A Letters Journal Exploring the Frontiers of Physics

September 2007

EPL, **79** (2007) 50001
doi: 10.1209/0295-5075/79/50001

www.epljournal.org

True and false forbidden patterns in deterministic and random dynamics

J. M. Amigó[1], S. Zambrano[2] and M. A. F. Sanjuán[2]

[1] *Centro de Investigación Operativa, Universidad Miguel Hernández - Avda. de la Universidad s/n. 03202 Elche, Spain*
[2] *Departamento de Fásica, Universidad Rey Juan Carlos - Tulipáns/n 28933 Madrid, Spain*

received 27 April 2007; accepted in final form 5 July 2007
published online 27 July 2007

PACS 05.45.Tp – Time series analysis
PACS 05.40.Ca – Noise
PACS 02.50.-r – Probability theory, stochastic processes, and statistics

Abstract – In this letter we discuss some properties of order patterns both in deterministic and random orbit generation. As it turns out, the orbits of one-dimensional maps have always forbidden patterns, *i.e.,* order patterns that cannot occur, in contrast with random time series, in which any order pattern appears with probability one. However, finite random sequences may exhibit "false" forbidden patterns with non-vanishing probability. In this case, forbidden patterns decay with the sequence length, thus unveiling the random nature of the sequence. Last but not least, true forbidden patterns are robust against noise and disintegrate with a rate that depends on the noise level. These properties can be embodied in a simple method to distinguish deterministic, finite time series with very high levels of observational noise, from random ones. We present numerical evidence for white noise.

Introduction. – Symbolic dynamics is a useful approach in the study of discrete-time dynamical systems, that consists of deriving sequences of symbol patterns via coarse-graining of the underlying state space [1]. If the state space is linearly ordered, one can derive sequences of order patterns as well [2–4], each encapsulating the up-and-down in the orbits of their elements (see below). It turns out that, under some mild mathematical assumptions, not all order patterns can be materialized by the orbits of a given one-dimensional map, not even if its dynamic is chaotic —contrarily to what happens with the symbol patterns. As a result, the existence of "forbidden" (*i.e.,* not occurring) order patterns is always a persistent dynamical feature, in opposition to properties such as proximity, correlation, etc., which die out with time in a chaotic dynamics. Moreover, if an order pattern is forbidden, its absence pervades all longer patterns in form of more missing order patterns, called *outgrowth forbidden patterns.*

Since random dynamics has no forbidden patterns with probability one, we conclude that their existence can be used as a fingerprint of deterministic orbit generation. Here and henceforth, "random" means generated by an unconstrained, stochastic process taking on real values in, say, an interval. However, when it comes to exploit this forbidden-pattern–based strategy to tell chaotic from random time series, two important practical issues arise: finiteness and noise contamination. Finiteness produces *false forbidden patterns* (*i.e.,* order patterns missing in a random sequence without constraints), whereas noise blurs the difference between deterministic and random time series. It is therefore interesting that the forbidden patterns themselves provide the remedy. First of all, the number of false forbidden patterns of a fixed length always decreases with the length of the time series, making it possible in turn that their outgrowth patterns become also visible. Secondly, order patterns are robust against experimental and numerical noise because they are defined by inequalities. We present numerical evidence that forbidden patterns persist in noisy deterministic data, even when the contamination is so high that other more traditional methods fail to uncover the underlying deterministic dynamics. This property can be capitalized on in practice by comparing the number of forbidden patterns in the sequence under scrutiny before and after randomizing it. For other approaches to the detection of determinism in time series, see, *e.g.,* [5,6] and the references therein.

What about higher dimensions? First of all, forbidden patterns are trivially invariant under order-isomorphisms (*i.e.,* isomorphisms which additionally preserve the order relations). Thus, higher-dimensional maps that are order-isomorphic to a map with forbidden patterns (say, a one-sided or two-sided shift), will also exhibit forbidden patterns; think, *e.g.*, of the baker's map and the two-sided $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift [7]. Furthermore, in the analysis of low dimensional chaotic flows it is well known that the so-called Lorenz maps (representing each local maxima of a continuous-time variable of a dynamical system against the next one) do often present a shape that is clearly reminiscent to unimodal maps [8]. Thus, the discrete time series resulting from sampling the maxima of a variable of this kind of systems will present forbidden patterns that would not appear for a random signal. Finally, the stretching and folding action of one-dimensional maps has been found to be deeply related with chaotic dynamics in higher-dimensional dynamical systems and flows [8]. Hence, we expect in general that higher-dimensional chaotic dynamical systems will also exhibit forbidden patterns.

Forbidden patterns. – Let $I \subset \mathbb{R}$ be a closed interval and $f: I \to I$ a map. If $x \in I$ is not periodic or its period is greater than $L \geqslant 2$, we associate with x an order pattern of length L as follows. We say that x *defines the order pattern* $\pi = \pi(x) = [\pi_0, \pi_1, \ldots, \pi_{L-1}]$, if

$$f^{\pi_0}(x) < f^{\pi_1}(x) < \ldots < f^{\pi_{L-1}}(x),$$

where $f^0(x) \triangleq x$ and $f^k(x) \triangleq f(f^{k-1}(x))$. We say also that π is realized by x. Thus, π is just a permutation of $\{0, 1, \ldots, L-1\}$ written between brackets, that condenses the order of the points $x_k = f^k(x)$, $0 \leqslant k \leqslant L-1$. A periodic point $x \in I$ of (minimal) period $T \geqslant 2$ can only define order patterns of lengths $L = 2, \ldots, T$.

The set of order patterns of length L will be denoted by $\mathcal{S}_L$. According to Stirling's formula, $|\mathcal{S}_L| = L! \propto \exp(L(\ln L - 1) + (1/2)\ln 2\pi L)$, where $|\cdot|$ denotes cardinality and $\propto$ means "asymptotically". Yet, numerical simulations support the conjecture that order patterns, like symbol patterns generated by coarse-graining partitions, grow only exponentially for "well-behaved" functions. In fact, if f is *piecewise monotone* (*i.e.,* there is a finite partition of I into intervals such that f is continuous and strictly monotone on each of those intervals), then one can prove [9] that

$$|\{\pi \in \mathcal{S}_L : \pi \text{ is realized by } x \in I\}| \propto e^{L h_{top}(f)},$$

where $h_{top}(f)$ is the topological entropy of f [1]. Since we can safely assume (*as we do henceforth*) that all scalar functions encountered in applications belong to this category, we conclude that time series generated by iteration of one-dimensional interval maps cannot realize all possible order patterns, but rather only a "small" part of them. Order patterns that do not appear in any orbit of f are called *forbidden patterns* for f (otherwise, they are *allowable* or "visible").

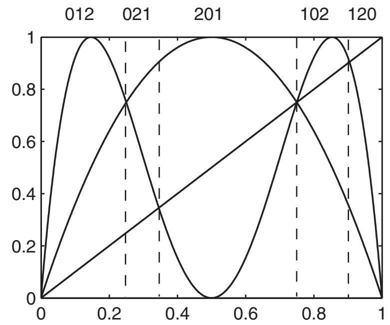

Fig. 1: The intervals of points defining the different order patterns are graphically obtained by raising vertical lines at the crossing points of the curves $y = f^0(x) \equiv x$, $y = f(x) = 4x(1-x)$ and $y = f^2(x) = -64x^4 + 128x^3 - 80x^2 + 16x$ (namely, 1/4, $(5 \mp \sqrt{5})/2$ and 3/4). The allowed order patterns are written on the top, centered over the corresponding interval. Note that $[2, 1, 0]$ is a forbidden pattern.

Example 1. – As a simple illustration, consider the logistic map $f(x) = 4x(1-x)$, $0 \leqslant x \leqslant 1$. For $L = 2$ we have

$$\{x \text{ defining } [1, 0]\} = (0, \tfrac{3}{4}), \quad \{x \text{ defining } [0, 1]\} = (\tfrac{3}{4}, 1).$$

But already for $L = 3$, the pattern $[2, 1, 0]$ is not realized (see fig. 1). The absence of $[2, 1, 0]$ triggers, in turn, an avalanche of longer forbidden patterns. To begin with, the pattern $[*, 2, *, 1, *, 0, *]$ (where the wildcard $*$ stands eventually for any other entries of the pattern) cannot be realized by any $0 \leqslant x \leqslant 1$ since the inequality

$$f^2(x) < f(x) < x \tag{1}$$

cannot occur. By the same token, the patterns $[*, 3, *, 2, *, 1, *]$, $[*, 4, *, 3, *, 2, *]$, and, more generally, $[*, 2+n, *, 1+n, *, n, *] \in \mathcal{S}_N$, $0 \leqslant n \leqslant N-3$, cannot be realized either for the same reason (substitute x by $f^n(x)$ in (1)).

Outgrowth forbidden patterns. – As discussed above, given a one-dimensional interval map $f: I \to I$, there exist $\pi \in \mathcal{S}_L$, $L \geqslant 2$, which is forbidden for f. Moreover, similarly to Example 1, the absence of π pervades all longer patterns in form of *forbidden outgrowth patterns.* Indeed, if $\pi = [\pi_0, \ldots, \pi_{L-1}]$, then all the patterns

$$[*, \pi_0 + n, *, \pi_1 + n, *, \ldots, *, \pi_{L-1} + n, *] \in \mathcal{S}_N \tag{2}$$

with $n = 0, 1, \ldots, N-L$, where $N - L \geqslant 1$ is the number of wildcards $* \in \{0, 1, \ldots, n-1, L+n, \ldots, N-1\}$ (with $* \in \{L, \ldots, N-1\}$ if $n = 0$ and $* \in \{0, \ldots, N-L-1\}$ if $n = N-L$), are also forbidden for f for the same arguments as in Example 1.

Denote now by $\mathcal{S}_N^{out}(\pi)$ the family of length-N outgrowth patterns of $\pi \in \mathcal{S}_L$. The fact that some of the outgrowth patterns of a given length will be the same (see, *e.g.,* the pattern $[3, 2, 1, 0]$ in (3) below) and that this depends on π, makes the analytical calculation of $|\mathcal{S}_N^{out}(\pi)|$ extremely complicated. Yet, it can be proven [10] that there exist constants $0<c,d<1$ such that $(1-d^N)N!<|\mathcal{S}_N^{out}(\pi)|<(1-c^N)N!$. Hence, the outgrowth forbidden patterns grow super-exponentially.

Example 2. – As a somewhat academic example, suppose that a black box generates a random-looking sequence $0\leqslant x_k\leqslant 1$ of length $N\gg 1$ according to the (for the observer unknown) recipe: $x_{k+1}=4x_k(1-x_k)$. If we go down the sequence with a sliding window of length $L=3$, we will find that the order pattern $[2,1,0]$ is missing (*i.e.*, $x_{k+2}<x_{k+1}<x_k$ never happens). What is the probability that this order pattern does not occur in a sequence of independent and uniformly distributed random variables? Due to stochastic dependences among overlapping windows, the exact calculation is not straightforward (nor specially illuminating), so we will content ourselves with a rough upper bound. Consider instead those windows of length L overlapping at a single point, so that their outcomes are independent; their number is $\lfloor N/(L-1)\rfloor$, where $\lfloor\cdot\rfloor$ stands for integer part. Hence, the sought probability is upper bounded by $p_0=(5/6)^{\lfloor N/2\rfloor}$. Given the "rejection threshold" $10^{-\varepsilon}$, then $p_0\leqslant 10^{-\varepsilon}$ if $N\geqslant 25.26\varepsilon$.

This probability can be lowered by considering the forbidden patterns of lengths 4, 5, etc. The forbidden pattern $[2,1,0]$ generates 7 different outgrowth patterns with $L=4$, see (2),

$$\begin{aligned}&(n=0)\ \ [3,2,1,0],[2,3,1,0],[2,1,3,0],[2,1,0,3]\\&(n=1)\ \ [0,3,2,1],[3,0,2,1],[3,2,0,1],([3,2,1,0])\end{aligned}\tag{3}$$

(the last is repeated) and 52 outgrowth patterns with $L=5$. A similar argument as before yields now the upper bound $p_1=(17/24)^{\lfloor N/3\rfloor}$ for the probability that none of the patterns (3) occurs in a long sequence of independent and uniformly distributed random variables and $p_1\leqslant 10^{-\varepsilon}$ if $N\geqslant 20.04\varepsilon$.

Practical issues. – We already mentioned that real time series may exhibit false forbidden patterns on account of being finite and noisy. It is thus clear that we cannot discriminate random from deterministic dynamics on the basis of output observation with absolute certainty, but only with a certain probability.

Before addressing these issues with more detail, note that a time series of length N allows only one window of length N, two (overlapping) windows of length $N-1$ and, in general, $N-L+1$ windows of length L, for $2\leqslant L\leqslant N$. Thus, in order to allow every possible order pattern of length L to occur in a time series of length N, the condition $L!\leqslant N-L+1$ must hold and, moreover,

$$N\gg L!+L-1\tag{4}$$

to avoid undersampling. For this reason, given a window of length L, we will choose $N\geqslant(L+1)!$ in the numerical simulations below.

Under this proviso, suppose now that the order pattern $\pi\in\mathcal{S}_L$ is missing in a time series of length N. Of course, if we dispose of many time series output by the same source or we can generate them at will, the chance that a false forbidden pattern persists in a randomly generated sequence will decrease with the number of the samples. But even if we have at our disposal one sufficiently long time series, the decay with N of the number of outgrowth patterns can make the difference. Indeed, if $\pi=[\pi_0,\ldots,\pi_{L-1}]$ is missing in a random sequence (thus, π is a false forbidden pattern), then all longer order patterns of the form (2) will be necessarily missing too. But now, at variance with the case of true forbidden patterns, the overall number of outgrowth forbidden patterns depends on N: if N increases, the probability that a false forbidden pattern becomes allowed increases. Once a forbidden pattern of length L disappears for probability reasons, their outgrowth patterns of length $L+1$ may also disappear and this chain process affects patterns of higher L as N increases.

Consider a fixed initial condition x and suppose that $\pi_{forb}=[\pi_0,\ldots,\pi_{L-1}]$ is a forbidden pattern for f. Suppose, furthermore, that we switch on now a discrete-time random perturbation η_k, $|\eta_k|\leqslant\eta_{\max}$, such that π_{forb} is still missing in the finite sequence $(f^k(x)+\eta_k)_{k=0}^{N-1}$ (due to robustness). Observe that the *noisy time series* $z_k=f^k(x)+\eta_k$ can be viewed both as a perturbation of an underlying deterministic dynamics and as a random process correlated with the deterministic dynamics f. If the orbit of x would be infinitely long, then the noisy time series had no forbidden patterns and π_{forb} would be allowed with probability 1. In the finite-length case we are considering, this is, in general, not the case; rather, there is a threshold $\theta=\theta(N)$ (the greater N, the smaller θ) such that π_{forb} will appear in $(z_k)_{k=0}^{N-1}$ only if $\eta_{\max}>\theta$. Again, once the pattern π_{forb} becomes visible, its outgrowth patterns may, in turn, become also visible with a higher θ or, alternatively, with a higher N. We conclude that amplifying a random perturbation destroys progressively the outgrowth patterns of the underlying deterministic dynamics but, as long as π_{forb} remains forbidden (*i.e.*, $\eta_{\max}\leqslant\theta$), all its outgrowth patterns will survive.

Numerical simulations. – Here we will numerically study only one of the properties discussed above, namely, the robustness of true forbidden patterns and their outgrowth patterns against observational random perturbations. In order to estimate the average number $\langle n(L,N)\rangle$ of forbidden patterns of length L in the finite, noisy sequence $z_k=x_k+\eta_k$, $0\leqslant k\leqslant N-1$, with $x_{k+1}=f(x_k)$ and η_k a random process, we generate 100 samples of length $N_{\max}=8000$ and normalize the corresponding count of missing patterns of lengths $3\leqslant L\leqslant 6$, complying with (4) for $N=N_{\max}$. We highlight next a few results obtained choosing f to be the logistic map and η_k being white noise uniformly distributed in the interval $[-\eta_{\max},\eta_{\max}]$, $0\leqslant\eta_{\max}\leqslant 1$.

Figures 2(a), (b) and (c) show $\langle n(L,N)\rangle$ for $(L+1)!\leqslant N\leqslant N_{\max}$, when i) $\eta_{\max}=0.25$, ii) $\eta_{\max}=0.50$ and

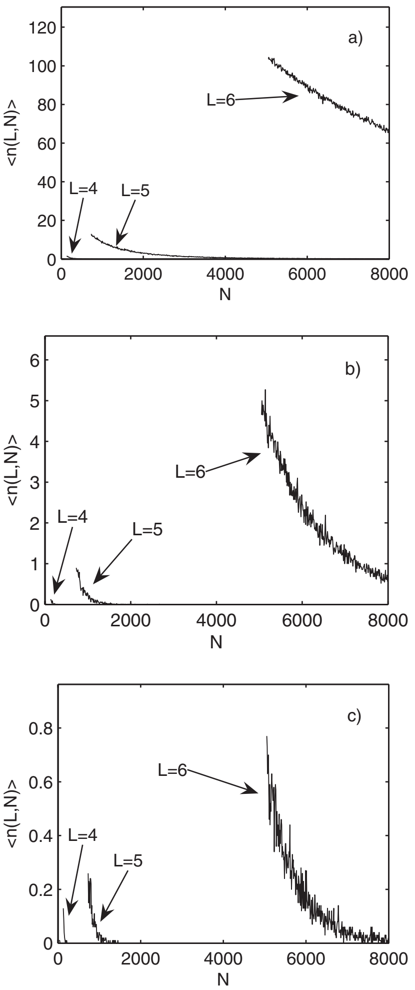

Fig. 2: Average number of forbidden patterns of length L found in a time series of length N, $\langle n(L,N)\rangle$, for noisy series of the logistic map with $\eta_{\max}=0.25$ (a), $\eta_{\max}=0.5$ (b) and for a series of uniformly distributed noise (c).

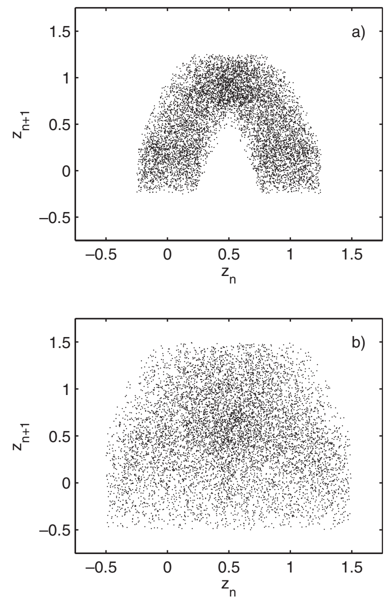

Fig. 3: Return map for noisy time series from the logistic map with $\eta_{\max}=0.25$ (a) and with $\eta_{\max}=0.5$ (b). In the latter case, the high noise level does not allow to recognize the underlying deterministic dynamics. However, we have shown that in this case the number of forbidden patterns is sensibly higher than in the purely random case.

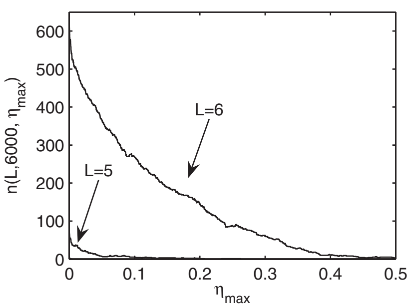

Fig. 4: Number of forbidden patterns of length L found in a noisy time series of the logistic map with length 6000 as a function of the uniform noise amplitude $\eta_{\max}$.

iii) $\eta_{\max}=1$ and $f^k(x)\equiv 0$ (only noise), respectively. Note the different order of magnitude of the vertical scales. Thus, for instance, we count on average some 2 or 3 forbidden patterns of length 6 in a 6000 point long sequence in the very noisy case ii), whereas the same count drops to 0 with high probability (or to be more specific, in about 75% of the samples) in the only-white-noise case iii). As a matter of fact, $\langle n(L,N)\rangle$ decays with increasing N because the greater N, the more unlikely that a length-L pattern is missing in a noisy or random sequence of length N; this is a statistical effect. The important features for us are the magnitude of $\langle n(L,N)\rangle$ and its decay rate with N, since these two properties are related to the true forbidden patterns and outgrowth patterns of the underlying deterministic dynamics via robustness: the smaller $\eta_{\max}$, the closer we are to the deterministic case and, therefore, the more missing order patterns and the slower their decrease with N.

Figure 3 shows z_{n+1} *vs.* z_n in the previous cases i) (fig. 3(a)) and ii) (fig. 3(b)). The higher order of magnitude of $\langle n(6,N)\rangle$ in fig. 2(b) as compared to fig. 2(c) signalizes an underlying iteration law, in spite of the fact that fig. 3(b) hardly gives any clue about this.

Finally, fig. 4 nicely illustrates the resistance of longer forbidden patterns to disappear with increasing noise levels due to their sheer number. In this figure, $N=6000$, $L=5,6$ and $0\leqslant\eta_{\max}\leqslant 0.5$.

These numerical simulations suggest the following simple-minded technique to discriminate noisy, deterministic, finite time series from the random ones, at least when the noise is white. a) Compute the number of forbidden patterns of adequate lengths (complying with eq. (4)) of the sequence in question; it is convenient to consider segments of variable lengths and to draw the corresponding curves, as in fig. 2. b) Randomize the sequence, *i.e.,* change the order of its elements in a random way. c) Proceed as in step a) with the randomized sequence. If the results concerning forbidden patterns are about the same, the sequence is very likely not deterministic (or the observational noise is so high that the deterministic component has been completely masked); otherwise, the observed sequence stems from a deterministic one. Needless to say, the method is more reliable if a statistically significant sample of sequences can be generated (for instance, by cutting a long sequence into shorter pieces).

Conclusions. – Exponential growth of allowable order patterns and, hence, the existence and super-exponential growth of forbidden order patterns are hallmarks of deterministic time series. But, in contrast to the former, the latter (whether true or false) have a well-defined and distinct structure (2) that allows, given a missing pattern of sufficiently small length, to elucidate the deterministic or random nature of a finite, noiseless time series, selectively examining a short range of longer pattern lengths. We have presented numerical evidence that forbidden patterns can also distinguish chaos from randomness in finite time series contaminated with observational white noise. In doing so, we have exploited the robustness of forbidden patterns against noise. Finally, what we have learnt from these numerical simulations has been used to implement a straightforward test to distinguish noisy, deterministic, finite time series from random ones. All this can be described as: count, randomize, count again and compare. The case of colored noise is currently under study.

* * *

We thank S. Elizalde (Dartmouth College) and U. Parlitz (University of Göttingen) for valuable comments. This research has been financed by the Spanish Ministry of Education and Science under Project No. MTM2005-04948 and FIS2006-08525, by the Spanish Ministry of Science and Technology under Project No. BFM2003-03081 and by Universidad Rey Juan Carlos and Comunidad de Madrid under Project No. URJC-CM-2006-CET-0643.

REFERENCES

[1] Ott E., *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge) 2002.
[2] Bandt C. and Pompe B., *Phys. Rev. Lett.*, **88** (2002) 174102.
[3] Amigó J. M., Kennel M. B. and Kocarev L., *Physica D*, **210** (2005) 77.
[4] Amigó J. M., Kocarev L. and Szczepanski J., *Phys. Lett. A*, **355** (2006) 27.
[5] Ortega G. J. and Louis E., *Phys. Rev. Lett.*, **81** (1998) 4345.
[6] Zhang J. and Small M., *Phys. Rev. Lett.*, **96** (2006) 238701.
[7] Choe G., *Computational Ergodic Theory* (Springer, New York) 2005.
[8] Ott E., *Rev. Mod. Phys.*, **53** (1981) 655.
[9] Bandt C., Keller G. and Pompe B., *Nonlinearity*, **15** (2002) 1595.
[10] Elizalde S., *Adv. Appl. Math.*, **36** (2006) 138.