AUTOCORRELATION TYPE FUNCTIONS FOR BIG AND DIRTY DATA SERIES

CHRISTOPH BANDT

One form of big data are signals - time series of consecutive values. In physical experiments, billions of values can now be measured within a second [1, 2]. Signals of heart and brain in intensive care, as well as seismic waves, are measured with 100 up to 1000 Hz over hours, days or even years [3, 4]. A song of 3 minutes on CD comprises 16 million values.

Music and seismic vibrations basically consist of harmonic oscillations for which classical tools like autocorrelation and spectrogram work well [5, 6, 7]. For non-Gaussian data with inaccuracies, outliers, and missing values, this note presents similar tools. Permutation entropy [8, 9, 10] has been used in physics [2, 11], medicine [12, 13, 14], and engineering [15, 16]. Big data allow a detailed analysis of ordinal patterns [17, 18, 19]. We present a new version of permutation entropy which consists of four components. The focus is on simple, universally applicable methods.

![Figure 1](image.png)

**Figure 1.** Nose breathing of a healthy volunteer in normal life. a: One minute of data without artefacts. b: Mean absolute flow for each 30 seconds of the 24 hours. c: New function $\tilde{\tau}(d)$ for each minute shows more structure, needs no calibration of data.

Classical time series from quarterly reports of companies, monthly unemployment figures, daily statistics of accidents etc. consist of 20 up to a few thousand values which were determined with great care. Machine-generated data amount to millions.

*Date:* November 17, 2014.
and are not double-checked. Usually it is easy to let the machine run faster and longer but processing goes the other way: data are smoothed out by downsampling. This note shows how high time resolution can be exploited with simple ordinal autocorrelation functions. To demonstrate the remarkable effect of these tools, they will be applied to raw data: no preprocessing, no filters, no detection of outliers, no removal of artefacts, and no polishing of the results.

Figure 1 comes from ongoing work with Achim Beule (Department ENT, Head and Neck Surgery, University of Greifswald) on respiration of healthy volunteers in everyday life. Sensors measuring air flow with sampling frequency 50 Hz were fixed in both nostrils, gently enough to ensure comfort for 24 hours measurement. Mouth breathing was not controlled, so the signals contain lots of artefacts. Traditional analysis takes averages over 30 seconds to obtain a better signal. More information is found in a function \( \tilde{\tau}(d) \), defined below, for each minute of the dirty signal. The collection of these functions, visualized like a spectrogram, shows phases of activity and sleep, various interruptions of sleep, inaccurate measurements around 8 am, a little nap after 2 pm. Frequency of respiration can be read from the lower dark red stripe which marks half of the wavelength, and the upper red and yellow stripe marks the full wavelength (4 sec in sleep, less than 3 in daily life).

1. **Up-down balance.** Only the order relation between the values of the time series \( x = (x_1, \ldots, x_T) \) will be used, not the values themselves. In the simplest case, the question is whether the time series goes more often upwards than downwards. For a delay \( d \) between 1 and \( T/2 \), let \( n_{12}(d) \) and \( n_{21}(d) \) denote the number of time points \( t \) for which \( x_t < x_{t+d} \) and \( x_t > x_{t+d} \), respectively. We determine the relative frequencies of increase and decrease over \( d \) steps: \( p_{12}(d) = n_{12}(d)/(n_{12}(d) + n_{21}(d)) \) and \( p_{21}(d) = 1 - p_{12}(d) \). Ties are disregarded. The up-down balance

\[
\beta(d) = p_{12}(d) - p_{21}(d) \quad \text{for} \quad d = 1, 2, \ldots
\]

is a kind of autocorrelation function. It reflects the dependence structure of the underlying process and has nothing to do with the size of the data.

For illustration we take water levels from the database of the National Ocean Service [20]. Since water tends to come fast and disappear more slowly, we could expect \( \beta(d) \) to be negative, at least for small \( d \). This is mostly true for water levels from lakes, like Milwaukee in Figure 2. For sea stations, tidal influence makes the data almost periodic, with a period slightly smaller than 25 hours. The data have 6 minute intervals and we measure \( d \) in hours, writing \( d = 250 \) as \( d = 25h \). A strictly periodic time series with period \( L \) fulfills \( \beta(L-d) = -\beta(d) \) which in the case \( L = 25 \) implies \( \beta(12.5) = 0 \), visible at all sea stations in Figure 2. Otherwise there are big differences: at Honolulu and Baltimore the water level is more likely to fall within the next few hours, at San Francisco it is more likely to increase, and at Anchorage there is a change at 6 hours. Each station has its specific \( \beta \)-profile, almost unchanged during 18 years, which characterizes its coastal shape. \( \beta \) can also change with the season, but these differences are smaller than those between stations.

Figure 2 indicates that \( \beta \), as well as related functions below, can solve some basic problems of statistics: describe, distinguish, and classify objects.
AUTOCORRELATION TYPE FUNCTIONS FOR BIG AND DIRTY DATA SERIES

Figure 2. Water levels at 6 minute intervals from [20]. Original data shown for Sep 1-2, 2014. Functions $\beta$ are given for September of 18 years 1997-2014, and also for January in case of Anchorage and San Francisco. $d$ runs from 6 min to 27 hours.

2. Sliding windows and statistical variation. The time series is divided into pieces, so-called windows, and $\beta$ is determined for every window. If we have some hundred windows, which may overlap, we can compose them in a map like Figure 1 or 4, where each time point corresponds to a window, the $\beta$-functions are represented by vertical lines, and the values $\beta(d)$ are color coded. Such maps show whether the regime of the underlying process changes in time, and indicate transitions between different regimes: activity and rest, or sleep stages.

If there are no systematic changes in the appearance of the functions, we can consider the underlying process as stationary and estimate statistical fluctuations. Mathematical arguments say that the standard deviation of $\beta(d)$ for fixed $d$ is $\sqrt{\frac{c}{n}}$ where $n$ is the window length and $c$ some constant. In our data, $c$ varied from $\frac{1}{2}$ to 4, and in most cases we had $c \approx 2$. For the data of Figure 2, the underlying tidal process is certainly not stationary - but the influence of the moon was reduced by calculating $\beta$ over one month, the annual component was minimized by taking only September. This gives a rough estimate for $\sigma$, since deviations still depend on $d$. For the medical data, errors can be obtained from stationary segments, e.g. deep sleep. Another error of order $\frac{d}{n}$ is observed for very large parameters $d$ (Methods 2,3).

The error estimate $\sigma \approx \frac{2}{\sqrt{n}}$ can be used to determine the appropriate window length. To have the $2\sigma$-radius of the 95% confidence interval for $\beta(d)$ smaller than 0.01, we must take $n \geq 160000$. For Figures 2 and 3 we have $n \approx 8000$ which gives a tolerance of $\pm 0.04$. Thus our method works only for big data.
A simple exact method can show that $\beta(d)$ significantly differs from zero, or the $\beta$-curves of two sites in Figure 2 significantly differ at some $d$. When windows do not overlap, the median test is applicable. Since $\beta(20) > 0$ for all 11 pieces in Figure 3b, one concludes that the median of $\beta(20)$ for the underlying process cannot be zero since the $p$-value for 11 heads in 11 coin tosses is less than 0.001. Thus, except for $\gamma$, most values of the functions in Figure 3 differ significantly from zero.

3. Patterns of length 3. Three equidistant values $x_t, x_{t+d}, x_{t+2d}$ without ties can realize six order patterns. 213 denotes the case $x_{t+d} < x_t < x_{t+2d}$.

For each pattern $\pi$ and $d = 1, 2, ..., T/3$ we count the number $n_\pi(d)$ of appearances in the same way as $n_{12}(d)$. In case $\pi = 312$ we count all $t = 1, ..., T - 2d$ with $x_{t+d} < x_{t+2d} < x_t$. Let $S$ be the sum of the six numbers. Patterns with ties are not counted. Next, we compute the relative frequencies $p_\pi(d) = n_\pi(d)/S$. For white noise it is known that all $p_\pi(d)$ are $\frac{1}{6}$ [8].

As autocorrelation type functions, we now define certain sums and differences of the $p_\pi$. The function

$$\tau(d) = p_{123}(d) + p_{321}(d) - \frac{1}{3}$$

is called persistence [17]. This function indicates the probability that the sign of $x_{t+d} - x_t$ persists when we go $d$ time steps ahead. The largest possible value of $\tau(d)$ is $\frac{2}{3}$, assumed for monotone time series. The minimal value is $-\frac{1}{3}$. The constant $\frac{1}{3}$ was chosen so that white noise has persistence zero. The letter $\tau$ indicates that this is one way to transfer Kendall’s tau to an autocorrelation function. Another version was studied in [22].

Similar to the classical autocorrelation function $\rho$, a period of length $L$ in a signal is indicated by minima of $\tau$ at $d = \frac{L}{2}, \frac{3L}{2}, \frac{5L}{2}, ...$. For these $d$ we have $x_t \approx x_{t+2d}$ so that the patterns 123 and 321 are rare. Near to $d = L, 2L, 3L, ...$ the function $\tau$ is large, as $\rho$. For a noisy sequence, however, a local minimum appears directly at $d = L, 2L, ...$, since patterns tend to have equal probability there. The larger the noise, the deeper the bump. In Figure 3b, $\tau$ at $d = 12$ and $d = 24$ shows exactly this appearance and proves the existence of a 24 hour rhythm better than $\rho$.

Beside $\tau$ and $\beta$ we define two other autocorrelation type functions. For convenience, we drop the argument $d$.

$$\gamma = p_{213} + p_{231} - p_{132} - p_{312}$$

is a measure of time irreversibility of the process, and

$$\delta = p_{132} + p_{213} - p_{231} - p_{312}$$

describes up-down scaling since it approximately fulfils $\delta(d) = \beta(2d) - \beta(d)$ (Methods 3). Like $\beta$, these functions measure certain symmetry breaks in the distribution of the time series.
Figure 3. Autocorrelation $\rho$ and ordinal functions $\Delta^2, \tau, \beta, \gamma, \delta$ for
a: the almost periodic series of water levels at Los Angeles (September 1997-2014, data from [20]), b: for the noisy series of hourly particulate values at nearby San Bernardino 2000-2011 from [21], with weak daily rhythm (Methods 5). The functions on the left are about three times larger, for $\Delta^2$ nine times, but fluctuations are of the same size.

All four ordinal functions remain unchanged when a nonlinear monotonous transformation is applied to the data. They are not influenced by low frequency components with wavelength much larger than $d$ which often appear as artefacts in data. Since $\tau, \beta, \gamma,$ and $\delta$ are defined by assertions like $X_{t+d} - X_t > 0$, they do not require full stationarity of the underlying process. Stationary increments suffice - Brownian motion for instance has $\tau(d) = -\frac{1}{6}$ for all $d$ [17]. Finally, the ordinal functions are not influenced much by a few outliers. One wrong value, no matter how large, can change $\tau(d)$ only by $\pm 2/S$ while it can completely spoil autocorrelation.

4. Permutation entropy is the Shannon entropy of the distribution of order patterns:

$$H = -\sum_\pi p_\pi \log p_\pi .$$

$H$ can be defined for any level $n$, using the vectors $(x_t, x_{t+d}, \ldots, x_{t+(n-1)d})$ and their $n!$ order patterns $\pi$ [8]. In practice we hardly go beyond $n = 7$. Used as a measure of complexity and disorder, $H$ can be calculated for time series of less than thousand values since statistical inaccuracies of the $p_\pi$ are smoothed out by averaging. Applications include EEG data [12, 13, 18, 14], optical experiments [2, 1, 11], river flow data [19], control of rotating machines [15, 16], economic applications (cf. [9]), and the theory of dynamical systems [23, 24]. Recent surveys on $H$ are [9, 10]. As a measure of disorder, $H$ assumes its maximum $\log n!$ for white noise. $D = \log n! - H$ is called divergence or Kullback-Leibler distance to the uniform distribution $p_\pi = \frac{1}{n!}$ of white noise.
In this note all functions, including autocorrelation, measure the distance of the data from white noise. For this reason, we take divergence rather than entropy, and we replace $-p_{\pi} \log p_{\pi}$ by $p_{\pi}^2$. As before, we drop the argument $d$. The function

$$\Delta^2 = \sum_{\pi} (p_{\pi} - \frac{1}{6})^2 = \sum_{\pi} p_{\pi}^2 - \frac{1}{6}$$

where the sum runs over the six order patterns $\pi$ of length 3, will be called the distance of the data from white noise. Of course, we can define $\Delta^2$ for any length $n$, replacing $\frac{1}{6}$ by $\frac{1}{n!}$. More precisely, $\Delta^2$ is the squared Euclidean distance between the observed order pattern distribution and the order pattern distribution of white noise. Considering white noise as complete disorder, $\Delta^2$ measures the amount of rule and order in the data. The minimal value 0 is obtained for white noise, and the maximum $\frac{5}{6}$ for a monotone time series.

From a practical viewpoint, $\Delta^2$ is just a rescaling of $H$, related to the quadratic Taylor approximation of $H$ around white noise parameters $p_{\pi} = \frac{1}{6}$

$$H \approx \log 6 - 3\Delta^2$$

For our data these two functions can hardly be distinguished by eyesight.

5. Partition of the distance to white noise. A Pythagoras type formula combines $\Delta^2$ with the ordinal functions:

$$4\Delta^2 = 3\tau^2 + 2\beta^2 + \gamma^2 + \delta^2$$

This holds for each $d = 1, 2, \ldots$ The equation is exact for random processes with stationary increments as well as for cyclic time series. The latter means that we calculate $p_{\pi}(d)$ from the series $(x_1, x_2, \ldots, x_T, x_1, x_2, \ldots, x_{2d})$ where $t$ runs from 1 to $T$. For real data we go only to $T - 2d$ and have a boundary effect which causes the equation to be only approximately fulfilled. The difference is negligible (Methods 4, Figure 5).

Statisticians will note that the partition formula is related to orthogonal contrasts in the analysis of variance. We use the equation to define new functions of $d$:

$$\tilde{\tau} = \frac{3\tau^2}{4\Delta^2}, \quad \tilde{\beta} = \frac{\beta^2}{2\Delta^2}, \quad \tilde{\gamma} = \frac{\gamma^2}{4\Delta^2}, \quad \tilde{\delta} = \frac{\delta^2}{4\Delta^2}.$$

By taking squares, we lose the sign of the values, but we gain a natural scale. $\tilde{\tau}, \tilde{\beta}, \tilde{\gamma}$ and $\tilde{\delta}$ lie between 0 and 1 = 100%, and they sum up to 1. For each $d$, they describe the percentage of order in the data which is due to the corresponding difference of patterns. Figure 1 shows $\tilde{\tau}$ for one-minute non-overlapping sliding windows as color code on vertical lines.

For Gaussian and elliptical symmetric processes, the functions $\beta, \gamma,$ and $\delta$ are all zero, and $\tilde{\tau}$ is 1, for every $d$ [17]. Thus the map of $\tilde{\tau}$ shows to which extent the data come from a Gaussian process and where the symmetry is broken. This may complement rigorous tests for elliptical symmetry based on the covariance matrix [25, 26, 27]. In contrast to simulated data from ARMA processes, long time series in reality seem never to fulfill $\tilde{\tau} \approx 1$ for all $d$ and all windows. There are some types of data for which $\tau$ dominates the other components. Music, seismic waves, and
EEG data belong to this class. For heart and respiration data and for hydrological series, the functions $\beta$ and $\delta$, and less frequently $\gamma$, represent essential parts of $\Delta^2$ in every window, but usually only for special values of $d$. See Figures 5, 7, and 8 in the supplement.

![Figure 4](image)

**Figure 4.** Sleep stages of a healthy subject, data from [28]. a: 10 seconds of data from EEG channel Fp2-F4, ECG and plethysmogram. b: Expert annotation of sleep depth from [28] agrees with distance $\Delta^2$. c: Function $\tilde{\beta}$ of the plethysmogram describes changes including REM.

### 6. Sleep data.

As an application, we study data from the CAP sleep database by Terzano et al. [28] available at physionet [3]. Sleep stages S1-S4 and R for REM sleep were already annotated by experts, mainly using the EEG channel Fp2-F4 and the oculogram. Figure 4 demonstrates that permutation entropy $\Delta^2$ of that EEG channel, averaged over $d = 2, ..., 20$ gives an almost identical estimate of sleep depth. In [13, 29] permutation entropy was already recommended as indicator of sleep stages. We verified this coincidence in all data of the database with a normal EEG, which seems magic since $\Delta^2$ was introduced as a simple quantity, without any regard to sleep data. For 512 Hz sampling frequency, $d = 2, ..., 20$ corresponds to the time scale between 4 and 40 ms. Thus our $\Delta^2$ is a measure of smoothness of data at small scales which increases when high-frequency oscillations disappear. This seems a complementary viewpoint to the classical R&K rules for sleep stage classification and their recent modification [30] which refer to the appearance of low-frequency waves and patterns.

REM phases are more difficult to detect. The oculogram was used for annotation. In Figure 4c, we illustrate $\tilde{\beta}$ for the plethysmogram which is measured with an optical sensor at the fingertip. We can see all interruptions of sleep and a lot of breakpoints which coincide with changes detected by the annotator. The REM phases are characterized by high values of $\tilde{\beta}(d)$, as well as an increase and a strongly increased variability of the heart rate: the characteristic wavelength $d$ goes down.
Since the variability also concerns the dependence structure of the signal, changes become more apparent than in a graph of heart rate alone. Oximeters measuring the plethysmogram are cheap and easy to apply, in bed at home and possibly in daily life. Many similar devices - armwrists, shirts, smartphone apps - are currently being developed. Our ordinal functions show some potential for evaluating and visualizing their high-frequency data.

7. Conclusion. On the basis of order patterns, error-resistent autocorrelation functions for big data were introduced. For patterns of length 3, the distance to white noise was divided into four interesting components. Various extensions seem possible. To develop a coherent theory of ordinal time series remains a challenge, despite groundbreaking work by Marc Hallin and co-authors [31, 22, 25] on rank statistics. On the practical side, a spectrogram-like visualization was introduced which allows to see at one glance the course of heart and respiration data over 24 hours (Figures 1,4). Simple as it is, this technique may apply to data of sensors on satellites, weather stations, factory chimneys etc. as well as to experiments on nanoscale [1, 11] which may lay the ground for future generations of sensors. In any application, the methods presented here have to be modified, combined with established techniques, specific machine-learning and imaging tricks. Mathematical extraction of information must keep track with the revolution in sensor and computer technology: this seems an emerging field of study.

REFERENCES

[1] A. Aragoneses, N. Rubido, J. Tiana-Aisina, M.C. Torrent & C. Masoller, Distinguishing signatures of determinism and stochasticity in spiking complex systems, Scientific Reports 3, Article 1778 (2012)
[2] M.C. Soriano, L. Zunino, O.A. Rosso, I. Fischer & C.R. Mirasso, Time scales of a chaotic semiconductor laser with optical feedback under the lens of a permutation information analysis, IEEE J. Quantum Electronics 47 issue. 2, 252-261 (2011)
[3] A.L. Goldberger et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215]; (2000). Data at: physionet.org/physiobank/database/capslpdb
[4] GFZ Seismological Data Archive: geofon.gfz-potsdam.de/waveform
[5] P.J. Brockwell & R.A. Davies, Time Series; Theory and Methods, 2nd ed., Springer, New York 1991
[6] R.H. Shumway & D.S. Stoffer, Time Series Analysis and Its Applications, 2nd ed., Springer, New York 2006
[7] A. Davis, M. Rubinstein, N. Wadhwa, G.J. Mysore, F. Durand & W.T. Freeman, The visual microphone: passive recovery of sound from video, SIGGRAPH 2014
[8] C. Bandt & B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88, 174102 (2002)
[9] M. Zunin, L. Zunino, O.A. Rosso & D. Papo, Permutation entropy and its main biomedical and econophysics applications: a review, Entropy 14, 1553-1577 (2012)
[10] J. Amigo, K. Keller & J. Kurths (eds.), Recent progress in symbolic dynamics and permutation entropy, Eur. Phys. J. Special Topics 222 (2013)
[11] J.P. Toomey & D.M. Kane, Mapping the dynamical complexity of a semiconductor laser with optical feedback using permutation entropy, Optics Express 22, issue 2, 1713-1725 (2014)
[12] M. Staniek & K. Lehnertz, Symbolic transfer entropy, Phys. Rev. Lett. 100, 158101 (2008)
[13] N. Nicolaou & J. Georgiou, The use of permutation entropy to characterize sleep encephalograms, Clin. EEG Neurosci 2011, 42:24
[14] E. Ferlazzo et al. Permutation entropy of scalp EEG: a tool to investigate epilepsies, Clinical Neurophysiology 125, 13-20 (2014)
[15] U. Nair, B.M. Krishna, V.N.N. Namboothiri & V.P.N. Nampoori, Permutation entropy based real-time chatter detection using audio signal in turning process, Int. J. Adv. Manuf. Technol. 46, 61-68 (2010)
[16] R. Yan, Y. Liu & R.X.Gao, Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines, Mechan. Syst. Signal Processing 29, 474-484 (2012)
[17] C. Bandt & F. Shiha, Order patterns in time series, J. Time Series Analysis 28, 646-665 (2007)
[18] G. Ouyang, C. Dang, D.A. Richards & X. Li, Ordinal pattern based similarity analysis for EEG recordings, Clinical Neurophysiology 121 (2010), 694-703
[19] H. Lange, O.A. Rosso & M. Hauhs, Ordinal pattern and statistical complexity analysis of daily stream flow time series, Eur. Phys. J. Special Topics 222, 535-552 (2013)
[20] The National Water Level Observation Network: www.tidesandcurrents.noaa.gov/nwlon.html
[21] California Environmental Protection Agency, Air Resources Board: www.arb.ca.gov/aqd/aqcdcd/aqcdcdld.htm
[22] T.S. Ferguson, C. Genest & M. Hallin, Kendall’s tau for serial dependence, Canadian J. Stat. 28, 587-604 (2000)
[23] J. Amigo, Permutation complexity in dynamical systems, Springer 2010
[24] A.M. Unakafov & K. Keller, Conditional entropy of ordinal patterns, Physica D 269, 94-102 (2014)
[25] M. Hallin & D. Paindaveine, Semiparametrically efficient rank-based inference for shape I. Ann. Statistics 34, 2707-2756 (2008)
[26] L. Sakharenko, Testing for ellipsoidal symmetry: a comparison study, Computational Stat. Data Analysis 53, 565-581 (2008)
[27] A. Onatski, M.J. Moreira & M. Hallin, Asymptotic power of sphericity tests for high-dimensional data, Ann. Statistics 41, 1204-1231 (2013)
[28] M.G. Terzano et al., Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep, Sleep Med. 2 (6), 537-553 (2001).
[29] C.-E. Kuo, S.-F. Liang, Automatic stage scoring of single-channel sleep EEG based on multiscale permutation entropy, Biomedical Circuits and Systems Conference (BioCAS), IEEE 2011, 448-451
[30] C. Iber, S. Anconi-Israel, A. Chesson & S.F. Quan, The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester: American Academy of Sleep Medicine (2007)
[31] M. Hallin & M.L. Puri, Aligned rank tests for linear models with autocorrelated error terms, J. Multivariate Analysis 50, 175-237 (1994)
Methods

1. The program code. A few lines of MATLAB code provide all our functions.

```matlab
x(1:T) denotes the given time series, d is between 1 and dmax.
y0=x(1:T-2*d); y1=x(d+1:T-d); y2=x(2*d+1:T);
s=2*((y0>y1)+(y0>y2))+(y1>y2);
s is a vector which contains the symbols 0, 1, ..., 5 representing the six order patterns of (x_t, x_{t+d}, x_{t+2d}) for d = 1, ..., T - 2d. We correct for missing values NaN and ties by assigning symbols 6 to 11:
h=isnan(y0)|isnan(y1)|isnan(y2)|(y0==y1)|(y0==y2)|(y1==y2);
s=s+6*h;
Here | means 'or'. If patterns with equality need be counted, as in [?], this line has to be modified. We now use the histogram function to determine the frequencies of the six patterns.
p=hist(s,12)/(T-2*d-sum(h)); q=p(1:6);
This will determine all our functions, for example:
beta=p(1)-p(6); tau=p(1)+p(6)-1/3; h2=(q-1/6)*(q-1/6)'; H=-log(q)*q';
```

On a PC, the algorithm obtains Figure 3 within a few seconds and Figure 1 within a few minutes. In some papers [10], permutation entropy of scale d uses patterns for sums of d consecutive terms of the x_t, which makes sense if the x_t represent a density function, like precipitation or workload on a server. This version is easily implemented using cumulative sums, just adding `x=cumsum(x)` as first line to the program.

2. Statistical accuracy. We consider a fixed d and windows of length n. Then p_{12}(d) is a random sum of n terms 1,0 which say whether x_t < x_{t+d} is true or false. According to the binomial model, the standard deviation of p_{12}(d) and β(d) = 2p_{12}(d) - 1 would be \( \frac{1}{\sqrt{n}} \) and \( \frac{1}{\sqrt{n}} \), respectively. In reality, the variation is bigger because the terms in the sum are correlated. The correlations between differences Y_t = X_{t+d} - X_t are relevant. As a simple model, we take the sum of \( \frac{n}{k} \) independent terms 1,0 each of which is repeated k times. For \( \beta(d) \) this gives the standard deviation \( \frac{c}{\sqrt{kn}} \) with \( c = k\sqrt{k} \). An estimate of c can only come from the data. As a rule we got c \approx 2. For τ the variation is a bit smaller, for γ and δ a bit larger. Correlations of neighboring terms often increase with increasing d.

3. Identities for pattern frequencies and boundary effect. We fix d and consider the p_ for the whole time series. The equation p_{12} = p_{123} + p_{231} + p_{132} holds since on the right we count all t = 1, ..., T - 2d with x_t < x_{t+d}. Similarly, p_{12} = p_{123} + p_{213} + p_{312} because on the right we count all t = d + 1, ..., T - d with x_t < x_{t+d}. Strictly speaking, both relations are not correct since the count for p_{12} is over t = 1, ..., T - d. The largest possible difference is \( \frac{d}{T-d} \) when we exclude ties for simplicity. This upper bound is negligible for large T and comparably small d, and random fluctuations usually make the difference much smaller than the upper bound. Taking the difference of the two equations we see that the quantity

\[ \epsilon = p_{231} + p_{132} - p_{213} - p_{312} \]
equals zero, with the same degree of accuracy. In other words, the numbers of local maxima and of local minima in a time series coincide. Adding both identities for $p_{12}$ and subtracting $1 = \sum p_{\pi}$ we obtain

$$\beta = p_{123} - p_{321}$$

which has been used throughout the paper to calculate $\beta$, see the code above. Another relation of this type is

$$p_{12}(2d) = p_{123}(d) + p_{132}(d) + p_{213}(d).$$

As a consequence, we get $\beta(2d) = 2p_{12}(2d) - 1 = \beta(d) + \delta(d)$ which says that the functions $\beta$ and $\delta$ are tightly connected. For the mathematical model of processes with stationary increments, as well as for the cyclic time series mentioned in section 5, all these equations are exact [17].

4. Proof of the entropy partition formula. As in the code, we set $q_1 = p_{123} - \frac{1}{6}, q_2 = p_{132} - \frac{1}{6}, q_3 = p_{213} - \frac{1}{6}, q_4 = p_{231} - \frac{1}{6}, q_5 = p_{312} - \frac{1}{6}, q_6 = p_{321} - \frac{1}{6}$. By high-school algebra

$$4 \sum q_k^2 = 2(q_1 + q_6)^2 + 2(q_4 - q_5)^2 + (q_2 + q_3 + q_4 + q_5)^2 + (q_2 - q_3 - q_4 + q_5)^2$$

$$+ (q_2 + q_3 - q_4 - q_5)^2 + (q_2 - q_3 + q_4 - q_5)^2$$

Since $\sum q_k = 0$, the third square on the right is the same as the first. Using the definitions of $\Delta^2, \tau, \gamma, \delta$ and the identities above for $\beta$ and $\epsilon$ we get

$$4\Delta^2 = 3\tau^2 + 2\beta^2 + \delta^2 + \gamma^2 + \epsilon^2$$

which implies the formula because $\epsilon = 0$. The equation is exact for cyclic time series, and for random processes with stationary increments. For ordinary time series there is a small error, as discussed above, and we must check the accuracy of the equation from the data, see Figure 5-8. Generally, the differences were smaller than 1%.

5. Supplementary material. As an extension of Figure 4, the components of $\Delta^2$ for the plethysmogram of another control n2 of [28] are visualized in Figure 5, providing also a visual verification of the partition formula. Autocorrelation and persistence are compared in Figure 6 for a simple AR2-process, for a song of the Beatles in Figure 7 and for a laser experiment of Sorriano et al. [2] in Figure 8.
Figure 5. The components of $\Delta^2$ for the plethysmogram of control n2 of [28]. The overall percentage is 48% for $\tilde{\tau}$, 24% for both $\tilde{\beta}$ and $\tilde{\delta}$, and 3% for $\tilde{\gamma}$. The remainder has 0.05% in average, concentrated on times with strong artefacts.

Figure 6. Autocorrelation and persistence for an AR2 process with additive noises. In each case 10 samples of 2000 points were processed, the fat line is the mean. a) Original signal. b) Gaussian white noise, signal-to-noise = 3:1. c) 1% outliers. d),e) Low- and high-frequency disturbances.
Figure 7. The first 12 seconds of the song "Hey Jude" of the Beatles in sliding window analysis. Autocorrelation and persistence were calculated for overlapping windows of 0.25 second length which means $n = 11025$. The $y$-axis is reversed to better recognize the melody, the range is $d = 49, \ldots, 294$ corresponding to 900 Hz down to 150 Hz. Persistence gives a very similar picture as autocorrelation, with the exception of the local minima at full periods. The function $\tilde{\tau}$ is shown at the bottom. It shows that squaring destroys some structure, and that there are quite a number of blue spots left where $\tilde{\beta}$ and $\tilde{\delta}$ yield the main contribution to $\Delta^2$. The overall percentage is 60% for $\tilde{\tau}$, 15% for $\tilde{\beta}$, 17% for $\tilde{\delta}$, and 8% for $\tilde{\gamma}$. The rest has mean 0.06%.
Figure 8. Data for this figure come from the laser experiment of Sorriano et al. [2]. A laser with wavelength 1542 nm and 11.7 mA threshold current was excited just above the threshold with 12 mA. The sampling frequency was 25ps which means that 240000 values were measured during 6µs. The optical feedback was arranged so that the signal has a period of L = 1544. We try to find this period with a window of 2.3 periods, or 88 ns, just the length of the series shown in a where periodicity can hardly be seen. In b and c we study autocorrelation and persistence around d = L. The vertical interruptions show low-frequency disturbances, the laser does not run smoothly. d and e show ρ and τ around d = \frac{L}{2}. While ρ is a bit better at the full wavelength, τ detects the periodicity reliably at the half wavelength.