

How does the quality of a prediction depend on the magnitude of the events under study?

S. Hallerberg and H. Kantz

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany

Received: 9 October 2007 – Revised: 30 January 2008 – Accepted: 17 March 2008 – Published: 16 April 2008

Abstract. We investigate the predictability of extreme events in time series. The focus of this work is to understand, under which circumstances large events are better predictable than smaller events. Therefore we use a simple prediction algorithm based on precursory structures which are identified via the maximum likelihood principle. Using these precursory structures we predict threshold crossings in autocorrelated processes of order one, which are either Gaussian, exponentially or Pareto distributed. The receiver operating characteristic curve is used as a measure for the quality of predictions we find that the dependence on the event magnitude is closely linked to the probability distribution function of the underlying stochastic process. We evaluate this dependence on the probability distribution function numerically and in the Gaussian case also analytically. Furthermore, we study predictions of threshold crossings in correlated data, i.e., velocity increments of a free jet flow. The velocity increments in the free jet flow are in dependence on the time scale either asymptotically Gaussian or asymptotically exponential distributed. If we assume that the optimal precursory structures are used to make the predictions, we find that large threshold crossings are for all different types of distributions better predictable. These results are in contrast to previous results, obtained for the prediction of large increments, which showed a strong dependence on the probability distribution function of the underlying process.

1 Introduction

Systems with a complex time evolution, which generate a great impact event from time to time, are ubiquitous. Examples include fluctuations of prices for financial assets in economy with rare market crashes, electrical activity of human brain with rare epileptic seizures, seismic activity of the earth with rare earthquakes, changing weather conditions with rare disastrous storms, and also fluctuations of on-line diagnostics

of technical machinery and networks with rare breakdowns or black-outs. Due to the complexity of the systems mentioned, a complete modeling is usually impossible, either due to the huge number of degrees of freedom involved, or due to a lack of precise knowledge about the governing equations.

This is why one applies the framework of prediction via precursory structures for such cases. The typical application for prediction with precursory structures is a prediction of an event which occurs in the very near future, i.e., on short timescales compared to the lifetime of the system. A classical example for the search for precursory structures is the prediction of earth-quakes (Jackson, 1996). A more recently studied example is the short term prediction of strong turbulent wind gusts, which can destroy wind turbines (Kantz et al., 2004, 2006).

In a previous work (Hallerberg et al., 2007), we studied the quality of predictions analytically via precursory structures for increments in an AR(1)-process and numerically in a long-range correlated ARMA process. The long-range correlations did not alter the general findings for Gaussian processes, namely, that larger increments are better predictable.

Furthermore, we found other works which report the same effect for the prediction of avalanches in SOC-models (Shapoval and Shrirman, 2006) and in multi-agent games (Lamper et al., 2002). In (Hallerberg and Kantz, 2008) we demonstrate that the quality of the prediction of increments is sensitively dependent on the probability distribution function (PDF) of the distribution of the underlying process. Furthermore we found, that increments are the better predictable, if the PDF of the process is Gaussian, that there is no significant dependence on the event magnitude, if this PDF is a symmetrised exponential, and that larger events are the harder to predict, if the PDF is a power law.

Since in earthquake prediction and systems which display self organized criticality the underlying distributions are definitely non-Gaussian, these results in comparison to our previous findings call for clarification.

Correspondence to: S. Hallerberg
(sarah@mpipks-dresden.mpg.de)

Indeed, the crucial distinction between our previous work and the observations in (Shapoval and Shrirman, 2006) is that we considered large increments as events, whereas in these examples the events were defined by the magnitude of some observable overcoming some predefined threshold. Consequently, we study now how the quality of a prediction depends on the event magnitude, if the events under study are not increments but threshold crossings.

Therefore we investigate predictions of threshold crossings in an autocorrelated process of order one [AR(1)] with Gaussian, exponential and power law distributions.

Furthermore, we compare the results for this short-range correlated processes with results for the prediction of threshold crossing in experimental data.

After defining the prediction scheme in Sect. 2.1 and the method for measuring the quality of a prediction in Sect. 2.2, we explain in Sect. 2.3 how to consider the influence on the event magnitude. In Sect. 2.4 we formulate a constraint, which has to be fulfilled in order to find a better predictability of larger (smaller) events. In the next section, we apply this constraint to compare the quality of predictions of threshold crossings within Gaussian (Sect. 3.1), exponential distributed (Sect. 3.2) and power law distributed AR(1) processes (Sect. 3.3). In the following we study the prediction of threshold crossings in free jet data in Sect. 4. In Sect. 5 we study the dependence on the event magnitude for a more realistic prediction procedure. Conclusions appear in Sect. 6.

2 Definitions and set-up

The considerations in this section are made for a time series (Box, Jenkins and Reinsel, 1994; Brockwell and Davis, 1998), 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 and $n \in \mathbb{N}$.

The recording should contain sufficiently many extreme events so that we are able to extract statistical information about them.

We assume here that the event of interest can be identified on the basis of the observations, more precisely, the value of the observation function exceeding some threshold. We express the presence (absence) of an event by using a binary variable Y_{n+1} .

$$Y_{n+1} = \begin{cases} 1 & \text{an event occurred at time } n + 1, \\ 0 & \text{no event occurred at time } n + 1. \end{cases} \quad (1)$$

2.1 The choice of the precursor

When we consider prediction via precursory structures (precursors, or predictors), we are typically in a situation, where we assume that the dynamics of the system under study has both, a deterministic and a stochastic part. The deterministic part allows to assume that there is a relation between the event and its precursory structure which we can use for predictive purposes. However, if the dynamic of the system

would be fully deterministic there would be no need to predict via precursory structures, but one could try to model the dynamical system.

In this contribution we focus on the influence of the stochastic part of the dynamics and assume therefore a very simple deterministic correlation between event and precursor. The presence of this stochastic part determines that we cannot expect the precursor to precede every individual event. That is why we define a precursor in this context as a data structure which is typically preceding an event, allowing deviations from the given structure, but also allowing events without preceding structure.

For reasons of simplicity the following considerations are made for precursors in real space, i.e., structures in the time series. However, there is no reason not to apply the same ideas for precursory structures, which live in phase space.

In order to predict an event Y_{n+1} occurring at the time $(n+1)$ we compare the last k observations, to which we will refer as precursory variable

$$\mathbf{x}_{(n-k+1,n)} = (x_{n-k+1}, x_{n-k+2}, \dots, x_{n-1}, x_n) \quad (2)$$

with a specific precursory structure

$$\mathbf{x}^{\text{pre}} = (x_{n-k+1}^{\text{pre}}, x_{n-k+2}^{\text{pre}}, \dots, x_{n-1}^{\text{pre}}, x_n^{\text{pre}}). \quad (3)$$

Once the precursory structure \mathbf{x}^{pre} is determined, we give an alarm for an event $Y_{n+1}=1$ when we find $\mathbf{x}_{(n-k+1,n)}$ in the volume

$$V^{\text{pre}}(\delta, \mathbf{x}^{\text{pre}}) = \prod_{j=n-k+1}^n \left(x_j^{\text{pre}} - \frac{\delta}{2}, x_j^{\text{pre}} + \frac{\delta}{2} \right), \quad (4)$$

where δ determines the magnitude of the precursory volume.

Upto here we did not specify how to obtain a suitable precursor \mathbf{x}^{pre} which provides optimal predictions. As we discussed in (Hallerberg et al., 2007), there are at least two natural choices. As one can argue using concepts from probabilistic forecasting (Hallerberg, Bröcker and Kantz, 2008), the following choice should be superior to all other choices: We define as \mathbf{x}^{pre} the vector for which the probability of an event to follow is maximal. More precisely, we consider the likelihood¹

$$L(Y_{n+1} = 1 | \mathbf{x}_{(n-k+1,n)}) = \frac{j(Y_{n+1} = 1, \mathbf{x}_{(n-k+1,n)})}{\rho(\mathbf{x}_{(n-k+1,n)})} \quad (5)$$

which provides the probability that an event $Y_{n+1}=1$ follows the precursor $\mathbf{x}_{(n-k+1,n)}$. It can be calculated numerically by determining the joint PDF $j(Y_{n+1}=1, \mathbf{x}_{(n-k+1,n)})$ and the

¹In this contribution we use the name likelihood for the probability that an event follows a precursor \mathbf{x} and the term aposterior PDF for the probability to find a precursor \mathbf{x} before of an already observed extreme event. Note that the names might be also used vice versa, if one refers to the precursor as the previously observed information, as it is the case in signal detection.

marginal PDF $\rho(\mathbf{x}_{(n-k+1,n)})$ of the process. Our prediction strategy consists in determining those values of each component x_i of $\mathbf{x}_{(n-k+1,n)}$ for which the likelihood is maximal. One can argue that this maximum of the likelihood is the optimal choice for a precursory variable, with respect to the measure for the quality of a prediction, which we are going to introduce in following section.

This strategy to identify the optimal precursor represents a rather fundamental choice. In more applied examples one looks for precursors which minimise or maximise 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., minimising the loss due to a false prediction. The strategy studied in this contribution is thus fundamental in the sense that it enters into many of the more sophisticated quantities which are used for predictions and decision making.

2.2 Testing for predictive power

A common method to verify a hypothesis or to test the quality of a prediction is the receiver operating characteristic curve (ROC curve) (Green and Swets, 1966; Egan, 1975; Pepe, 2003). 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 rate of correct predictions r_c and the rate of false alarms r_f can be obtained by integrating the aposterior PDFs $\rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1}=1)$ and $\rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1}=0)$ on the precursory volume.

$$r_c(\delta, \mathbf{x}^{\text{pre}}) = \int_{V(\delta, \mathbf{x}^{\text{pre}})} \rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1} = 1) d\mathbf{x}_{(n-k+1,n)} \tag{6}$$

$$r_f(\delta, \mathbf{x}^{\text{pre}}) = \int_{V(\delta, \mathbf{x}^{\text{pre}})} \rho(\mathbf{x}_{(n-k+1,n)}|Y_{n+1} = 0) d\mathbf{x}_{(n-k+1,n)} \tag{7}$$

Note that these rates are defined with respect to the total numbers of events $Y_{n+1}=1$ and non-events $Y_{n+1}=0$. Thus the relative frequency of events has no direct influence on the ROC curve, unlike on other measures of predictability, as e.g., the Brier score or the ignorance.

Plotting r_c vs. r_f for increasing values of δ one obtains a curve in the unit-square of the r_f - r_c plane (see, e.g., Fig. 5). The curve approaches the origin for $\delta \rightarrow 0$ and the point (1, 1) in the limit $\delta \rightarrow \infty$, where δ accounts for the magnitude of the precursor volume $V_{\text{pre}}(\delta)$. The shape of the curve characterises 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 characterised by the diagonal. Furthermore we are interested in curves which converge as fast as possible to 1, since this scenario tells us that we reach the highest possible rate of correct prediction without having a large rate of false alarms.

That is why we use the so called likelihood ratio as a summary index, to quantify the ROC curve. For our inference problems the likelihood ratio is identical to the slope m of the ROC curve in the vicinity of the origin which implies $\delta \rightarrow 0$.

This region of the ROC plot is particularly interesting, since it corresponds to a low rate of false alarms. The ‘‘likelihood ratio’’ is in our notation a ratio of aposterior PDFs.

$$m = \frac{\Delta r_c}{\Delta r_f} \sim \left. \frac{\rho(\mathbf{x}^{\text{pre}}|Y_{n+1} = 1)}{\rho(\mathbf{x}^{\text{pre}}|Y_{n+1} = 0)} \right|_{r_f \approx 0, \delta \approx 0} + \mathcal{O}(\delta). \tag{8}$$

For other problems the name likelihood ratio is also used for the slope at every point of the ROC curve.

Since we apply the likelihood ratio as a summary index for ROC curves, we specify that for our purposes the term likelihood ratio refers only to the slope of the ROC-plot at the vicinity of the origin as in Eq. (8).

2.3 Addressing the dependence on the event magnitude

We are now interested in learning how the predictability depends on the event magnitude η which is measured in units of the standard deviation of the time series under study. Thus the event variable Y_{n+1} becomes dependent on the event magnitude

$$Y_{n+1}(\eta) = \begin{cases} 1, & \text{an event of magnitude } \eta \text{ or larger} \\ & \text{occurred at time } n + 1, \\ 0, & \text{no event of magnitude } \eta \text{ or larger} \\ & \text{occurred at time } n + 1. \end{cases} \tag{9}$$

Via Bayes’ Theorem the likelihood ratio can be expressed in terms of the likelihood $L(Y_{n+1}(\eta)=1|\mathbf{x}_{\text{pre}})$ and the total probability to find events $P(Y_{n+1}(\eta)=1)$. Inserting the technical details of the calculation of the likelihood and the total probability one finds that the likelihood ratio depends sensitively on the joint PDF $j(\mathbf{x}_{(n-k+1,n)}, Y_{n+1}(\eta)=1)$ of precursor and event.

Hence once the precursor is chosen, the dependence on the event magnitude η enters into the likelihood ratio, via the joint PDF of event and precursor. This implies that the slope of the ROC-plot is fully characterised by the knowledge of the joint PDF of precursor and event.

Thus, in the framework of statistical predictions all kind of (long-range) correlations which might be present in the time series influence the quality of the predictions only through their influence on the joint PDF.

Exploiting the likelihood ratio we can then determine the dependence of the likelihood ratio and the ROC curve on the events magnitude η , via the dependence of the joint PDF of the process under study. This approach is motivated by the work of Egan (Egan, 1975), who already discovered that a family of likelihoods can in many cases lead to an ordered family of ROC curves, parameterized by the same parameter, which parameterizes the likelihood.

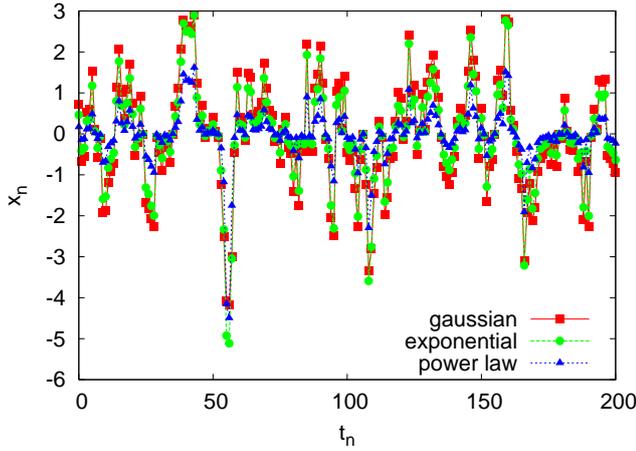


Fig. 1. Samples from the AR(1) processes with Gaussian, exponential and power-law distributions.

2.4 Constraint for increasing quality of predictions with increasing event magnitude

In order to study the dependence of the likelihood ratio on the event magnitude we are going to introduce a constraint which the likelihood and the total probability to find events have to fulfil in order to find a better predictability of larger (smaller) events.

In order to improve the readability of the paper, we will first introduce the following notations for the aposterior PDFs, the likelihood and the total probability to find events

$$\rho_c(\eta, \mathbf{x}_{(n-k+1,n)}) = \rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1}(\eta) = 1), \quad (10)$$

$$\rho_f(\eta, \mathbf{x}_{(n-k+1,n)}) = \rho(\mathbf{x}_{(n-k+1,n)} | Y_{n+1}(\eta) = 0), \quad (11)$$

$$L(\eta, \mathbf{x}_{(n-k+1,n)}) = L(Y_{n+1}(\eta) = 1 | \mathbf{x}_{(n-k+1,n)}), \quad (12)$$

$$P(\eta) = P(Y_{n+1}(\eta) = 1). \quad (13)$$

We can then ask for the change of the likelihood ratio with changing event magnitude η .

$$\frac{\partial}{\partial \eta} m(Y_{n+1}(\eta), \mathbf{x}_{(n-k+1,n)}) \stackrel{\geq}{\leq} 0. \quad (14)$$

The derivative of the likelihood ratio is positive (negative, zero), if the following sufficient condition $c(\eta)$ is fulfilled.

$$c(\eta, \mathbf{x}_{(n-k+1,n)}) = \frac{\partial}{\partial \eta} \ln L(\eta, \mathbf{x}_{(n-k+1,n)}) - \frac{(1 - L(\eta, \mathbf{x}_{(n-k+1,n)}))}{(1 - P(\eta))} \frac{\partial}{\partial \eta} \ln P(\eta) \stackrel{\geq}{\leq} 0. \quad (15)$$

Hence one can tell for an arbitrary process, if extreme events are better predictable, by simply testing, if the condition in Eq. (15) is positive for the respective marginal PDF of the events and the likelihood of event and precursor.

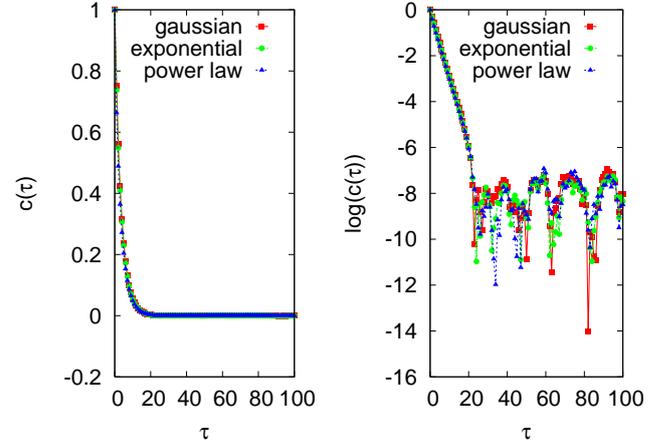


Fig. 2. The autocorrelation function $c(\tau) = \sum_n (x_n - \mu)(x_{n+\tau} - \mu) / ((n - \tau)\sigma^2)$, with the mean μ and the standard deviation σ is evaluated on the AR(1) correlated data.

3 Predictions of Threshold Crossing in short range correlated stochastic processes

In this section we test the condition $c(\eta, \mathbf{x}_{(n-k+1,n)})$ as given in Eq. (15) for threshold crossing in AR(1) processes, which have a Gaussian, power-law and exponential distribution.

The most popular example for an extreme event, which consist in a threshold crossing is probably the level of water in a river, which can exceed the height of a levee and then flood an area inhabited by humans. However one can easily find other examples, in which it would be desirable to predict the exceeding of a threshold. Inspired by this motivation, we study the prediction of threshold exceedances in simple short range correlated processes. We define our extreme event by a value x_{n+1} of the time series exceeding a given threshold η

$$Y_{n+1} = \begin{cases} 1, & x_{n+1} \geq \eta, \\ 0, & x_{n+1} < \eta. \end{cases} \quad (16)$$

where the event magnitude η is again measured in units of the standard deviation.

Due to the correlation of the AR(1) process we use the present value x_n of a time series as a precursory variable for the event happening at time $n+1$.

The short range correlated processes in focus are generated by an autoregressive model of order 1 [AR(1)] (see, e.g., (Box, Jenkins and Reinsel, 1994))

$$x_{n+1} = ax_n + \xi_n, \quad (17)$$

where ξ_n are uncorrelated random numbers with mean zero. The value and the sign of the coupling strength a determines whether successive values of x_n are clustered or spread. Since we are not in particular interested in the influence of the coupling strength we will set $a=0.75$ in all following

considerations. For $a \neq 0$ the process is exponentially correlated, $\langle x_n x_{n+k} \rangle = a^k$.

Typically the random numbers ξ_n are chosen to be Gaussian distributed. In this case the data generated by the AR(1) model is as well Gaussian distributed. However, due to the summation of random numbers in Eq. (17) also non-Gaussian random numbers might lead to a process with an approximately Gaussian distribution. That is why one has to apply other methods in order to obtain an AR(1) correlated process with a non-Gaussian distribution. We create the non-Gaussian distributed AR(1) processes by replacing the data of the Gaussian AR(1) process by random numbers, which follow the desired distribution function. This is done by ordering the data of the Gaussian AR(1) process and the random numbers according to their magnitude and then replacing the n -th largest value of the data set by the n -th largest random number. This procedure lead of course to local fluctuations in the value of the correlations strength a . However, the characteristic behaviour of the process is still preserved, as one can see in Figs. 1 and 2.

For the Gaussian AR(1) process all quantities which enter into the prediction can be evaluated analytically. Since in most cases the structure of the PDF is not known analytically, we evaluate $c(\eta, x_n)$ also numerically. In this case the approximations of the total probability and the likelihood are obtained by "binning and counting" and their numerical derivatives are evaluated via a Savitzky-Golay-filter (Savitzky and Golay, 1964; Press, 1992). The numerical evaluation is done within 10^7 data points. In order to check the stability of this procedure, we evaluate $c(\eta, x_n)$ also on 20 bootstrap samples, which are generated from the original data set such that choosing with repetition is allowed. These bootstrap samples consist of 10^7 pairs of event and precursor, which were drawn randomly from the original data set. Thus their PDFs are slightly different in their first and second moment and they contain different numbers of events. Evaluating $c(\eta, x_n)$ on the bootstrap samples thus shows, how sensitive our numerical evaluation procedure is towards changes in the numbers of events. This is especially important for large and therefore rare events.

In order to check the results obtained by the evaluation of $c(\eta, x_n)$, we compute also the corresponding ROC curves analytically and numerically.

Note that for both, the numerical evaluation of the condition and the ROC-plots, we used only event magnitudes η , for which we found at least 100 events, so that the observed effects are not due to a lack of statistics of the large events.

3.1 AR(1) process with a Gaussian distribution

As it is well known, the marginal PDF of the time step x_n in an AR(1) process is a Gaussian,

$$\rho(x_n, a) = \sqrt{\frac{1-a^2}{2\pi}} \exp\left(-\frac{1-a^2}{2} x_n^2\right). \tag{18}$$

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

For $a \neq 0$ the process is exponentially correlated $\langle x_n x_{n+k} \rangle = a^k$ and the joint PDF of two successive values $\rho(x_n, x_{n+1})$ is a bivariate Gaussian.

From this we derive the joint PDF $j(x_n, Y_{n+1}=1)$ by a simple integration using the Heaviside function Θ as a filter (see e.g. (Hallerberg et al., 2007) for details),

$$j(x_n, Y_n = 1) = \int dx_n \Theta(x_n - \eta\sigma) \rho(x_n, x_{n+1})$$

The a posteriori PDFs to find or not to find events are then given by

$$\rho_c(\eta, x_n, a) = \frac{\sqrt{1-a^2} \exp\left(-\frac{1-a^2}{2} x_n^2\right)}{2\sqrt{2\pi} \rho^\Theta(a, \eta)} \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right), \tag{19}$$

$$\rho_f(\eta, x_n, a) = \frac{\sqrt{1-a^2} \exp\left(-\frac{1-a^2}{2} x_n^2\right)}{2\sqrt{2\pi} (1 - \rho^\Theta(a, \eta))} \left(1 + \operatorname{erf}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)\right). \tag{20}$$

The corresponding likelihood reads

$$L(\eta, x_n, a) = \frac{1}{2} \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right). \tag{21}$$

We recall that the optimal precursor is given by x_{pre} which maximises the likelihood and hence $x_{\text{pre}} = \infty$. In this case the alarm volume is the interval $[\delta, \infty]$. From the mean value of the a posterior PDF $\langle x_n \rangle$ we can obtain the analytic structure of the total PDF to find events

$$P(\eta, a) = \frac{a}{\sqrt{2(1-a^2)}} \frac{1}{\langle x_n \rangle} \exp\left(-\frac{\eta^2}{2}\right). \tag{22}$$

Using

$$\operatorname{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi}z} \left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \cdot 3 \dots (2m-1)}{(2z^2)^m}\right), \tag{23}$$

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

which can be found in (Abramovitz and Stegun, 1972) and approximating the mean value $\langle x_n \rangle$ with the maximum x_n^* of

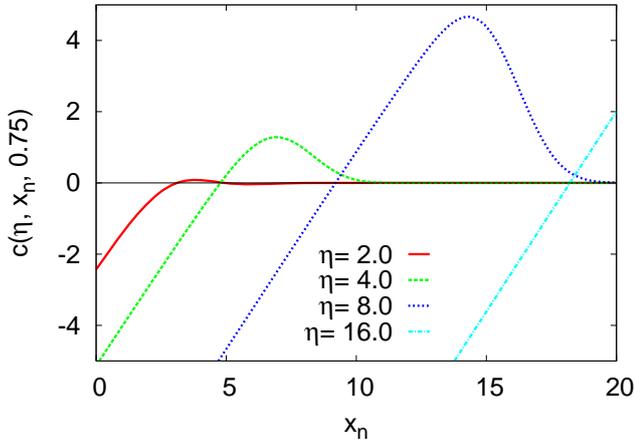


Fig. 3. The condition $c(\eta, x_n, 0.75)$ for the Gaussian distributed AR(1) process as given by Eq. (26).

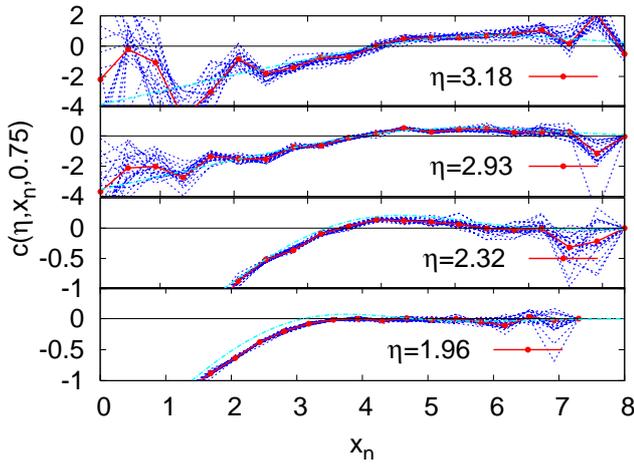


Fig. 4. The analytical results (dashed light blue line) for the Gaussian distributed AR(1) process are given by Eq. (26). The analytical and the numerical results for the evaluation of the condition $c(\eta, x_n, 0.75)$ are compared. The results for the numerical evaluation of $c(\eta, x_n, 0.75)$ are indicated by the symbols connected with lines. The results for the bootstrap samples are plotted in dashed lines.

the aposterior PDF

$$x_n^* = \sqrt{\frac{2}{\pi}} \frac{a}{1-a^2} \frac{\exp\left(-\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n^*}{\sqrt{2}}\right)^2\right)}{\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n^*}{\sqrt{2}}\right)} \propto \frac{a\eta}{\sqrt{1-a^2}\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad \eta \rightarrow \infty, \quad (24)$$

we obtain the following approximation of the total probability to find events

$$P(\eta, a) \propto \frac{\exp\left(-\frac{\eta^2}{2}\right)}{\sqrt{2}\eta} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right), \quad \eta \rightarrow \infty \quad (25)$$

Note that this expression is only valid in the limit of large η . In particular it does not hold for $\eta=0$. Using Eq. (21) and Eq. (25) the constraint $c(x_n, a, \eta)$ reads

$$c(\eta, x_n, a) \propto -\sqrt{\frac{2}{\pi(1-a^2)}} \frac{\exp\left(\frac{-1}{2}\left(\frac{\eta}{\sqrt{1-a^2}} - ax_n\right)^2\right)}{\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)} + \left(\eta + \frac{1}{\eta}\right) \frac{\left(1 - \frac{1}{2}\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)\right)}{1 - \frac{\exp(-\eta^2/2)}{\sqrt{2}\eta}\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \quad (26)$$

Using again Eq. 23 we obtain the following asymptotic behaviour for large values of η

$$c(\eta, x_n, a) \rightarrow \eta \left(\frac{\left(1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)\right)}{1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)} - \frac{1}{\sqrt{1-a^2}} \right) + \frac{ax_n}{\sqrt{1-a^2}} \frac{1}{1 + \mathcal{O}(1/\eta^2)} + \frac{1}{\eta} \left(\frac{1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)}{1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)} \right), \quad \eta \rightarrow \infty. \quad (27)$$

This expression is larger than zero, if terms of the order $\mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)$ are negligible. Hence we can conclude, that $c(\eta, x_n, a)$ is positive for large values of η and arbitrary values of x_n .

However for finite values of η we observe a dependence on the precursory variable. Figures 3 and 4 display that $c(\eta, x_n, a)$ is positive for larger values, i.e., values, which are closer to the ideal precursor $x_{\text{pre}}=\infty$. Hence we should expect larger events to be better predictable, if our alarm interval is situated in this region, i.e., if the alarm interval $[\delta, \infty]$ is small.

The ROC curves in Fig. 5 support this result. In the region of low rates of false alarms which corresponds to a small alarm interval we find a strong dependence on the event magnitude in the sense, that larger events are better predictable.

Finally one can discuss the case of the ideal precursor $x_{\text{pre}}=\infty$. Inserting this value of the precursory variable into Eq. 26 one obtains $c(\eta, x_n, a)=0$. This ideal precursor corresponds to the idea of an ideal ROC curve, which is identical to the axis of the ROC-plot. Thus no further improvement of the ROC curve, e.g. by a change of the event magnitude, is possible.

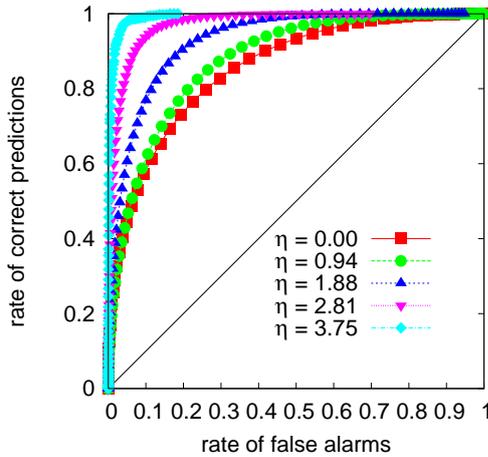


Fig. 5. ROC curves for the Gaussian distributed AR(1) process with correlation coefficient $a=0.75$. The ROC curves were made via predicting threshold crossings of magnitude η within 10^7 data points. The predictions were made according to the prediction strategy described in Sect. 2.1. Note that the quality of the prediction increases with increasing event magnitude.

3.2 AR(1) Process with Symmetrised Exponential Distribution

The AR(1) data with exponential distribution were created via replacing the values of the Gaussian distributed AR(1) data with exponentially distributed i.i.d. random variables, as explained in Sect. 3. The exponential distributed AR(1) process has the following PDF

$$\rho(x) = \frac{\lambda}{2} \exp(-\lambda|x_n|)$$

with $\lambda=1$ and was generated by transformation from uniformly distributed random numbers. (The uniformly distributed random numbers were generated by using the Mersenne twister algorithm (Matsumoto and Nishimura, 1998).) Numerically we find the maximum of the likelihood also in the region of large values of x_n , similar to the Gaussian case with an alarm interval $[\delta, -\infty]$. We compute the condition according to Eq. (15) and the ROC curves numerically by using 10^7 exponentially distributed AR(1) correlated data.

Figure 6 compares the results of the numerical evaluation of the condition $c(\eta, x_n, \lambda)$. In the vicinity of the larger values of the data set, the condition $c(\eta, x_n, \lambda)$ is positive as in the Gaussian case.

The ROC curves in Fig. 7 support the qualitative results from Fig. 6, that larger events are better predictable. The numerical ROC curves were made via predicting threshold crossings in 10^7 AR(1) correlated exponentially distributed data points according to the prediction strategy described in Sect. 2.1.

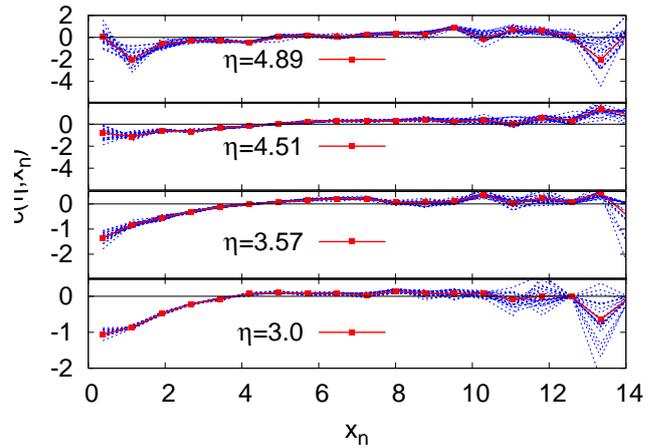


Fig. 6. The condition according to Eq. (15) evaluated on 10^7 exponentially distributed AR(1) correlated data.

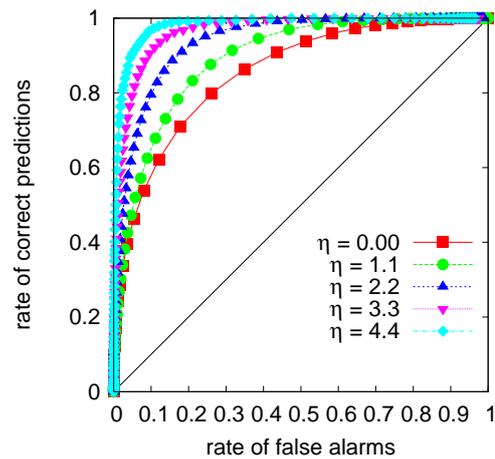


Fig. 7. The ROC curves were made via predicting threshold crossings in 10^7 exponentially distributed AR(1) correlated data and the predictions were made according to the prediction strategy described in Sect. 2.1.

This result is qualitatively different from the results of prediction of increments in sequences of exponentially distributed i.i.d. random numbers in (Hallerberg et al., 2007). In this previous work we found that the event magnitude has no influence on the prediction of large increments in sequences of exponentially distributed i.i.d. random numbers. This difference can probably be understood by the fact that the condition $c(\eta, x_n)$ is not only a function of the event magnitude η , but also a function of the event class and of the precursor values x_n .

3.3 Power-law distributed random variables

The AR(1) data with Power-law distribution were created via replacing the values of the Gaussian distributed AR(1) data

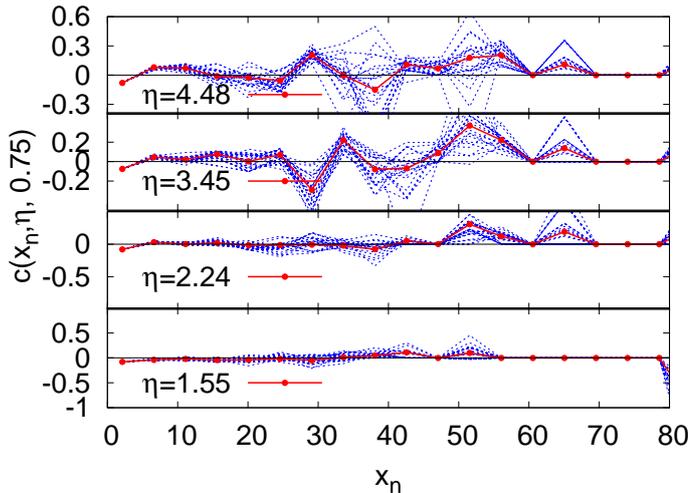


Fig. 8. The condition $c(\eta, x_n, \alpha, \sigma)$ evaluated on 10^7 power-law distributed AR(1) correlated data with variance $\sigma=1$, mean zero and power-law coefficient $\alpha=3$. x_n .

with symmetrised power-law distributed i.i.d. random variables, as explained in Sect. 3.

The symmetrised power-law distributed i.i.d. random variables follow the following distribution

$$\begin{aligned} \rho(x) &= \alpha x_{\min}^{\alpha} x^{-(\alpha+1)}, & x > x_{\min} > 0 \\ \rho(x) &= \alpha |x_{\max}|^{\alpha} |x|^{-(\alpha+1)}, & x < x_{\max} < 0, \end{aligned} \quad (28)$$

with $x_{\min}=|x_{\max}|=0.01$ and power-law coefficient $\alpha=3$ were generated by transformation from uniformly distributed random numbers. Since the distributions with $x_{\min} = |x_{\max}|=0.01$ would allow no values in the interval $]x_{\min}, x_{\max}[$, the resulting random numbers were shifted to the left (right) by subtracting (adding) x_{\min} . The result is a symmetrised power law distribution with mean zero and variance $\sigma=0.01$, see Figs. 1 and 2. Finally, the values of the AR(1) process were amplified by multiplication with a constant $c_a=100$, so that the data set of the power law distributed AR(1) process has a variance of $\sigma=1$, as the Gaussian and the exponential AR(1) process.

Figures 8 and 9 show the numerical results for $c(\eta, x_n, \alpha, \sigma)$ and the ROC curves.

Although $c(\eta, x_n, \alpha, \sigma)$ is less regular than in the Gaussian or the exponential case, its values are mainly above zero which corresponds to the ROC curves in Fig. 9.

Hence large threshold crossings are also within the Pareto distributed AR(1) process better predictable than smaller. As in the exponential case, this result for threshold crossings in AR(1) correlated data is qualitatively different from the results for the prediction of increments in sequences of Pareto distributed random numbers.

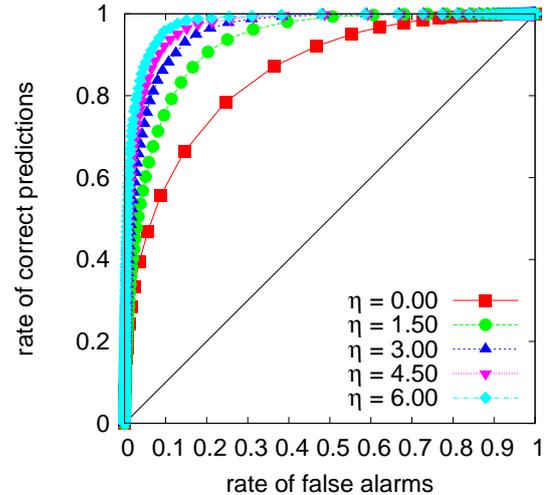


Fig. 9. ROC-plot for the power-law distribution. The ROC curves were made via predicting increments in 10^7 data points of the AR(1) process with power-law distribution.

4 Predicting Threshold Crossings in Free Jet Data

In this section, we apply the method of statistical inference to predict threshold crossings of the acceleration in a free jet flow. Therefore we use a data set of 1.25×10^7 samples of the local velocity measured in the turbulent region of a round free jet (Renner, Peinke and Friedrich, 2001). The data were sampled by a hot-wire measurement in the central region of an air into air free jet. One can then calculate the PDF of velocity differences $a_{n,k}=v_{n+k}-v_n$, where v_n and v_{n+k} are the velocities measured at time step n and $n+k$. The Taylor hypothesis allows to relate the time-resolution to a spatial resolution (Renner, Peinke and Friedrich, 2001). One observes that for large values of k the PDF of the velocity differences is essentially indistinguishable from a Gaussian, whereas for small k , the PDF develops approximately exponential wings (Van Atta and Park, 1972; Gagne et al., 1990; Frisch, 1995). Figure 10 illustrates this effect using the data set under study.

Thus the incremental data sets $a_{n,k}$ provides us with the opportunity to test the results for statistical predictions within Gaussian and exponential distributed AR(1) correlated processes on a data set, which exhibits correlated structures.

We are now interested in predicting larger values of the acceleration $a_{n+j,k} \geq \eta$ in the incremental data sets $a_{n,k}=v_{n+k}-v_n$. In the following we concentrate on the data set $a_{n,10}$, which has an asymptotically exponential PDF and the data set $a_{n,144}$, which has an asymptotically Gaussian PDF. Since the free jet data are correlated, predicting the next time step would be equivalent to predicting persistence. That is why we choose a prediction horizon of seven time steps, i.e., $j=7$ in the Gaussian regime ($k=144$) and a smaller time horizon $j=2$ in the less correlated exponential distributed increments ($k=10$). Note, that not only the event magnitude

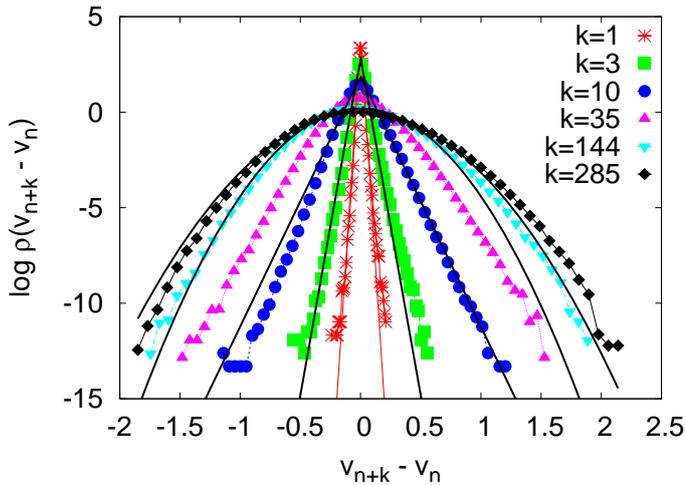


Fig. 10. PDF of the accelerations $a_{n,k} = v_{n+k} - v_n$ with $k=1, 3, 10, 35, 144, 285$.

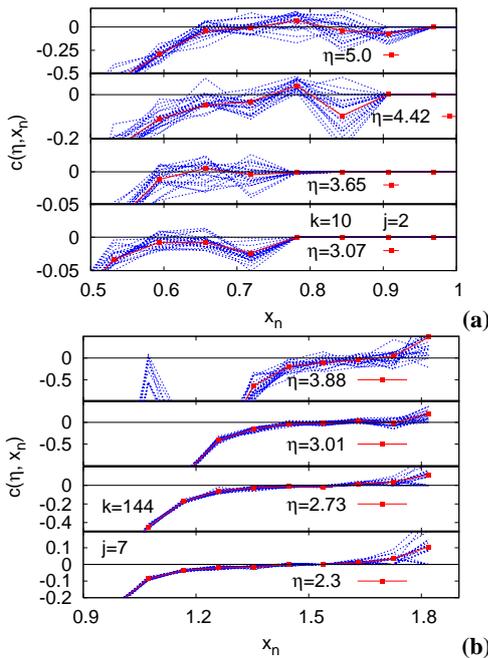


Fig. 11. The condition $c(\eta, a_{n,k})$ evaluated in the exponential regime ($k=10$) (a) and in the Gaussian regime ($k=144$) (b).

and the precursor, but also the prediction horizon influences the resulting ROC curves.

As in the previous sections we are hence exploiting the conditional probabilities of the time series to make predictions. We can now use the algorithm which was tested on the previous examples to evaluate the condition for these data sets. The results are shown in Fig. 11. In both examples the evaluation of the condition $c(\eta, a_{n,k})$ reflects the behaviour of the ROC curves. This example of the free jet data-set shows, that the specific dependence of the ROC curve on the

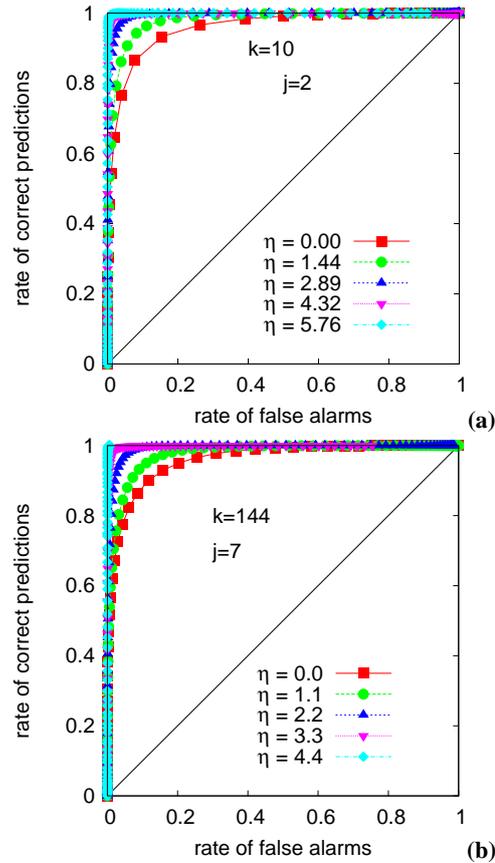


Fig. 12. ROC curves in the exponential ($k=10$) (a) and in the Gaussian regime ($k=144$) (b).

event magnitude can also in the case of long-range correlated data sets be characterised by the PDF of the underlying process.

5 A more realistic prediction procedure

Predicting an above-threshold event when the current observation itself is already above the threshold might not be really relevant in most applications.

We therefore modify here the sample on which predictions are to be made: We define as events the subset of previous events, where not only the future value is above threshold, but simultaneously the current value is below threshold.

$$Y_{n+1} = \begin{cases} 1 & : x_{n+1} \geq \eta, \quad x_n < \eta \\ 0 & : x_{n+1} < \eta, \quad x_n < \eta \end{cases} \quad (29)$$

The events $Y_n=1$ according to this definition are a subset of the previously discussed events. Hence, this modification reduces the number of events in the time series and might render the prediction task more difficult. The corresponding ROC-curves in Figs. 13–15 show qualitatively the same dependence on the event magnitude as the ROC

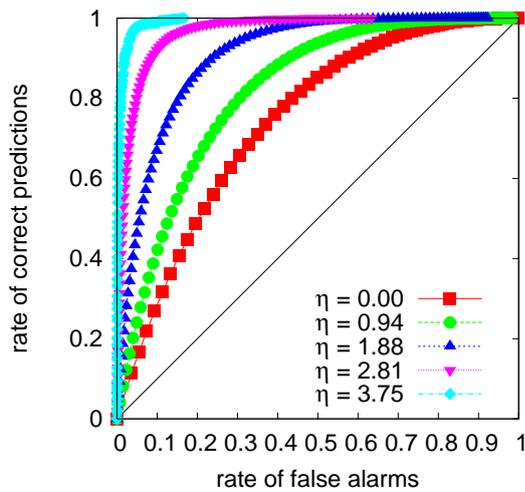


Fig. 13. ROC curves for the Gaussian AR(1) process made according to the more realistic prediction procedure described in Sect. 5.

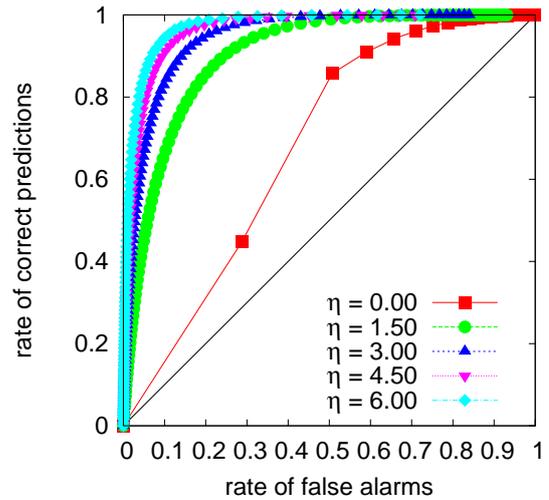


Fig. 15. ROC curves for the power law distributed AR(1) process made according to the more realistic prediction procedure described in Sect. 5.

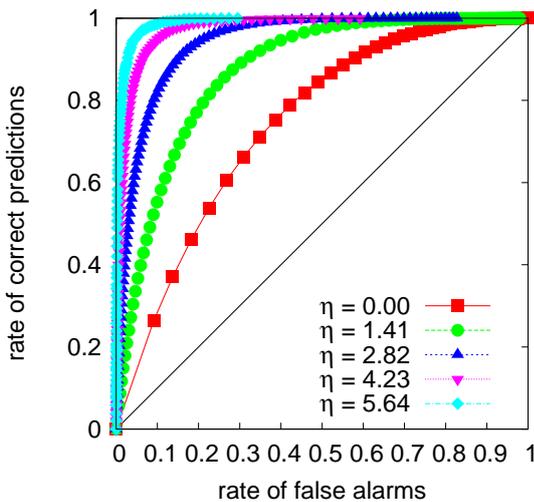


Fig. 14. ROC curves for the exponential distributed AR(1) process made according to the more realistic prediction procedure described in Sect. 5

curves obtained in the previous section: Threshold crossings in Gaussian, approximately exponential distributed, and approximately power-law distributed AR(1) processes are better predictable, the larger they are.

6 Conclusions

We study the magnitude dependence of the quality of predictions for threshold crossings in autocorrelated processes of order one and in measured accelerations in a free jet flow. Using the present value x_n as a precursory variable we predict threshold crossings at a future time step x_{n+j} via statistical considerations. In order to measure the quality of the

predictions we use ROC curves. Furthermore, we introduce a quantitative criterion which can determine, whether larger or smaller events are better predictable.

We are especially interested in the influence of the probability distribution of the underlying process on changes in the quality of the predictions, which are evoked by focusing on different event magnitudes. For Gaussian, exponential and power law distributed AR(1) processes we find, that larger threshold crossings are better predictable, the higher the threshold. In all cases studied the behaviour of the ROC curves was reasonably well reflected by the condition $c(\eta, x_n)$, which is an expression that depends on the total probability to find events and the likelihood to observe an event after a given value of x_n . This theoretical results could in principle help to understand the effects reported for avalanches in systems, which display self organized criticality (Shapoval and Shrirman, 2006).

The velocity measurements in the free jet flow provide us with the opportunity to redo the predictions in data sets, which inhibits correlated structures, and have either asymptotically Gaussian or asymptotically exponential distributions.

In both cases larger threshold crossings are also in the free jet data set better predictable, the higher the threshold.

These results are in contrast to a previous study (Hallerberg and Kantz, 2008) on the prediction of increments in Gaussian, exponential and power law distributed i.i.d. random numbers, in which we found a qualitatively different behavior for exponentially and power law distributed random numbers: In the exponential case we did not find any significant dependence on the event magnitude, whereas in the power-law case larger increments were much harder to predict than smaller.

This difference to the recent results for threshold crossings can be explained by taking into account the different regimes in which we find the optimal precursors:

When predicting increments, the optimal precursors are typically among the smallest values in the data set, while for the prediction of threshold crossings, large values are optimal. Furthermore threshold crossings form a different class of events. Hence both PDFs which contribute to the value of $c(\eta, x_n)$, namely the likelihood and the total probability to find events are different. Hence we should not be surprised to find different results for the predictability of larger events. In summary we find, that threshold crossings in AR(1) processes and also in the correlated free jet flow data are the better predictable, the larger they are.

Acknowledgements. We thank J. Peinke and his group for supplying us with their excellent free jet data.

Edited by: S. Vannitsem

Reviewed by: two anonymous referees

References

- Jackson, D. J.: Hypothesis testing and earthquake prediction, *Proc. Natl. Acad. Sci. USA*, 93, 3772–3775, 1996.
- Kantz, H., Holstein, D., Ragwitz, M., and Vitanov, N. K.: Markov chain model for turbulent wind speed data, *Physica A*, 342, 315–321, 2002.
- Kantz, H., Holstein, D., Ragwitz, M., and Vitanov, N. K.: Short time prediction of wind speeds from local measurements, in: *Wind Energy—Proceedings of the EUROMECH Colloquium*, edited by: Peinke, J., Schaumann, P., and Barth, S., Springer, 2006.
- Hallerberg, S., Altmann, E. G., Holstein, D., and Kantz, H.: Precursors of Extreme Increments, *Phys. Rev. E*, 75, 016706, 2007.
- Shapoval, A. B. and Shirman M. G.: How size of target avalanches influence prediction efficiency, *Int. J. Mod. Phys. C*, 17, 1777–1790, 2006.
- Hallerberg, S. and Kantz, H.: Influence of the Event Magnitude on the Predictability of Extreme Events, *Phys. Rev. E*, 77, 011108, 2008.
- Lamper, D., Howison, S. D., and Johnson N. F.: Predictability of large future changes in a competitive evolving population, *Phys. Rev. Lett.*, 88, 1, 2002.
- Box, G. E. P., Jenkins G. M., and Reinsel G. C.: *Time Series Analysis*, Prentice-Hall, Inc., 1994.
- Brockwell, P. J. and Davis, R. A.: *Time Series: Theory and Methods*, Springer, 1998.
- Hallerberg, S., Bröcker, J., and Kantz, H.: Prediction of Extreme Events, accepted for publication in *Nonlinear Time Series Analysis in the Geosciences—Applications in Climatology, Geodynamics, and Solar-Terrestrial Physics*, edited by: Donner, R. and Barbosa, S.: *Lecture Notes in Earth Sciences*, Springer, accepted, 2008.
- Egan, J. P.: *Signal detection theory and ROC analysis*, Academic Press, New York, 1975.
- Green, D. M. and Swets J. A.: *Signal detection theory and psychophysics*, Wiley, New York, 1966.
- Pepe M. S.: *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, 2003.
- Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.*, 36, 1627–1639, 1964.
- Press, W. H.: *Numerical recipes in C*, Cambridge University Press, Cambridge, 1992.
- Abramowitz, M. and Stegun, I. A.: *Handbook of Mathematical Functions*, Dover, New York, 1972.
- Matsumoto, M. and Nishimura, T.: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, in *ACM Transactions on Modeling and Computer Simulation*, 8, 1, 3–30, 1998.
- Renner, C., Peinke, J. and Friedrich, R.: Experimental indications for Markov properties of small-scale turbulence, *J. Fluid. Mech.*, 433, 383–09, 2001.
- Van Atta, C. W. and Park, J.: Statistical self-similarity and inertial subrange turbulence, in *Statistical Models and Turbulence*, *Lect. Notes in Phys.*, 12, 402–426, edited by: Rosenblatt, M. and Van Atta, C. W., Springer Berlin, 1972.
- Gagne, Y., Hopfinger, E., and Frisch, U.: A new universal scaling for fully developed turbulence: the distribution of velocity increments, in *New Trends in Nonlinear Dynamics and Pattern-Forming Phenomena*, NATO ASI, 237, 315–319, edited by: Coulet, P. and Huerre, P., Plenum Press, New York, 1990.
- Frisch, U.: *Turbulence*, Cambridge University Press, Cambridge, 1995.