Statistical Analysis of Stochastic Resonance in a Thresholded Detector

Priscilla E. Greenwood, Ursula U. Müller,
Lawrence M. Ward, and Wolfgang Wefelmeyer

Arizona State University, Universität Bremen,
University of British Columbia, and Universität Siegen

Abstract: A subthreshold signal may be detected if noise is added to the data. The noisy signal must be strong enough to exceed the threshold at least occasionally; but very strong noise tends to drown out the signal. There is an optimal noise level, called stochastic resonance. We explore the detectability of different signals, using statistical detectability measures.

In the simplest setting, the signal is constant, noise is added in the form of i.i.d. random variables at uniformly spaced times, and the detector records the times at which the noisy signal exceeds the threshold. We study the best estimator for the signal from the thresholded data and determine optimal configurations of several detectors with different thresholds.

In a more realistic setting, the noisy signal is described by a nonparametric regression model with equally spaced covariates and i.i.d. errors, and the detector records again the times at which the noisy signal exceeds the threshold. We study Nadaraya–Watson kernel estimators from thresholded data. We determine the asymptotic mean squared error and the asymptotic mean average squared error and calculate the corresponding local and global optimal bandwidths. The minimal asymptotic mean average squared error shows a strong stochastic resonance effect.

Keywords: Stochastic Resonance, Partially Observed Model, Thresholded Data, Noisy Signal, Efficient Estimator, Fisher Information, Nonparametric Regression, Kernel Density Estimator, Optimal Bandwidth.

1 Introduction

Stochastic resonance 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 Gammaitoni et al. (1998) and the recent monograph of Anishchenko et al. (2002) for reviews. The term *stochastic resonance* was introduced by Benzi et al. (1981) in the context of a model describing the periodic recurrence of ice ages. Major climate changes leading to ice ages on the earth were modeled as transitions in a double-well potential system pushed by a signal, the earth's orbital eccentricity, which causes small variations of the solar energy influx. Since this periodic forcing for switching from one climate state to the other is very weak, it must be assisted by other factors such as short term climate fluctuations which were modeled as noise. There is not so much noise that

the transitions become independent of the frequency of the periodic signal, but sufficient noise to assist the periodic transitions, i.e. stochastic resonance occurs.

Classical stochastic resonance in physical systems such as the earth's climate or electrical circuits 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 models of excitable systems with one stable state and a threshold to an unstable excited state, such as model neurons, exhibit this generalized form of stochastic resonance, as observed by Collins et al. (1996b) and Longtin (1993, 1997). With suitable tuning of the noise, 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 (Stemmler, 1996). Moreover, the generalized form of stochastic resonance has also been demonstrated to exist in a variety of living systems, including networks of neurons (Gluckman et al., 1996). Some researchers have speculated that stochastic resonance is a fundamental and general principle of biological information processing (e.g., Plesser and Tanaka, 1996).

The study of systems that exhibit stochastic resonance has been made easier by the demonstration that a simple thresholded detector also exhibits stochastic resonance (e.g., Wiesenfeld et al., 1994). Such a detector has simple dynamics and fires a pulse each time its input exceeds a threshold. If the noise is very large, the probability of an exceedance is close to one half. On the other hand, if the noise is small, the probability of an exceedance is close to zero. Clearly, there will be an optimal amount of noise at which the exceedance times will most closely reflect the variations in the signal intensity. This noise level is called the stochastic resonance point.

Since stochastic resonance is an aspect of the detectability of a signal in thresholded data, not of the dynamics that cause it (Moss et al., 1994), studying simple thresholded detectors is sufficient for many purposes. Indeed, the simplest model of a neuron is just such a simple thresholded detector (McCulloch and Pitts, 1943). Note that the threshold creates a significant nonlinearity in the response of the system to inputs. This nonlinearity mimics the nonlinearities characteristic of bistable and resonant systems and allows stochastic resonance to be studied separately from the dynamics.

In this paper we review three studies of simple thresholded detectors that show stochastic resonance. In Section 2, based on Greenwood et al. (1999), we calculate the Fisher information about a constant signal that can be obtained from the time series of occasions at which the noisy signal exceeds the threshold of the detector. In Section 3, based on Müller (2000), we describe a generalization of the approach to the case of time-varying noisy signals, written as a nonparametric regression model with independent errors. In Section 4, based on Müller and Ward (2000), we describe computer simulations in which the efficacy of the nonparametric regression approach is explored across various signal types.

2   Fisher Information of Constant Signals

Stochastic resonance can be exhibited using almost any method of detecting or reconstructing a subthreshold signal from the information contained in the exceedances of the

threshold of the detector by the signal plus noise. If the signal is *periodic* and observed over a relatively long time interval, then it is common to do a Fourier analysis of the exceedance times and to measure the information thus gained about the signal as the ratio of the power spectral density at the signal frequency to that generated by noise at nearby frequencies (the signal-to-noise ratio) (see Wiesenfeld and Moss, 1995; Gingl et al., 1995; Jung, 1995; Loerincz et al., 1996). Another way to analyse threshold data, related to the way neural activity is analysed, is to investigate the (empirical) residence-time probability distribution, or interspike interval histogram (see Zhou et al., 1990; Longtin et al., 1991; Bulsara et al., 1994). If an *aperiodic* signal is observed over a relatively long time interval, then goodness of signal reconstruction has been measured by a correlation measure; see Collins et al. (1995a,b, 1996a) and Chialvo et al. (1997). Unfortunately, if a signal is to be reconstructed without much delay, then the identification must be based on observations over a relatively short time interval, in which the signal may be nearly constant. In this situation, the model reduces to a parametric one, and the signal-to-noise ratio and correlation measures break down. Fortunately, for such models information measures such as the Fisher information can still be used to measure signal reconstruction. See Levin and Miller (1996); Bulsara and Zador (1996); Collins et al. (1996b); Heneghan et al. (1996); Neiman et al. (1996), and in particular Stemmler (1996) and Chapeau-Blondeau (1997).

The inverse of the Fisher information is the same as the minimal asymptotic variance of any regular estimator (see Bickel et al., 1998, Sections 2.1 to 2.3). In this section we show how optimal estimators of a constant signal can be constructed in several simple situations. We then explore the Fisher information of these signal estimators in a system with one or more detectors. In the case of several detectors, we assume that the same noise is fed into each detector. This is always true for external noise but may also happen if the noise is internal, e.g., when neurons receive background noise from other neurons. Different detectors may well have different thresholds, or a detector may have more than one threshold (see Gammaitoni, 1995a,b; Chapeau-Blondeau and Godivier, 1997). We determine optimal configurations of detectors, varying the distances between the thresholds and the signal, as well as the noise level. We study the simplest possible model of signal plus noise. The signal s is constant over some time interval, say $[0, 1]$. At uniformly spaced times $t_i = i/n$, independent and identically distributed ε_i are introduced. The noisy signal is $s + \varepsilon_i$, $i = 1, \ldots, n$.

If the signal is observed over a longer time interval, or if the noise has "higher frequency" in the sense that the times t_i are more densely spaced, or if there are several detectors each of which receives internal noise independently of the others, then the number n of observations is increased, and the variance of the estimator for the signal is reduced correspondingly. For large n, the signal can be estimated well for a large range of noise variances. This effect of the law of large numbers was first observed in a different setting by Collins et al. (1995b) as *stochastic resonance without tuning*; see also Chialvo et al. (1997); Neiman et al. (1997); Gailey et al. (1997). We come back to this point for varying signals in Section 3.

Our approach differs from the previous literature on stochastic resonance in that we study estimation of the signal from a statistical point of view. In particular, this means that we are concerned with optimal reconstruction of the signal from the data in terms of the variance of *rescaled* estimators for the signal, i.e. of $n^{1/2}(\hat{s} - s)$ rather than of

$\hat{s}$. By the central limit theorem, the variance of $n^{1/2}(\hat{s} - s)$ is about the same for all (sufficiently large) n, whereas the variance of $\hat{s}$ tends to zero as $1/n$. This is why we see stochastic resonance for arbitrarily large n in the present context, whereas stochastic resonance diminishes with increasing n in the previous treatments. Stochastic resonance without tuning is an aspect of the diminishing effect. A similar deficiency of the correlational approach to measuring reconstruction of a varying signal is mentioned in Section 3. Thus we see that the stochastic resonance effect depends heavily on the choice of measure of information transmission. A discussion is in Tougaard (2000, 2002) and Ward et al. (2002).

In this section, we compare four different types of observation of the noisy signal:

1. The noisy signal $X_i = s + \varepsilon_i$ is fully observed. We need the information in the noisy signal itself to measure how much information is lost when the noisy signal is not completely observed but only causes "firing" of the detector.

2. Those times t_i are recorded at which the noisy signal $s + \varepsilon_i$ exceeds a single threshold, $a > 0$. The observations are then the indicators $X_i^a = 1(s + \varepsilon_i > a)$. This scheme was proposed by McCulloch and Pitts (1943) as a minimal model of a neuron. It corresponds to the standard thresholded detector.

3. It is recorded when and which of a finite number of thresholds $0 < a_1 < \cdots < a_r$ are exceeded. Let $A = \{a_1, \ldots, a_r\}$ denote the set of thresholds. The observations can then be written as

$$X_i^A = \begin{cases} 0, \ s + \varepsilon_i \leq a_1, \\ j, \ a_j < s + \varepsilon_i \leq a_{j+1} \quad \text{for } j = 1, \ldots, r-1, \\ r, \ s + \varepsilon_i > a_r. \end{cases}$$

Such observations arise with r detectors with different thresholds, and common background or internal noise.

4. Whenever the single threshold a is exceeded, the noisy signal itself is observed. Then the observations are

$$X_i^{>a} = (s + \varepsilon_i)1(s + \varepsilon_i > a).$$

Case 4 is approximated by case 3 for a large number of closely spaced thresholds above a.

2.1 One Threshold

Let a be a threshold and s a constant signal. We think of s as being nonnegative and below the threshold, but the calculations will not depend on this assumption. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent with distribution function F. Write P_s for the distribution of $X_i = s + \varepsilon_i$. We assume that the only information we have about the signal is whether it exceeds the threshold a. Equivalently, we observe

$$X_i^a = 1(s + \varepsilon_i > a), \quad i = 1, \ldots, n. \tag{1}$$

The observations are independent Bernoulli random variables with probabilities

$$p_s = P(X_i^a = 1) = P_s(a, \infty) = 1 - F(a - s). \tag{2}$$

In this section, we consider a single threshold a and suppress a in the notation. Indeed, by choosing an appropriate scale, we may take a equal to 1.

We can write the signal as a function of p_s,

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

The usual estimator for p_s is the empirical estimator

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i^a = \frac{\hat{n}}{n} \tag{3}$$

with $\hat{n} = \#\{i : X_i^a = 1\}$. The estimator $\hat{p}$ is unbiased and consistent for p_s. The standardized error $n^{1/2}(\hat{p} - p_s)$ is asymptotically normal with variance $p_s(1 - p_s)$. The estimator for the signal as a function of the empirical estimator is

$$\hat{s} = a - F^{-1}(1 - \hat{p}). \tag{4}$$

This estimator is not unbiased. Since $\hat{s}$ is a continuous function of $\hat{p}$, however, the estimator $\hat{s}$ is a consistent estimator for s. Since $\hat{s}$ is a continuously differentiable function of $\hat{p}$, it follows that $n^{1/2}(\hat{s} - s)$ is also asymptotically normal, with variance

$$v_s = \frac{p_s(1 - p_s)}{f(F^{-1}(1 - p_s))^2} = \frac{F(a - s)(1 - F(a - s))}{f(a - s)^2}, \tag{5}$$

where f is the probability density function corresponding to the distribution function F. It is well known and easy to check that $\hat{p}$ is regular and efficient for p_s. Since continuously differentiable functions of regular and efficient estimators are again regular and efficient, the estimator $\hat{s}$ is regular and efficient for the signal, and v_s is the minimal asymptotic variance of regular estimators of s.

As mentioned above, the minimal asymptotic variance v_s can be calculated as the inverse of the Fisher information for s,

$$I_s^a = \frac{f(a - s)^2}{F(a - s)(1 - F(a - s))} = v_s^{-1}.$$

This Fisher information is also given in Stemmler (1996), relation (5.1). The Fisher information has been used as a measure of the transmitted information in related models; see Paradiso (1988); Seung and Sompolinsky (1993); Stemmler (1996).

In order to ascertain whether the expression for the Fisher information exhibits stochastic resonance, we allow the standard deviation σ of the noise distribution to vary. Then the noise density and distribution function become $f((a-s)/\sigma)/\sigma$ and $F((a-s)/\sigma)$, respectively. For normal noise distribution with variance σ^2, we have

$$I_{s\sigma}^a = \frac{\varphi(\frac{a-s}{\sigma})^2}{\sigma^2 \Phi(\frac{a-s}{\sigma})(1 - \Phi(\frac{a-s}{\sigma}))}, \tag{6}$$

with $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and $\Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy$ denoting the density and distribution functions of the standard normal $N(0, 1)$ distribution. $I_{s\sigma}^a$ is a unimodal function of σ with a very pronounced stochastic resonance point. The function is symmetric
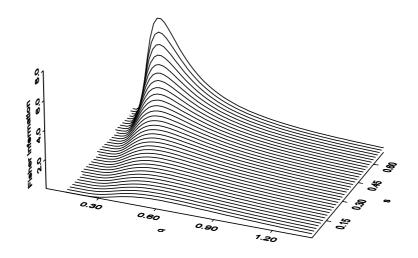
Figure 1: Fisher information $I_{s\sigma}^1$.

in $a - s$. Hence a superthreshold signal produces the same stochastic resonance property as a subthreshold signal. Figure 1 shows $I_{s\sigma}^1$ as a function of s and σ. The optimal σ decreases with the distance from the signal to the threshold; at the same time the maximal Fisher information goes to infinity. For example, if $a = 1$ and the signal is low, $s = 0$, then the optimal σ is 0.63500, and the maximal value of $I_{0\sigma}^1$ is 0.60842.

How much information is lost by observing the threshold exceedances only, rather than the noisy signal $s + \varepsilon_i$? Greenwood et al. (1999) derive the expression for the ratio of Fisher information in the exceedances, $I_{s\sigma}^a$, to that in the noisy signal itself, I_σ. When the noise distribution is normal with variance σ^2, the ratio is

$$ I_{s\sigma}^a / I_\sigma = R\Big(\frac{a - s}{\sigma}\Big), $$

where

$$ R(u) = \frac{\big(\int_u^\infty x\varphi(x)dx\big)^2}{\Phi(u)(1 - \Phi(u))} = \frac{\varphi(u)^2}{\Phi(u)(1 - \Phi(u))}. $$

The function R is unimodal and symmetric, and $R(0) = 0.636620$. Hence X_i^a retains almost two thirds of the available information if the signal is at the threshold, and considerably less if it is above or below and σ is small. Figure 2 shows $I_{s\sigma}^1 / I_\sigma = R((1 - s)/\sigma)$ as a function of s and σ. Notice the discontinuity where $s = a = 1$ in the figure; this is where $R(0) = 0.636620$.

2.2   Several Thresholds

Consider a system in which there are r thresholds, $0 < a_1 < \cdots < a_r$, a constant signal s, and a noisy signal $s + \varepsilon_i$, with $\varepsilon_1, \ldots, \varepsilon_n$ independent with distribution function F and density f. This could be, for example, a neural network in which several neurons with different thresholds are exposed to the same signal and noise and then converge to drive a single higher-order neuron. The output of the higher-order neuron depends on how many of the neurons converging on it are activated above their thresholds; it simply sums
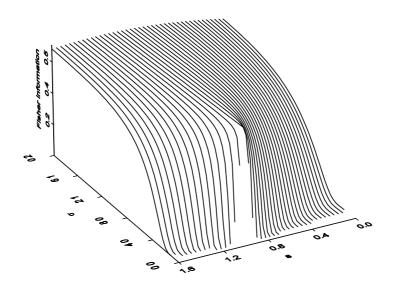
Figure 2: Proportion of information, $I_{s\sigma}^1/I_\sigma$, retained by X_i^1.

the activities of the converging neurons. Thus, its firing indicates which thresholds are exceeded by the noisy signal. Equivalently, we observe

$$X_i^A = \begin{cases} 0, \ s + \varepsilon_i \leq a_1, \\ j, \ a_j < s + \varepsilon_i \leq a_{j+1} \quad \text{for } j = 1, \ldots, r-1, \\ r, \ s + \varepsilon_i > a_r. \end{cases}$$

Here A stands for the set of thresholds, $\{a_1, \ldots, a_r\}$. The observations $X_1^A, \ldots, X_n^A$ are independent, with probabilities

$$\begin{aligned} p_{s0} &= P(X_i^A = 0) = F(a_1 - s), \\ p_{sj} &= P(X_i^A = j) = F(a_{j+1} - s) - F(a_j - s) \quad \text{for } j = 1, \ldots, r-1, \\ p_{sr} &= P(X_i^A = r) = 1 - F(a_r - s). \end{aligned}$$

The observations follow a distribution on $\{0, \ldots, r\}$, with a one-dimensional parameter s. For $r = 1$, the family of distributions consists of *all* distributions on $\{0, 1\}$, and an efficient estimator for s is obtained as a function of the empirical estimator for $p_s = P_s(a_1, \infty)$; see Subsection 2.1. For $r > 1$, we do not get such a simple efficient estimator, but the maximum likelihood estimator is, of course, still efficient.

Again, Greenwood et al. (1999) derive an expression for the maximum likelihood estimator of the signal for any set of thresholds A. Here we will consider only the simple case of multiple thresholds in which we have two thresholds, $0 < a < b$ and a signal $s < a$. As in Subsection 2.1, we assume that the noise distribution is normal with variance σ^2. In this case, the Fisher information in observing which of the two thresholds is exceeded by the noisy signal is

$$I_{s\sigma}^{ab} = \frac{1}{\sigma^2} \left(\frac{\varphi(\frac{a-s}{\sigma})^2}{\Phi(\frac{a-s}{\sigma})} + \frac{\left(\varphi(\frac{b-s}{\sigma}) - \varphi(\frac{a-s}{\sigma}) \right)^2}{\Phi(\frac{b-s}{\sigma}) - \Phi(\frac{a-s}{\sigma})} + \frac{\varphi(\frac{b-s}{\sigma})^2}{1 - \Phi(\frac{b-s}{\sigma})} \right).$$

Suppose in particular that $a = 1$. We have seen in Subsection 2.1 that X_i^1 retains the most information, as a function of s, at $s = 1$, where $R(0) = 0.636620$. The value does
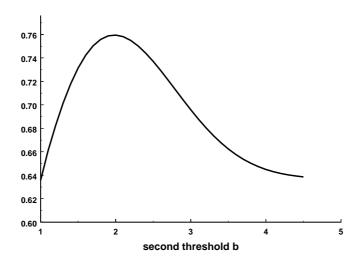
Figure 3: Proportion of information, I_{11}^{1b}/I_1, retained by two thresholds, at 1 and b, for noise variance 1.

not depend on the noise variance σ^2, so we may take $\sigma = 1$. Now we add a second threshold, $b > 1$. We can calculate the ratio of $I_{s\sigma}^{ab}$ to I_σ under these assumptions, giving the proportion of information retained by two thresholds. Figure 3 shows the results of such calculations for a range of second threshold values. The maximum is 0.75957 which is attained for $b = 1.98$.

Let us return to the situation where we have only a single threshold a. So far we have studied the situation where we observe only whether or not the noisy signal exceeds the threshold. Suppose now that we also observe the *size* of the noisy signal whenever it exceeds the threshold. The observations are then $X_i^{>a} = (s + \varepsilon_i)1(s + \varepsilon_i > a)$. They contain more information about the signal than the indicators $X_i^a = 1(s + \varepsilon_i > a)$. Greenwood et al. (1999) derive an expression for the Fisher information, $I_{s\sigma}^{>a}$, in these observations; it is of course true that $I_{s\sigma}^{>a} < I_\sigma$, the maximum amount of information in a fully observed noisy signal. For a normal noise distribution with variance σ^2, the proportion of information retained by $X_i^{>a}$ is

$$\frac{I_{s\sigma}^{>a}}{I_\sigma} = R^> \left(\frac{a-s}{\sigma} \right),$$

where

$$R^>(u) = \frac{\left(\int_{-\infty}^u x\varphi(x)dx \right)^2}{\Phi(u)} + \int_u^\infty x^2 \varphi(x)dx$$

$$= \frac{\varphi(u)^2}{\Phi(u)} + 1 - \Phi(u) + u\varphi(u).$$

If the signal is at the threshold, $s = a$, then $I_{s\sigma}^{>a}/I_\sigma = R^>(0)$, which is independent of the noise variance σ. If the signal is below the threshold, $s < a$, we expect $I_{s\sigma}^{>a}/I_\sigma$ to be large for large σ because the $X_i^{>a}$ are most informative if the noisy signal is with high probability above the threshold. For the same reason, $I_{s\sigma}^{>a}/I_\sigma$ is large for *small* σ if $s > a$. Under these conditions we have $R^>(0) = 0.818310$. Hence $X_i^{>a}$ retains about
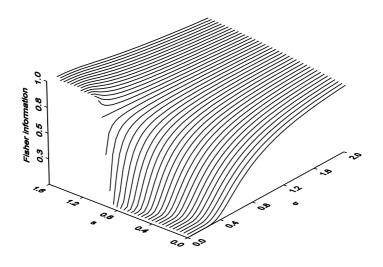
Figure 4: Proportion of information, $I_{s\sigma}^{>1}/I_\sigma$, retained by $X_i^{>1}$.

four fifths of the information if the signal is at the threshold, considerably less if it is below and σ is small, and most of the information if $s > a$ and σ is small. Figure 4 shows $I_{s\sigma}^{>1}/I_\sigma = R^>((1-s)/\sigma)$ as a function of s and σ.

Suppose we fix a lowest threshold a and add more and more thresholds above a such that in the limit they become dense above a. We expect that the information in observing which thresholds are exceeded by the noisy signal converges to the information in seeing the noisy signal above the threshold. To see this, choose thresholds $a_1, \ldots, a_n > a$ such that the gaps between them tend to zero and their maximum tends to infinity with n. The thresholds partition (a, ∞) into $n + 1$ intervals $B_1, \ldots, B_{n+1}$. Greenwood et al. (1999) show that $I_{s\sigma}^{A_n}$, the Fisher information for exceedances of the n thresholds, converges to $I_{s\sigma}^{>a}$, the Fisher information in the part of the noisy signal that exceeds the threshold. Thus, in the limit of dense thresholds, the model of a set of simple thresholded detectors is the same as a model of thresholded detector that accurately represents the entire above-threshold part of the noisy signal.

3   Nonparametric Regression Model for Varying Signals

The model of a simple thresholded detector outlined above works well for a constant signal. In fact, it has been applied in a biological setting as a minimal model of the paddlefish using its electrosense to capture *Daphnia* plankton (Greenwood et al., 2000). A version of this model is also used for edge detection in images observed with vibrating vision systems (Hongler et al., 2002). Greenwood et al. (1999) suggest that signal estimation from thresholded data is also possible if the noisy signal is described by a nonparametric regression model with independent errors. Such a model is more realistic in many biological systems. This suggestion is realized in Müller (2000). The noisy signal is written as $X(t_i) = s(t_i) + \varepsilon(t_i)$, with independent mean zero error variables $\varepsilon(t_i)$. We observe

$$X^a(t_i) = 1(s(t_i) + \varepsilon(t_i) > a) = 1(X(t_i) > a), \quad i = 1, \ldots, n, \tag{7}$$

as in (1). The exceedance probabilities are now

$$E(X^a(t_i)) = p(t_i) = P(X^a(t_i) = 1) = P(s(t_i) + \varepsilon(t_i) > a). \qquad (8)$$

Assuming that the distribution function of $\varepsilon(t_i)$, say F_{t_i}, is known and has an inverse, we use the transformation

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

to estimate the signal $s(t_i)$ employing a kernel estimator for $p(t_i)$.

  If the noise is artificially generated, then it is reasonable to assume that the distribution function F_{t_i} is completely known. If the noise is background noise, we may at least know the *form* of the distribution. It may, for example, be plausible to assume that the errors are normally distributed. However, we cannot identify the noise amplitude from the $X^a(t_i)$. One way of dealing with this problem is to get information about the noise from elsewhere, for example by using a second detector with a different threshold, as in Subsection 2.2 for constant signal. A second is to note that even if the noise distribution is known up to a scale parameter, the signal can still be identified up to a one-parameter family of transformations. Hence most of the information is retained, unless, of course, the signal is constant.

  Our criterion for bandwidth selection is the asymptotic mean average squared error of the estimator of s, which we derive. A formula for the asymptotically optimal bandwidth can then be written. It contains unknown quantities which must be estimated. Our approach is similar to that of Ruppert et al. (1995), who derive an asymptotically optimal bandwidth for the classical setting, i.e. fully observed data. The main difference between the present derivation and their article is the nonlinear link occurring here, in particular in the mean squared error expression. Through the linearization of this expression, the calculation of the optimal bandwidth becomes similar to the standard case, and familiar results can be utilized.

  We use the Nadaraya-Watson (kernel) estimator

$$\hat{p}_h(t) = \frac{\sum_{i=1}^n \frac{1}{h} K(\frac{t-t_i}{h}) \cdot X^a(t_i)}{\sum_{i=1}^n \frac{1}{h} K(\frac{t-t_i}{h})} \qquad (9)$$

for the exceedance probability $p(t)$ in (7), and estimate the signal $s(t)$ by $\hat{s}_h(t) = a - F_t^{-1}(1 - \hat{p}_h(t))$. In nonparametric regression theory, generally accepted measures of the goodness of the estimation are the mean squared error $E(\hat{s}_h(t) - s(t))^2$ and the mean average squared error $n^{-1} \sum_{i=1}^n E(\hat{s}_h(t_i) - s(t_i))^2$. We take these quantities as criteria for an asymptotically optimal local bandwidth and an asymptotically optimal global bandwidth. The general approach taken by Müller (2000) is to derive a Taylor approximation for the mean squared error, which immediately gives the approximation for the mean average squared error. The bandwidth h that minimizes the leading term of the expansion is then called *optimal*. The distinguishing characteristic of this model is that it involves the nonlinear transformation $s(t) = a - F_t^{-1}(1 - p(t))$ of the exceedance probability $p(t)$.

  The main theorem proved by Müller (2000) is about the signal estimator $\hat{s}_h$. It gives the asymptotic mean squared error (local—for each time point), the asymptotic mean average squared error (global—over all time points) and the respective optimal bandwidths

as $n \to \infty$ and $h \to 0$, $nh^3 \to \infty$:

$$AMSE(h,t) = \frac{1}{f(F^{-1}(p(t)))^2}\left(\frac{1}{nh}R(K)p(t)(1-p(t)) + \frac{h^4}{4}\mu_2(K)^2 p''(t)^2\right), \quad (10)$$

$$AMASE(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{nh}R(K)\frac{p(t_i)(1-p(t_i))}{f(F^{-1}(p(t_i)))^2} + \frac{h^4}{4}\mu_2(K)^2\frac{p''(t_i)^2}{f(F^{-1}(p(t_i)))^2}\right), \quad (11)$$

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

$$h_{opt}(t) = n^{-1/5}\left(\frac{R(K)p(t)(1-p(t))}{\mu_2(K)^2 p''(t)^2}\right)^{1/5}, \quad (12)$$

$$h_{opt} = n^{-1/5}\left(\frac{R(K)\sum_{i=1}^{n}\frac{p(t_i)(1-p(t_i))}{f(F^{-1}(p(t_i)))^2}}{\mu_2(K)^2\sum_{i=1}^{n}\frac{p''(t_i)^2}{f(F^{-1}(p(t_i)))^2}}\right)^{1/5}. \quad (13)$$

The approximate mean squared error (10) has two terms, the approximate variance (first term) and squared bias (second term) of $\hat{p}_h(t)$. In particular, the characteristic variance-bias trade-off is evident.

With the optimal asymptotic bandwidth h_{opt} at hand, we may now write the minimal value of $AMASE(h)$ as

$$AMASE(h_{opt}) = \frac{5}{4}n^{-4/5}\left(\mu_2(K)^2\frac{1}{n}\sum_{i=1}^{n}\frac{p''(t_i)^2}{f(F^{-1}(p(t_i)))^2}\right)^{1/5}$$
$$\left(R(K)\frac{1}{n}\sum_{i=1}^{n}\frac{p(t_i)(1-p(t_i))}{f(F^{-1}(p(t_i)))^2}\right)^{4/5}.$$

This value depends on the squared second derivatives $p''(t)^2$. Smooth signals will, in general, lead to small values of $p''(t)^2$ and thus to small values of $\inf_{h>0} AMASE(h)$. Furthermore, this formula shows the influence of the kernel K, which appears only in the expression $\mu_2(K)^{2/5} \cdot R(K)^{4/5}$. The (second order) kernel with support $[-1, 1]$ that minimizes this term and thus the approximate mean average squared error under the further constraint $K \geq 0$, is the Epanechnikov kernel

$$K(u) = \frac{3}{4}(1 - u^2)1_{[-1,1]}(u).$$

The optimal kernel is not much better than other kernels, for example the Gaussian kernel (see Wand and Jones, 1995). What is really crucial is the correct choice of the bandwidth h. A method of data-driven bandwidth selection is necessary since we do not know what the optimal bandwidth is for an unknown signal.

In data from simulations or experiments, the stochastic resonance effect is often assessed in simulated data, and in data from experimental systems, by comparing the estimator for the exceedance probabilities with the signal using the Pearson product-moment correlation coefficient or modifications thereof, usually called the "normalized power norm" or "cross-correlation coefficient", namely

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}(s(t_i) - \bar{s})(\hat{p}(t_i) - \bar{\hat{p}})}{\left(\frac{1}{n}\sum_{i=1}^{n}(s(t_i) - \bar{s})^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{p}(t_i) - \bar{\hat{p}})^2\right)^{1/2}}.$$
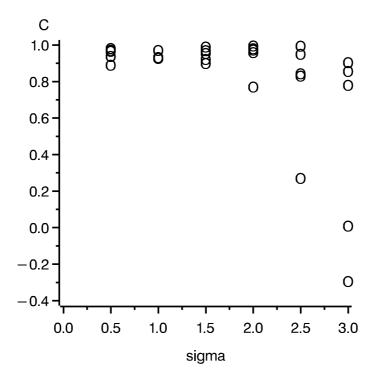
Figure 5: Realizations of the Pearson product-moment correlation coefficient C (circles) between the sinusoid $s(t) = \sin(2\pi t)$ and its estimates $\hat{s}_h(t)$, for $\sigma = 0.5, 1, \ldots, 3$ and $n = 100$.

Here $\hat{p}(t)$ is an estimator of the exceedance probability $p(t)$, and $\overline{s} = \frac{1}{n}\sum_{i=1}^{n} s(t_i)$ and $\overline{\hat{p}} = \frac{1}{n}\sum_{i=1}^{n} \hat{p}(t_i)$ denote averages. In several papers (Collins et al., 1995b,a; Heneghan et al., 1996; Chialvo et al., 1997) a box kernel was used to estimate the probability of an exceedance. This kernel corresponds to the ones used in our more sophisticated estimator $\hat{p}_h(t)$ but with a fixed bandwidth. Clearly, $\hat{p}_h$ is not consistent if h is kept fixed and n tends to infinity.

In studies such as those just cited, researchers usually report the results of elaborate simulation studies and obtain empirical estimates of the mean and standard error of C. In particular, they often plot a curve of the estimated mean as a function of noise variance. This curve is often concave downward, indicating stochastic resonance. However, as mentioned in Section 2, there are some problems with the use of C as a measure of goodness of estimation. As an example, Figure 5 shows, for normal noise distribution with standard deviation $\sigma = 0.5, 1, \ldots, 3$, five realizations of the Pearson product-moment correlation coefficient between s and our estimator $\hat{s}_h$ (estimated as described above for a sinusoid) for $n = 100$. Stochastic resonance appears in the figure since for $\sigma = 1$ the five values of C are all close to one whereas for smaller and larger σ, the correlation is not always high. However, when the same simulation is done for $n = 1,000$ time points, the values of C are close to one for all realizations; no stochastic resonance is apparent. This observation illustrates well a phenomenon first observed by Collins et al. (1995b) in a different setting, called "stochastic resonance without tuning", which we mentioned in Section 2: Since the variance of the signal estimator and hence the variance of C decreases with n, the correlation is high for a broad range of σ's. Although there is stochastic resonance,

i.e. an optimal level of noise, the Pearson product-moment correlation coefficient does not reflect it if the time points are too densely spaced.

4   Stochastic Resonance for Varying Signals

In this section we explore further the applicability of the thresholded nonparametric regression model as a simple model of a neuron. We consider the neuron to be embedded in a neural network. It is exposed to "noise" consisting of rapidly varying, unsynchronized inputs from perhaps a thousand other neurons. The signal is assumed to have longer time scale variations in the input of one specific neuron (or a few synchronized ones), and to be by itself insufficient to drive this neuron. Nonetheless, the output of the neuron could become synchronized with the variations in the input of the subthreshold signaling neurons through the mechanism of stochastic resonance, with the noise "amplifying" the signal.

Although a simple thresholded detector such as that described in Section 2 can exhibit stochastic resonance, 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., Collins et al., 1996a). This model integrates its inputs over a moving time window of length 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 (Seamans et al., 1997). Neurons with larger time windows act as integrators, while those such as apical tufts, with smaller time windows, act as coincidence detectors (König et al., 1996). 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, and constant diffusion of ions caused by electrical and concentration gradients and active pumps. The time-varying instantaneous firing rate of the neuron ($1,000/\Delta t$, 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.

A kernel estimator can mimic temporal integration of input. Previous studies implementing numerical smoothing of model excitable system outputs with a box-type kernel or of thresholded data with a Gaussian kernel were done by Collins et al. (1996a) and Fakir (1998), respectively. However, both previous studies computed the correlation between smoothed system outputs and signals rather than applying the mean squared error criterion to estimators of the signal, as we do here. The limitations of the correlational approach were described at the end of Section 3.

4.1   Model and Estimation

The model is that described in (8). In the case of modeling a neuron, the bandwidth (13) of the kernel can be considered to represent the window of temporal integration of the neuron. Although the Epanechnikov kernel is optimal, other kernels would do nearly as well, for example the Gaussian kernel. In particular, an asymmetric kernel would be more

appropriate to model the time integration behavior of actual neurons. We can speculate, however, on the basis of some pilot work, that as long as the kernel covers sufficiently many 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. In the case of neurons, we expect that the asymmetric kernel might be a negative exponential function with rate such that inputs over the previous 10–100 *ms* or so would receive noticeable weight in the output. This is in the range of temporal integration times observed in actual neurons, as mentioned earlier.

The bandwidth formulas (12) and (13) are both of the form $n^{-1/5}$ times a factor which depends on the exceedance probability. For good performance in the finite sample situation the factor should be estimated suitably. Müller and Ward (2000) use plug-in methods, i.e. they estimate the exceedance probability and substitute into the bandwidth formulas. This approach has been shown, in the classical setting with fully observed data, to perform more reliably than standard methods such as cross-validation (see, for example, Chiu, 1991).

Our procedure is applicable to all signals of arbitrary shape. Here we demonstrate it 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 Figures 6a – 6c), for optimal noise levels from a normal error distribution and $n = 10,000$ time points. In the examples we use threshold $a = 1$ throughout.
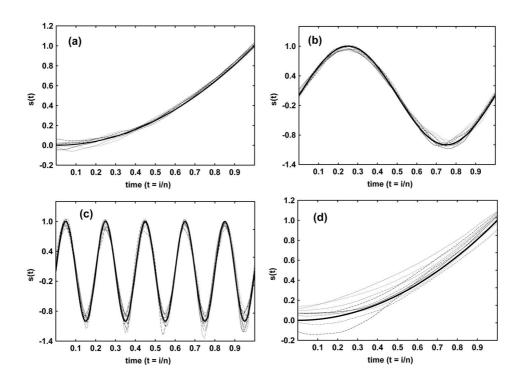

Figure 6: Ten realizations of the estimation procedure for (a) $s_1(t) = t^2$ with $\sigma = 0.59$, average estimated bandwidth $h = 0.16$, and theoretically optimal bandwidth (13) $h_{opt} = 0.16$; (b) $s_2(t) = \sin(2\pi t)$ with $\sigma = 1.08$, average bandwidth $h = 0.09$, and $h_{opt} = 0.08$; (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.
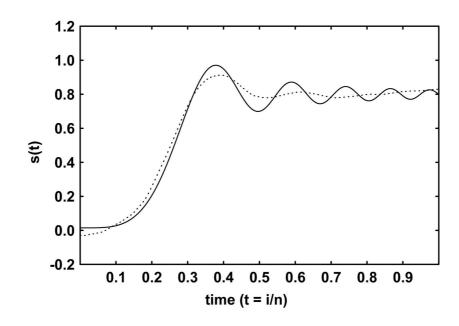
Figure 7: Signal estimation of $s_4(t) = 0.8 - 0.025\sin(10\pi t^2)/t^2$ with $\sigma = 0.76$ and estimated bandwidth $h = 0.10$. Signal and estimated signals are shown by continuous and broken lines respectively.

These examples illustrate several properties of the estimation technique. One is that estimation is noticeably worse for parts of the signal that are far from threshold than for parts near threshold. This is because the bandwidth used is globally optimal, and can 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.

We also consider 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 Figure 7. 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 signal is a mixture of several types of signal. It resembles signals that might be of biological importance, for example that of an approaching predator who "holds" at an attack launch point. The estimation technique does a good job of capturing the low frequency content, but fails to estimate the signal accurately at the high frequency end. Again, this is because we use a globally optimal bandwidth, $h = 0.10$, which is near that of $s_2(t)$ and does well for the components near 1 Hz.

4.2   Stochastic Resonance

So far we have described estimation of a noisy subthreshold signal without mentioning stochastic resonance. This section discusses the stochastic resonance effect in the context of the nonparametric kernel regression model.

Consider the global criterion $AMASE(h)$, equation (11), as a function of noise variance. With the optimal bandwidth h_{opt}, equation (13), inserted, it can be rewritten in terms of the signal function s. Restricting attention to the normal $N(0, \sigma^2)$ error distribution and
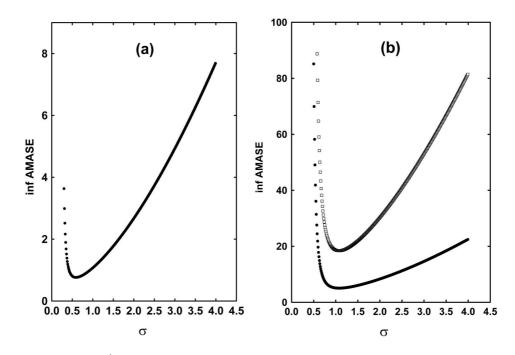
Figure 8: Plot of $n^{4/5}\inf_h AMASE(h)$ versus σ for (a) $s_1(t) = t^2$, and (b) $s_2(t) = \sin(2\pi t)$ (dots) and $s_3(t) = \sin(10\pi t)$ (open squares).

the Epanechnikov kernel, we obtain $\mu_2(K) = 1/5$ and $R(K) = 3/5$, and hence

$$AMASE(h_{opt}) = \frac{5}{4}n^{-4/5}\left(\frac{1}{25n}\sum_{i=1}^{n}\left(\frac{a - s(t_i)}{\sigma^2}s'(t_i)^2 + s''(t_i)\right)^2\right)^{1/5}$$

$$\cdot\left(\frac{3}{5n}\sum_{i=1}^{n}\frac{\sigma^2\Phi\left(\frac{s(t_i)-a}{\sigma}\right)\Phi\left(\frac{a-s(t_i)}{\sigma}\right)}{\varphi\left(\frac{s(t_i)-a}{\sigma}\right)^2}\right)^{4/5}. \tag{14}$$

Notice that (14) consists of the product of two factors. The first factor 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.

A noise level σ that is optimal with respect to $AMASE$ depends on the signal function s. An explicit general formula obviously cannot be given. Instead, we compute the optimal σ for each signal by minimizing $AMASE$ numerically.

In Figure 8 we have plotted $AMASE$ as given in (14), multiplied by the convergence rate $n^{4/5}$, for $s_1(t) = t^2$ (Figure 8a), and $s_2(t) = \sin(2\pi t)$ and $s_3(t) = \sin(10\pi t)$ (Figure 8b). For our calculations we chose a threshold $a = 1$ and $n = 10,000$. Since the sums in (14) approximate integrals, the curves are the same for all sufficiently large n. In particular, they have a sharp minimum at their respective stochastic resonance points, namely $\sigma = 0.59$ for signal s_1 and $\sigma = 1.08$ for both signal s_2 and s_3. We already used the optimal noise levels for our simulated estimation examples in Subsection 4.1.

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 estimator obtained using the procedure of Subsection 4.1. To demonstrate this, we consider again signal s_1. We estimate the signal for 10 additional realizations of signal plus noise but this time we use the theoretically bad value $\sigma = 4.0$. These are
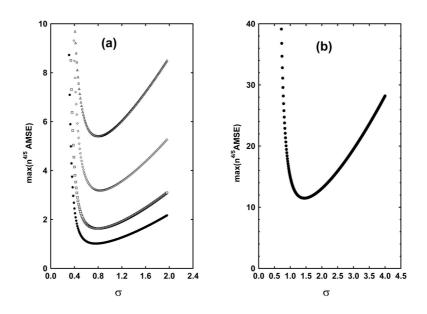
Figure 9: (a) The local asymptotic mean squared error maximized over s' and s'' both having the range $[-1, 1]$ (dots), $[-2, 2]$ (open squares), $[-5, 5]$ (open diamonds) and $[-10, 10]$ (open triangles) respectively. (b) Determination of the globally uniformly optimal noise level $\sigma_{**} = 1.45$ for the ranges of $s_2(t) = \sin(2\pi t)$ and its first and second derivative.

plotted in Figure 8d (compare with Figure 8a). As expected, the estimates are significantly worse. In particular, a large increase of the variance over realizations becomes evident.

4.3 Choosing the Noise Level

As seen in the previous section, the optimal noise level depends on the signal and its derivatives and cannot be determined without prior knowledge. In this section we determine noise levels σ_* and σ_{**} that are locally and globally uniformly optimal in the sense that the estimator for the signal behaves well for a given range of the signal and its first and second derivatives near a fixed time point t and in a fixed time interval, respectively. Again we take $N(0, \sigma^2)$ as error distribution and use the Epanechnikov kernel.

Fix t. Write $AMSE(h_{opt}(t), t)$ as a function $AMSE(s, s', s'', \sigma)$ of $s = s(t)$, $s' = s'(t)$, $s'' = s''(t)$, and the error variance σ^2:

$$AMSE(s, s', s'', \sigma) = \frac{5}{4}n^{-4/5}\left(\frac{1}{25}\left(\frac{a-s}{\sigma^2}s'^2 + s''\right)^2\right)^{1/5}\left(\frac{3}{5}\frac{\sigma^2\Phi\left(\frac{s-a}{\sigma}\right)\Phi\left(\frac{a-s}{\sigma}\right)}{\varphi\left(\frac{s-a}{\sigma}\right)^2}\right)^{4/5}.$$

Determine the error variance $\sigma^2 = \sigma_*^2$ that minimizes

$$\max_{s'\in[c_1,d_1],s''\in[c_2,d_2]} AMSE(s, s', s'', \sigma).$$

In Figure 9a this maximum, multiplied by the convergence rate $n^{4/5}$, is plotted as a function of σ for both s' and s'' in the ranges $[-1, 1]$, $[-2, 2]$, $[-5, 5]$ and $[-10, 10]$, respectively, with $s = 0$ and $a = 1$. For these intervals we obtain the locally uniformly optimal 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 some time interval, say $[0, 1]$, one can, analogously, derive a globally uniformly optimal noise level σ_{**}, given a lower bound c_0 for the signal. This is the error variance $\sigma^2 = \sigma_{**}^2$ that minimizes

$$\max_{s \in [c_0, a], s' \in [c_1, d_1], s'' \in [c_2, d_2]} AMSE(s, s', s'', \sigma).$$

As an illustration we derive such a globally uniformly optimal 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 Figure 9b) we obtain a globally uniformly optimal noise level $\sigma_{**} = 1.45$ which is larger than the noise level $\sigma = 1.08$ that is optimal for the signal $\sin(2\pi t)$ (see Subsection 4.2, Figure 8b). We expect this phenomenon since "worse" signals fit into these ranges, for example $s(t) = -1$ which clearly needs more noise in order to guarantee sufficiently many threshold crossings.

Acknowledgment

References

V.S. Anishchenko, V. Astakhov, A.B. Neiman, T. Vadivasova, and L. Schimansky-Geier. *Nonlinear Dynamics of Chaotic and Stochastic Systems. Tutorial and Modern Developments.* Springer, New York, 2002.

R. Benzi, A. Sutera, and A. Vulpiani. The mechanism of stochastic resonance. *J. Phys. A: Math. Gen.*, 14:L453–L457, 1981.

P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York, 1998.

A.R. Bulsara, S.B. Lowen, and C.D. Rees. Cooperative behavior in the periodically modulated Wiener process: Noise-induced complexity in a model neuron. *Phys. Rev. E (3)*, 49:4989–500, 1994.

A.R. Bulsara and A. Zador. Threshold detection of wideband signals: A noise-induced maximum in the mutual information. *Phys. Rev. E (3)*, 54:R2185–R218, 1996.

F. Chapeau-Blondeau. Input-output gains for signal in noise in stochastic resonance. *Phys. Lett. A*, 232:41–48, 1997.

F. Chapeau-Blondeau and X. Godivier. Noise-enhanced capacity via stochastic resonance in an asymmetric binary channel. *Phys. Rev. E (3)*, 55:2016–2019, 1997.

D.R. Chialvo, A. Longtin, and J. Müller-Gerking. Stochastic resonance in models of neuronal ensembles. *Phys. Rev. E (3)*, 55:1798–1808, 1997.

S.-T. Chiu. Some stabilized bandwidth selectors for nonparametric regression. *Ann. Statist.*, 19:1883–1905, 1991.

J.J. Collins, C.C. Chow, A.C. Capela, and T.T. Imhof. Aperiodic stochastic resonance. *Phys. Rev. E (3)*, 54:5575–5584, 1996a.

J.J. Collins, C.C. Chow, and T.T. Imhof. Aperiodic stochastic resonance in excitable systems. *Phys. Rev. E (3)*, 52:R3321–R3324, 1995a.

J.J. Collins, C.C. Chow, and T.T. Imhof. Stochastic resonance without tuning. *Nature*, 376:236–238, 1995b.

J.J. Collins, T.T. Imhof, and P. Grigg. Noise-enhanced information transmission in RAT SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance. *J. Neurophysiology*, 76:64–645, 1996b.

R. Fakir. Nonstationary stochastic resonance in a single neuron-like system. *Phys. Rev. E (3)*, 58:5175–5178, 1998.

P.C. Gailey, A. Neiman, J.J. Collins, and F. Moss. Stochastic resonance in ensembles of nondynamical elements: The role of internal noise. *Phys. Rev. Lett.*, 23:4701–4704, 1997.

L. Gammaitoni. Stochastic resonance and the dithering effect in threshold physical systems. *Phys. Rev. E (3)*, 52:4691–4698, 1995a.

L. Gammaitoni. Stochastic resonance in multi-threshold systems. *Phys. Lett. A*, 208: 315–322, 1995b.

L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Rev. Modern Phys.*, 70:223–288, 1998.

Z. Gingl, L.B. Kiss, and F. Moss. Non-dynamical stochastic resonance: theory and experiments with white and arbitrarily coloured noise. *Europhys. Lett.*, 29:191–19, 1995.

B.J. Gluckman, T.I. Netoff, E.J. Neel, W.L. Ditto, M.L. Spano, and S.J. Schiff. Stochastic resonance in a neuronal network from mammalian brain. *Phys. Rev. Lett.*, 77:4098–4101, 1996.

P.E. Greenwood, L.M. Ward, D. Russell, A. Neiman, and F. Moss. Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture. *Phys. Rev. Lett.*, 84:4773–4776, 2000.

P.E. Greenwood, L.M. Ward, and W. Wefelmeyer. Statistical analysis of stochastic resonance in a simple setting. *Phys. Rev. E (3)*, 60:4687–4695, 1999.

C. Heneghan, C.C. Chow, J.J. Collins, T.T. Imhoff, S.B. Lowen, and M.C. Teich. Information measures quantifying aperiodic stochastic resonance. *Phys. Rev. E (3)*, 54: R2228–R2231, 1996.

M.-O. Hongler, Y. López de Meneses, A. Beyeler, and J. Jaquot. The resonant retina: Exploiting vibration noise to optimally detect edges in an image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002. To appear.

P. Jung. Stochastic resonance and optimal design of threshold detectors. *Phys. Lett. A*, 207:93–104, 1995.

P. König, A.K. Engel, and W. Singer. Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.*, 19:130–137, 1996.

J. Levin and J.P. Miller. Stochastic resonance in the cricket cercal system. *Nature*, 380: 165–168, 1996.

K. Loerincz, Z. Gingl, and L.B. Kiss. A stochastic resonator is able to greatly improve signal-to-noise ratio. *Phys. Lett. A*, 224:63–67, 1996.

A. Longtin. Stochastic resonance in neuron models. *J. Stat. Phys.*, 70:309–327, 1993.

A. Longtin. Autonomous stochastic resonance in bursting neurons. *Phys. Rev. E (3)*, 55: 868–876, 1997.

A. Longtin, A. Bulsara, and F. Moss. Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons. *Phys. Rev. Lett.*, 67: 656–659, 1991.

W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.

F. Moss, D. Pierson, and D. O'Gorman. Stochastic resonance: Tutorial and update. *Int. J. Bifurcation Chaos*, 4:1383–1397, 1994.

U.U. Müller. Nonparametric regression for threshold data. *Canad. J. Statist.*, 28:301–310, 2000.

U.U. Müller and L.M. Ward. Stochastic resonance in a statistical model of a time-integrating detector. *Phys. Rev. E (3)*, 61:4286–4294, 2000.

A. Neiman, L. Schimansky-Geier, and F. Moss. Linear response theory applied to stochastic resonance in models of ensembles of oscillators. *Phys. Rev. E (3)*, 56:R9–R12, 1997.

A. Neiman, B. Shulgin, V.S. Anishchenko, W. Ebeling, L. Schimansky-Geier, and J. Freund. Dynamic entropies applied to stochastic resonance. *Phys. Rev. Lett.*, 76:4299–4302, 1996.

M.A. Paradiso. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.*, 58:35–49, 1988.

H.E. Plesser and S. Tanaka. Stochastic resonance in a model neuron with reset. *Phys. Lett. A*, 225:228–234, 1996.

D. Ruppert, S.J. Sheather, and M.P. Wand. An effective bandwidth selector for local least squares regression. *J. Amer. Stat. Assoc.*, 90:1257–1270, 1995.

J.K. Seamans, N. Gorelova, and C.R. Yang. Contributions of voltage-gated Ca2+ channels in the proximal versus distal dendrites to synaptic integration in prefrontal cortical neurons. *J. Neurosci.*, 17:5936–5948, 1997.

H.S. Seung and H. Sompolinsky. Simple models for reading neuronal population codes. *Proc. Nat. Acad. Sci. U.S.A.*, 90:10749–10753, 1993.

M. Stemmler. A single spike suffices: the simplest form of stochastic resonance in model neurons. *Network: Computation in Neural Systems*, 7:687–716, 1996.

J. Tougaard. Stochastic resonance and signal detection in an energy detector — implications for biological receptor systems. *Biol. Cybern.*, 83:471–480, 2000.

J. Tougaard. Signal detection theory, detectability and stochastic resonance effects. *Biol. Cybern.*, 87:79–90, 2002.

M. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, London, 1995.

L.M. Ward, A. Neiman, and F. Moss. Stochastic resonance in psychophysics and in animal behavior. *Biol. Cybern.*, 87:91–101, 2002.

K. Wiesenfeld and F. Moss. Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. *Nature*, 373:33–36, 1995.

K. Wiesenfeld, D. Pierson, E. Pantazelou, C. Dames, and F. Moss. Stochastic resonance on a circle. *Phys. Rev. Lett.*, 72:2125–2129, 1994.

T. Zhou, F. Moss, and P. Jung. Escape time distributions of periodically modulated bistable systems with noise. *Phys. Rev. A (3)*, 42:3161–3169, 1990.

Authors' addresses:

Priscilla E. Greenwood
Department of Mathematics and Statistics
Arizona State University
Tempe, AZ, 85287-1804, USA

E-mail: pgreenw@math.la.asu.edu
http://www.math.ubc.ca/people/faculty/pgreenw/pgreenw.html

Ursula U. Müller
Fachbereich 3: Mathematik und Informatik
Universität Bremen
28334 Bremen, Germany

E-mail: uschi@math.uni-bremen.de
http://www.math.uni-bremen.de/~uschi/

Lawrence M. Ward
Department of Psychology
University of British Columbia
Vancouver, B.C., Canada, V6T 1Z4

E-mail: lward@cortex.psych.ubc.ca
http://www.psych.ubc.ca/∼lward/index.html

Wolfgang Wefelmeyer
Fachbereich 6 Mathematik
Universität Siegen
Walter-Flex-Str. 3
57068 Siegen, Germany

E-mail: wefelmeyer@mathematik.uni-siegen.de
http://www.math.uni-siegen.de/statistik/wefelmeyer.html