Peer-reviewed version available at *Entropy* **2019**, *21*, 613; doi:10.3390/e21060613

Article

SMALL ORDER PATTERNS IN BIG TIME SERIES: A PRACTICAL GUIDE

Christoph Bandt

Institute of Mathematics, University of Greifswald, 17487 Greifswald, Germany; bandt@uni-greifswald.de

Abstract: The study of order patterns of three equally spaced values x_t, x_{t+d}, x_{t+2d} in a time series is a powerful tool. The lag d is changed in a wide range so that differences of frequencies of order patterns become autocorrelation functions. Similar to a spectrogram in speech analysis, four ordinal autocorrelation functions are used to visualize big data series, as for instance heart and brain activity over many hours. The method applies to real data without preprocessing, outliers and missing data do not matter. On the theoretical side, we study properties of order correlation functions and show that the four autocorrelation functions are orthogonal in a certain sense. An analysis of variance of a modified permutation entropy can be performed with four variance components associated with the functions.

Keywords: permutation entropy; autocorrelation; time series; order pattern

1. Order patterns fit big data

Few decades ago, when time series analysis was created, it did suffer from a lack of data. Now the challenge is an oversupply of data. Old methods and models have to be modified, and new ones have to be developed. Let us state the main requirements to new methods for big time series.

- Basic methods should be simple and transparent.
- Few assumptions should be made on the underlying process.
- Algorithms should be resilient with respect to outliers and artefacts.
- Computations should be very fast.

Classical probabilistic methods have problems with the second property. They often require conditions like multivariate normal distribution which are not fulfilled in practice, or formulate their conditions in terms of limits so that they cannot be checked. On the other hand, machine learning procedures are in fashion, and very successful for specific purposes, but they do not have the first property. They do not provide clear insight into the observed phenomena.

The analysis of order patterns in a signal fulfils all four requirements. It has been neglected until recently when the tedious work of counting 'larger' and 'smaller' in a large series could be delegated to a computer. Today, use of permutation entropy is well-established [1–3], and pattern frequencies have been studied in particular for chaotic dynamical systems and heartbeat [4,5]. Here we continue our efforts [6–8] to supply a theoretical foundation of ordinal time series analysis, concentrating on the simplest patterns of length 3, depicted in Figure 2 below.

After years of doubt, we are now convinced that the study of order patterns in signals and functions is interesting for its own sake, even from a theoretical point of view. It does not fit the dominating analysis of smooth functions, and in [6] we found only few connections with the established theory of stochastic processes. However, the emerging modern combinatorics of permutations [9–12] will probably provide appropriate models.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

This note will show that for many practical applications, order patterns are comparable or even more appropriate than classical methods. Section 2 presents key concepts of the paper. In Section 3 we explain the basic calculation of pattern frequencies, including treatment of ties and missing values. Properties of order correlation functions are discussed in Section 5, leading to an analysis of variance of permutation entropy. The main purpose of this paper, however, is a proof of concept by application to biomedical signals, to speech and music, and to environmental and weather signals.

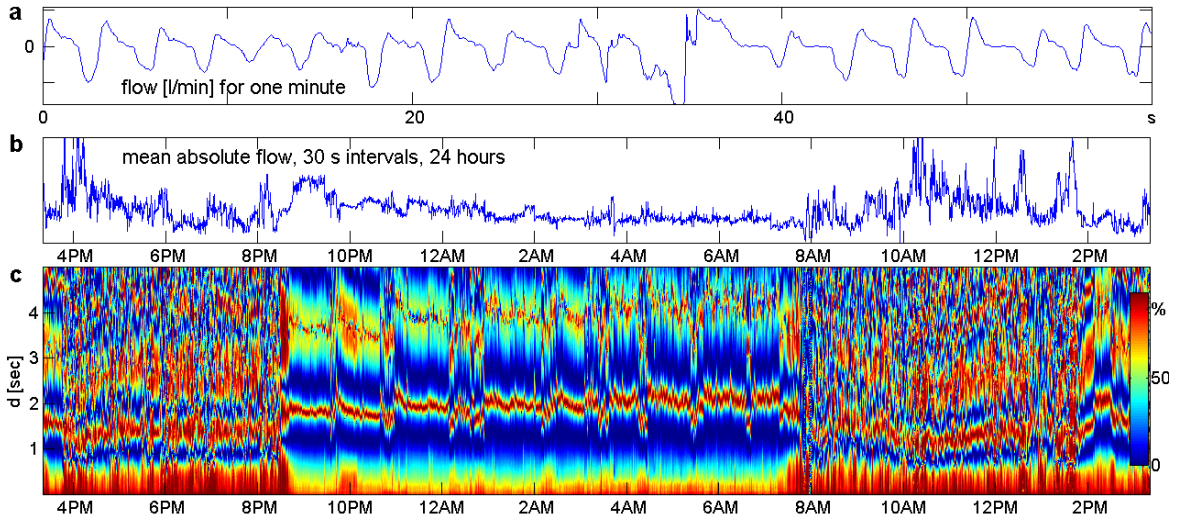


Figure 1. Respiration of a healthy volunteer during 24 hours of normal life. a: One minute of clean data. b: Mean flow for intervals of 30 seconds. c: Order correlation function $\tilde{\tau}(d)$ for each minute.

Figure 1 gives a first example of a typical application. Data come from a cooperation with Achim Beule (University of Münster and University of Greifswald, Department ENT, Head and Neck Surgery). To study respiration in everyday life, two sensors measuring air flow intensity with sampling frequency 50 Hz were fitted to the nostrils of a healthy volunteer. This experimental setting was by no means perfect. For example, mouth breathing could not be controlled. As a result, the data contain lots of artefacts, and even the 3000 values of a 'clean minute' shown in Figure 1a look pretty irregular. Traditional analysis takes averages over 30 seconds to obtain a better signal, as shown in Figure 1b. Much more information is contained in a function $\tilde{\tau}(d)$, defined below. Figure 1c shows the collection of these functions for all minutes of the signal, visualized like a spectrogram. We see 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. The upper light stripe marks the full wavelength: 4 s in sleep, 3 s or less in daily life.

2. Key concepts and viewpoints

We assume we have a time series $x_1, x_2, ..., x_T$. To any three equally spaced values x_t, x_{t+d}, x_{t+2d} we can assign one of the six order patterns in Figure 2. To the pattern $\pi = 123$ we determine its relative frequency $p_\pi(d) = p_{123}(d)$ in the time series, and this is done for each pattern π.. Details are described in Section 3. Here we explain how to arrange and combine the six pattern frequencies. Patterns of length 4 or longer are more difficult to understand and will not be studied in this paper.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

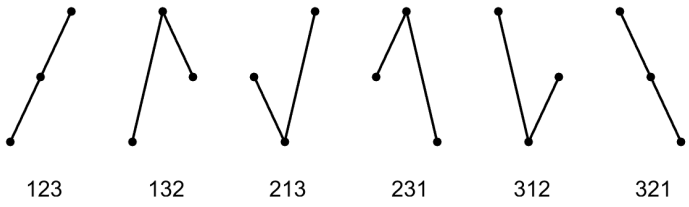


Figure 2. The six order patterns of length 3

The important point is that we vary the lag or delay d - between 1 and 1000, say. Thus $p_\pi(d)$ is a function like the autocorrelation function. Both are obtained by a kind of averaging over a certain time period. Our hope is that *such a function will express essential features of the process which generates the data, and suppress unimportant individual properties of the observed series.* Thus $p_\pi(d)$ is considered as an estimate of a probability which belongs to the underlying process. To justify this viewpoint, we must make a stationarity assumption for the process: the probability of a pattern does not change during time. This is a weak condition, for instance stationary increments in the usual sense will be sufficient [6]. A stronger condition is already needed when we define an average value $\overline{x}$ of the x_t. In practice, stationarity is not fulfilled for a long series as in Figure 1. In such biomedical series it can be assumed for small intervals of 1 minute, for which we determined the $p_\pi(d)$. There may be a few minutes where respiration drastically changes, but on the whole the stationarity assumption is natural and appropriate.

It turns out that the pattern frequencies themselves are not so informative, but they can be combined to form better descriptions of the underlying process. The *permutation entropy* is

$$H(d) = -\sum_\pi p_\pi(d) \log p_\pi(d) \,, \tag{1}$$

where the sum is taken over the six patterns in Figure 2, or all $m!$ permutations of some lengths m. In other words, we have the probability space of order patterns and take its Shannon entropy. The permutation entropy is a measure of complexity of the underlying process [13] and has found lots of applications: distinguishing chaotic and noisy dynamics, classifying sleep stages and detecting epileptic activity from brain signals, etc.

A similar complexity measure introduced in [7,14] is the *distance to white noise*

$$\Delta^2(d) = \sum_\pi \left(p_\pi(d) - \frac{1}{m!} \right)^2 . \tag{2}$$

White noise is a series of independent random numbers from a fixed distribution, and it is well known that for this process, all pattern probabilities $p_\pi(d)$ are equal to $1/m!$ for length m, in particular $1/6$ for length 3. We just take the quadratic Euclidean distance between the vectors of observed pattern frequencies and the frequencies of white noise. There is an even simpler interpretation: the average pattern frequency is always $1/m!$, so the *variance of pattern frequencies* equals $\Delta^2/m!$. In [8] it was pointed out that Δ^2 can be considered as a rescaled Taylor approximation of H, and it has a more convenient scale than H.

Now let us come back to our six patterns of length $m = 3$. It turns out that four differences of pattern frequencies provide meaningful autocorrelation functions [6,7,14].

$$\text{up-down balance} \qquad \beta(d) = p_{123}(d) - p_{321}(d) = p_{12}(d) - p_{21}(d) \,, \tag{3}$$

$$\text{persistence} \qquad \tau(d) = p_{123}(d) + p_{321}(d) - \frac{1}{3} \,, \tag{4}$$

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

$$\text{rotational asymmetry} \qquad \gamma(d) = p_{213}(d) + p_{231}(d) - p_{132}(d) - p_{312}(d)\ , \text{ and} \tag{5}$$

$$\text{up-down scaling} \qquad \delta(d) = p_{132}(d) + p_{213}(d) - p_{231}(d) - p_{312}(d)\ . \tag{6}$$

The names sound a bit clumsy, and the interpretation of the functions in Section 4 is not straightforward. However, it will be shown that these four functions include all information of the six pattern frequencies, that they are orthogonal in a certain sense, and form a variance decomposition of Δ^2 given by the Pythagoras type formula

$$4\Delta^2 = 3\tau^2 + 2\beta^2 + \gamma^2 + \delta^2\ . \tag{7}$$

For the equal pattern probabilities of a white noise process all terms of this equation are zero. Thus the definitions were arranged so that white noise is a good null hypothesis for statistical tests. This aspect will not be worked out here. However, the *variance components* of Δ^2 are considered as some other ordinal autocorrelation functions and used in some applications, as Figure 1c.

$$\tilde{\tau} = \frac{3\tau^2}{4\Delta^2}\ , \quad \tilde{\beta} = \frac{\beta^2}{2\Delta^2}\ , \quad \tilde{\gamma} = \frac{\gamma^2}{4\Delta^2}\ , \quad \tilde{\delta} = \frac{\delta^2}{4\Delta^2}\ . \tag{8}$$

A stationary Gaussian process is completely determined by its mean, variance, and autocorrelation function $\rho(d)$, and τ can be directly calculated from ρ [6]. For stationary Gaussian processes, and more general for all reversible processes, β, γ, and δ are zero for every d. Thus our correlation functions can be considered as a tool for non-Gaussian and irreversible processes. Actually, calculation of β, γ, and δ in our examples will show that processes in practice are mostly irreversible and non-Gaussian.

To conclude this introduction, let us explain what we call a 'big time series'.. We assume that *we observe a continuous quantity in continuous time*, like temperature, wind speed, blood pressure, intensity of an electric or acoustic signal. Current sensor technology makes it possible to measure with big accuracy as fast and as long as we want. On the one hand this means that values x_t are rarely equal, so cases of equality can be discarded. On the other hand, we have no smallest lag d. Although $d = 0$ makes no sense, we can even think about taking the limit $d \to 0$. In practice, we must always discretize the measurement, so we have a minimal $d = 1$. However, we could also measure with double speed and thus work with $d = \frac{1}{2}$.

The situation here is different from discrete dynamical systems, like logistic or Henon maps, ARMA models, or RR-intervals of heartbeat where the state of the system changes step by step. All the above concepts work well for those systems. However, such systems usually have a short memory so that it suffices to consider a few small values of d. Figure 1c, where d runs from 1 to 300, would not show any structure. On the other hand, there are well-established methods for such systems. It also makes sense to consider single $p_\pi(d)$, even for long patterns [2,4].

The challenge and bottleneck in time series analysis are the great continuous systems, where we may observe interacting periodicities and dynamical effects over several scales. Below we mention tides, with a daily, monthly and yearly scale, and biomedical signals with scales for heart, respiration, slow biorhythms, day and night. Such kind of data must now be addressed.

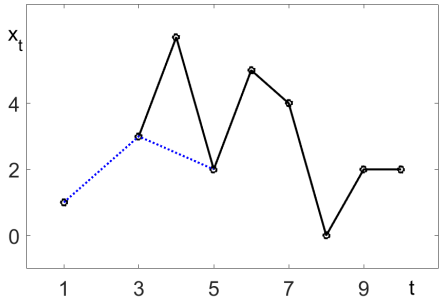
3. Calculation of pattern frequencies

We consider a time series $x_1, x_2, ..., x_T$ with T values. For a brain signal measured with 500 Hz for 8 hours, T amounts to 15 millions. However, since we divide the series into smaller parts where the underlying process is assumed to be stationary, the typical length T will vary between 500 and 20000. Any three consecutive values x_t, x_{t+1}, x_{t+2} form one of the six order patterns, or permutations, shown in Figure 2. We also consider three values x_t, x_{t+d}, x_{t+2d} with a time lag $d > 1$. The points represent pattern 231, for instance, if $x_{t+2d} < x_t < x_{t+d}$. If there are ties $x_s = x_t$ or missing values, the pattern is not defined. The initial time point t runs from 1 to $T - 2d$. The delay parameter d can vary between 1 and $d_{\max} \le T/6$, say, and has the same meaning as in classical autocorrelation. For fixed d, let n be the

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

number of time points t for which a pattern is defined, and n_π the number of time points where we have pattern π. Then the pattern frequency is $p_\pi(d) = n_\pi/n$.

In Figure 3 we have $T = 10$, and the value x_2 is missing. Thus $t = 1$ cannot be considered for $d = 1$, but for all $d > 1$. Moreover, there are three equal values $x_5 = x_9 = x_{10}$ which exclude $t = 8$ for $d = 1$, and $t = 5$ for $d = 2$. The resulting n_π are listed in the table. We obtain $p_{312}(1) = 2/5$, for instance.



π	123	132	213	231	312	321	n
$d = 1$	-	1	1	-	2	1	5
$d = 2$	-	1	1	-	1	1	4
$d = 3$	-	1	-	1	-	1	3

Figure 3. Example time series and pattern frequencies n_π. The dotted line indicates $d = 2$.

Since we have only six patterns, the statistics of pattern frequencies is excellent, even for short time series like $T = 300$. For $m = 6$, for instance, we have $6! = 720$ patterns and need a very long time series to estimate all those pattern frequencies accurately. Permutation entropy would still work, however, since it is an average over all patterns. Accuracy of estimates is one more reason to restrict ourselves to patterns of length 3.

We can also determine frequencies of patterns 12 and 21 of length 2, for $x_t < x_{t+d}$ and $x_t > x_{t+d}$. For $d = 1$, we have $n_{12} = n_{21} = 3$, and for $d = 2$ we get $n_{12} = 3, n_{21} = 4$. When we determine $\beta(d) = p_{12}(d) - p_{21}(d)$, we obtain $\beta(1) = 0, \beta(2) = -1/7$. If we use the definition $\beta(d) = p_{123}(d) - p_{321}(d)$ in (3), we get $\beta(1) = -1/5, \beta(2) = -1/4$. What is wrong?

There are several equations for the $p_\pi = p_\pi(d)$ for fixed d. The basic equation is always accurate:

$$\sum_\pi p_\pi = 1 \qquad \text{for fixed } d \text{ and } m\,. \tag{9}$$

The following equation is also accurate for probabilities in a stationary process [6].

$$p_{123} + p_{132} + p_{231} = p_{12} = p_{123} + p_{213} + p_{312}\,. \tag{10}$$

To show equality on the left, divide the probability $P(x_t < x_{t+d})$ into three terms, depending on whether x_{t+2d} is above, between or below x_t and x_{t+d}. For the equality on the right, we consider $P(x_{t+d} < x_{t+2d})$ and sum up three cases for x_t.

Now if we calculate the frequencies for a time series with $d = 1$, then p_{12} is an average over $t = 1, ..., T-1$. To determine p_{12} from the 3-patterns on the left, we average over $t = 1, ..., T-2$. When we take the 3-patterns on the right, we average over $t = 2, ..., T-1$. There is a marginal difference. Moreover, the exclusion of missing and equal values can be different for patterns of different length, as demonstrated above. Such marginal effects are the reason for the difference in the two expressions for β in (3). Actually, (3) is a corollary of (10) and $p_{321} + p_{312} + p_{213} = p_{21}$, see [6]

For $d = 1$, such marginal effects are really harmless. For large d, however, they can become larger since the margins of (10) will involve d time points on the right and left end of the time series. This can be a problem when we have a downward trend at the beginning of the series and an upward trend at the end. For EEG brain data we helped ourselves defining the parts of the large series not by equal length, but by zero crossings of the time series from the positive to the negative side [8]. Other data may require other solutions for improved estimates of pattern frequencies.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

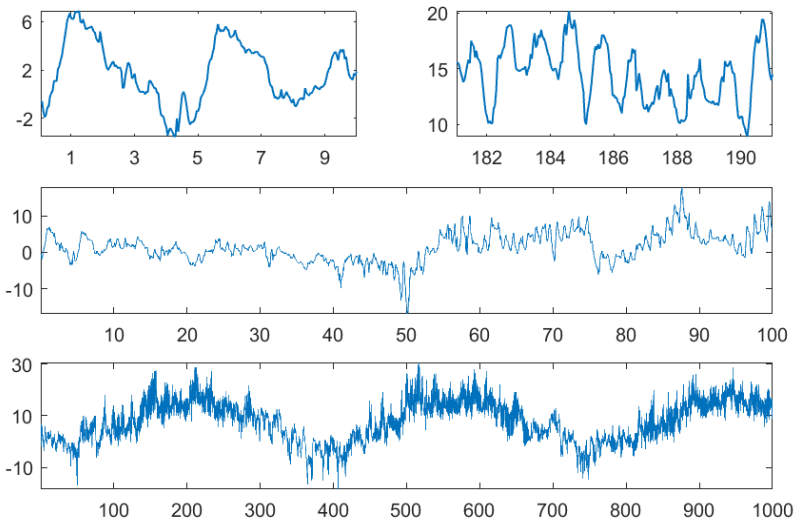


Figure 4. Temperature in ^{0}C for the first 10 days in January and July 1978, for 100 and 1000 days

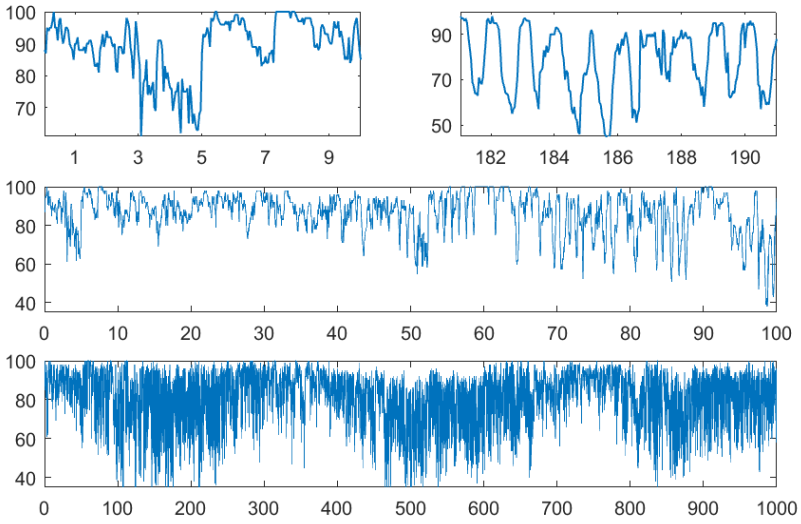


Figure 5. Relative humidity in % for the first 10 days in January and July 1978, for 100 and 1000 days

4. First examples: weather data

To demonstrate the use of the above functions, we take hourly measurements of air temperature and relative humidity from the author's town: Greifswald, North Germany. The German weather service DWD [15]. offers such data for hundreds of stations, where results are expected to be similar. Measurements of our data started 1 January 1978. Figures 4, 5 show temperature and humidity for the first 10, 100, and 1000 days, and also for the first 10 days of July 1978. While in summer there is an obvious day-night rhythm, this need not be so in winter when bright sunshine often comes together with cold temperature. This effect is also visible in the bottom panel with data of almost three years.

We now look for correlation functions which describe the underlying weather process. We calculate them bimonthly, and draw the curves for 35 years for the same season into one plot. When they agree, we have a nice description of the underlying process - but this can never be perfect since

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

weather and calendar are not in one-to-one correspondence. Figure 6 compares classical autocorrelation with our persistence function.

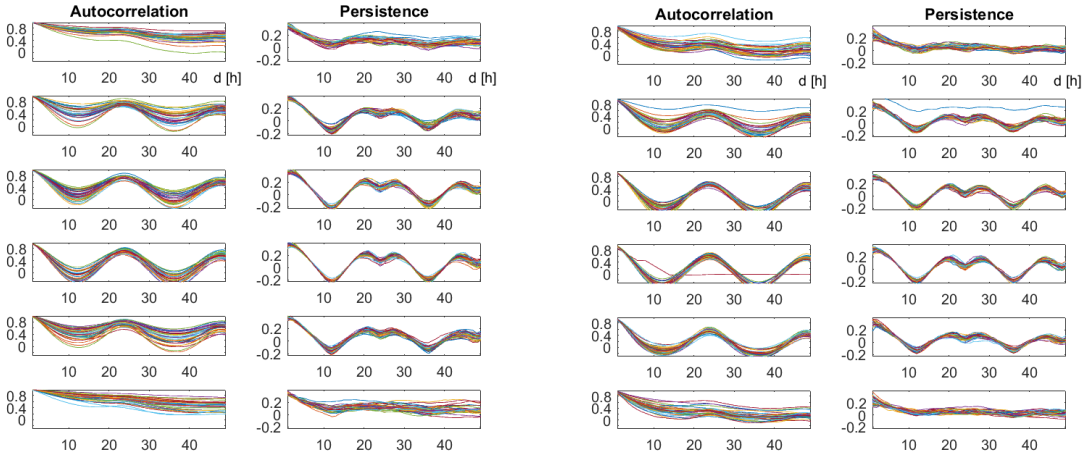


Figure 6. Autocorrelation and persistence for temperature (left part) and relative humidity (right part). The curves correspond to 35 consecutive years. The lag d runs from 1 to 49 hours. Each row describes a two-month period, from January/February up to November/December.

In winter (November to February) a day-night rhythm is not found, and all correlation functions differ a lot over the years. From March until October, the day-night rhythm is well recognized, both by autocorrelation and persistence, and both for temperature and humidity. Classical autocorrelation curves coincide best at the full period maximum at $d = 24$ hours while persistence accurately shows the minimum at the half period of 12 hours. At the full period, there is a local minimum between two maximum points, which will be explained in the next section. Both functions succeed in defining the basic period within each window of 2 months (1461 hourly values). The coincidence over years seems even better for persistence.

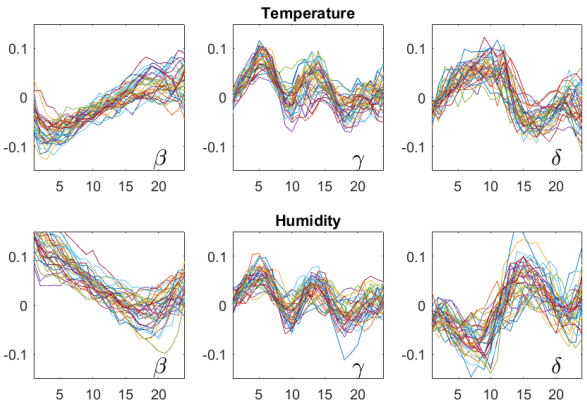


Figure 7. The functions β, γ, and δ, for $d = 1, ..., 24$ hours, for July-August of 35 consecutive years. The upper row corresponds to temperature, the bottom row to relative humidity. Although there is considerable variation over the years, some common structure can be seen.

Is there more structure than day-night rhythm? Figure 7 shows the functions β, γ, and δ for the summer season (July-August). In summer, the air warms up fast and cools down slowly which explains why β is negative for small d. Humidity increases slowly and decreases fast, so β is positive

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

for small d. In view of equation (11) below, the δ functions also show opposite behavior. This was found for all seasons. The structure of γ, which is similar for temperature and humidity, was found only in spring and summer (April to August), and we have no interpretation for it. Of course, the functions will not improve weather prediction, but we think that any mathematical description of the underlying process is of some value. For instance, the functions can be used to characterize or cluster sites.

5. Properties of correlation functions

The classical autocorrelation function ρ (see [16,17]) measures the degree of coincidence of the time series with a copy shifted by d time steps. Thus autocorrelation is large for small d. It will decrease, and the rate of decrease reflects the memory of the underlying process. If we have a periodic dynamics, autocorrelation will be large at the period, and small, mostly negative, at the half period. Moreover, the sign of ρ can indicate an increasing or decreasing trend. The last point is not quite correct, because a trend will exclude stationarity, and then, strictly speaking, we cannot estimate autocorrelation.

The direction of a trend is always indicated by β, and this can be consistent with the weak stationarity assumptions to estimate order pattern frequencies, for instance in the case of stationary increments with non-zero mean. The main purpose of β, as well as of γ and δ, is the description of certain asymmetries in the process. The up-down balance β is positive for small d if the process spends more time with increase than with decrease. When the process is stationary, this means that values will increase more slowly, and decrease faster. The functions γ and δ are more difficult to interpret, but we can at least say that δ is tightly connected to β since

$$\delta(d) = \beta(2d) - \beta(d)\ . \tag{11}$$

This follows from $\beta(2d) = p_{123}(d) + p_{132}(d) + p_{213}(d) - p_{321}(d) - p_{312}(d) - p_{231}(d)$ and from the definition of β. As in Section 3, the equation holds for pattern probabilities of a process, and may have a marginal error for frequencies of a concrete time series. This equation justifies the name up-down scaling for δ.

Persistence is the most common and most important of our functions. It says how often a relation $x_t < x_{t+d}$ or $x_t > x_{t+d}$ will persist in the next comparison of x_{t+d} with x_{t+2d}. Up to the sign, it can replace autocorrelation, and can be interpreted similarly. For small d, persistence measures the degree of smoothness of the time series. When we have a smooth function, then $\tau(d)$ tends to the maximal value $\frac{2}{3}$ for $d \to 0$. We can also consider the complementary quantity

$$\text{turning rate} \qquad TR = p_{132}(d) + p_{213}(d) + p_{231}(d) + p_{312}(d) = \frac{2}{3} - \tau(d)\ , \tag{12}$$

which counts the frequencies of turning points (local maxima or minima) in the series. Obviously, TR is a measure of roughness, or variation [18].

In this paper, however, the functions are chosen so that they are all zero for white noise - and for the slightly more general case of an exchangeable process where all patterns of Figure 2 have the same probability $\frac{1}{6}$. In other words, white noise is the origin point of our coordinate system. This viewpoint is especially useful when we deal with noisy signals, for instance EEG brain data. All our functions are differences of pattern frequencies. For τ we have

$$\tau(d) = \frac{1}{3}\left(2p_{123}(d) + 2p_{321}(d) - p_{132}(d) - p_{213}(d) - p_{231}(d) - p_{312}(d)\right)\ . \tag{13}$$

Both (12) and (13) can be easily checked with (9). In the following, we systematically summarize the properties of autocorrelation functions. For the sake of completeness, the next tables include Spearman's rank correlation, but not Kendall's τ [19]. We think that for big time series, persistence

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

better reflects Kendall's idea to measure correlation by order comparison. This is why we used the letter τ.

Table 1. Range and extreme cases for various correlation functions

Function	Range	min assumed for	max assumed for
Autocorrelation ρ	$[-1,1]$	linear decreasing series	linear increasing series
Spearman rank autocorr.	$[-1,1]$	decreasing series	increasing series
up-down balance β	$[-1,1]$	decreasing series	increasing series
persistence τ	$[-\frac{1}{3},\frac{2}{3}]$	alternating series	monotone series
turning rate TR	$[0,1]$	monotone series	alternating series
up-down scaling δ	$[-1,1]$	$-(t+(-1)^t)$	$t+(-1)^t$
rotational asymmetry γ	$[-1,1]$	$(-2)^t$	$(-\frac{1}{2})^t$

Table 1 shows the range of values for all correlation functions, adding for which series minimal and maximal values are attained. An alternating series changes up and down at each step, like $x_t = (-1)^t + 0.1 \cdot$ (uniform white noise) where the noise is added to avoid equal values. This works for $d = 1$, not for continuous d. For γ and δ we just gave an example series which may give a feeling for what these functions measure.

Another viewpoint is symmetry. A function $f(d)$ is even if $f(-d) = f(d)$ and odd if $f(-d) = -f(d)$. For autocorrelation functions, even functions are those which assume the same values for a time series $x_1, x_2, ..., x_T$ and its time-reversed series $x_T, x_{T-1}, ..., x_1$. For odd functions, time reversal will change the sign of the correlation function. We can also ask for invariance under change of sign of the series, which gives $-x_1, -x_2, ..., -x_T$, or for the combined change of sign and time reversal, $-x_T, -x_{T-1}, ..., -x_1$. The latter can be interpreted as a 180^o rotation of the graph of the series.

Table 2. Invariance of correlation functions under symmetries

Function	Time reversal	Negative function	Rotation
ρ, Spearman, τ	$+$	$-$	$-$
β and δ	$-$	$-$	$+$
rotational asymmetry γ	$-$	$+$	$-$

All autocorrelation behave well under these symmetries. They either remain invariant, in which case we write a $+$ in Table 2, or they change sign in which case we note $-$. Persistence is an even function like ρ, but our functions β, γ, and δ are odd. This causes a discontinuity at $d = 0$ unless the value is zero there. Actually, none of our correlation functions is defined for $d = 0$. We shall simplify matters by restricting ourselves strictly positive d. However, the problem will return when we consider periodic series.

A time series $x_1, ..., x_T$ is periodic with period L if $x_{t+L} = x_t$ for $t = 1, ..., T - L$. Theoretically, none of our correlation functions is then defined for $d = L$ because there are only equal values to compare. In reality, however, we have only approximate periodicity and can always determine values for $d = L$. Nevertheless, some phenomena occur at $d = L, 2L, 3L, ...$ and at half periods $d = \frac{L}{2}, \frac{3L}{2}, ...$ For a periodic time series, all autocorrelation functions are periodic with the same period - in practice only approximately since only part of the generating mechanism works periodically. So it suffices to consider $d = L$ and $d = \frac{L}{2}$.

For an even function $f(d)$ with period L we have $f(d) = f(L - d)$ and hence $f(\frac{L}{2} + k) = f(\frac{L}{2} - k)$. That is, an even correlation function is mirror-symmetric with respect to the vertical line $d = L$ as well as $d = \frac{L}{2}$. This is the case for ρ and τ. For an odd function $g(d)$ with period L we have $g(L - d) = -g(d)$ and hence $g(\frac{L}{2} + k) = -g(\frac{L}{2} - k)$. In particular $g(\frac{L}{2}) = 0$. The graph of g then has a symmetry center at $(L, 0)$ and at $(\frac{L}{2}, 0)$. Both symmetry centers will be seen as zero crossings in the graphs of β, γ, and δ

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

even though theoretically we may have a discontinuity at L. For noisy series, the study of symmetries of correlation functions can be quite helpful to recognize periodicities.

Table 3. Behavior of correlation functions of periodic series

Function	Half period $\frac{L}{2}$	Period L	Symmetry type
ρ, Spearman,	minimum	maximum	vertical line
persistence τ	minimum	bumped maximum	vertical line
β, γ, and δ	zero	zero or discontinuity	symmetry center

Let us now explain the behavior of persistence which we have already seen in Figure 6. For $d = \frac{L}{2}$, we always have $x_t \approx x_{t+2d}$ in a series with approximate period L. Thus if x_{t+d} is only a bit larger or smaller, we cannot have pattern 123 or 321. This means that τ has a clear minimum at $\frac{L}{2}$. If the noise level goes to zero, the value will approach the absolute minimum $-\frac{1}{3}$. This is the best way to determine a period with τ.

At $d = L$ the behavior is more complicated, and different from the maximum of ρ. Assume that much of the time series consists of monotone pieces, and x_t is on an increasing branch. If we have the exact period L, and take $d = L - \varepsilon$ for a small $\varepsilon > 0$ then $x_t > x_{t+d} > x_{t+2d}$ since the shifted values are further downwards on the repeating increasing branch. However, for $d = L + \varepsilon$ with small positive ε we have $x_t < x_{t+d} < x_{t+2d}$ since now the shifted values come further upwards on the repeating branch. Only for $d = L$ we theoretically have $x_t = x_{t+d} = x_{t+2d}$ and can say nothing. In practice there will be some noise which disturbs the monotone branches and changes our conclusion for small ε. Anyway, left and right of $d = L$ the patterns 123 and 321 dominate and we have two maxima of persistence there. But for $d = L$ itself, the patterns 123 and 321 do not dominate anymore, and we get a local minimum of τ. This is what we call a bumped maximum. The height and width of the bump decreases when the noise level goes to zero. See Figure 6.

One can ask why among our four correlation functions, only τ is even and the other three are odd. Actually we have six order patterns, so we could have three even and three odd functions. However, the patterns fulfil the sum condition (9) and another condition which directly follows from (10):

$$p_{132} + p_{231} - p_{213} - p_{312} = 0\,. \tag{14}$$

Both of these conditions are expressed by even functions of d, since the frequencies of a permutation and its reversal, like 213 and 312, have the same sign in the equation. Thus we are left with four degrees of freedom for our pattern frequencies, and it is natural to obtain three odd correlation functions.

The following theorem shows that τ, β, γ, and δ are an optimal choice of correlation functions which explain in some way all the difference between our observed time series and white noise, a series of independent random numbers. The information given by these functions is largely independent, with the exception of dependence between β and δ which cannot be avoided. This will be made precise in a subsequent mathematical paper. Here we argue that certain vectors corresponding to the functions are orthogonal, which can be the basis for ANOVA (analysis of variance) techniques.

Theorem 1 (Pythagoras theorem for order patterns of length 3)**.** *For a process with stationary increments and an arbitrary lag d, the quadratic distance Δ^2 of pattern probabilities to white noise uniform pattern frequencies $\frac{1}{6}$ defined in (2) has the following representation:*

$$4\Delta^2 = 3\tau^2 + 2\beta^2 + \gamma^2 + \delta^2\,. \tag{15}$$

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

Proof. For arbitrary numbers $q_1, q_2, ..., q_6$ we have the following identity.

$$4\sum_{k=1}^{6} q_k^2 = \quad 2(q_1+q_6)^2 + 2(q_1-q_6)^2 \quad +(q_2+q_3+q_4+q_5)^2 + (q_2-q_3-q_4+q_5)^2$$
$$+(q_2+q_3-q_4-q_5)^2 + (q_2-q_3+q_4-q_5)^2$$

If $\sum q_k = 0$, the third square on the right is the same as the first. Now let $q_1 = p_{123} - \frac{1}{6}, q_2 = p_{132} - \frac{1}{6}, q_3 = p_{213} - \frac{1}{6}, q_4 = p_{231} - \frac{1}{6}, q_5 = p_{312} - \frac{1}{6}$, and $q_6 = p_{321} - \frac{1}{6}$. Then $\sum q_k = 0$ by (9), and the last term on the right is zero by (14). According to the definitions of $\Delta^2, \tau, \beta, \gamma$, and δ, the identity turns into the equation (15) of the theorem. This completes the theoretical part of the paper. □

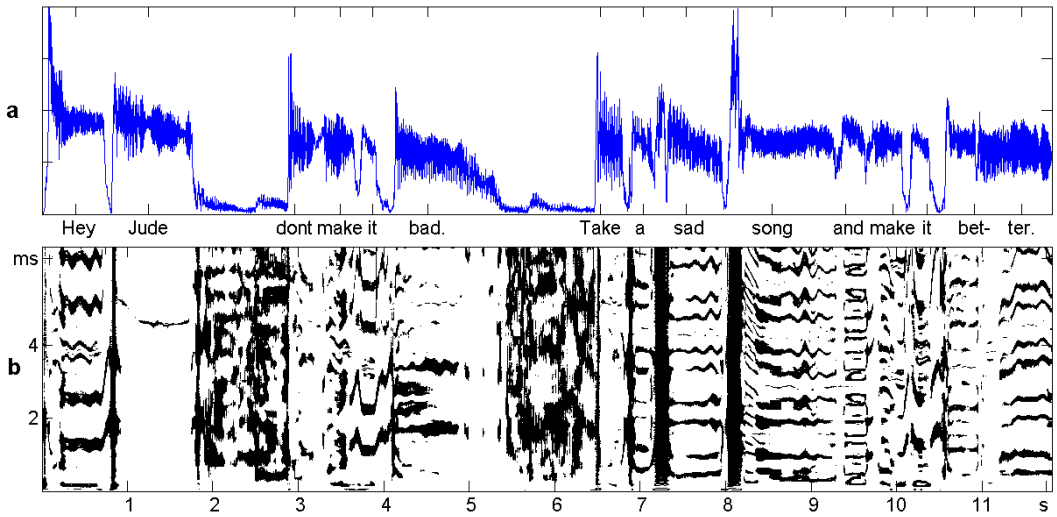


Figure 8. 12 seconds of the song "Hey Jude" of The Beatles. **a.** The signal - mean of absolute amplitude over non-overlapping windows of 50 ms. **b.** The noisy places (x, d) for which $T\Delta^2 < 15$, drawn in black. The vertical axis represents the lag $d = 1, ..., 30$, considered as wavelength which ranges from 0 to 7 ms. Each column of the matrix corresponds to one window x.

6. Case study: speech and music

One potential field of appplication of our correlation functions is speech and music: speech recognition, speaker identification, emotional content of speech sounds etc. This field is dominated by spectral techniques and by machine learning, and additional information on speech processes is certainly welcome. We are not going into detail since we know this is hard work! Here we just analyze the first 12 seconds of the song "Hey Jude" of The Beatles, to indicate what is possible.

The intensity of the signal is shown in Figure 8. Since music is sampled with 44 kHz, there are half a million amplitude values. For a rough analysis, we divide the large time series into 240 non-overlapping pieces, called windows, of length 50 ms. Thus there are 240 time series of length $T = 2200$ for which we can determine mean absolute amplitude and correlation functions. The delay d will run from $1 = 1/44$ ms to $d = 300$ which is 7 ms. The functions will not be drawn as curves, but as color-coded vertical sections of an image which describes the whole data set. The windows, or columns of the matrix, are numbered $x = 1, ..., 240$, the rows $d = 1, ..., 300$, written as 0..7 ms.

As a first experiment, we choose the points (x, d) for which $T\Delta^2 < 15$. They are colored black in Figure 8b. At these places we could perhaps still confuse the signal with white noise - according to simulations in [8], the p-value is larger than 10^{-15}. These places occupy 26% of the matrix, notably the 'k' of 'take' and's' of 'sad' and 'song' with almost any d. For all the other (x, d) Theorem 1 says that some of our correlation functions are significant.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

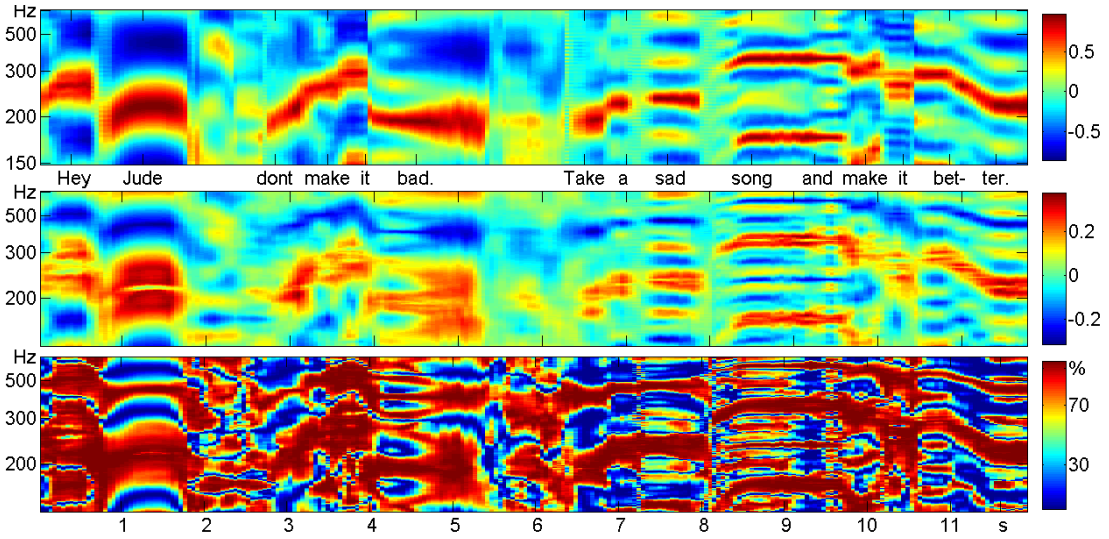


Figure 9. Correlogram (upper panel) and persistence (middle panel) of 12 s of "Hey Jude". The scale of d was reverted and written as frequencies so that the melody can be read like musical notes. The bottom panel shows the percentage $\tilde{\tau}$ of Δ^2 which is due to persistence.

Sliding window analysis of speech and music and visualization in a matrix is well-known from the spectrogram where columns correspond to the Fourier spectrum of the windows. Since here, we are satisfied with the melody - the so-called pitch frequency of the singer - we take the correlogram instead, and compare with persistence in Figure 9. For convenience, the d-scale is transformed into frequencies - 500 Hz means $d = 2$ ms and 200 Hz means $d = 5$ ms. The maxima of ρ and the bumped maxima of τ are very clear for all voiced sounds. They describe the melody, and they do coincide.

The bottom panel of Figure 9 shows $\tilde{\tau} = \frac{3\tau^2}{4\Delta^2}$, the percentage of δ^2 which is due to persistence. Clearly, persistence is the dominating function, with 80% of Δ^2 at most places. But there also many places with 30% and less where the functions β, γ, δ form the larger part of Δ^2. Figure 10 shows some phonemes which may be characterized this way.

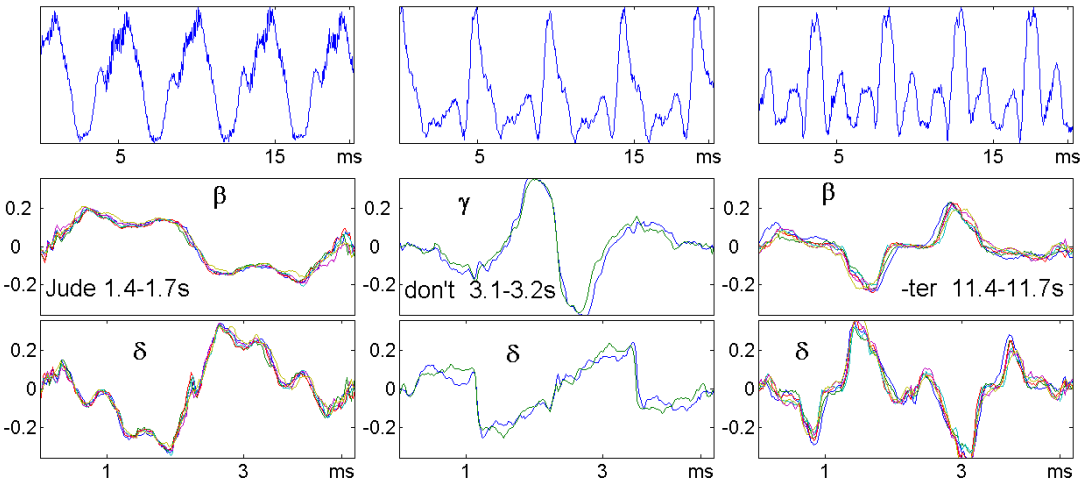


Figure 10. Detail from Figure 9. The vowels of 'Jude', 'don't', and the second syllable in 'better' provide stationary parts of the time series lasting for 0.3, 0.1, and 0.3 s. Only 20 ms of each signal are shown in the top panel. Order correlation functions were calculated for six resp. two disjoint windows of length 50ms and drawn for one pitch period, which equals 4.5 ms for all three sounds.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

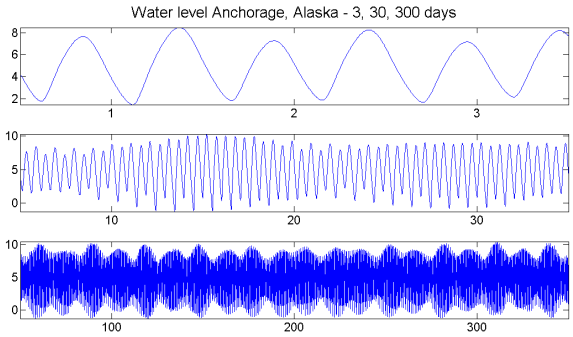


Figure 11. Tides form an almost deterministic process with periodicities on the scale of days, months and years. Data from [20].

7. Case study: tides

Tides of the oceans are a well-studied phenomenon, and tidal physics is a science which has developed over centuries. They form a good testbed for our correlation functions for two reasons. On the one hand, there are excellent data series of water levels at many stations, provided for the USA by the National Water Level Obersvation Network [20] over many years with 6 minute intervals and few missing values. On the other hand, tides can be considered as an almost deterministic process driven by the interaction of moon, sun, and earth, with geographical site and coastal topography of the station as parameters. As Figure 11 shows, there are daily, monthly and yearly scales of periodicities: tides at our selected places come two times a day, spring tides two times a month etc. Classical time series will consider the data as a sum of sine waves, and there is an important theorem which says that the 'spectral measure' for the frequencies of such a process is obtained as Fourier transform of the autocorrelation function. From this point of view, our asymmetry functions β, γ, and δ are negligible and inappropriate. We shall show, however, that they do contain structural information on the process.

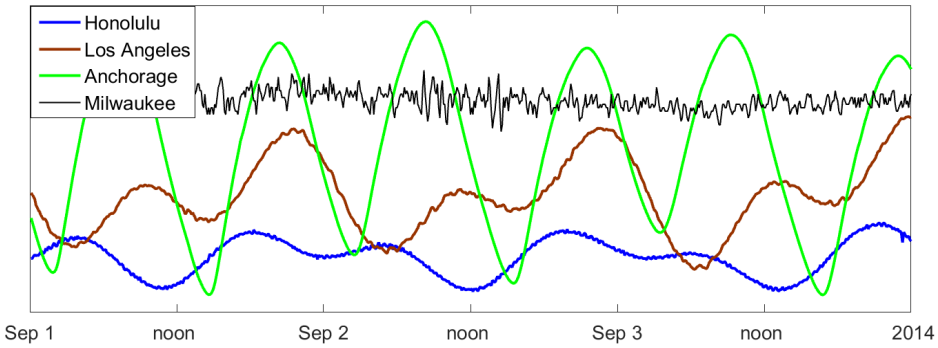


Figure 12. Water levels for three days in 2014, measured every six minutes at different stations in the USA. Data taken from [20], shifted and scaled for better visibility.

Since Anchorage, Alaska is situated at the end of a narrow bay, the tidal range (difference between high tide and low tide) is 8 metres. In Los Angeles, California it is less than 2 m, and in Honolulu, Hawaii, it is only half a metre. Order patterns cannot distinguish size of waves, we are only interested in their structure! Moreover, it makes no sense to compare zero levels of stations, so they were shifted for better visibility. For sake of completeness, Figure 12 includes Milwaukee at Lake Michigan where there are no tides and differences are only 10 cm, caused by changes of wind and barometric pressure. This series shows what fluctuations arise when data represent random weather change.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

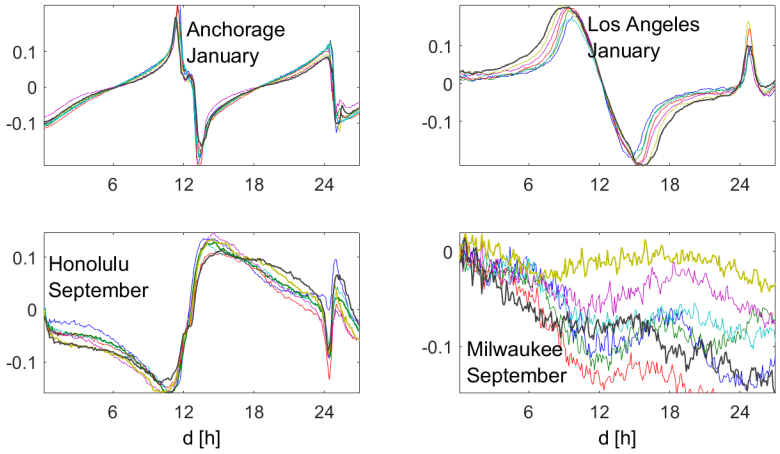


Figure 13. Correlation functions β, for one month in seven consecutive years, at the four places of Figure 12. For a given month, each ocean station has its specific β-profile.

The functions β given in Figure 13 show that each of the ocran stations has its own profile which did not change much in seven consecutive years. It did depend a bit on the season, however, so it was taken for one month in each year. Since we have 240 values per day, we could afford even shorter periods.

Experience shows that water comes fast and goes slowly, so β should be negative for small d. This is clearly true for Milwaukee where no periodicity was found, and for Honolulu. For Anchorage and Los Angeles, it is the other way round. For the three ocean stations, β has a zero-crossing for $d \approx 12.5$, and a 'discontinuity' at $d = 25$. As explained in Section 5, this corresponds to the basic daily periodicity of 24.8 hours (the difference to 24 is caused by the moon).

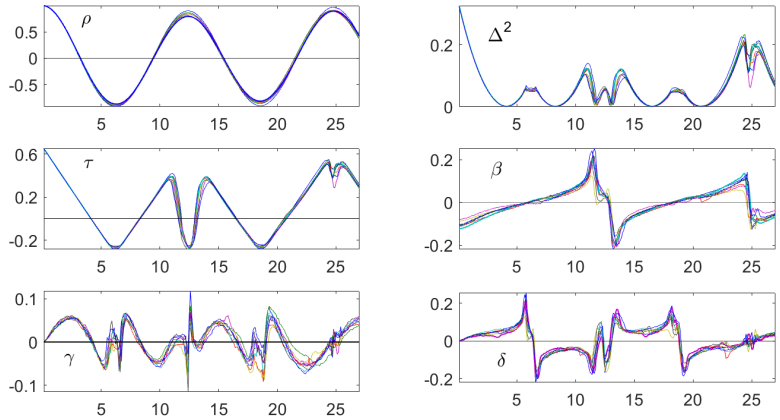


Figure 14. Classical autocorrelation and order correlation functions for station Anchorage in January. Persistence reflects the diurnal rhythm. The asymmetry functions β, γ, δ all show a very specific structure which remains stable through seven consecutive years. In contrast, ρ does not seem to contain much structural information.

Figure 14 shows all order correlations for Anchorage for the month January. Both the autocorrelation and persistence express the fact that there are two periodicities with periods of 12.5 and 25 hours. For ρ, the maximum at 25 is larger than the first maximum at 12.5 while the minimum at 19 is not smaller than the first minimum at 6. For persistence, the maximum at 25 has a small bump while the maximum at 12.5 has a very large bump since it is a also a minimum corresponding to 25.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

The functions β, γ, δ all show a sharp structure which is preserved thorough consecutive years. This is no proof that they contain more information than ρ but it is an argument for further study of these functions.

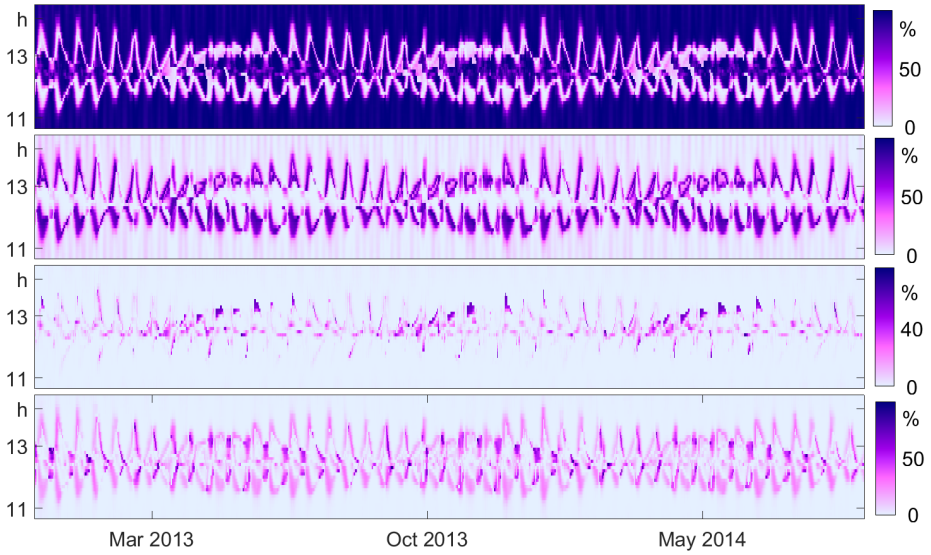


Figure 15. The division of Δ^2 into components described in Theorem 1, illustrated for the tides at Anchorage in the years 2013 and 2014 in sliding window analysis. From top to bottom, the four panels correspond to the function $\tilde{\tau}$ which takes the largest part of Δ^2, to $\tilde{\beta}, \tilde{\gamma}$ and $\tilde{\delta}$.

Finally we check Theorem 1 with the tides data of Anchorage in Figure 15. Windows of length 1242 (5 days and 4.2 hours - just 5 main periods) were used for the years 2013 and 2014. For good resolution, the step size is half a day, so successive windows overlap by 9/10 of their length. The lag d was restricted to the time between 10.5 and 14.5 hours, around the first the important period 12.5 hours where Δ^2 is large. Roughly speaking, the picture shows how the position and width of the main bump of τ in Figure 14 changes with time. There are small waves which describe the bimonthly period, and long waves describing biyearly periodicity.

It can be seen how the different components fit together, with $\tilde{\tau}$ as dominant component. The average value of $\tilde{\tau}$ was 76% , followed by 15% for $\tilde{\beta}$, 6% for $\tilde{\delta}$, and 2% for $\tilde{\gamma}$. Since Theorem 1 holds for probabilities of the process and here we have frequencies from data, there is a marginal error, as explained in Section 3. Since d is between 105 and 145, this error could be rather large. Here the average error is 0.5% . There are very few places (x, d) where the error is more than 1% , the maximum was 2% .

This example shows that the order correlation functions $\tilde{\tau}, \tilde{\beta}, \tilde{\gamma}$ and $\tilde{\delta}$ can also detect structure in data, even though they are squares and lack the information of the sign of the original functions. In the present case, as well as in Figure 1c these relative quantities are more informative than the original ones.

8. Case study: particulates

In the previous section, we had an almost deterministic process with excellent data. Now we consider the opposite situation: measurements of particulates PM10 (aerodynamic diameter smaller than $10\,\mu$m). Such data, measured by a kind of vacuum cleaner with a light sensor for the dust, are notoriously noisy. Dust will not come uniformly, it will rather form clusters of various size. Although measurements can be taken several times in a second, they are averaged automatically at least over several minutes. In Figure 16 we consider hourly values from the public database [21] for the station

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

3215 at Trona-Athol in San Bernardino, California. They were transformed to logarithmic scale to reduce the influence of large values. Such a transform does not change order patterns.

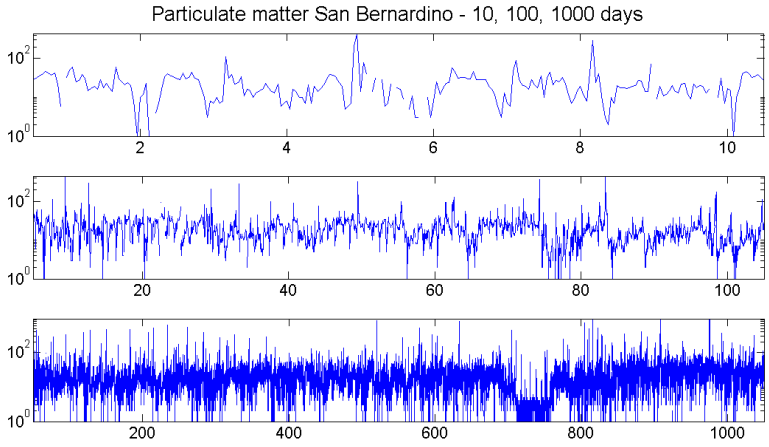


Figure 16. Particulate measurements are notoriously noisy. They show a weak daily and yearly rhythm which can hardly be detected from the data. The PM10 measurements for station 3215 Trona -Athol at San Bernardino, California are from the public database [21].

There are 13% of missing values, some of which can be detected in the upper panel of Figure 16. They affect the estimation of autocorrelation much more than estimation of order correlation functions. This can be seen in Figure 17 where all functions where determined for 12 successive years. The variation is of course a bit larger than in the corresponding Figure 14 for the tides. The values of the order correlations are about three times smaller than those for the tides. The important point, however, is the consistent structure of the order correlation functions. It is much better than classical autocorrelation or Δ^2 in the upper row. Persistence shows a loss of intensity from one up to six days, though not as strong as autocorrelation. However, the structure of β, δ, and even γ remains almost unchanged from day 1 to day 6.

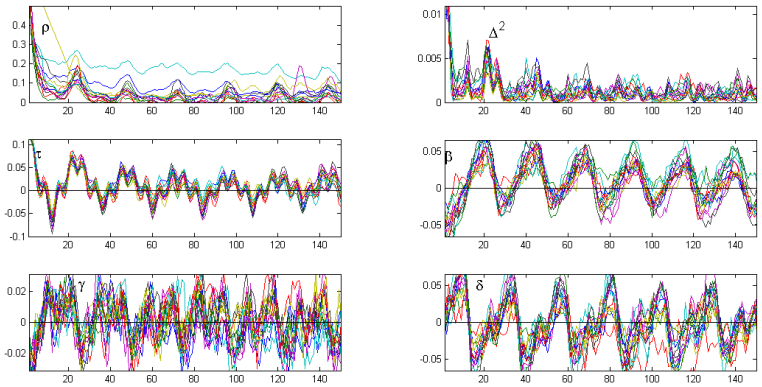


Figure 17. Correlation functions for hourly particulate measurements at San Bernardino [21] in 2000-2011. Each of the 12 curves corresponds to one year, in order to check consistency of correlation structure over the years. The functions β, δ, and γ show similar structure from day 1 to day 6.

To perform a sliding windows analysis with these data, we take windows of length 1200, that is 50 days. Smaller windows do not provide sufficiently accurate estimates. Probably the situation would be better if the data were measured every 6 minutes like tides above, even though there were

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

larger variations. Figure 18 shows that the daily rhythm appears mainly in summer, not in winter, similar to temperature in Section 4. Thus the curves of Figure **??** are obtained mainly from the summer measurements. The strength of daily rhythm and the 'length of summer season' varies over the years.

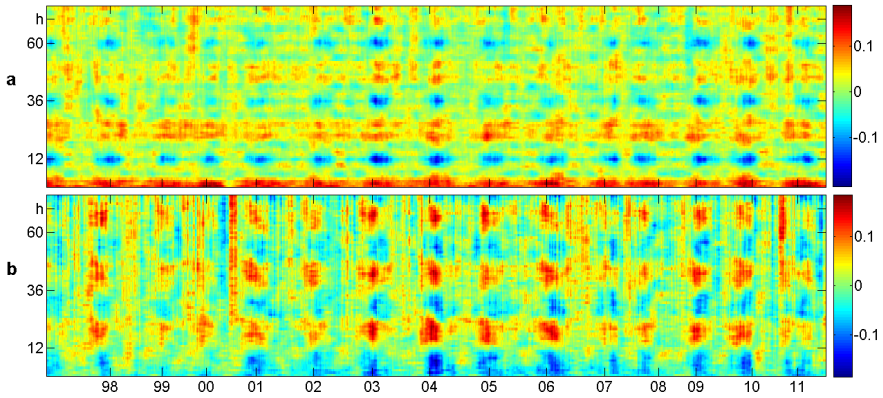


Figure 18. Sliding window analysis of **a** persistence $\tau(d)$ and **b** up-down balance $\beta(d)$ for hourly particulate measurements at San Bernardino 1997-2011 [21]. The lag d runs from 1 to 72, that is 3 days, on the vertical axis. Overlapping windows of 50 days were used. Daily rhythm is present mainly in summer, in both τ and β.

9. Case study: brain and heart signals

The examples in Sections 4,7, and 8 were presented as proof of concept, with the idea to encourage readers to work out the applications in a more careful way. For biomedical data, however, the author has done a careful study [8,14,18] of data published by Terzano et al [22] at physionet [23] which seems very promising. Three types of data were considered, as indicated in Figure 19: the noisy EEG brain data (electroencephalogram), the well-known heart ECG (electrodardiogram), and the rather smooth plethysmogram which measures the blood flow at the fingertip.

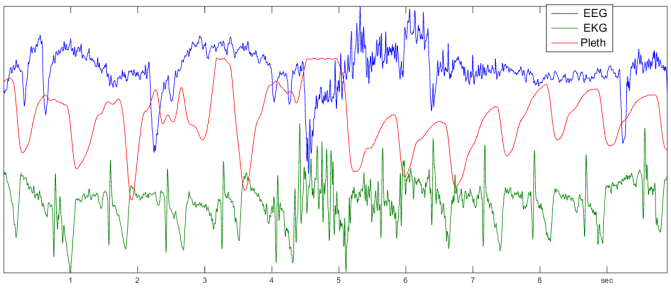


Figure 19. Biomedical signals: 8 seconds of a electroencephalogram, an electrocardiogram and a plethysmogram. Order correlation functions seem to apply to all of them.

In [8] it was found that the function Δ^2 can distinguish sleep stages from the EEG data without any further processing. In [18] it was noted that it is actually the persistence, or turning rate, the main component of Δ^2, which can be taken as measure of sleep depth. For all healthy control persons and a number of patients with different sleep problems, the results are as impressive as Figure 20 where manual annotation by a medical expert and turning rate are almost parallel. The definition of sleep depth is a complicated and controversial issue among physicians, and a pure mathematical definition, even if it is not perfect, could be helpful.

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

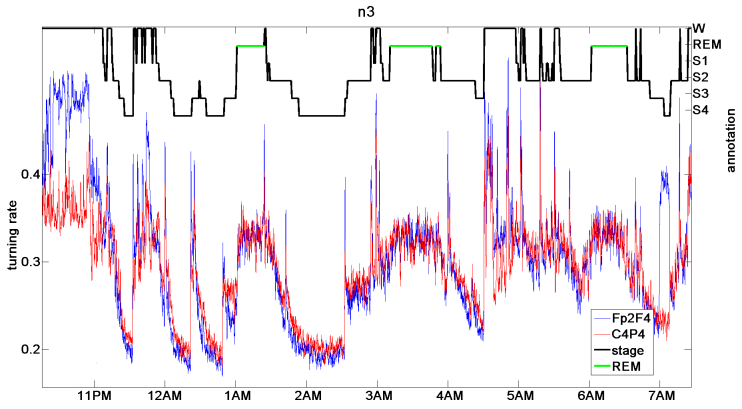


Figure 20. Coincidence of sleep stage annotation by a medical expert and turning rate for $d = 8$ ms of two EEG channels [18]. Original data from Terzano et al. [22].

So far, the author could not reproduce this result with more recent data. Actually, there are huge databases of sleep data in different countries. But these data are not of the same quality as those of [22] for two reasons. On the one hand, since there are so many data, they are sampled not at 500 Hz but only at 200 Hz or less. On the other hand, EEG data are contaminated by the field of the electrical power net with 50-60 Hz frequency. It is quite an effort to avoid this influence, especially in a hospital environment. Since conventional evaluation of EEG data considers wavelength of at least 50 ms, experts are content with this type of data. The standard preprocessing is a low-pass filtering with a cutoff of 40 Hz, say. For determining the persistence or turning rate at $d \leq 8$ ms, or frequency ≥ 130 Hz, however, the 50 Hz noise of the power net is a real obstacle. This also holds for the ECG.

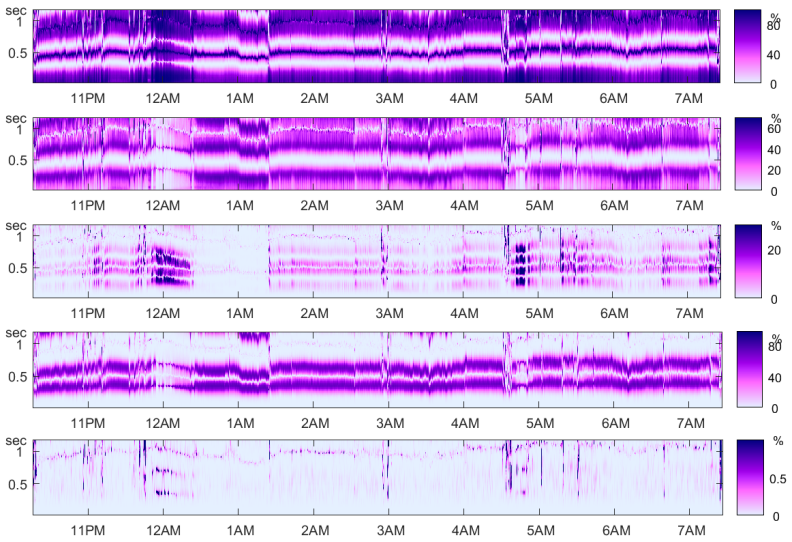


Figure 21. Theorem 1 for the plethysmogram over a whole night. From top to bottom: $\tilde{\tau}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}$. The bottom panel contains the error of the equation of Theorem 1 on a scale from 0 up to 1% . Data of person n3 from Terzano et al. [22].

The plethysmogram is determined by a light sensor, and can be measured without power noise contamination. Nevertheless, it is mostly measured at low frequencies like 30 Hz, because of its smooth appearance. Perhaps there is also a fine structure for higher resolution which can be studied by order

Peer-reviewed version available at Entropy 2019, 21, 613; doi:10.3390/e21060613

methods. We conclude this paper with a Figure 21 which shows the validity of Theorem 1 for a plethysmogram of [22].

Acknowledgments: The author thanks Bernd Pompe for discussions over many years, and for his opinion on the unpublished paper [14] on which part of this work is based. Costs to publish in open access were covered by Deutsche Forschungsgemeinschaft, project Ba1332/11-1.

Conflicts of Interest: The author declares no conflict of interest.

Bibliography

1. Amigo, J.; Keller, K.; Kurths, J. Recent progress in symbolic dynamics and permutation complexity. Ten years of permutation entropy. *Eur. Phys. J. Special Topics* **2013**, *222*, 247–257.
2. Amigo, J.M. *Permutation complexity in dynamical systems*; Springer Seires in Synergetics, Springer: Berlin, 2010.
3. Zanin, M.; Zunino, L.; Rosso, O.; Papo, D. Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy* **2012**, *14*, 1553–1577.
4. Parlitz, U.; Berg, S.; Luther, S.; Schirdewan, A.; Kurths, J.; Wessel, N. Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Computers in biology and medicine* **2012**, *42*, 319–327.
5. McCullough, M.; Small, M.; Iu, H.; Stemler, T. Multiscale ordinal network analysis of human cardiac dynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2017**, *375*, 20160292.
6. Bandt, C.; Shiha, F. Order patterns in time series. *J. Time Series Analysis* **2007**, *28*, 646–665.
7. Bandt, C. Permutation Entropy and Order Patterns in Long Time Series. In *Time Series Analysis and Forecasting*; Rojas, I.; Pomares, H., Eds.; Contributions to Statistics, Springer, 2015.
8. Bandt, C. A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure. *Entropy* **2017**, *19*, 197.
9. Bóna, M. *Combinatorics of permutations*; Chapman & Hall/ CRC, 2004.
10. Elizalde, S. A survey of consecutive patterns in permutations. arXiv150411.07265.
11. Elizalde, S.; Noy, M. Consecutive patterns in permutations. *Adv. Appl. Math.* **2003**, *30*, 110–123.
12. Glebov, R.; C., C.H.; Klimosova, T.; Kohayakawa, Y.; Král, D.; Liu, H. Densities in large permutations and parameter testing. arXiv1412.5622v3.
13. Bandt, C.; Pompe, B. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **2001**, *88*, 174102.
14. Bandt, C. Autocorrelation type functions for big and dirty data series. arXiv1411.3904.
15. DeutscherWetterdienst. Climate Data Center. ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany.
16. Brockwell, P.; Davies, R. *Time Series, Theory and Methods*, 2 ed.; Springer: New York, 1991.
17. Shumway, R.; Stoffer, D. *Time Series Analysis and Its Applications*, 2 ed.; Springer: New York, 2006.
18. Bandt, C. Crude EEG parameter provides sleep medicine with well-defined continuous hypnograms. arXiv1710.00559.
19. Ferguson, S.; Genest, C.; Hallin, M. Kendall's tau for serial dependence. *Canadian J. Stat.* **2000**, *28*, 587–604.
20. NationalOceanicandAtmosphericAdministration. National Water Level Obersvation Network. https://www.tidesandcurrents.noaa.gov/nwlon.html.
21. California. Air Resources Board. www.arb.ca.gov/adam.
22. Terzano, M.; Parrino, L.; Sherieri, A.; Chervin, R.; Chokroverty, S.; Guilleminault, C.; Hirshkowitz, M.; Mahowald, M.; Moldofsky, H.; Rosa, A.; Thomas, R.; Walters, A. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. *Sleep Med.* **2001**, *2*, 537–553.
23. Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P.; Mark, R.; Mietus, J.; Moody, G.; Peng, C.K.; Stanley, H. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2000**, *101*, e215–e220. database at www.physionet.org.