Clustering of extreme and recurrent events in deterministic chaotic systems

Thomas Schweigler and Jörn Davidsen*
Complexity Science Group, Department of Physics & Astronomy, University of Calgary, Calgary, Alberta T2N 1N4, Canada
(Received 16 October 2010; published 6 July 2011)

We study the nontrivial clustering properties of extreme or recurrent events in the context of deterministic chaotic systems. We find that correlations between return times of such events can depend nonmonotonically on the threshold used to define the events, which leads to counterintuitive behavior. In particular, the distribution of the conditional return intervals can indicate clustering as well as repelling of extreme events for the same condition but different thresholds—in sharp contrast to what has been observed for stochastic processes with long-range correlations as well as for independent and identically distributed random variables. This has important implications for the time-dependent hazard assessment of extreme events, indicating that possible threshold dependencies should always be taken into account.

DOI: 10.1103/PhysRevE.84.016202 PACS number(s): 05.45.−a, 02.50.−r, 89.75.Da

I. INTRODUCTION

Extreme events are an important theme in various areas of science because of their typically devastating effects on society and their scientific complexities [1,2]. In particular, the ever-increasing economic and human losses from natural hazards underscore the urgency for improving understanding of extreme events to develop effective strategies to reduce their impact. The classic approach to studying the probability of extreme events has been to assume independent and identically distributed (IID) event sizes. This has led to a powerful statistical theory (see, e.g., Refs. [3–5]), which has been successfully applied in many cases. The latter is also related to the fact that some parts of the theory—for example, the limit distributions of block maxima—can be extended to a wide class of dependent stationary time series [6,7].

Aside from the distribution of block maxima, another very useful and practical indicator for hazard assessment is the distribution of return intervals between well-defined extreme events. For time series of IID events, the Poisson process associated with events above a fixed threshold gives rise to the well-known exponential return interval distribution and, in particular, independent return intervals [8]. For (long-range) correlated stochastic processes frequently encountered in nature, however, the return intervals are typically not independent. Besides, the distribution is often no longer exponential and a number of distributions have been put forward, depending on the data set, such as γ distributions and power laws with stretched exponential tails [9–15]. Artificially generated long-range correlated stochastic signals have been studied in detail [16–19], and some progress has been made toward a theoretical understanding of return intervals in (long-range) correlated series [20–23], but the overall picture is far from complete.

Here, we focus on the statistical properties of return intervals in *deterministic* chaotic systems like the logistic map. In contrast to the IID case or stochastic processes with long-range correlations, it has been recently shown that the principal signatures of such deterministic dynamics in the statistics of extreme or recurrent events and their return intervals give rise to novel and distinct properties [24,25]. We extend these findings here by showing that the clustering of extreme events due to nontrivial correlations in the return intervals can depend sensitively on the threshold used to define extreme events, even switching from clustering to repelling of extreme events and back. Here, nontrivial clustering corresponds to a tendency—which is higher than expected from the distribution of return intervals alone—that return intervals are similar to the previous one, while repelling corresponds to an alternating behavior in the return times; namely, large values have a tendency to be followed by small values and vice versa. In particular, we show that the conditional return interval distributions, i.e., the distributions of those return intervals that follow immediately a preceding return interval of a given size in the series of the return intervals, vary qualitatively with threshold. Due to these variations, the conditional mean return intervals exhibit nonmonotonic effects that render it impossible to make a time-dependent hazard assessment independent of the applied threshold to define extreme events as recurrences to a finite interval. Thus, our findings clearly indicate that one should always investigate the influence of the applied threshold on the properties of the return times and on measures used for time-dependent hazard assessment.

II. FULLY DEVELOPED CHAOS

A prime example of an iterated map system $x_{n+1} = f(x_n)$ that exhibits fully developed chaos and is particularly simple is the one-dimensional tent map, which we will consider here. It is defined on the unit interval $0 \leqslant x \leqslant 1$ as

$$f(x) = 1 - |1 - 2x|. \tag{1}$$

The corresponding natural density is given by $\rho(x) = 1$. Note that a large class of maps conjugate to the tent map exists. One particularly important example is the logistic map [26].

In the context of such an iterated map system, we consider extreme events as threshold exceedances for a fixed threshold $x_{\rm th}$ or equivalently as recurrences to the interval $[x_{\rm th}, 1]$. This leaves us with two possible variants of a precise definition of extreme events: Either all extreme events only consist of a single extreme value, or extreme events are defined as a continuous sequence of values above $x_{\rm th}$. The former case is particularly suitable in the context of intrinsically discrete

*davidsen@phas.ucalgary.ca

processes, for which it is more appropriate to consider two subsequent values above x_{th} as separate events. In the latter case, extreme events are simply peak over threshold events, which is typically more suitable for continuous processes. While we focus here on the first convention—also used in Ref. [17]—we obtain qualitatively similar results for the second convention. In terms of the return intervals r_k between subsequent extreme events, one direct consequence of this choice is that *any* positive integer value is possible.[1]

To test for the existence of nontrivial clustering or repelling of extreme events,[2] we numerically analyzed the iterated map system given by Eq. (1) and estimated the autocorrelation function $C_{x_{th}}(l)$ of the series of return intervals r_k for various threshold values x_{th}.[3] For stationary processes as the one considered here, the autocorrelation function is defined by

$$C_{x_{th}}(l) = \frac{E_{x_{th}}\left[\left(r_k - \bar{r}_{x_{th}}\right)\left(r_{k+l} - \bar{r}_{x_{th}}\right)\right]}{\sigma^2_{x_{th}}}, \tag{2}$$

where $E_{x_{th}}[\cdots]$ stands for the expectation value for a given threshold value x_{th}. The variable l is commonly called "lag" and refers to the (discrete) separation between the return intervals whose relationship should be investigated. As shown in Figs. 1 and 2, $C_{x_{th}}(l)$ displays nonmonotonic dependence on both the threshold value and the lag l. This behavior is fundamentally different from the behavior observed for stochastic systems with long-range correlations [17] or IID signals. In particular, Figs. 1 and 2 imply that one can observe clustering or repelling of extreme events as well as independent occurrences of extreme events depending on the chosen threshold value. For example, the alternating behavior of the autocorrelation function for $x_{th} = 0.698$ indicates that large return intervals tend to be followed by small ones and vice versa, which corresponds to repelling of extreme events.

By knowing the possible Markov partitions of the tent map, we can predict that the autocorrelation function must be identically zero for certain threshold values. This follows from the fact that if $[x_{th}, 1]$ is a single cell of any Markov partition, i.e., if the occurrences of extreme events are recurrences to a Markov cell, then successive recurrence times must be independent of each other. It is known that a partitioning of the unit interval into K equal intervals ($[(n-1)/K, n/K], n = 1, \ldots, K$) is Markovian [27]. Therefore, the intervals $[(K-1)/K, 1]$ represent cells of a Markov partition, which implies that the autocorrelation function for threshold values of the form $(K-1)/K, K \in \mathbb{N}$, must be zero. Some of these values are represented by vertical lines in Fig. 1. Note that the lines would get dense for threshold values close to 1.

It is also important to realize that the cells of the Markov partition are in general not "lumpable," or in other words, the union of two or more cells of one given Markov partition will in general not be a single cell of another Markov partition [25]. This means that successive return intervals will in general not be independent of each other even for rational threshold values of the form $(K-n)/K, K \in \mathbb{N}, n \in \mathbb{N}, (K-1) > n > 1$. The nonvanishing autocorrelation function for some of these values is evident from Fig. 1.

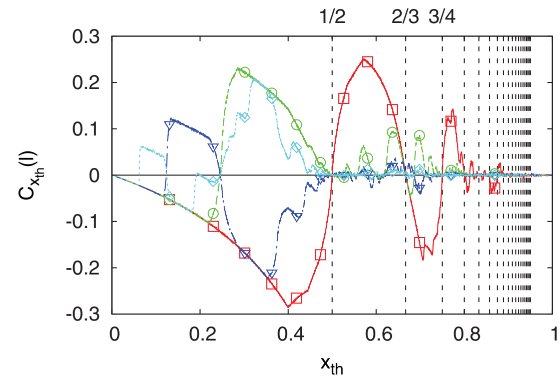

FIG. 1. (Color online) Autocorrelation coefficient of the sequence of return intervals of extreme events as a function of threshold values x_{th} for different lags l. The curves belonging to the different lags are marked by different symbols [$l = 1$: (red) squares; $l = 2$: (green) circles; $l = 3$: (blue) reverse triangles; $l = 4$: (cyan) diamonds]. Nonmonotonic dependence on both threshold and lag can be observed. The vertical lines represent some of the threshold values for which all coefficients must be zero (see text for a discussion).

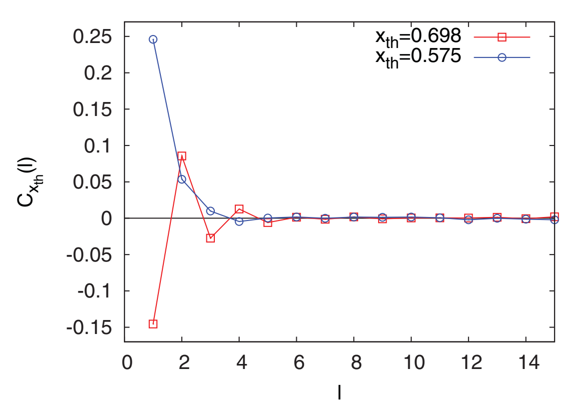

FIG. 2. (Color online) Autocorrelation function of the sequence of return intervals of extreme events for two different threshold values representing clustering and repelling of return intervals, respectively.

[1]For the alternative definition of extreme events, the minimum return interval is two.

[2]As already mentioned above, clustering corresponds to a tendency—which is higher than expected from the distribution of return intervals alone—that return intervals are similar to the previous one, while repelling corresponds to an alternating behavior in the return times; namely, large values have a tendency to be followed by small values and vice versa. Extreme events are defined as threshold exceedances; each single exceedance is interpreted as a separate extreme event.

[3]For the actual simulation, the logistic map was used. The logistic map given by $f(y) = 1 - 2y^2, \ -1 \leqslant y \leqslant 1$ is conjugate to the tent map. The transformation is given by $y = -\cos(\pi x)$. Simulating the logistic map instead of the tent map allows us to generate longer sequences. In a direct simulation of the tent map, a fast convergence to a periodic orbit or a fixed point would occur (every rational number finally converges to a periodic orbit; clearly we can only represent rational numbers with a computer). After the simulation of the logistic map had been performed, a back-transformation was applied so that the results are valid for the tent map. The statistical results for the autocorrelation function have been obtained numerically by averaging over 10^6 return intervals.

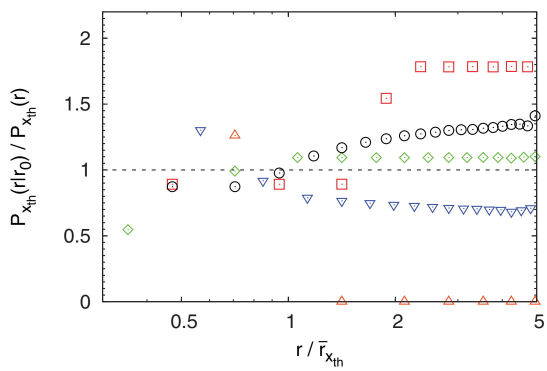

FIG. 3. (Color online) The figure shows the conditional probability $P_{x_{th}}(r|r_0)$ divided by the unconditional probability $P_{x_{th}}(r)$ as a function of $r/\bar{r}_{x_{th}}$ (return interval in units of the mean return interval). For all data sets we chose $r_0/\bar{r}_{x_{th}} = \sqrt{2}$. Data for the threshold values $x_{th} = 0.293$ (triangles), $x_{th} = 0.529$ (squares), $x_{th} = 0.646$ (diamonds), $x_{th} = 0.717$ (reverse triangles), and $x_{th} = 0.764$ (circles) are presented. The results have been obtained numerically from a data set with 10^9 return intervals per threshold value.

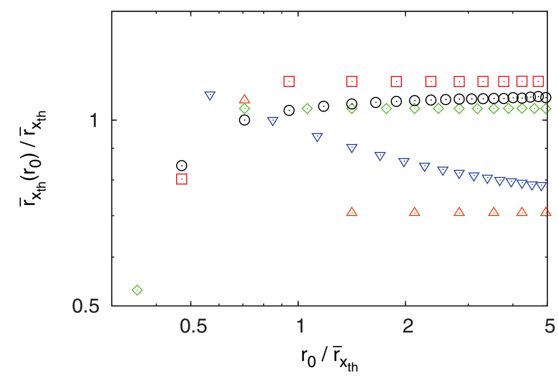

FIG. 4. (Color online) The figure shows the conditional mean return interval $\bar{r}_{x_{th}}(r_0)$ in units of the unconditional mean return interval $\bar{r}_{x_{th}}$ vs the condition $r_0/\bar{r}_{x_{th}}$. The same threshold values as in Fig. 3 have been used. The results have been obtained numerically from a data set with 10^9 return intervals per threshold value.

The autocorrelation function of the series of return intervals has a direct influence on the probability, $P_{x_{th}}(r|r_0)$, of finding a return interval r conditioned that the preceding return interval was of a certain length r_0. This distribution is a convenient tool often used for time-dependent hazard assessment. In Fig. 3 the ratio of the conditional to the unconditional return interval distribution $P_{x_{th}}(r|r_0)/P_{x_{th}}(r)$ is shown for various threshold values. If this ratio is identically 1, then the observation of a return interval of length r_0 has no predictive power for the direct subsequent return interval. In all cases shown, we have fixed r_0 such that $r_0/\bar{r}_{x_{th}} = \sqrt{2}$, where $\bar{r}_{x_{th}}$ is the average return time for a given x_{th}. In agreement with the results obtained for the autocorrelation coefficient $C_{x_{th}}(1)$ (see Fig. 1), we get qualitatively different curves for the different threshold values. The condition $r_0/\bar{r}_{x_{th}} = \sqrt{2}$ means that we condition on a return interval r_0 that is larger than the mean return interval. Therefore, in the case that repelling of extreme events occurs, the conditional probability should be bigger than the unconditional for small return intervals and smaller than the unconditional for large return intervals (i.e., we should get a decaying curve in Fig. 3). On the other hand, when clustering of extreme events occurs—large (small) intervals tend to be followed by large (small) ones—the curve should be increasing. Consequently, if the threshold value corresponds to a negative $C(1)$ (repelling of extreme events), then a falling curve is expected. For threshold values corresponding to a positive $C(1)$, on the other hand, we should get a rising curve. This is fully confirmed by the results presented in Fig. 3.

The behavior shown in Fig. 3 is fundamentally different from what has been observed for stochastic processes with long-range correlations [17] or IID processes. In particular, the figure illustrates that no universal scaling law of the form $P_{x_{th}}(r|r_0) = (1/\bar{r}_{x_{th}}) f_{r_0/\bar{r}_{x_{th}}}(r/\bar{r}_{x_{th}})$ exists, which has been proposed for stochastic processes [17]. The existence of such a scaling law would imply that all data points in the graph would lie on a single curve, which is clearly not the case.

The absence of such a scaling law is actually a consequence of the properties of the *unconditional* distribution of return intervals $P_{x_{th}}(r)$ and the form of the autocorrelation coefficient $C_{x_{th}}(l = 1)$. Similar to what has been recently shown for related quantities[4] in deterministic systems [25,28], $P_{x_{th}}(r)$ is not smooth as a function of the threshold value but undergoes abrupt changes at a discrete set of values (not shown). This prevents the existence of a scaling law of the form $P_{x_{th}}(r) = (1/\bar{r}_{x_{th}}) f(r/\bar{r}_{x_{th}})$, which was also proposed for long-range correlated stochastic processes [17]. Moreover, $P_{x_{th}}(r|r_0)$ together with $P_{x_{th}}(r)$ fully determines the autocorrelation coefficient, since by definition [Eq. (2)],

$$C_{x_{th}}(l = 1) = \frac{\sum_{r_0} P_{x_{th}}(r_0)\big(r_0 - \bar{r}_{x_{th}}\big) \sum_r P_{x_{th}}(r|r_0)\big(r - \bar{r}_{x_{th}}\big)}{\sigma^2_{x_{th}}}. \quad (3)$$

As can be clearly seen from Eq. (3), the existence of a unique scaling law for $P_{x_{th}}(r|r_0)$ is incompatible not only with the lack of smoothness of $P_{x_{th}}(r)$, but also with the nonmonotonic threshold dependence of the autocorrelation function (see Figs. 1 and 2), assuming that finite size effects can be neglected.

A related tool often used for time-dependent hazard assessment is the *conditional* mean return interval $\bar{r}_{x_{th}}(r_0)$, which is defined as $\bar{r}_{x_{th}}(r_0) = \sum_{r=1}^{\infty} r P_{x_{th}}(r|r_0)$. In Fig. 4, $\bar{r}_{x_{th}}(r_0)$ is shown for five different threshold values. In all cases, the dependence on the previous return interval r_0 agrees with the results for the autocorrelation function (see Fig. 1). For the threshold value of 0.717, for example, the autocorrelation function exhibits a repelling of extreme events and in particular the autocorrelation coefficient for lag 1 is negative. Therefore, short return intervals should on average be followed by long return intervals, and long return intervals should on average be followed by short return intervals. This is indeed confirmed by the results for the conditional mean return

[4] In Ref. [25], first exceedence times instead of return intervals were investigated, while [28] used the second convention to define extreme events discussed above.

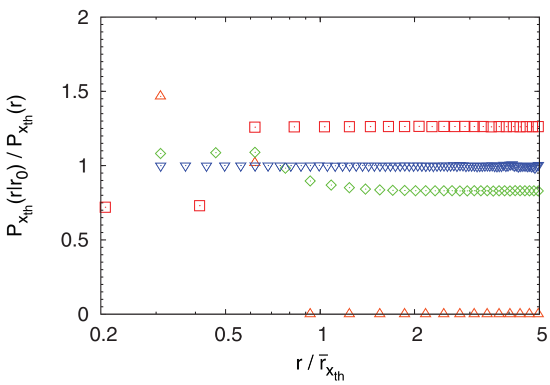

FIG. 5. (Color online) Like Fig. 3 but for the cusp map and $r_0/\bar{r}_{x_{\rm th}} = 0.6$. Data for the threshold values $x_{\rm th} = -0.0954$ (triangles), $x_{\rm th} = 0.1056$ (squares), $x_{\rm th} = 0.2254$ (diamonds), and $x_{\rm th} = 0.5101$ (reverse triangles) are presented. The results have been obtained numerically from a data set with 2×10^9 return intervals per threshold value.

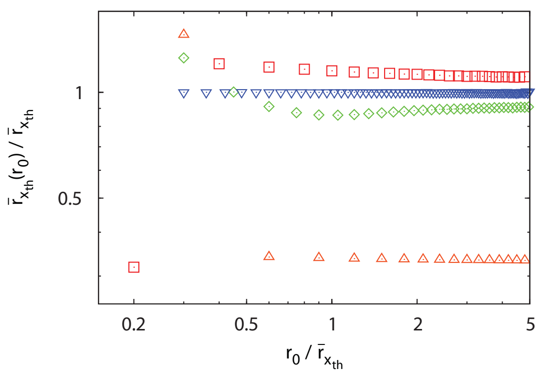

FIG. 6. (Color online) Like Fig. 4 but for the cusp map. The same threshold values as in Fig. 5 have been used. The results have been obtained numerically from a data set with 2×10^9 return intervals per threshold value.

interval, since it decreases monotonically with r_0 (reverse triangles in Fig. 4). For the threshold value of 0.764 on the other hand, the autocorrelation coefficient $C_{x_{\rm th}}(1)$ is positive, which indicates that clustering of extreme events should occur. Indeed, we find that the conditional mean return interval increases monotonically with r_0 (circles in Fig. 4).[5]

III. INTERMITTENT CHAOS

The results in Figs. 3 and 4 show that estimating conditional occurrence probabilities of extreme events in deterministic systems can depend sensitively on the choice of the threshold value used to define extreme events. To get a complete picture, one has to explicitly consider different threshold values. The importance and generality of these findings is further emphasized by the fact that they are not specific to the tent map and maps conjugate to it. In particular, we obtain qualitatively similar results for maps exhibiting a different form of deterministic chaos namely intermittency. This is evident from Figs. 5 and 6, which show the conditional occurrence probability and the conditional mean return interval of extreme events, respectively, for the cusp map. This map is defined as $f(x) = 1 - 2\sqrt{|x|},\ -1 \leqslant x \leqslant 1$. In contrast to the tent map, the autocorrelation function of the return intervals for the cusp map is not well-defined. This is due to the fact that the variance of the distribution of return intervals in the cusp map is infinite for all threshold values $-1 < x_{\rm th} < 1$ [29]. Despite this, clustering and repelling of extreme events can occur, depending on the threshold. This follows from Figs. 5 and 6. Thus, we expect that many deterministically chaotic systems exhibit strong qualitative changes in statistical measures typically used for time-dependent hazard assessment if the threshold used to define extreme events is varied. In particular for the distribution of conditional return intervals, the same condition can have qualitatively different effects depending on the threshold.

IV. CONCLUSIONS

It is important to realize that our findings are to a certain extent specific to the definition of extreme events as recurrences to *finite* intervals. In the limit of vanishing intervals corresponding to asymptotically rare events, analytical results exist for a large number of dynamical systems. For example, it is known that for piecewise expanding maps with certain properties, the unconditional distribution of return intervals approaches an exponential form [30,31]. Moreover, successive return intervals are independent from each other such that asymptotically rare events can be treated within the classic framework of IID processes. The main objective of our study, the tent map, belongs to this class of dynamical systems. However, many other chaotic systems, including some intermittent maps [32], show a different asymptotic behavior. In practical situations, one is always interested in extreme events, which have a finite probability of occurrence. Thus, deviations from the behavior expected for asymptotically rare events—if known at all—can become important. As we showed here, such deviations can be significant and indeed lead to counterintuitive behavior.

To summarize, our findings prove that some of the standard tools used for time-dependent hazard assessment of extreme events can sensitively depend on the exact threshold values used to define extreme events if the underlying deterministic dynamics is chaotic. In particular, this sensitive dependence can lead to diametrically opposed results, namely, from a clustering to a repelling of extreme events and vice versa. This is in sharp contrast to idealized situations like IID random variables and stochastic processes with long-range correlations, which are often considered in the context of extreme events and which do not exhibit any nonmonotonic threshold dependencies. Expectations built on such simple stochastic models can, thus, be misleading for time-dependent hazard assessment in real

[5] Note that even though we get a monotonic dependence of the conditional mean return interval on r_0 for all considered threshold values in case of the tent map, this property does not hold true for all deterministic systems. For the cusp map, for example, a nonmonotonic behavior was observed (see Fig. 6), which should be related to nonlinear effects.

world situations. Our results constitute a proof of principle that one should always investigate the influence of the applied threshold on the properties of the return times and on measures used for time-dependent hazard assessment of extreme events.

ACKNOWLEDGMENTS

The authors thank A. Schumann and P. Grassberger for helpful discussions. This project was financially supported by the Alberta Ingenuity Fund (now Alberta Innovates).

[1] *The Science of Disasters: Climate Disruptions, Heart Attacks and Market Crashes*, edited by A. Bunde, J. Kropp, and H. J. Schellnhuber (Springer, Berlin, 2002).
[2] *Extreme Events in Nature and Society*, edited by S. Albeverio, V. Jentsch, and H. Kantz (Springer, Berlin, 2006).
[3] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer, Berlin, 2004).
[4] S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer, New York, 2001).
[5] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, New York, 2006).
[6] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer, New York, 1983).
[7] M. R. Leadbetter and H. Rootzén, Ann. Probab. **16**, 431 (1988).
[8] D. R. Cox and V. Isham, *Point Processes* (Chapman and Hall, London, 1980).
[9] J. Davidsen and H. G. Schuster, Phys. Rev. E **65**, 026120 (2002).
[10] J. Davidsen and C. Goltz, Geophys. Res. Lett. **31**, L21612 (2004).
[11] A. Corral, Phys. Rev. Lett. **92**, 108501 (2004).
[12] J. Davidsen, S. Stanchits, and G. Dresen, Phys. Rev. Lett. **98**, 125502 (2007).
[13] M. Baiesi, M. Paczuski, and A. L. Stella, Phys. Rev. Lett. **96**, 051103 (2006).
[14] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005).
[15] A. Vázquez, J. a. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, and A.-L. Barabási, Phys. Rev. E **73**, 036127 (2006).
[16] E. G. Altmann and H. Kantz, Phys. Rev. E **71**, 056106 (2005).
[17] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Phys. Rev. E **75**, 011128 (2007).
[18] N. R. Moloney and J. Davidsen, Phys. Rev. E **79**, 041131 (2009).
[19] B. Kaulakys and M. Alaburda, J. Stat. Mech. (2009) P02051.
[20] A. Corral, Phys. Rev. Lett. **95**, 028501 (2005).
[21] P. Olla, Phys. Rev. E **76**, 011122 (2007).
[22] D. Sornette, S. Utkin, and A. Saichev, Phys. Rev. E **77**, 066109 (2008).
[23] M. S. Santhanam and H. Kantz, Phys. Rev. E **78**, 051113 (2008).
[24] C. Nicolis, V. Balakrishnan, and G. Nicolis, Phys. Rev. Lett. **97**, 210602 (2006).
[25] C. Nicolis and S. C. Nicolis, Europhys. Lett. **80**, 40003 (2007).
[26] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems* (Springer, New York, 1996).
[27] V. Balakrishnan, G. Nicolis, and C. Nicolis, J. Stat. Phys. **86**, 191 (1997).
[28] C. Nicolis and G. Nicolis, Phys. Rev. E **80**, 061119 (2009).
[29] V. Balakrishnan and M. Theunissen, Stochastics and Dynamics **1**, 339 (2001).
[30] P. Collet and A. Galves, Dynamical Systems and Applications, World Sci. Ser. Appl. Anal. **4**, 139 (1995).
[31] Z. Coelho, Summer School on Symbolic Dynamics, Temuco, Chile, 1997.
[32] X. Bressaud and R. Zweimüller, Ann. Henri Poincaré **2**, 501 (2001).