_2 and s_3 from Sec. II [Fig. 3(b)], which we already used for our simulated estimation examples there.

The stochastic resonance effect should, of course, occur not only in reference to the theoretical function AMASE. In addition the noise level should strongly influence the quality of the estimate obtained using the procedure of Sec. II. To demonstrate this, we considered again signal s_1 from Sec. II. We estimated the signal for ten additional realizations of signal plus noise but this time we used the theoretically bad value $\sigma = 4.0$ [see Fig. 3(a)]. These are plotted in Fig. 1(d) [compare with Fig. 1(a)]. As expected, the estimates are significantly worse. In particular, a large increase of the variance over realizations becomes evident.

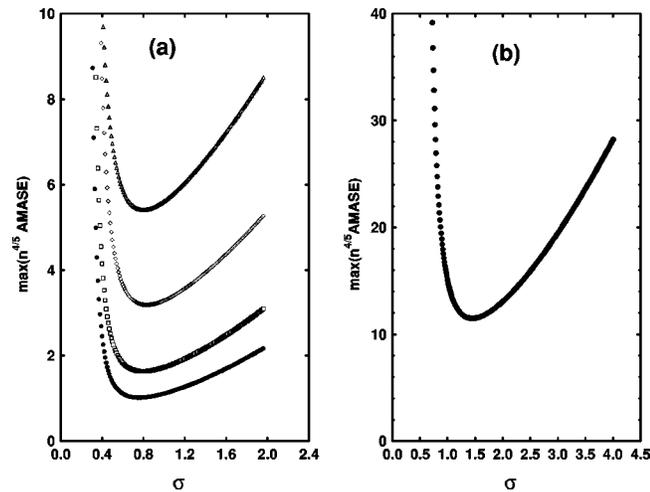


FIG. 4. (a) The local asymptotic mean-squared-error maximized over s' and s'' both having the range $[-1,1]$ (dots), $[-2,2]$ (open squares), $[-5,5]$ (open diamonds), and $[-10,10]$ (open triangles), respectively. (b) Determination of the robust noise level $\sigma = 1.45$ for the ranges of $s_2(t) = \sin(2\pi t)$ and its first and second derivative.

IV. ROBUST ESTIMATION

As seen in Sec. III, a noise level that is optimal with respect to the mean-squared error of estimation depends on

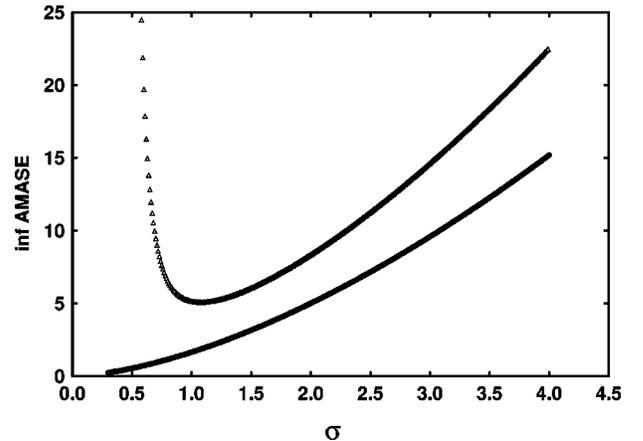


FIG. 5. Plot of $n^{4/5} \inf_h \Theta^{\text{AMASE}}(h)$ for fully observed (dots) and thresholded (open triangles) data for $s_2(t) = \sin(2\pi t)$.

the signal and its derivatives and cannot be determined without prior knowledge. In this section we discuss a noise level σ_r that is robust in the sense that the estimator for the regression function behaves well for a given range of the signal and its first and second derivatives. It is possible that neural systems with noise in this robust range, being better able to detect a wide range of important subthreshold signals via stochastic resonance, would be selected by evolution and would occur in living brains.

We determined a value σ_r with the following local minimax property: For fixed $s(t)$, say $s(t)=0$, the mean-squared error at t maximized over a certain range of $s'(t)$ and $s''(t)$ is minimized in $\sigma=\sigma_r$. For this purpose we considered $\Psi^{\text{AMSE}}(h,t)$ with the optimal bandwidth inserted, i.e., $\Psi^{\text{AMSE}}(h_{\text{opt}}(t),t)=\inf_h \Psi^{\text{AMSE}}(h,t)$, now regarded as a function of $s'(t)$, $s''(t)$, and σ . Since t and $s(t)$ are fixed and the optimal bandwidth is used we suppress the dependence on t and write briefly $\Psi^{\text{AMSE}}(s',s'',\sigma)$. With this notation the robust σ is

$$\begin{aligned} \sigma_r &= \operatorname{argmin}_{\sigma} \max_{\substack{s' \in [c_1, d_1] \\ s'' \in [c_2, d_2]}} \Psi^{\text{AMSE}}(s',s'',\sigma) \\ &= \operatorname{argmin}_{\sigma} \max_{\substack{s' \in [c_1, d_1] \\ s'' \in [c_2, d_2]}} \frac{5}{4} n^{-4/5} \left[\mu_2(K)^2 \left(\frac{a-s}{\sigma^2} s'^2 + s'' \right)^2 \right]^{1/5} \left(\frac{\sigma^2 \Phi\left(\frac{s-a}{\sigma}\right) \Phi\left(\frac{a-s}{\sigma}\right)}{\phi^2\left(\frac{s-a}{\sigma}\right)} R(K) \right)^{4/5} \\ &= \operatorname{argmin}_{\sigma} \max_{\substack{s' \in [c_1, d_1] \\ s'' \in [c_2, d_2]}} \frac{5}{4} n^{-4/5} \left[\left(\frac{1}{5}\right)^2 \left(\frac{1}{\sigma^2} s'^2 + s'' \right)^2 \right]^{1/5} \left(\frac{3}{5} \frac{\sigma^2 \Phi\left(\frac{-1}{\sigma}\right) \Phi\left(\frac{1}{\sigma}\right)}{\phi^2\left(\frac{-1}{\sigma}\right)} \right)^{4/5}, \end{aligned}$$

where we used normal error, $s(t)=s=0$, $a=1$, and the Epanechnikov kernel K with kernel constants $\mu_2(K)=1/5$ and $R(K)=3/5$. In Fig. 4(a) we have plotted $\max_{s',s''} \Psi^{\text{AMSE}}$ multiplied by the convergence rate $n^{4/5}$ as a function of σ for s' and s'' both in the range $[-1,1]$, $[-2,2]$, $[-5,5]$, and $[-10,10]$, respectively. For these intervals we obtained the robust noise levels 0.76, 0.79, 0.83, and 0.85, respectively.

If restrictions on s' and s'' can be assumed to hold uniformly for $t \in [0,1]$ one can, analogously, derive a global minimax σ , provided a lower bound for the signal, say c_0 , is given. Additionally, an argument $s \in [c_0,1]$ maximizing

$$\begin{aligned} & \max_{\substack{s' \in [c_1, d_1] \\ s'' \in [c_2, d_2]}} \left[\left(\frac{1-s}{\sigma^2} s'^2 + s'' \right)^2 \right]^{1/5} \\ & \times \left(\frac{\Phi\left(\frac{s-1}{\sigma}\right) \Phi\left(\frac{1-s}{\sigma}\right)}{\phi^2\left(\frac{s-1}{\sigma}\right)} \right)^{4/5} \end{aligned}$$

has to be determined. The resulting $\max_{s,s',s''} \Psi^{\text{AMSE}}$ then will, analogously, be minimized with respect to σ . As an illustration we derived such a global minimax σ under the constraints $s \in [-1,1]$, $s' \in [-6.28, 6.28]$, $s'' \in [-39.48, 39.48]$. These are the ranges of $s_2(t)=\sin(2\pi t)$ and its derivatives. For this example [see Fig. 4(b)] we obtained a robust noise level $\sigma=1.45$ that is larger than the optimal one, $\sigma=1.08$ [see Sec. III, Fig. 3(b)]. This phenomenon is ex-

plained by the fact that a large class of signals fits into the given range, especially the constant signal $s \equiv -1$, which clearly needs more noise in order to guarantee sufficiently many threshold crossings.

V. COMPARISON WITH FULLY OBSERVED DATA

In the classical setting we have fully observed data $Y(t) = s(t) + \epsilon(t)$, corresponding to above-threshold signals. For such signals the relevant signal estimator would be the kernel estimator from Eq. (2.1) with the observations $Y(t)$ instead of $X(t)$ inserted. In general we expect that this estimator should perform better than our estimator for the threshold detector that only uses indicators. The minimum value of the asymptotic mean-average-squared error of the estimator in the full data situation is known to be

$$\frac{5}{4} n^{-4/5} \left(\mu_2(K)^2 \frac{1}{n} \sum_{t_i \in T} s''(t_i)^2 \right)^{1/5} [R(K)\sigma^2]^{4/5}. \quad (5.1)$$

Here K is again the Epanechnikov kernel since it has analogous optimality properties in this setting.

In order to compare the two estimators for the same signal we considered the simple sinusoid $s_2(t)=\sin(2\pi t)$. In Fig. 5 we have plotted $n^{4/5} \inf \Theta^{\text{AMASE}}$ for fully observed and for threshold data ($a=1, n=10000$). For this example, the classical estimator using full observations is clearly superior to our estimator using threshold data, especially when σ is small, i.e., it should produce better estimates. The same applies to the other examples of this paper.

Theoretically, in terms of $\inf_h \Theta^{\text{AMASE}}(h)$, the estimator based on full observations is better if

$$\left(\sum_{t \in T} s''(t)^2 \right)^{1/5} (\sigma^2)^{4/5} < \left[\sum_{t \in T} \left(\frac{-f'(a-s(t))s'(t)^2 + f(a-s(t))s''(t)}{f(a-s(t))} \right)^2 \right]^{1/5} \left(\frac{1}{n} \sum_{t \in T} \frac{F(a-s(t))[1-F(a-s(t))]}{f^2(a-s(t))} \right)^{4/5}$$

[see (3.1) and (5.1)]. If the noise is identically $N(0, \sigma^2)$ distributed, the formula can equivalently be rewritten as follows (ϕ and Φ denote the standard normal density and distribution function, respectively):

$$\sum_{t \in T} s''(t)^2 < \sum_{t \in T} \left(s'(t)^2 \frac{a-s(t)}{\sigma^2} + s''(t) \right)^2 \left[\frac{1}{n} \sum_{t \in T} \frac{\Phi\left(\frac{s(t)-a}{\sigma}\right)\Phi\left(\frac{a-s(t)}{\sigma}\right)}{\phi^2\left(\frac{s(t)-a}{\sigma}\right)} \right]^4.$$

The relation does not hold uniformly for all values of s and its derivatives, particularly not if $s''(t) < 0$, which occurs, for example, near points of inflection for a wide class of signals. Hence, we cannot make a general statement favoring one estimator over the other. Although this seems to be surprising, we can explain it by the fact that the estimators are not really comparable: Our estimator for threshold data uses the distribution function of the noise F , whereas the classical estimator does not profit by this additional information. More importantly, in the situation of a neuron detecting a signal, our inability to assert that the classical estimator is uniformly superior does not imply that sometimes a subthreshold signal might be estimated better than a superthreshold signal. In most situations we would expect the result to be similar to the one displayed in Fig. 5.

VI. CONSTANT SIGNAL

Finally, consider the situation when the signal is constant $s(t) = s$, which was studied elaborately by Greenwood *et al.* [15]. In this case our approach based on kernel methods breaks down, especially when we assume i.i.d. noise. Then the exceedance probabilities are also constant, $p(t) = p$ for all t , and the second derivatives of both functions, s and p , are zero. Under these conditions our global criterion coincides with the local one:

$$\begin{aligned} \Theta^{\text{AMASE}}(h) &= \frac{1}{nh} R(K) \frac{1}{n} \sum_{t \in T} \frac{1}{f(a-s(t))^2} p(t)(1-p(t)) \\ &\quad + \frac{h^4}{4} \mu_2(K)^2 \frac{1}{n} \sum_{t \in T} \frac{1}{f(a-s(t))^2} p''(t)^2 \\ &= \frac{1}{nh} R(K) \frac{1}{f(a-s)^2} F(a-s)(1-F(a-s)) \\ &= \Psi^{\text{AMSE}}(h, t) \end{aligned} \quad (6.1)$$

for all t . A bandwidth formula applicable for a kernel estimator cannot be written because the optimal bandwidth with respect to (6.1) would be infinitely large. Although our method turns out to be inappropriate for a constant signal, our findings do accord well with the results of Greenwood *et al.*: They suggested the mean average of the exceedance

data $\bar{X} = 1/n \sum_{i=1}^n X(t_i)$ as an estimator for the probability p , which is the (efficient) maximum likelihood estimator. In particular, it corresponds to a kernel estimator with infinite bandwidth.

For their investigations of stochastic resonance Greenwood *et al.* [15] considered the asymptotic variance of the estimator $\hat{s} = a - F^{-1}(1 - \bar{X})$, which is the inverse Fisher information and, up to the factor $R(K)/h$, coincides with our formula (6.1) for AM(A)SE in the constant signal situation.

Greenwood *et al.* argued that their approach is valid not only for constant signals but also for smooth non-constant signals. This is only the case if the bias term involving p'' in (6.1) is negligible, which is not always true. If we want to determine the stochastic resonance point, which we will illustrate later, this point becomes especially important.

To explore the differences, first consider $\Psi^{\text{AMSE}}(h, t)$. Let the noise be normally distributed and the signal be not necessarily constant. Then

$$\begin{aligned} \Psi^{\text{AMSE}}(h, t) &= \frac{1}{nh} \frac{\sigma^2 \Phi\left(\frac{s(t)-a}{\sigma}\right)\Phi\left(\frac{a-s(t)}{\sigma}\right)}{\phi^2\left(\frac{s(t)-a}{\sigma}\right)} R(K) \\ &\quad + \frac{h^4}{4} \left(\frac{a-s(t)}{\sigma^2} s'(t)^2 + s''(t) \right)^2 \mu_2(K)^2. \end{aligned}$$

The first term corresponds to the inverse Fisher information considered by Greenwood *et al.* [15]. They showed that it exhibits typical stochastic resonance behavior. However, the second (bias) term will usually be nonzero and, regarded as a function of σ , not necessarily have a minimum. In general AMSE will exhibit stochastic resonance. For certain constellations of values, however, for example, if the second term dominates the first, this will not be the case.

More interesting for comparison to our estimation procedure is probably the asymptotic mean-average-squared error [Eq. (3.2)] of the signal estimator using the optimal bandwidth. We used this to derive the stochastic resonance points for some characteristic signals and to compare them with the optimal noise levels computed by Greenwood *et al.* for a constant signal. For $s \equiv 0$ and threshold $a = 1$ they determined an optimal noise level $\sigma = 0.64$. Following their rea-

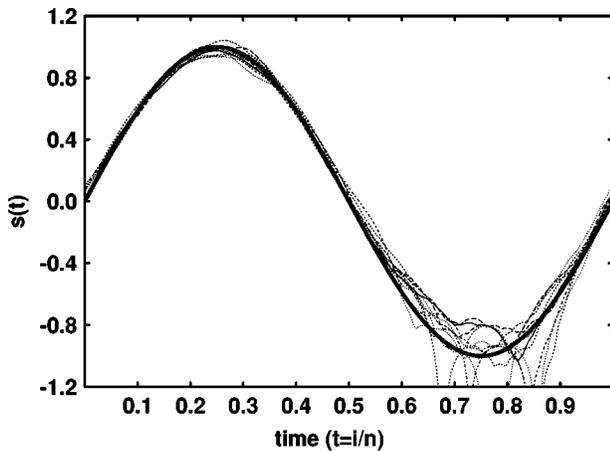


FIG. 6. Ten realizations of the estimation procedure for $s_2(t) = \sin(2\pi t)$ with $\sigma = 0.64$. The average bandwidth used was $h = 0.06$, the theoretically optimal bandwidth [Eq. 2.5] is $h_{opt} = 0.07$. Signal and estimated signals are shown by thick continuous and thin broken lines, respectively.

soning, this noise level should guarantee good performance of estimates of smooth nonconstant signals varying around zero. To check this conjecture, we computed $\inf_h \Theta^{AMASE}$ for three sinusoids $s(t) = A \sin(2\pi t)$ with different amplitudes, $A = 1, 0.5$, and 0.1 . As expected, the optimal noise levels σ_{opt} decrease as the signal degenerates to the constant function $s = 0$ ($\sigma_{opt} = 1.08, 0.83$, and 0.65 , respectively). Correspondingly, the optimal bandwidths increase: $h_{opt} = 0.08$,

0.10 , and 0.20 . As the last case suggests, the parametric approach of Greenwood *et al.* seems to be adequate if the signal is almost constant. Their approach is not adequate, however, if the signal departs significantly from being constant as in the sinusoid example with amplitude 1. In particular, the noise level $\sigma = 0.64$ is much smaller than our optimal value ($\sigma = 1.08$) in this case. For this σ not only is AMASE very large (see Fig. 5), but also estimation sometimes works badly as one can see in the simulation example in Fig. 6. In particular, the estimation procedure fails completely at some time points for some of the realizations. In this region the combination of small σ and large distance $[s(t) - a]$ often results in an undefined estimator because of too few exceedances. Thus, our data-driven procedure provides a more generally useful technique for unknown, nonconstant signals. Whether an approximation of this procedure has been adopted by real neurons remains to be determined. At the least, our procedure provides a normative baseline against which the performance of real systems can be assessed.

ACKNOWLEDGMENTS

This research was supported by the Peter Wall Institute for Advanced Study at the University of British Columbia and by grants from NSERC of Canada to U.U.M. through P. Greenwood and to L.M.W. We thank P. Greenwood, N. Heckman, and W. Wefelmeyer for helpful discussions and comments, and S. Marion for assistance with the logistic fitting algorithm.

-
- [1] L. Gamaitoni, P. Hänggi, P. Jung, and F. Marchesoni, *Rev. Mod. Phys.* **70**, 223 (1998).
 - [2] J. J. Collins, C. C. Chow, A. C. Capela, and T. T. Imhoff, *Phys. Rev. E* **54**, 5575 (1996).
 - [3] A. Longtin, *J. Stat. Phys.* **70**, 309 (1993).
 - [4] A. Longtin, *Phys. Rev. E* **55**, 868 (1997).
 - [5] M. Stemmler, *Network Comput. Neural Syst.* **7**, 687 (1996).
 - [6] B. J. Gluckman, T. I. Netoff, E. J. Neel, W. L. Ditto, M. L. Spano, and S. J. Schiff, *Phys. Rev. Lett.* **77**, 4098 (1996).
 - [7] H. E. Plesser and S. Tanaka, *Phys. Lett. A* **225**, 228 (1996).
 - [8] K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss, *Phys. Rev. Lett.* **72**, 2125 (1994).
 - [9] F. Moss, D. Pierson, and D. O’Gorman, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **4**, 1383 (1994).
 - [10] W. S. McCulloch and W. Pitts, *Bull. Math. Biophys.* **5**, 115 (1943).
 - [11] J. K. Seamans, N. Gorelova, and C. R. Yang, *J. Neurosci.* **17**, 5936 (1997).
 - [12] P. König, A. K. Engel, and W. Singer, *Trends Neurosci.* **19**, 130 (1996).
 - [13] U. U. Müller, *Can. J. Stat.* (to be published).
 - [14] R. Fakir, *Phys. Rev. E* **58**, 5175 (1998).
 - [15] P. E. Greenwood, L. M. Ward, and W. Wefelmeyer, *Phys. Rev. E* **60**, 4687 (1999).
 - [16] S. Link, *The Wave Theory of Similarity and Difference* (Lawrence Erlbaum Associates, Mahwah, NJ, 1993).
 - [17] S.-T. Chiu, *Ann. Stat.* **19**, 1883 (1991).
 - [18] T. Gasser, in *Robust Statistics, Data Analysis, and Computer Intensive Methods*, edited by H. Rieder, *Lecture Notes in Statistics* Vol. 109 (Springer, New York, 1996), pp. 173–184.
 - [19] D. Ruppert, S. J. Sheather, and M. P. Wand, *J. Am. Stat. Assoc.* **90**, 1257 (1995).
 - [20] W. Härdle and J. S. Marron, *Comput. Stat. Data Anal.* **20**, 1 (1995).
 - [21] H.-G. Müller, *Nonparametric Regression Analysis of Longitudinal Data*, *Lecture Notes in Statistics* Vol. 46 (Springer, New York, 1988).
 - [22] C. L. Mallows, *Technometrics* **15**, 661 (1973).