Precursors of extreme increments

Sarah Hallerberg, Eduardo G. Altmann, Detlef Holstein, and Holger Kantz
Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, D-01187 Dresden, Germany
(Received 22 December 2005; revised manuscript received 12 September 2006; published 17 January 2007)

We investigate precursors and the predictability of extreme increments in a time series. The events we are focusing on consist in large increments within successive time steps. We are especially interested in understanding how the quality of the predictions depends on the strategy to choose precursors, on the size of the event, and on the correlation strength. We study the prediction of extreme increments analytically in an autoregressive process of order 1, and numerically in wind speed recordings and long-range correlated autoregressive moving average processes data. We evaluate the success of predictions via receiver-operator characteristics (ROC curves). Furthermore, we observe an increase of the quality of predictions with increasing event size and with decreasing correlation in all examples. Both effects can be understood by using the likelihood ratio as a summary index for smooth ROC curves.

DOI: 10.1103/PhysRevE.75.016706 PACS number(s): 95.75.Wx, 02.50.−r, 05.45.Tp.

I. INTRODUCTION

Extreme value statistics [1] is a well-established approach to predict the relative frequency of rare extreme events, but does not include forecasts of when the next event will occur. There have been many attempts to employ time series strategies for the latter purpose. These strategies usually investigate a record of historical data about the phenomenon under study and try to infer knowledge about the future. A standard approach is to search for precursors, i.e., typical signatures preceding an extreme event. Such precursors have been discussed, e.g., in the literature about earthquakes [2], epileptic seizures [3], and stock market crashes [4–6]. As the above-listed examples illustrate, the definitions of what an extreme event is depends on the context. Frequently, one encounters extremely large values of some observable, or some drastic changes. It is the latter that is the focus of this paper where we discuss large increments motivated by stock markets or by a turbulent wind gusts in wind speed data.

One might expect that the more extreme an event is, the more difficult it is to predict it, simply because more extreme events are usually also much rarer. However, it has been reported in the literature of wind speed predictions [7], precipitation forecasts [8], multiagent games [11], and earthquakes [12] that more extreme events are better predictable than small events. Therefore one particular goal of this contribution is to investigate how the predictability of large increments depends on the size of the increment.

In this contribution we study predictions in a simple autoregressive process of order 1 [9,10] analytically in order to obtain a detailed understanding of some questions on precursors and predictions. The autoregressive process of order 1 is a simple stationary stochastic model process, which might not reflect all features of more complex processes occurring in nature, but it admits a fully analytic treatment. Additionally, we study similar prediction procedures numerically in long-range correlated data and in wind speed data, verifying the same quantitative results. The questions, which we intend to answer, are the following:

(Q1) How do you choose a precursor in order to obtain good predictions?

(Q2) Are extreme increments the better predictable, the more extreme they are?

(Q3) How does the correlation of the data influence the predictability of extreme increments?

The paper is organized as follows. In Sec. II A we discuss two strategies which can be used to choose precursory structures and in Sec. II B we introduce a method to evaluate the predictive power of precursors. The extreme events we discuss in this contribution are defined in Sec. II C and we show how to obtain their joint PDFs analytically in Sec. II D. We apply these procedures to autoregressive stochastic processes of order 1 in Sec. III, to wind speed measurements in Sec. IV and to long-range correlated data in Sec. V. Conclusions appear in Sec. VI.

II. DEFINITIONS AND SETUP

The considerations in this introductory section are made for general dynamical systems with a complex time evolution. They might be purely deterministic, then high-dimensional and chaotic, or they might be stochastic. In any case we assume that the time evolution of the system cannot be easily modeled and hence one tries to extract information about the future from time series data. This means that through some experimental observation one can record a usually univariate time series, i.e., a set of measurements x_n at discrete times t_n, where $t_n = t_0 + n\Delta$ with a sampling interval Δ. The recording should contain sufficiently many extreme events so that we are able to extract statistical information about them. We also assume that the event of interest can be identified on the basis of the observations, e.g., by the value of the observation function exceeding some threshold, by a sudden increase, or by its variance exceeding again some threshold.

A. Choice of the precursor

Ideally, a precursor is a typical signature in the data preceding *every* individual event. Unfortunately the time evolution of most systems is usually too irregular to demand this, so one would call a precursor a data structure, which is *typi-*

cally preceding an event, allowing deviations from the given structure, but also allowing events without preceding structure. This interpretation of a precursor allows us to determine the specific values of the precursory structure by statistical considerations.

In order to predict an event occurring at the time $(n+1)$ we compare the last k observations $\mathbf{x}_{(n,k)} = (x_{n-k+1}, x_{n-k+2}, \ldots, x_{n-1}, x_n)$ with a specific precursory structure $\mathbf{x}_{pre} = (x^{pre}_{n-k+1}, x^{pre}_{n-k+2}, \ldots, x^{pre}_{n-1}, x^{pre}_n)$.

This precursory structure can be chosen according to different strategies. The two possible strategies, which we address here, represent the most fundamental choices. They consist in using either the maximum of the *a posteriori PDF* or the maximum of the *likelihood* [13]. In more applied examples one looks for precursors that minimize or maximize more sophisticated quantities, e.g., discriminant functions or loss matrices. These quantities are usually functions of the posterior PDF or the likelihood, but they take into account the additional demands of the specific problem, e.g., minimizing the loss due to a false prediction [14]. The two strategies studied in this contribution are thus fundamental in the sense that they enter into most of the more sophisticated quantities, which are used for predictions and decision making.

The *a posteriori* PDF $\rho(\mathbf{x}_{(n,k)}|X)$ takes into account all events of size X and provides the probability density to find a specific precursory structure before an observed event.

(I) Hence strategy I consists in defining the precursors in a retrospective or *a posteriori* way: once the extreme event X has been identified, one asks for the signals right before it. Formally, this implies that the precursory structure consists of the global maxima in each component $(x^*_{n-k+1}, x^*_{n-k+2}, \ldots, x^*_{n-1}, x^*_n)$ of the *a posteriori* PDF.

The likelihood $\rho(X|\mathbf{x}_{(n,k)})$ takes into account all possible values of precursory structures, and provides the probability density that an event of size X will follow them. Note that the likelihood is thus not a density function with respect to the precursory structure, but with respect to the event size X. The precursory structure enters into the likelihood only as a parameter.

(II) Strategy II consists in determining those values of each component x_i of the condition $\mathbf{x}_{(n,k)}$ for which the likelihood has a global maximum.

Note that the *a posteriori* PDF and the likelihood are linked via Bayes's theorem

$$\rho(\mathbf{x}_{(n,k)}, X) = \rho(\mathbf{x}_{(n,k)})\rho(X|\mathbf{x}_{(n,k)}) = \rho(\mathbf{x}_{(n,k)}|X)\rho(X),$$

where $\rho(\mathbf{x}_{(n,k)})$ represents the marginal PDF to find the precursory structure $\mathbf{x}_{(n,k)}$ and $\rho(X)$ represents the marginal PDF to find events of size X.

In summary the possible values of precursors are given by

$$\mathbf{x}_{pre} = \begin{cases} \mathbf{x}_{\mathrm{I}}, \\ \mathbf{x}_{\mathrm{II}}, \end{cases}$$

$$\text{where } \mathbf{x}_{\mathrm{I}} := (x^*_{n-k+1}, x^*_{n-k+2}, \ldots, x^*_{n-1}, x^*_n),$$

$$\text{and } \mathbf{x}_{\mathrm{II}} := (x^\dagger_{n-k+1}, x^\dagger_{n-k+2}, \ldots, x^\dagger_{n-1}, x^\dagger_n), \tag{1}$$

where x^*_i are the points in which $\rho(\mathbf{x}_{(n,k)}|X)$ has a global maximum and $x^\dagger_i$ are the points in which $\rho(X|\mathbf{x}_{(n,k)})$ has a global maximum, with $n-k+1 \leq i \leq n$. In both cases the event size X is assumed to be fixed. Once the precursory structure $\mathbf{x}_{pre}$ is determined, we give an alarm for an extreme event when we find the last k observations $\mathbf{x}_{(n,k)}$ in the volume

$$V_{pre}(\delta) = \left(x^{pre}_{n-k+1} - \frac{\delta}{2}, x^{pre}_{n-k+1} + \frac{\delta}{2}\right) \otimes \left(x^{pre}_{n-k+2} - \frac{\delta}{2}, x^{pre}_{n-k+2} + \frac{\delta}{2}\right) \otimes \cdots \otimes \left(x^{pre}_{n} - \frac{\delta}{2}, x^{pre}_{n} + \frac{\delta}{2}\right). \tag{2}$$

This method of determining the precursor is especially useful if the PDF of a process has one clearly defined maximum. For multimodal PDFs the strategy of using only the global maxima can surely be improved by considering also the influence of smaller maxima of the PDF. In this case the precursory volume could, e.g., consist of $\mathbf{x}_{(n,k)}$ for which the PDFs have values above a certain threshold. In this case $V_{pre}(\delta)$ might not be simple connected, but apart from this the procedure of predicting should not be different. However, we restrict ourselves to unimodal PDFs in this contribution.

B. Testing for predictive power

A common method to verify a hypothesis or test the quality of a prediction is the receiver operating characteristic curve (ROC curve) [15,16]. The idea of the ROC curve consists simply in comparing the rate of correctly predicted events r_c with the rate of false alarms r_f by plotting r_c vs r_f. The resulting curve in the unit square of the r_f-r_c plane approaches the origin for $\delta \to 0$ and the point (1,1) in the limit $\delta \to \infty$, where δ accounts for the size of the precursor volume $V_{pre}(\delta)$ [see Eq. (2)].

The shape of the curve characterizes the significance of the prediction. A curve above the diagonal reveals that the corresponding strategy of prediction is better than a random prediction, which is characterized by the diagonal. Furthermore we are interested in curves that converge as fast as possible to $r_c = 1$, since this scenario tells us that we reach the highest possible rate of correct prediction without having a large rate of false alarms.

There are various so-called *summary indices* [17] that quantify the behavior of the ROC. In this contribution we use the so-called *likelihood ratio* [16] in order to quantify the ROC curve. The likelihood ratio is identical to the slope m of the ROC curve. For the usage as a summary index, we consider the slope in the vicinity of the origin, which implies $\delta \to 0$.

The term likelihood ratio results from signal detection theory in which context the term "*a posteriori* PDF" refers to the PDF, which we call likelihood in the context of predictions, and vice versa. This is due to the fact that the aim of signal detection is to identify a signal that was already observed in the past, whereas predictions are made about future events. Thus the "likelihood ratio" is in our case in fact a ratio of the posterior PDFs, as defined by

$$m = \frac{\Delta r_c}{\Delta r_f} \sim \left. \frac{\rho(\mathbf{x}_{(n,k)}|X)}{\rho(\mathbf{x}_{(n,k)}|\bar{X})} \right|_{\delta \approx 0} + \mathcal{O}(\delta), \quad (3)$$

where $\rho(\mathbf{x}_{(n,k)}|\bar{X})$ denotes the *a posteriori* PDF for non-events. However, we will use the common name likelihood ratio throughout the text.

The likelihood ratio can be expressed in terms of the likelihood $\rho(X|\mathbf{x}_{(n,k)})$ and the total probability to find events $\rho(X)$,

$$m(\mathbf{x}_{(n,k)},X) \sim \frac{[1-\rho(X)]}{\rho(X)} \frac{\rho(X|\mathbf{x}_{(n,k)})}{[1-\rho(X|\mathbf{x}_{(n,k)})]}. \quad (4)$$

If we assume that the events we are observing are quite rare and hence $\rho(X), \rho(X|\mathbf{x}_{(n,k)}) \ll 1$, the likelihood ratio is approximately given by

$$m(\mathbf{x}_{(n,k)},X) \sim \frac{\rho(X|\mathbf{x}_{(n,k)})}{\rho(X)} = \frac{\rho(\mathbf{x}_{(n,k)}|X)}{\rho(\mathbf{x}_{(n,k)})}. \quad (5)$$

Equation (5) already suggests answers to questions (Q1) and (Q2), by considering $m(\mathbf{x}_{(n,k)},X)$ as a summary index.

Addressing (Q1): This asymptotic form of the likelihood ratio allows us to compare different strategies of prediction. Looking for the maximum of $\rho(\mathbf{x}_{(n,k)}|X)$ in $\mathbf{x}_{(n,k)}$, according to strategy I, there is always the influence of the denominator $\rho(\mathbf{x}_{(n,k)})$, which will keep the likelihood ratio small, even if $\rho(\mathbf{x}_{(n,k)}|X)$ in $\mathbf{x}_{(n,k)}$ is maximized. This is due to the fact that $\rho(\mathbf{x}_{(n,k)}|X)$ cannot be large without $\rho(\mathbf{x}_{(n,k)})$ being large. Strategy II, which uses the maximum of $\rho(X|\mathbf{x}_{(n,k)})$ in $\mathbf{x}_{(n,k)}$ should thus be superior, since the denominator $\rho(X)$ is independent of the chosen precursor. The examples that are studied in Sec. III, Sec. IV, and Sec. V support this idea.

Addressing (Q2): According to Eq. (5), the likelihood ratio is larger than unity, if $\rho(\mathbf{x}_{(n,k)},X) > \rho(\mathbf{x}_{(n,k)})\rho(X)$, i.e., if $\mathbf{x}_{(n,k)}$ and X are correlated. This condition can be also written as $\rho(X|\mathbf{x}_{(n,k)}) > \rho(X)$ or as $\rho(\mathbf{x}_{(n,k)}|X) > \rho(\mathbf{x}_{(n,k)})$ using Bayes's theorem. The latter expression states that the *a posteriori* PDF $\rho(\mathbf{x}_{(n,k)}|X)$, i.e., the probability to find the precursor prior to an event should be larger than the probability to find the precursor prior to an arbitrary value. Thus, the condition is fulfilled by choosing the precursor in a reasonable way, e.g., using the maximum of $\rho(\mathbf{x}_{(n,k)}|X)$ in $\mathbf{x}_{(n,k)}$ or the maximum of $\rho(\mathbf{x}_{(n,k)}|X)$.

C. Definition of extreme increments

In this contribution we will concentrate on extreme events that consist in a sudden increase (or decrease) of the observed variable within a few time steps. Examples of this kind of extreme events are the increases in wind speed in Refs. [7,18], but also stock market crashes [4,5] that consist in sudden decreases.

We define our extreme event by an increment $x_{n+1}-x_n$ exceeding a given threshold d,

$$x_{n+1} - x_n \geq d, \quad (6)$$

where x_n and x_{n+1} denote the observed values at two consecutive time steps.

D. Obtaining the analytic expression of the posterior PDFs

A mathematical expression for a filter, which selects the PDF of our extreme events out of the PDFs of the underlying stochastic process, can be obtained through the Heaviside function $\Theta(x_{n+1}-x_n-d)$. This filter is then applied to the joint PDF of a stochastic process.

Since only the time steps (x_n, x_{n+1}) are of relevance for the filtering, we can neglect all previous time steps and apply the filter simply to the joint PDF for (x_n, x_{n+1}), which has then the form $\rho(x_n, x_{n+1}) = \rho(x_n)\rho(x_{n+1}|x_n)$. This implies that we can regard all previous time steps $x_0, x_1, \ldots, x_{n-1}$, on which ρ_n and ρ_{n+1} might depend, as parameters.

The joint PDF of the extreme events $\rho^{\Theta}(x_{n+1}, x_n, d)$ can then be obtained by multiplication with $\Theta(x_{n+1}-x_n-d)$. If the resulting expression is nonzero, the condition of the extreme event (6) is fulfilled and for x_{n+1} and x_n the following relation holds:

$$x_{n+1} = x_n + d + \gamma \quad (\gamma \in \mathbb{R}, \gamma \geq 0). \quad (7)$$

Hence it is possible to express the joint probability density in terms of x_n or x_{n+1} with the new random variable γ. We can then use the integral representation of the Heaviside function with appropriate substitutions to obtain

$$f^{\Theta}(x_{n+1},x_n,d) = \rho(x_n) \int_0^{\infty} \rho(x_n + d + \gamma | x_n) \times \delta[(x_{n+1} - x_n - d) - \gamma] d\gamma. \quad (8)$$

By normalizing this expression with the total probability $\rho_{\Theta}(d)$ to find extreme events of size d or larger we obtain the joint PDF $\rho^{\Theta}(x_n, x_{n+1}, d)$ of all values of x_n and x_{n+1}, which are part of an extreme event. Integrating the resulting joint PDF $\rho^{\Theta}(x_n, x_{n+1}, d)$ over x_{n+1} we find the following expression for the marginal distribution, i.e., the *a posteriori* PDF:

$$\rho(x_n|X(d)) = \frac{\rho(x_n)}{\rho_{\Theta}(d)} \int_0^{\infty} d\gamma \rho(x_n + d + \gamma | x_n). \quad (9)$$

Analogously $\rho(x_n|\overline{X(d)})$ denotes the *a posteriori* PDF to observe the value x_n before a nonevent, i.e., before an increment which is smaller than d.

$$\rho(x_n|\overline{X(d)}) = \frac{\rho(x_n)}{[1-\rho_{\Theta}(d)]} \int_{-\infty}^{\infty} dx_{n+1} \times [1 - \Theta(x_{n+1} - x_n - d)] \rho_{n+1}(x_{n+1}|x_n). \quad (10)$$

If for a given process the joint PDF of two consecutive events is known, we can hence analytically determine $\rho(x_n|X(d))$, $\rho(x_n|\overline{X(d)})$, and $\rho_{\Theta}(d)$.

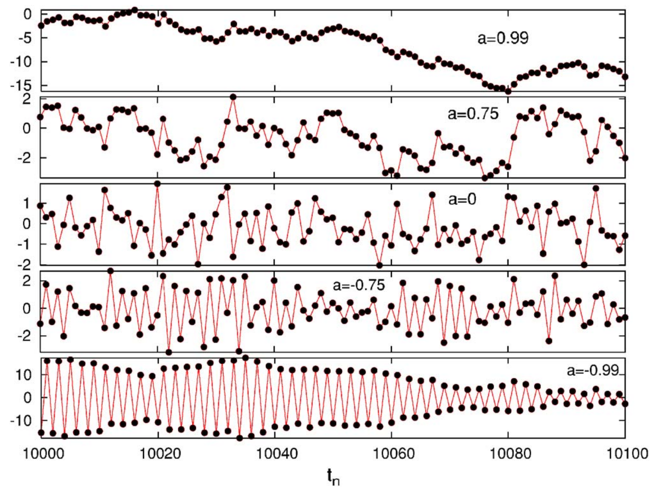

FIG. 1. (Color online) Parts of the time series of the autoregressive process of order 1 for different values of a.

III. EXTREME INCREMENTS IN THE AUTOREGRESSIVE PROCESS OF ORDER 1

A. Autoregressive process of order 1

We assume that the time series $\{x_n\}$ is generated by an autoregressive process of order 1 [AR(1)] (see, e.g., Ref. [9,10])

$$x_{n+1} = a x_n + \xi_n, \tag{11}$$

where ξ_n are uncorrelated Gaussian random numbers with unit variance and $-1 < a < 1$ is a constant that represents the coupling strength. The size and the sign of the coupling strength sets whether successive values of x_n are clustered or spread, as illustrated in Fig. 1.

In the case $a=0$ the process reduces to uncorrelated random numbers with mean $\mu=0$ and variance $\sigma^2=1$, whereas generally the process is exponentially correlated $\langle x_n x_{n+k}\rangle = a^k < 1$ and has the marginal PDF

$$\rho(x_n) = \sqrt{\frac{1-a^2}{2\pi}} \exp\left(-\frac{1-a^2}{2} x_n{}^2\right). \tag{12}$$

Since the size of the events is naturally measured in units of the standard deviation $\sigma(a)$ we introduce a new scaled variable $\eta = \frac{d}{\sigma(a)} = d\sqrt{1-a^2}$.

Applying the filter mechanism developed in Sec. II D we obtain the following expressions for the posterior PDF of extreme events and the posterior PDF of nonextreme events:

$$\rho(x_n|X(\eta),a) = \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}\rho^{\Theta}(a,\eta)} \exp\left(-\frac{1-a^2}{2} x_n^2\right) \times \operatorname{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right), \tag{13}$$

$$\rho(x_n|\overline{X(\eta)},a) = \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}(1-\rho^{\Theta}(a,\eta))} \exp\left(-\frac{1-a^2}{2} x_n^2\right) \times \left[1 + \operatorname{erf}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)\right]. \tag{14}$$

B. Determining the precursor value

Because of the Markov property of the AR(1) process the probability for an event at time $n+1$ depends only on the last value x_n, hence $k=1$ in Eq. (1). Thus, we give an alarm for an extreme event when an observed value x_n is in an interval $V_{pre} = [x_{pre} - \delta/2, x_{pre} + \delta/2]$, around the precursor value x_{pre}. We compute the precursor values x_{I} and x_{II} defined by Eq. (1) according to the strategies described in Sec. II A.

The maximum x_{I} of $\rho(x_n|X(\eta),a)$ is given by the solution of the transcendental equation

$$x_{\mathrm{I}}(\eta) = \frac{\sqrt{2}}{\sqrt{\pi}(1+a)} \frac{\exp\left[-\frac{1}{2}\left((1-a)x_{\mathrm{I}} + \frac{\eta}{\sqrt{1-a^2}}\right)^2\right]}{\operatorname{erfc}\left(\frac{(1-a)x_{\mathrm{I}}}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}. \tag{15}$$

Inserting the asymptotic expansion for large arguments of the complementary error function

$$\operatorname{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi} z}\left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \times 3 \cdots (2m-1)}{(2z^2)^m}\right),$$

$$\left(z \to \infty, |\arg z| < \frac{3\pi}{4}\right), \tag{16}$$

which can be found in Ref. [19], we obtain

$$x_{\mathrm{I}}(\eta) \sim \frac{-\eta}{2\sqrt{1-a^2}\left[1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right]} \quad (\eta \to \infty). \tag{17}$$

Figure 2 shows the posterior PDFs $\rho(x_n|X(\eta),a)$ according to Eq. (13) for different values of a and η. One can see that the maximum of $\rho(x_n|X(\eta),a)$ moves towards $-\infty$ with increasing size of η and $a \to 1$. Although we can always formally define the maximum x_{I} and the mean $\langle x_n\rangle$ as precursor values, one can argue that the maximum of the distribution has no predictive power if $a \to 1$. Since the variance of the posterior PDF increases immensely in this limit, the value of $\rho(x_n|X(\eta),a)$ in its maximum does not considerably differ from the values in any other point.

For large values of η we can also assume that the maximum and the mean of $\rho(x_n|X(\eta),a)$ nearly coincide, i.e.,

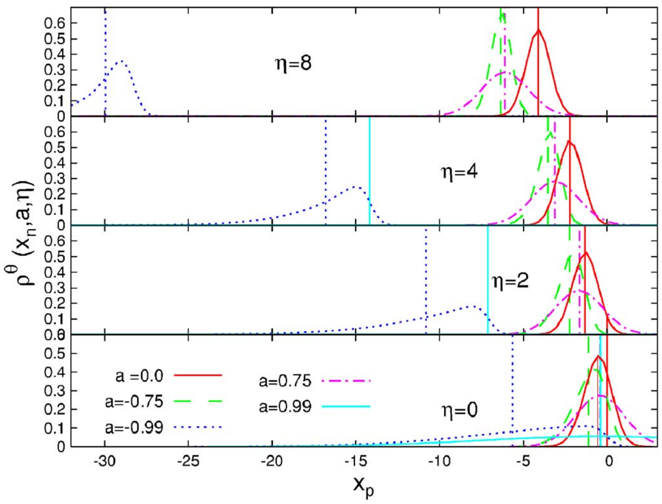

FIG. 2. (Color online) The *a posteriori* PDFs for the autoregressive process of order 1 for different values of $a<0$ and η. The vertical lines represent the means. The PDFs become asymmetric for $a\to-1$. (For $a=-0.99$ and $\eta\to\infty$ the marginal PDFs become very flat and hence cannot be distinguished from the x axis in this figure.)

$$\langle x_n\rangle \simeq x_{\mathrm{I}} \sim \frac{-\eta}{2\sqrt{1-a^2}\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]} \quad (\eta\to\infty), \qquad (18)$$

provided that $\rho(x_n|X(\eta),a)$ is not too asymmetric (i.e., a is not close to -1). In the numerical tests in Sec. III C we will hence use the mean of the posterior PDF as a precursor for strategy I, since it can be calculated explicitly by evaluating the corresponding integral.

In order to determine x_{II}, the precursor for strategy II, we have to find the maximum in x_n of the likelihood

$$\rho(X(\eta)|x_n,a) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a}}\right). \qquad (19)$$

Since the complementary error function is a monotonously decreasing function of x_n we see that we do not have a well-defined maximum x_{II}, (we will thus denote $x_{\mathrm{II}}=-\infty$) and that the interval $V_-=(-\infty,x_-]$ with the upper limit x_- represents the interval for raising alarms according to strategy II.

C. Testing the performance of the precursors

In order to test for the predictive power of the precursors specified above, we used two different methods to create ROC curves (see Sec. II B). The first method consists in evaluating the integrals that lead to the rate of correct and false predictions,

$$r_c(x_{pre},\eta,\delta) = \int_{V(\delta)} dx_n \rho(x_n|X(\eta),a), \qquad (20)$$

$$r_f(x_{pre},\eta,\delta) = \int_{V(\delta)} dx_n \rho(x_n|\overline{X(\eta)},a). \qquad (21)$$

The second method consists in simply performing predictions on a time series of 10^7 AR(1) data and counting the number of extreme increments, which could be predicted by using the precursors specified above. For different values of the correlation coefficients the data sets contained the following numbers of extreme increments:

	Number of increments of size			
a	$\eta\geq 0$	$\eta\geq 2$	$\eta\geq 4$	$\eta\geq 8$
−0.99	5000059	1579103	222858	310
−0.75	5000563	1425146	162405	107
0	5000417	786355	23370	0
0.75	5000818	23377	0	0
0.99	5001081	0	0	0

In all cases, where the AR(1)-correlated data sets contain increments, the empirically determined rates comply very well with the rates obtained via the evaluation of Eqs. (20) and (21). For those values of a and η, which were not accessible for the numerical test, we evaluated the integrals in Eqs. (20) and (21).

In the numerical tests for both strategies and also for the evaluation of the integrals in Eqs. (20) and (21) according to strategy I, the size of the precursory volume ranged from 10^{-6} to 4, measured in the size of the standard deviation of the marginal PDF of the AR(1) $\sigma(a)=1/\sqrt{1-a^2}$. As precursors according to strategy I we used the means of the *a posteriori* PDF. For the empirically created ROC plots according to strategy II we used the smallest values of the data sets as precursors.

The evaluation of the integrals in Eqs. (20) and (21) was done in a slightly different way for strategy II. Since there were no events in the data sets for certain values of a and d (as indicated in the table above), one could argue that the data sets also did not contain any precursor. From the previous section, we know that the theoretical precursor value according to strategy II should be $x_{\mathrm{II}}=-\infty$. Thus, we used a sufficiently small value as a precursor and adjusted the size of the prediction interval in order to capture all events. However, the resulting ROC curves for strategy II coincided with the curves obtained empirically, as far as they were available.

The resulting ROC curves in Fig. 3 display the following properties:

Addressing (Q1): The predictions according to strategy II are better than the predictions according to strategy I for all values of a and η.

Addressing (Q2): The ROC curves display an increase of the quality of our prediction with the increasing size of the events η.

Addressing (Q3): The ROC curves in Fig. 3 show that the quality of the predictions increases with decreasing correlation strength a. Especially for $a=0$, when the predictions were made within completely uncorrelated random numbers,

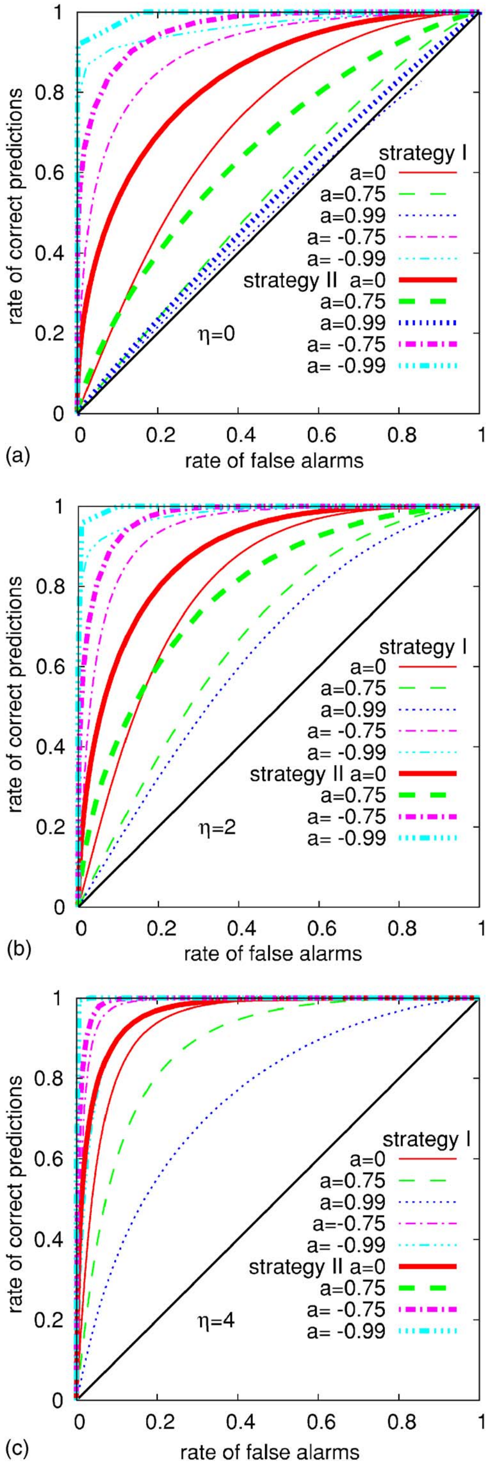

FIG. 3. (Color online) The ROC curves made for the precursors of strategies I and II. The thin lines represent the results of strategy I; the bold lines correspond to predictions made according to strategy II. In both cases the predictions were made within 10^7 AR(1)-correlated data. For the values of η and a, where the data sets contained no increments, we created the ROC curves by evaluating the integrals in Eqs. (20) and (21).

the ROC curves are far better than ROC curves for any random prediction. This is in agreement with results reported in Ref. [21] for the prediction of signs of increments in uncorrelated random numbers, i.e., the case ($a=0$, $\eta=0$).

Intuitively, the result for (Q3) can be understood easily by considering that increments are not independent from the last observation. More precisely, $x_{n+1}-x_n=(a-1)x_n+\xi_n$, so that the known part of the increment $(a-1)x_n$ is the larger, the smaller a. In other words, if we consider a very small value of x_n (small compared to the mean) in an uncorrelated process, the probability that the next value will be closer to the mean and hence lead to a large increment is high. Positive correlation hinders this effect, since it causes successive values to be closer to each other.

A formal explanation of the results (Q1)–(Q3) is also given by an asymptotic expression for the slope $m(a,\eta,x_{pre})$ in the following section.

D. Analytical discussion of the precursor performance

In this section, we will try to understand the effects shown by the ROC curves in the previous section in more detail. Thus, we evaluate the asymptotic structure of the likelihood ratio as defined by Eq. (3) for different scenarios.

In the case of the AR(1) process the slope of the ROC curve in the vicinity of the origin is given by

$$m(a,\eta,x_{pre}) \sim \frac{[1-\rho_\Theta(\eta)]}{\rho_\Theta(\eta)} r(x_{pre},\eta), \tag{22}$$

$$\text{with } r(x_{pre},\eta) = \frac{\operatorname{erfc}\left(\frac{(1-a)x_{pre}}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}{1+\operatorname{erf}\left(\frac{(1-a)x_{pre}}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}. \tag{23}$$

Addressing (Q1): We will first consider the behavior of the precursor according to strategy II. As we saw in Sec. III B, the optimal precursor value of strategy II is the limiting case $x_{\mathrm{II}}=-\infty$.

Since $\lim_{x_{pre}\to-\infty} r(x_{pre},\eta)=\infty$ we find $\lim_{x_{pre}\to-\infty} m(a,\eta,x_{\mathrm{II}})=\infty$. Thus, we should expect ROC curves made with $x_{\mathrm{II}}=-\infty$ to be tangent to the vertical axis of the curve and hence represent an ideal predictability for all sizes of events and all possible correlation strengths. However, for any finite precursor value of strategy I and strategy II we find nonideal ROC curves.

Another way to understand the superiority of strategy II is to analyze the asymptotic behavior of the rate of correct predictions $\rho(x_n|X(\eta),a)$ and the rate of false alarms, $\rho(x_n|\overline{X(\eta)},a)$ at the precursor value of strategy I. For the following calculations we use an approximation for the total probability to observe extreme events:

$$\rho_\Theta(\eta,a) \sim \frac{\sqrt{1-a}}{\sqrt{\pi}}\frac{1}{\eta}\exp\left(-\frac{\eta^2}{4(1-a)}\right)\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right] \quad (\eta\to\infty), \tag{24}$$

which is derived in Appendix A.

Inserting the asymptotic expression for $\rho_\Theta(\eta,a)$, the approximation of x_{I} in Eq. (A3) and the asymptotic expansion

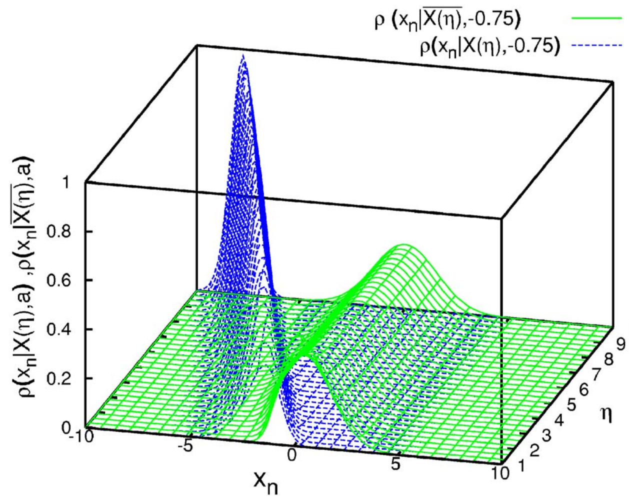

FIG. 4. (Color online) $\rho(x_n|X(\eta),a)$ and $\rho(x_n|\overline{X(\eta)},a)$ for $a=-0.75$. The maximum of the posterior PDF to observe extreme events $\rho(x_n|X(\eta),a)$, which is used as a precursor, moves towards $-\infty$ with increasing η since $x_\mathrm{I}\sim-\eta/[2(1-a^2)]$. Because the maximum of the failure posterior PDF $\rho(x_n|\overline{X(\eta)},a)$ remains at the origin, the values of $\rho(x_n|\overline{X(\eta)},a)$, which are observed at the precursor value x_I decrease according to the decrease of $\rho(x_n|\overline{X(\eta)},a)$ as $x_n\to-\infty$.

of the complementary error function Eq. (16) into Eqs. (13) and (14), we find the following expressions:

$$\rho(x_\mathrm{I}|X(\eta),a)\sim\frac{\sqrt{1-a^2}\sqrt{1+a}}{\sqrt{\pi}}\frac{\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]}{\left[1+a+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]}\quad(\eta\to\infty). \tag{25}$$

$$\rho(x_\mathrm{I}|\overline{X(\eta)},a)\sim\frac{\sqrt{1-a^2}}{\sqrt{2\pi}}\exp\left(-\frac{\eta^2}{8}\frac{1}{\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]}\right)\quad(\eta\to\infty). \tag{26}$$

Hence the value of $\rho(x_n|X(\eta),a)$ at the precursor value approaches a constant for large η, whereas the values of $\rho(x_n|\overline{X(\eta)},a)$ decrease exponentially in this limit. Figure 4 illustrates this effect for the case $a=-0.75$. The maximum of the failure PDF remains at the origin for $\eta\to\infty$. Thus the values of this PDF, which are observed at the decreasing precursor value $x_\mathrm{I}\propto\frac{-\eta}{2\sqrt{1-a^2}}$, decrease according to the shape of the distribution. This explains also the success of strategy II. Since the precursor value obtained by strategy II is the smallest possible value, strategy II seems to focus on the minimization of the failure rate. Note that by "minimization of the failure rate," we understand here a minimization of the integrand in Eq. (21), while the alarm interval of size δ remains constant. The fact that in this point the corresponding value of $\rho(x_n|X(\eta),a)$ is also far away from the maximum of $\rho(x_n|X(\eta),a)$ does not apparently influence the outcome of the prediction.

Addressing (Q2): In the following calculation we will obtain the asymptotic form of the likelihood ratio for large events. Inserting the asymptotic form of the probability $\rho_\Theta(\eta,a)$ provided by Eq. (A4), and using the asymptotic expansion of the complementary error function in Eq. (16), the likelihood ratio reads

$$m(a,\eta,x_{pre})\sim\frac{1}{2\sqrt{1-a}}\frac{\eta\exp\left(\frac{\eta^2}{4(1-a)}-z(\eta,a)^2\right)\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]}{z(\eta,a)\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]+\mathcal{O}[\exp(-z(\eta,a)^2)]}+\mathcal{O}\left(\frac{\exp(-z(\eta,a)^2)}{z}\right)(\eta\to\infty),\quad[z(\eta,a)\to\infty]$$

$$\text{with}\quad z(\eta,a)=\frac{(1-a)}{\sqrt{2}}x_{pre}+\frac{\eta}{\sqrt{2}\sqrt{1-a^2}}. \tag{27}$$

Note that the limit $z(\eta,a)\to\infty$ corresponds to the limit $\eta\to\infty$ in the context of (Q2), but we can also interpret it as the limit $a\to\pm1$ in the context of (Q3) if $\eta\neq0$.

The expression in Eq. (27) tends to infinity in the limit $\eta\to\infty$, if the argument of the exponential function in Eq. (27)

$$f(x_{pre},a,\eta)=\frac{\eta^2}{4(1-a)}-\left(\frac{(1-a)x_{pre}}{\sqrt{2}}+\frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)^2 \tag{28}$$

is positive. This is indeed the case for every precursor value $x_{pre}<0$. Therefore, for both strategies of prediction, the slope $m(x_{pre},a,\eta)$ increases as a squared exponential with increasing size of the events η according to Eq. (27). Hence, the considerations of Sec. II B hold for our example, according to which an event is the better predictable the more rare it is.

Addressing (Q3): One can also calculate the asymptotic behavior of the likelihood ratio for $a\to\pm1$. The limit $z(\eta,a)\to\infty$, which is relevant for the asymptotic form in Eq. (27), can also be interpreted as the limit $a\to\pm1$. We assume that η is big enough, e.g., $\eta>2$, such that Eq. (A4), which enters into Eq. (27), is a useful approximation. One can now discuss again the argument of the exponential function in Eq. (28).

Inserting the precursor of strategy I [as given by Eq. (17)], one obtains $f(x_\mathrm{I},a,\eta)=\frac{\eta^2}{8}$, hence

$$m(a,\eta,x_\mathrm{I})\to\sqrt{\frac{2}{1+a}}\exp\left(\frac{\eta^2}{8}\right)\quad[z(\eta,a)\to\infty]. \tag{29}$$

As $a\to1$, this expression converges to $\exp(\eta^2/8)$. As $a\to-1$, this expression approaches infinity as $m(1,\eta,x_\mathrm{I})\sim1/\sqrt{1+a}$. Figure 5(a) illustrates this behavior. Figure 5(b) shows that the asymptotic expression in Eq. (29) becomes better in the limit $\eta\to\infty$, since in this limit the higher-order terms of the approximation vanish even faster.

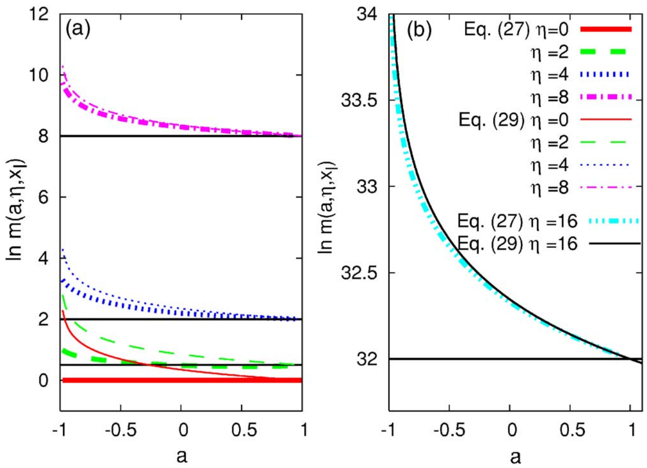

FIG. 5. (Color online) The bold lines show the dependence of the slope $m(x_I, a, \eta)$ on the coupling strength according to Eq. (27). The thin lines display the asymptotic behavior given by Eq. (29). The constant lines represent the values, obtained from Eq. (29) in the limit $a \to 1$. (b) illustrates that this asymptotic expression becomes better in the limit $\eta \to \infty$, since in this limit the higher-order terms in the approximation vanish even faster.

For the theoretical precursor of strategy II, $x_{II} = -\infty$, the slope would be independent of the value of the coupling strength if the exact precursor of strategy II could be used. For any real precursor value of strategy II, $x_{II} = \text{const} < 0$, Eq. (28) reads

$$f(x_{II}, a, \eta) \sim \frac{\eta^2}{2(1-a)} \left(\frac{1}{2} - \frac{1}{1+a} \right) + \mathcal{O}[(1-a)] \quad (a \to 1). \tag{30}$$

This expression approaches a small negative value close to zero in the point $a=1$. Hence, we find $m(a, \eta, x_{II}) \sim 1$, as $a \to 1$.

In the limit $a \to -1$ and for any finite precursor value $x_{II} = \text{const} < 0$, Eq. (28) reads

$$f(x_{II}, a, \eta) \sim \frac{\eta^2}{4} \left(\frac{1}{2} - \frac{1}{1-a^2} \right) - \frac{2x_{II}\eta}{\sqrt{1-a^2}} - 2x_{II}^2$$
$$\sim -\frac{1}{1-a^2}\frac{\eta^2}{4} - \frac{2x_{II}\eta}{\sqrt{1-a^2}} - 2x_{II}^2. \tag{31}$$

If the precursor is sufficiently small, e.g., $x_{II} < -\eta/(4\sqrt{1-a^2})$, this expression is positive and hence $m(a, \eta, x_{II}) \to \infty$, as $a \to -1$.

Hence, the asymptotic expressions of the likelihood ratio are able to describe the behavior of the ROC curves, shown in the previous section. Figure 6 combines the dependence of the likelihood ratio on the event size and the correlation strength. One can see that the influence of the event size on the likelihood ratio is dominating, as long as one does not approach the singularity at $a \to -1$.

IV. APPLICATION: WIND SPEED MEASUREMENTS

As an illustration of the preceding considerations and also in order to demonstrate the usefulness of the benchmarks

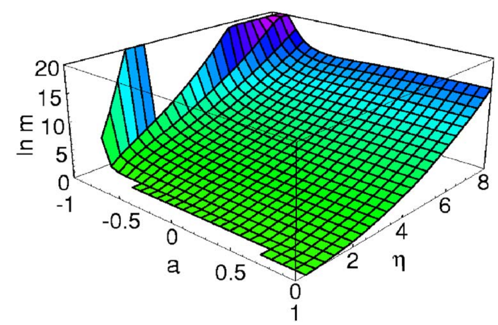

FIG. 6. (Color online) The asymptotic dependence of the slope $m(x_I, a, \eta)$ on the coupling strength and the event size, if the precursor of strategy I is used.

derived for AR(1) processes, we study here time series data of wind speed measurements. The data are recorded at 30 m above ground by a cup anemometer with a sampling rate of 8 Hz in the Lammefjord site of the Risø research center [22]. Wind speed data are evidently nonstationary and strongly correlated, so that, e.g., the principle of persistence yields surprisingly accurate forecasts: the very simple prediction scheme $\hat{x}_{n+1} = x_n$ is almost as accurate as an AR(20) process fitted on moving windows (in order to take nonstationarity into account) or order-ten Markov chains [18]. The amplitude of the fluctuations around a time local mean value are proportional to this mean value, i.e., there is statistical evidence that the noise in this process is multiplicative. However, when subtracting the time local mean (more precisely, performing a high-pass filtering with a Gaussian kernel with a standard deviation of 75 time steps), we receive data for which it is reasonable to fit an AR(1) process. When doing so, we find a coefficient $a \approx 0.94$.

Turbulent gusts, i.e., sudden increases of the wind speed, are relevant events, e.g., for the save operation of wind turbines, for aircrafts during takeoff and landing, and for all wind-driven sports activities. In previous work [7] we were therefore concerned with their prediction, where we were studying the performance of a Markov chain model. Here, we will restrict ourselves to the simpler (and less appropriate) AR(1) philosophy: The current state of the process generating the wind time series is assumed to be fully specified by the last observation x_n, and the event is assumed to be characterized by the upward jump of the wind speed in a single time step by more than g m/s.

A. Determining the precursor value

If we extract from the data set all subsequences of data where such a jump is present, then we can, in principle, construct empirically the distribution $p(x_n|g)$, which corresponds to $\rho(\mathbf{x}_{(n,k)}|X)$ of strategy I.

In Fig. 7 we show instead the mean value of $p(x_{n+k}|g)$ for $k = -20, \ldots, 20$, i.e., we show the mean profile of gusts of strength g. Otherwise said, this is an average of all those time series segments, which (in shifted time) fulfill $x_1 - x_0 > g$, so that the part of these segments with $k \leq 0$ is what one would call naively a precursor of a gust event. This has to be compared to the values x_{n+k}, which we find when we focus on the

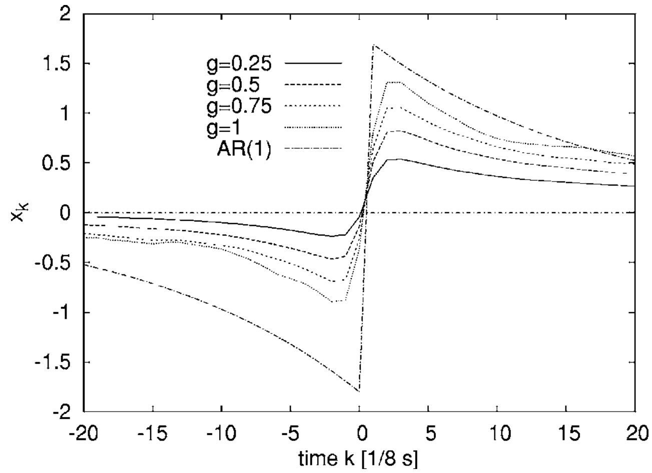

FIG. 7. The profiles obtained from the mean of $p(x_{n+k}|g)$ for gust events of amplitude g. Also shown is the theoretical profile for an AR(1) process with $a=0.94$.

maximum x_{II} in x_n of $p(g|x_n)$, which corresponds to the conditional probability $\rho(X|x_n)$ of strategy II. More specifically, in Fig. 8 we show the profiles $\langle x_{n+k}\rangle|_{x_n=x_{II}}$, where x_{II} is defined by $p(g|x_{II})=\max_{x_n}$. In even different words, the value plotted at $k=0$ is the value x_n for which $p(g|x_n)$ is maximal, and at the preceding and succeeding time steps we show the average over all time series segments that fulfill $x_n=x_{II}$ is some precision. These profiles differ from the precursors shown before, as we have to expect for an AR(1) process: In a perfect autoregressive process of order 1, the precursors equivalent to those in Fig. 7 would show a jump larger than g from $k=0$ to $k=1$, with $x_0=-x_1$, and with $x_k=a^k x_0$ for $k<0$, and $x_k=a^k x_1$ for $k>1$. For the same idealized process, one expects Fig. 8 to show curves given by $x_k=a^{|k|}x_{II}$ for all k. Evidently, the wind data show a qualitatively very similar behavior, whereas, however, additional correlations are visible.

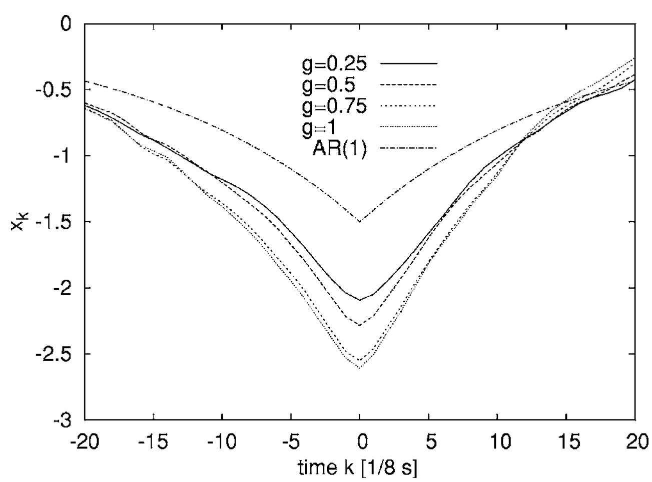

FIG. 8. The profiles obtained from the maxima of $p(g|x_{n+k})$ for gust events of amplitude g. Also shown is the theoretical profile for an AR(1) process with $a=0.94$.

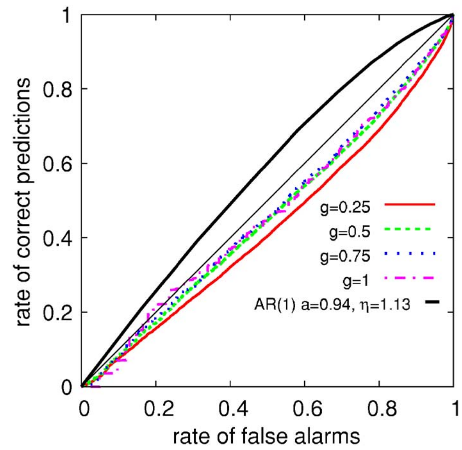

FIG. 9. (Color online) The ROC curves using strategy I, exploiting $p(x_n|X)$ and maximizing the hit rate. Evidently, the rate of false alarms exceeds the hit rate.

B. Testing for predictive power

The ROC curves for the two prediction strategies are shown in Figs. 9 and 10. As expected, the minimization of false alarms (strategy II) is here superior, as strategy I has no predictive power. The latter is consistent with the observed value $a\approx 0.94$ and the results for the autoregressive process of order 1.

In order to compute the ROC curves we use the following numerically expensive but theoretically best justified algorithm: In theory, we want to generate an alarm if the current observation x_n lies in an interval V, which is defined by the subset of the $\mathbb{R}$ where either $p(g|x_n)$ or $p(x_n|g)$ exceeds some threshold $0\leqslant p_c\leqslant 1$. We assume that both conditional PDFs are smooth in x_n.

We can locally approximate $p(g|x_n)$ by searching all similar states x_j, with $|x_n-x_j|<\epsilon$ and counting the relative number of events in this set of states. When this number exceeds p_c, we give the alarm and can see whether it is a hit or a false alarm.

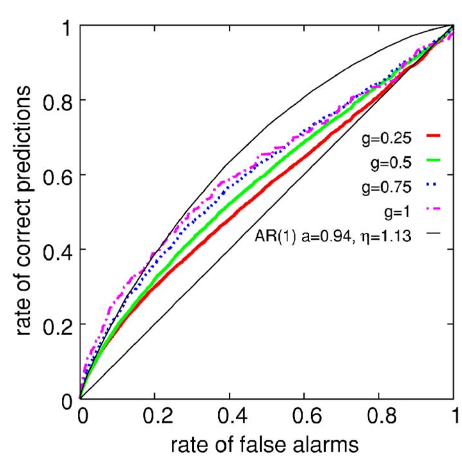

FIG. 10. (Color online) The ROC curves for the prediction of jumps of amplitude larger than g for the wind data. Strategy II exploits $p(X|x_n)$, which minimizes the false alarm rate and performs the better the larger g.

In order to evaluate $p(x_n|g)$ we first create the set of all states x_e, which are preceding an event, and then compute the fraction of these, which is ϵ close to the current state x_n. Since this fraction evidently depends on the value of ϵ, we should introduce a normalization. However, in order to create the ROC statistics we just have to introduce a threshold, which runs from zero to the largest value thus found. Both schemes can be straightforwardly generalized to situations where the current state of the process is defined by a sequence $\mathbf{x}_{(n,k)}$ of k past measurements $(x_{n-k+1},x_{n-k+2},\ldots,x_{n-1},x_n)$, e.g., for an AR(2) process $k=2$, whereas in Ref. [7] we were using $k=10$ for a Markov chain of order ten.

Since the wind speed data are strongly correlated, $a \approx 0.94$, it is not possible to predict the increments of the data sufficiently well. This corresponds to the previously derived results for the AR(1) process in the limit $a \rightarrow 1$. However, we also find deviations from the theoretical ROC curve for $a = 0.94$, which is additionally plotted in Figs. 9 and 10. These deviations show that the AR(1) process is not able to describe the wind data completely.

The wind data also show the increase of predictability with increasing event size. This suggests that this effect is more general and not limited to the class of AR(1) processes. Again, we also observe that strategy II is superior to strategy I.

V. EXTREME INCREMENTS IN LONG-RANGE CORRELATED PROCESSES

We studied the same questions, which are described before, in long-range correlated processes. Since the precursors we were interested in live on a very short time scale (one step before the event), one should not expect long-range correlations to lead to qualitatively different results for the aspects we were interested in. The results obtained in this section support this assumption.

There are various definitions of long-range correlation. Typically long-range correlation in a time series is characterized by the exponent $0<\gamma_c<1$ of the power-law decay of the autocorrelation function as a function of the time t,

$$C_x(t) = \langle x_n x_{n+t}\rangle = \frac{1}{N-t}\sum_{n=1}^{N-t} x_n x_{n+t} \sim t^{-\gamma_c}. \quad (32)$$

The correlation coefficient γ_c is controlling how fast the correlations decay.

We study the predictability of increments numerically by applying the prediction strategies described in Sec. II A. The data used for this numerical study were generated as described in Ref. [23] and used in Ref. [24]: Imposing a power-law decay on the Fourier spectrum,

$$f_x(k) \propto k^{-\beta}, \quad (33)$$

with $0<\beta<0.5$ and choosing phase angles at random one obtains through an inverse Fourier transform the long-range correlated time series in x with $\gamma_c=1-2\beta$. The data are Gaussian distributed with $\langle x\rangle=0$, $\sigma=1$. Having specified the power spectrum or, correspondingly, the autocorrelation function for sequences of Gaussian random numbers means to have fixed all parameters of a linear stochastic process. Hence, in principle the coefficients of an autoregressive or moving average process can be uniquely determined, where, due to the power-law nature of the spectrum and autocorrelation function, the order of either of these models has to be infinite [9,10]. Thus, the effects that we observed for this autoregressive moving average process of order ∞ and ∞ [ARMA(∞,∞) process] model should be valid for the whole class of linear long-term correlated processes.

The ROC curves in Fig. 11, which are generated from the long-range correlated data are very similar to the ones for the autoregressive process in order 1 in terms of the question we want to study.

Addressing (Q1): The ROC curves obtained by using strategy II are superior to the curves resulting from strategy I.

Addressing (Q2) and (Q3): The quality of the prediction also increases with increasing event size and decreasing correlation.

Hence we observe the same effects that we described before for the autoregressive process in order 1 and the wind speed data in a long-range correlated ARMA(∞,∞) process.

VI. CONCLUSIONS

We studied the predictability of extreme increments in an AR(1)-correlated process, in wind speed data, and in a long-range correlated ARMA process. To measure the quality of the prediction we used the ROC curve and additionally the slope of the ROC curve in the vicinity of the origin as a summary index. This so-called *likelihood ratio* characterizes particularly the behavior in the limit of low false-alarm rates.

In the case of the AR(1) process we could construct the posterior PDF and the likelihood analytically from a given joint PDF and hence we were able to obtain the asymptotic behavior of the likelihood ratio analytically. In the case of the two other examples, we constructed the posterior PDFs numerically. The resulting distributions were then used to determine precursors according to two different strategies of prediction.

In all examples we studied the aspects: (Q1) Which is the best strategy to choose precursors? (Q2) How does the predictability depend on the event size? (Q3) And how does the predictability depend on the correlation? The results can be summarized as follows:

Addressing (Q1): Strategy I, the *a posteriori* approach, maximizes the rate of correct predictions, while strategy II focuses on the minimization of the rate of false alarms. For the example of the AR(1) process one can show that strategy II is the optimal strategy to make predictions. For other stochastic processes, it is not in general clear which of the two strategies leads to a better predictability. However, the application to the prediction of wind speeds and the numerical study within long-range correlated data reveals that also for these examples better results are obtained by predicting according to strategy II.

Addressing (Q2): For all examples studied, we observe an increase of predictability with increasing size of the events.

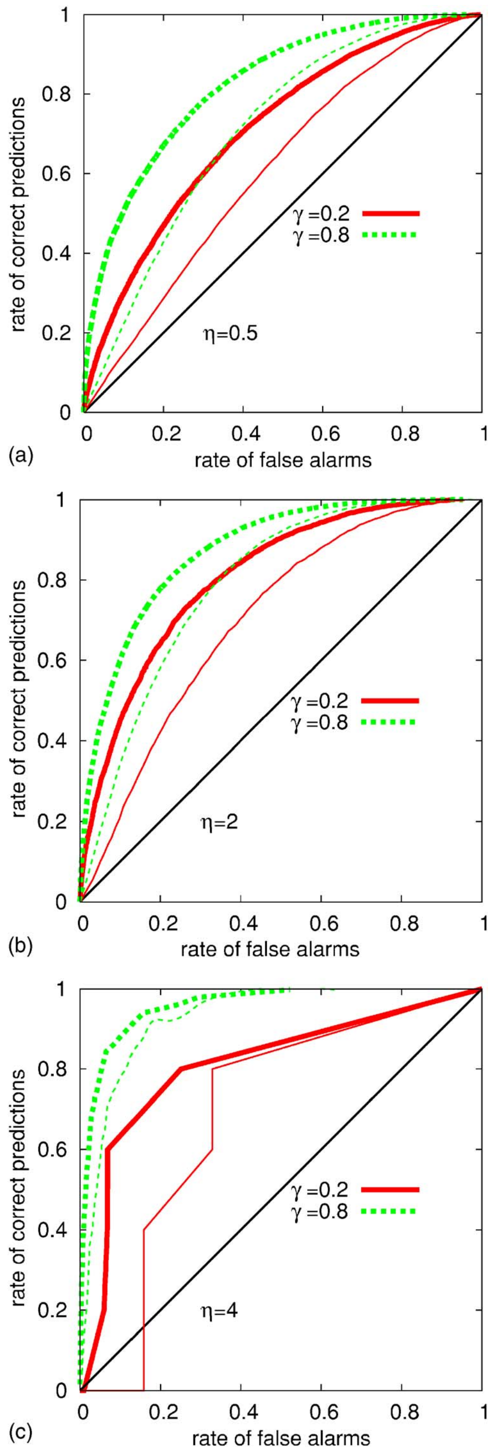

FIG. 11. (Color online) ROC curves for the ARMA(∞,∞) processes with γ_c=0.2 and γ_c=0.8. For η=4 and γ_c=0.2 only few data points were available.

This phenomenon, which is also reported in the literature [7,8,11], can be better studied by investigating the asymptotic behavior of our summary index. In the case of the AR(1) process we showed explicitly that the likelihood ratio increases as a squared exponential with increasing event size. In Sec. II B we discussed for a general stochastic process that this effect appears, if the PDFs of the studied process fulfill certain conditions.

Addressing (Q3): For the AR(1) process and the long-range correlated data we observe that the correlation of the data is inversely proportional to the quality of the predictions. The ROC curves for the wind data, which we assume to be a strongly correlated AR(1) process with correlation strength a=0.94, display also a bad predictability. This effect is due to the special definition of the events as increments. The asymptotic expression for the likelihood ratio in Eq. (27) provides us also with a formal understanding of the a dependence.

All the considerations made in this contribution are made for a very simple but general method. In order to make predictions, we use the largest maximum of the *a posteriori* PDF or the likelihood. For multimodal distributions, one can think about more sophisticated methods, which take into account also other maxima of the distribution. Furthermore, we investigate only stationary processes in these contributions. It remains to be studied, whether the answers obtained to the questions (Q1)–(Q3) are also valid for nonstationary processes or multimodal distributions.

ACKNOWLEDGMENT

E.G.A. was supported by CAPES (Brazil).

APPENDIX A: OBTAINING AN ASYMTOTIC FORM OF THE TOTAL PROBABILITY TO FIND INCREMENTS OF SIZE η

The total probability $\rho_\Theta(\eta,a)$ to find increments of size η can be obtained by integrating the preform of the posterior probability in Eq. (8). For the example of the AR(1) process the corresponding integral reads

$$\rho_\Theta(\eta,a) = \int_{-\infty}^{\infty} \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}} \exp\left(-\frac{1-a^2}{2}x_n^2\right) \times \operatorname{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right). \qquad \text{(A1)}$$

In the special case η=0 one can find the analytical form of the total probability $\rho_\Theta(0,a)$ using again an integral identity from Ref. [20]. The resulting value $\rho_\Theta(0,a)=1/2$ corresponds to the intuitive expectation one would have, since for η=0 the condition of our extreme event is always fulfilled if x_{n+1} is larger than x_n. This special case of predicting the sign of increments in uncorrelated data is discussed in Ref. [21].

For $\eta \neq 0$, we can find an asymptotic form of the total probability $\rho_\Theta(\eta,a)$ via evaluating the mean of the posterior PDF. An analytic expression of the mean can be obtained using an integral representation from [20]

$$\langle x_n \rangle = \frac{-\exp\left(-\frac{\eta^2}{4(1-a)}\right)}{2\sqrt{\pi}\sqrt{1+a}\,\rho_\Theta(\eta,a)}. \qquad \text{(A2)}$$

For large values of η we can also assume that the maximum and the mean of $\rho(x_n|X(\eta),a)$ nearly coincide, i.e.,

$$\langle x_n \rangle \simeq x_{\rm I} \sim \frac{-\eta}{2\sqrt{1-a^2}\left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right]} \quad (\eta \to \infty), \quad \text{(A3)}$$

provided that $\rho(x_n | X(\eta), a)$ is not too asymmetric (i.e., a is not close to -1). Using this approximation, we find the following asymptotic form of the total probability to find increments of size η:

$$\rho_\Theta(\eta, a) \sim \frac{\sqrt{1-a}}{\sqrt{\pi}} \frac{1}{\eta} \exp\left(-\frac{\eta^2}{4(1-a)}\right) \times \left[1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right] \quad (\eta \to \infty). \quad \text{(A4)}$$

APPENDIX B: TRANSFORMATION OF EXTREME INCREMENTS INTO EXTREME VALUES

We show how to relate the results obtained using the definition of extreme events as extreme increments [$x_{n+1} - x_n \geq d$, as in Eq. (6)] to the case when extreme events are defined as extreme values ($y_{n+1} \geq d$), which exceed a certain threshold d, for autoregressive moving average processes of orders p and q [ARMA(p,q) processes]. An ARMA(p,q) process is defined as [9]

$$\Phi(B) x_n = \theta(B) \xi_n, \quad \text{(B1)}$$

where $\{\xi\}$ correspond to white noise and

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p,$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q,$$

with $B^j x_n = x_{n-j}$. Searching for extreme increments in a time series $\{x\}$ is equivalent to search for extreme values in the time series $\{y\}$, defined through the transformation

$$y_{n+1} = x_{n+1} - x_n. \quad \text{(B2)}$$

Assuming that $\{x\}$ is described by an ARMA(p,q) process defined by Eq. (B1), and inserting Eq. (B2) in Eq. (B1), one obtains that $\{y\}$ is described by an ARMA($p,q+1$) model with the following transformed coefficients:

$$\Phi_i^\dagger = \Phi_i, \quad i = 1,2, \ldots p,$$

$$\theta_i^\dagger = \theta_i - \theta_{i-1}, \quad i = 1,2, \ldots q,$$

$$\theta_{q+1}^\dagger = \theta_q. \quad \text{(B3)}$$

Due to the transformation (B2) the precursory structure equivalent to the one used in Sec. III is obtained choosing [25]

$$y_{pre} = \sum_{j=0}^{n} y_j - x_0 = x_n. \quad \text{(B4)}$$

With this choice of precursory structure and the corresponding transformation of the process [Eq. (B2)], the results obtained for extreme increments can be transferred to the case of extreme values. In particular, for the case of autoregressive processes in order 1 [which corresponds to an ARMA(1,0)] discussed in Sec. III, all results are also valid for an ARMA(1,1) process with the precursor given by Eq. (B4) and events defined as extreme values. For example, the alarm strategies consist in this case in raising an alarm whenever y_{pre} falls near the precursor values given in Eq. (1).

[1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, New York, 2001).

[2] David D. Jackson, Proc. Natl. Acad. Sci. U.S.A. **93**, 3772 (1996).

[3] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, Clin. Neurophysiol. **116**, 569 (2005).

[4] A. Johansen and D. Sornette, Eur. Phys. J. B **1**, 141 (1998).

[5] N. Vandewalle, M. Ausloos, and P. Boveroux *et al.*, Eur. Phys. J. B **4**, 139 (1998).

[6] D. Sornette, Proc. Natl. Acad. Sci. U.S.A. **99**, 2522 (2002).

[7] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, Physica A **342**, 315 (2004).

[8] M. Göber, C. A. Wilson, S. F. Milton, and D. B. Stephenson, J. Hydrol. **288**, 225 (2004).

[9] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1994).

[10] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer, New York, 1998).

[11] D. Lamper, S. D. Howison, and N. F. Johnson, Phys. Rev. Lett. **88**, 1 (2002).

[12] M. Reza Rahimi Tabar, M. Sahimi, K. Kaviani, M. Allamehzadeh, J. Peinke, M. Mokhtaru, M. Vesaghi, M. D. Niry, F. Ghasemi, A. Bahraminasab, S. Tabatabai, S. Fayazbakhsh, and M. Akbari, Dynamics of the Markov Time Scale of Seismic Activity May Provide a Short-Term Alert for Earthquakes, in *Modelling Critical and Catastrophic Phenomena in Geoscience: A Statistical Physics Approach*, edited by P. Bhattacharyya and B. K. Chakrabarti (Springer Verlag, Berlin, 2006), p. 281–301.

[13] J. M. Bernado and A. F. M. Smith, *Bayesian Theory* (Wiley, New York, 1994).

[14] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New York, 1995).

[15] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).

[16] J. P. Egan, *Signal Detection Theory and ROC Analysis* (Academic Press, New York, 1975).

[17] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction* (Oxford University Press, New

York, 2003).

[18] Holger Kantz, Detlef Holstein, Mario Ragwitz, and Nikolay K. Vitanov, *Short Time Prediction of Wind Speeds from Local Measurements*, in Wind Energy—Proceedings of the Euromech Colloquium, edited by J. Peinke, P. Schaumann, and S. Barth (Springer, New York, 2006).

[19] M. Abramowitz, and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[20] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series Vol. II. Special Functions* (Gordon and Breach Science Publishers, New York, 1992).

[21] D. Sornette and J. V. Andersen, Int. J. Mod. Phys. C **11**, 713720 (2000).

[22] The wind-speed data were recorded at the Risø Research National Research Laboratory in Denmark; http://www.risoe.dk/vea; see also http://winddata.com

[23] S. Prakash, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. A **46**, R1724 (1992).

[24] E. G. Altmann and H. Kantz, Phys. Rev. E **71**, 056106 (2005).

[25] We assume $x_0=0$, which is the mean value of $\{x\}$. The assumption is irrelevant for large values of n.