A Physicist's Approach to Multidisciplinary Time Series Analysis
Finalist Article for Finnish Physical Society Young Physicist Prize 2019

Matti Molkkari
Computational physics laboratory, Tampere University[*]
(Dated: February 12, 2020)

What do heat conduction, quantized energy levels, stock prices, DNA sequences, and the beating of the heart have in common? Their behavior may be modeled by diffusion-like relations. While some of these processes are already discrete in their nature, diffusion is conventionally considered to be continuous in time. However, in practice any measurements must be carried out at discrete time intervals of finite length, yielding time series of measurements. This lends itself to the atomistic viewpoint of diffusion where "particles", whatever they may be, can be understood to undergo the steps of a random walk within the time intervals. The details of any interactions are hidden inside the probability distributions of the steps. Uncorrelated normally distributed steps lead to ordinary Brownian motion, whose expected end-to-end distances, or root-mean-squared displacements, Δx scale proportionally to the square root of the number of steps s taken. Correlations, on the other hand, introduce memory effects with positive and negative correlations leading to super- and subdiffusion, respectively. The expected displacements of this anomalous diffusion are characterized by power law relations $\langle \Delta x \rangle \propto s^{\alpha}$ with scaling exponents α greater or less than $1/2$.

The straightforward determination of these exponents from the displacements has several shortcomings. For example, measuring the literal displacements is not very robust, as these are plagued by huge variations and usually by measurement noise as well. Many phenomena are also influenced by other non-diffusive effects, such as convection, which are composed into the random walk as external trends. These trends are problematic, as they lead to spurious detection of correlations. *Detrended fluctuation analysis* (DFA) attempts to overcome these difficulties in a systematic manner [1, 2].

The DFA algorithm takes the steps of a random walk as input, from which the walk is constructed as their cumulative sum. To estimate the displacement $\Delta x(s)$ after s steps, the walk is divided into windows of length s. Within each window local trends are determined by least-squares fits of low-degree polynomials, which are then subtracted to obtain the detrended walk. Their variances are averaged over all the windows, yielding the displacement as its square root. In the context of DFA, the estimated expectation value of this root-mean-squared detrended displacement is called the fluctuation function $F(s)$. Power laws $F(s) \propto s^{\alpha}$ are transformed into linear relationships on a logarithmic scale, where the scaling exponent α is conveniently determined by linear regression.

An important consequence of the detrending step is that it increases the maximal detectable exponent from 1 to $n+1$, where n is the degree of the detrending polynomial. This is necessary for the analysis of complex signals that could be comprised of both stationary ($\alpha < 1$) and non-stationary ($\alpha > 1$) regions. While the detrending also acts akin to a regularizer, increasing the robustness of the method [3], it also introduces bias to the estimates, particularly at the shortest scales.

In practice, however, most phenomena deviate from exact power law scaling, or the behavior may change when observed at different scales. This is also evident in some models, such as Lévy flights or autoregressive processes, whose long scale asymptotic behavior differs from shorter scale details. Therefore, it is instructive to consider systematic methods for determining scale-dependent scaling exponents, which is a major topic considered in my thesis [4]. This could be accomplished by attempting to partition the fluctuation function into approximately scaling regions yielding piecewise defined exponents. Another approach, that is focused on here, considers a full spectrum of exponents $\alpha(s)$ as a function of the scale s. This method is based on the notion that the exponent can be defined as the local slope of the logarithmic fluctuation function: $\alpha(s) = \mathrm{d}\log F(s)/\mathrm{d}\log s$. Direct differentiation is not usually feasible, as the fluctuation function can be very noisy.

I have attempted to tackle this problem in a parameter-free manner by proposing an estimator based on Kalman filter and smoother [4, 5]. The Kalman filter [6] provides an efficient, recursive, solution to a linear state-space estimation problem where both the state and measurements are disturbed by Gaussian noise. In this particular approach the state consists of the logarithmic fluctuation function and its the derivatives at a particular scale. The values of the fluctuation function at each scale are considered the measurements, whose error estimates are based on the standard error of the mean. The model assumes that the function remains constant, except the highest order derivative is disturbed by noise whose strength is derived from the data. The Kalman smoother [7] improves the estimate by utilizing all the measurements for the state estimates at each scale.

Instead of focusing on asymptotics or scale-averaged results, the full scaling spectra may yield new insights about the underlying mechanics of the studied processes. It is also possible to compute the exact theoretical scale-dependent response of DFA to different models, if the

[*] matti.molkkari@tuni.fi

processes or their increments are stationary [4, 8, 9]. An example is shown in Fig. 1(a), where the theoretical scaling exponent is shown for an AR1 process as a function of the scale s and the autoregressive parameter ϕ_1. Together with the robustness of DFA, this could be fruitful for, e.g., parameter estimation or choosing between different models for anomalous diffusion. Furthermore, the short-scale bias from detrending could be accounted for.

The scaling may also exhibit temporal variations due to changes in external conditions, or the process itself may be comprised of distinct intrinsic modes. Straightforward segmentation of the time series suffers from the following limitation: The length of the segments, and hence the temporal resolution, is dictated by the largest scale to guarantee sufficient statistical accuracy. This led to the development of dynamic DFA that adopts a scale-dependent segmenting procedure to keep the accuracy approximately constant and letting the temporal resolution vary [10, 11]. The method yields point estimates $\alpha(t, s)$ for the scaling exponent as a function of both time and scale.

To maximize the temporal resolution, it is necessary to accurately estimate the scaling exponent with minimal data. Conventionally the DFA fluctuation function is estimated by computing the variances of the detrended walk in non-overlapping windows at each scale. However, performing the calculations in maximally overlapping windows improves the statistical accuracy at increased computational cost [12]. Another advantage is that fluctuation functions computed from these overlapping windows are generally smooth enough to permit direct numerical differentiation [11], but to keep the computations feasible, particularly for longer data and larger scales, specialized algorithms must be utilized [13].

The described methodology is applicable to any series of measurements in wide array of disciplines, but so far my research on this topic has mostly focused on heart rate variability (HRV). The beating of the heart constitutes a complex system with consecutive heart beats, or RR-intervals (RRIs), showing fractal-like patterns. In a healthy individual the RRIs exhibit long-range correlations, which are altered by disease, sleep and exercise [14–16]. The addition of scale-dependence into the analysis enhances the detection and classification of different heart diseases [4, 5]. This scale-dependent behavior is illustrated in Fig. 1(b), where it is clear that particularly in congestive heart failure the behavior would not be adequately described the conventional (in HRV analysis) short- (4–16 RRIs) and long-range (16–64 RRIs) scaling exponents. Ongoing research studies whether these results can be further improved by considering also temporal variations in the correlations, and in particular if there exist characteristic transient changes due to disease. This could especially allow distinguishing healthy individuals from those suffering from ST episodes.

The high temporal resolution attained by the dynamic DFA also permits the accurate determination of the distributions of the scaling spectra under various circumstances. This was found to be useful in classifying different sleep phases based on the RRI correlations, as also the variance in the scaling spectra proved to be a descriptive feature [10]. An excerpt of a correlation landscape during sleep is depicted in Fig. 1(c). There appears to be a clear trend according to the sleep phases, but the intra-phase variance complicates the analysis.

The dynamics are also important in constantly changing fast-paced activities, such as sports. Recently we have demonstrated that the correlations in the RRIs display complex changes as a function of exercise intensity [11]. Figure 1(d) graphs the dynamic correlations as a function of the heart rate, averaged over many running exercises. An example of an individual marathon run (for a different person) is shown in Fig. 1(e). It is again evident that the complex correlations benefit from a more complete description than is provided by the conventional two-range HRV scaling exponent model. These results are anticipated to facilitate the optimization of training programs and provide more accurate real-time feedback about exercise intensity without the knowledge of the maximal heart rate. Furthermore, it could be possible to deduce quantities such as the anaerobic threshold or maximal oxygen uptake from these dynamical RRI correlations.

This article is in part motivated by introducing potentially useful computational tools into the physicist's toolkit that may not be well known. The multidisciplinary nature of my research is also manifested in the collaborative work I am involved within the research group. The methods are utilized for analyzing the quantized energy spectra of classically chaotic systems [17], stock market analysis, and gene informatics to name just a few examples.

---

[1] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, Physical Review E **49**, 1685 (1994).

[2] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin, and A. Bunde, Detecting long-range correlations with detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications **295**, 441 (2001).

[3] M. Höll, K. Kiyono, and H. Kantz, Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average, Physical Review E **99**, 033305 (2019).

[4] M. Molkkari, *Advanced Methods in Detrended Fluctuation Analysis with Applications in Computational Cardiology*, Master's thesis, Tampere University (2019).

[5] M. Molkkari and E. Räsänen, Robust estimation of the scaling exponent in detrended fluctuation analysis of beat rate variability, *Computing in Cardiology*, Computing in Cardiology **45**, 219 (2018).

[6] R. E. Kalman, A new approach to linear filtering and prediction problems, Journal of Basic Engineering **82**, 35 (1960).

[7] H. E. Rauch, F. Tung, and C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, American Institute of Aeronautics and Astronautics Journal **3**, 1445 (1965).

[8] M. Höll and H. Kantz, The relationship between the detrendend fluctuation analysis and the autocorrelation function of a signal, The European Physical Journal B **88**, 327 (2015).

[9] O. Løvsletten, Consistency of detrended fluctuation analysis, Physical Review E **96**, 012141 (2017).

[10] M. Molkkari, M. Tenhunen, A. Tarniceriu, A. Vehkaoja, S.-L. Himanen, and E. Räsänen, Non-linear heart rate variability measures in sleep stage analysis with photoplethysmography, Computing in Cardiology **46**, 287 (2019).

[11] M. Molkkari, G. Angelotti, T. Emig, and E. Räsänen, Dynamical heart beat correlations as a measure of exercise intensity, submitted, preprint available: `https://arxiv.org/abs/1911.04218`.

[12] K. Kiyono and Y. Tsujimoto, Nonlinear filtering properties of detrended fluctuation analysis, Physica A: Statistical Mechanics and its Applications **462**, 807 (2016).

[13] M. Molkkari and E. Räsänen, Efficient computation of detrended fluctuation analysis in maximally overlapping windows, in preparation.

[14] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, Fractal dynamics in physiology: Alterations with disease and aging, Proceedings of the National Academy of Sciences **99**, 2466 (2002).

[15] F. Chouchou and M. Desseilles, Heart rate variability: a tool to explore the sleeping brain?, Frontiers in Neuroscience **8**, 402 (2014).

[16] T. Gronwald and O. Hoos, Correlation properties of heart rate variability during endurance exercise: A systematic review, Annals of Noninvasive Electrocardiology **25**, e12697 (2020).

[17] J. Keski-Rahkonen, M. Molkkari, and E. Räsänen, Detrended fluctuation analysis meets quantum chaos: Exploring long-range correlations in quantum spectra, in preparation.

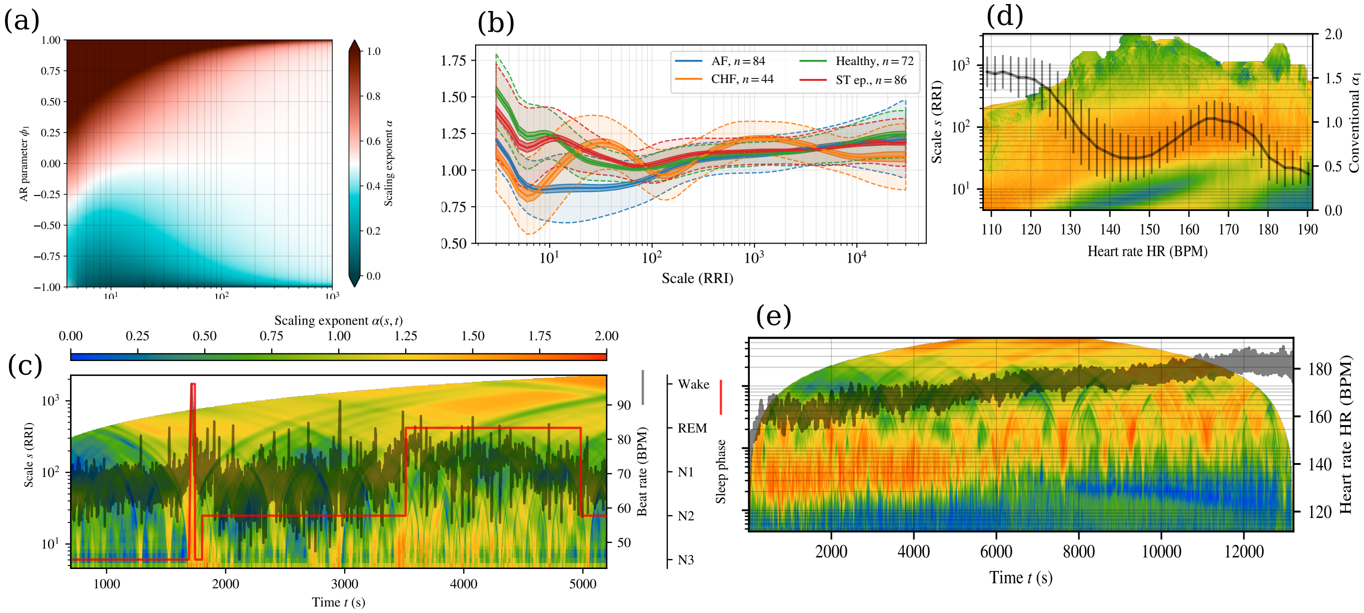

Figure 1. Examples of results. (a) The theoretical DFA scaling exponent in AR1-process as a function of the scale and the autoregressive parameter. (b) The mean scaling spectra in different heart conditions with the standard error of the mean (solid bounds) and standard deviation (dashed bounds). AF = Atrial Fibrillation, CHF = Congestive Heart Failure, ST ep. = episodic ST segment variations. (c) Dynamic correlation landscape of RRIs during sleep with the instantaneous beat rate overlaid on the data. The sleep phases were determined by a sleep physician from polysomnography. (d) Dynamic RRI correlations of an experienced runner as a function of the heart rate averaged over many running exercises. The conventional short-range (4–16 RRIs) scaling exponent α_1 is overlaid on the data, along with its standard error (thick error bars, hardly visible) and standard deviation (thin error bars). (e) Dynamic RRI correlation landscape during a marathon run with the instantaneous heart rate overlaid on the data.