The Structure of Chaos: An Empirical Comparison of Fractal Physiology Complexity Indices using NeuroKit2

Dominique Makowski[1], An Shu Te[1], Tam Pham[1], Zen J. Lau[1], & S.H. Annabel Chen[1]

[1] Nanyang Technological University

Author Note

The authors made the following contributions. Dominique Makowski: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft; An Shu Te: Software, Project administration, Writing – review & editing; Tam Pham: Software, Writing – review & editing; Zen J. Lau: Software, Writing – review & editing; S.H. Annabel Chen: Project administration, Writing – review & editing.

Correspondence concerning this article should be addressed to Dominique Makowski, HSS 04-18, 48 Nanyang Avenue, Singapore. E-mail: dmakowski@ntu.edu.sg

Abstract

Complexity quantification, through entropy, information and fractal dimension indices, is gaining a renewed traction in psychopsyiology, as new measures with promising qualities emerge from the computational and mathematical advances. Unfortunately, few studies compare the relationship and objective performance of the plethora of existing metrics, in turn hindering reproducibility, replicability, consistency, and restults clarity in the field. In this study, we systematically compared 125 indices of complexity by their computational weight, their representativeness of a multidimensional space of latent dimensions, and empirical proximity with other indices. We propose that a selection of indices, including *ShanEn (D)*, *MSWPEn*, *CWPEn*, *FuzzyMSEn*, *AttEn*, *NLDFD*, *Hjorth*, *MFDFA (Width)*, *MFDFA (Max)*, *MFDFA (Mean)*, *SVDEn*, *MFDFA (Increment)*, might offer a complimentary choice in regards to the quantification of the complexity of time series.

Keywords: chaos, complexity, fractal, physiology, noise

Word count: 2353

The Structure of Chaos: An Empirical Comparison of Fractal Physiology Complexity Indices using NeuroKit2

Introduction

Complexity is an umbrella term for concepts derived from information theory, chaos theory, and fractal mathematics, used to quantify unpredictability, entropy, and/or randomness. Using these tools to characterize signals (a subfield commonly referred to as "fractal physiology," Bassingthwaighte et al., 2013) has shown promising results in physiology in the assessment and diagnostic of the state and health of living systems Ehlers (1995).

There has been a large and accelerating increase in the number of complexity indices in the past few decades. These new procedures are usually mathematically well-defined and theoretically promising. However, few empirical evidence exist to understand their differences and similarities. Moreover, some can be very expensive in terms of computation power and thus, time, which can become an issue in some applications such as high sampling-rate techniques (e.g., M/EEG) or real-time settings (brain-computer interface). As such, having a general view depicting the relationship between the indices with information about their computation time would be useful, for instance to guide the indices selection in settings where time or computational power is limited.

One of the contributing factor of this lack of empirical comparison is the lack of free, open-source, unified, and easy to use software for computing various complexity indices. Indeed, most of them are described mathematically in journal articles, and reusable code is seldom made available, which limits their further application and validation. *NeuroKit2* (Makowski et al., 2021) is a Python package for physiological signal processing that aims at providing the most comprehensive, accurate and fast pure Python implementations of complexity indices.

Leveraging this tool, the goal of this study is to empirically compare a vast number of complexity indices, inspect how they relate to one another, and extract some recommendations for indices selection, based on their added-value and computational efficiency. Using *NeuroKit2*, we will compute more than a hundred complexity indices on various types of signals, with varying degrees of noise. We will then project the results on a latent space through factor analysis, and report the most interesting indices in regards to their representation of the latent dimensions.

Methods

The script to generate the data can be found at **github.com/neuropsychology/NeuroKit/studies/complexity_benchmark**

We started by generating 5 types of signals, one random-walk, two oscillatory signals made (one made of harmonic frequencies that results in a self-repeating - fractal-like - signal), and two complex signals derived from Lorenz systems (with parameters $(\sigma = 10, \beta = 2.5, \rho = 28)$; and $(\sigma = 20, \beta = 2, \rho = 30)$, respectively). Each of this signal was iteratively generated at ... different lengths (). The resulting vectors were standardized and each were added 5 types of $(1/f)^\beta$ noise (namely violet $\beta = -2$, blue $\beta = -1$, white $\beta = 0$, pink $\beta = 1$, and brown $\beta = 2$ noise). Each noise type was added at 48 different intensities (linearly ranging from 0.1 to 4). Examples of generated signals are presented in **Figure 1**.

The combination of these parameters resulted in a total of 6000 signal iterations. For each of them, we computed 128 complexity indices, and additionally basic metric such as the standard deviation (*SD*), the *length* of the signal and its dominant *frequency*. We also included a *random* number to make sure that our our clustering / dimensionality analyses accurately discriminate this unrelated feature. The parameters used (such as the time-delay τ or the embedding dimension) are documented in the data generation script.

For a complete description of the various indices included, please refer to NeuroKit's documentation (https://neuropsychology.github.io/NeuroKit).

Results

The data analysis script, the data and the code for the figures is fully available at **github.com/neuropsychology/NeuroKit/studies/complexity_benchmark**. The analysis was performed in R using the *easystats* collection of packages (Lüdecke et al., 2021; Lüdecke et al., 2020; Makowski et al., 2020/2022, 2020).

Computation Time. Despite the relative shortness of the signals considered (a few thousand points at most), the fully-parallelized data generation script took 12h to run on a 48-cores machine. After summarizing and sorting the indices by computation time, the most striking feature are the orders of magnitude of difference between the fastest and slowest indices. Some of them are also particularly sensitive to the data length, a property which combined with computational expensiveness leads to indices being 100,000 slower to compute than other basic metrics.

Multiscale indices are among the slowest, due to their iterative nature (a given index is computed multiple times on coarse-grained subseries of the signal). Indices related to Recurrence Quantification Analysis (RQA) are also relatively slow and don't scale well with signal length.

For the subsequent analyses, we removed statistically redundant indices, such as *PowEn* - identical to *SD*, *CREn (100)* -identical to *CREn (10)*, and *FuzzyRCMSEn* - identical to *RCMSEn*.

Correlation. The Pearson correlation analysis revealed that complexity indices, despite their multitude, their unicities and specificities, do indeed share similarities. They form clusters, with two major ones easily appearing to the naked eye (the blue and the red groups). These two anti-correlated groups are driven by the fact that some indices, by design, index the “predictability”, whereas other the “randomness”, and thus are negatively related to one-another (see **Figure 2**).

Factor Analysis. The agreement procedure for the optimal number of factors suggested that the 125 indices can be mapped on a multidimensional space of 14 orthogonal latent factors, that we extracted using a *varimax* rotation. We then took interest in the loading profile of each indices, and in particular the latent dimension that maximally related to each index (see **Figure 3**).

The first factor is the closest to the largest amount of indices. Many indices with positive and strong loadings are particularly sensitive to the deviation of consecutive differences (e.g., *ShanEn - D*, *NLDFD*, *PFD - D*). It was negatively loaded by indices related to Detrended Fluctuation Analysis (DFA), which tend to index the presence of long-term correlations. This latent factor might encapsulate the predominance of short-term vs. long-term unpredictability. Indices that are the most representative (positively and negatively) and have a relatively low computational cost include *ShanEn - D*, *NLDFD*, *PFD - D*, and *AttEn*, *PSDFD*, *FuzzyMSEn*. The second factor was loaded maximally by signal *length* and *SD*, and thus might not capture features of complexity *per se*. Indices the most related to it were indices known to be sensitive to signal length, such as *ApEn*. The third factor included multiscale indices, such as *MSWPEn*. The fourth factor included indices that quantified the diversity of the tendency of a signal to revisit a past state (within a certain tolerance threshold). It was positively loaded by *ShanEn - r* and negatively by *RQA - Reccurence Rate*. The fifth factor was loaded by permutation entropy indices, such as *WPEn*. The sixth factor was driven by indices that were based on converting the signal into a number of bins. The seventh factor was loaded positively by the amount of noise, and negatively by multifractal indices such as *MFDFA - Increment*, suggesting a sensitivity to regularity. The last notable result is that indices based on a symbolization (discretization) of the time series do tend to create factors alongside the symbolization method. Finally, as a manipulation check of our factorization method, the random vector does indeed form its own factor, and doesn't load unto anything else.

Hierarchical Clustering and Connectivity Network. For illustration purposes, we represented the correlation matrix as a connectivity graph (see **Figure 4**). We then ran a hierarchical clustering (with a Ward D2 distance) to provide additional information or confirmation about the groups discussed above. This allowed us to fine-grain our recommendations of complimentary complexity indices (see **Figure 5**).

Indices Selection. The selection of a subset of indices was based on the following considerations: 1) high loadings on one predominant latent dimension, with additional attention to the pattern of secondary loadings. For instance, an index with a positive factor 1 loading and a negative factor 2 loading could complement another index with a similar factor 1 loading, but a positive factor 2 loading. This was helped by 2) the hierarchical clustering dendrogram, with which we attempted to indices from each (meaningful) higher order clusters. Items related to clusters that we know were related to noise, length or other artifacts were omitted. 3) A preference for indices with relatively shorter computation times. This yielded a selection of 12 indices. Next, we computed the cumulative variance explained of this selection in respect to the entirety of indices, and derived the optimal order to maximize the variance explained (see **Figure 6**). The included indices were:

- *ShanEn (D)*: The Shannon Entropy of the symbolic times series obtained by the "D" method described in Petrosian (1995) used traditionally in the context of the Petrosian fractal dimension (Esteller et al., 2001). The successive differences of the time series are assigned to 1 if the difference exceeds one standard deviation or 0 otherwise. The Entropy of the probabilities of these two events is then computed.
- *MSWPEn*: The Multiscale Weighted Permutation Entropy is the entropy of weighted ordinal descriptors of the time-embedded signal computed at different scales obtained by a coarsegraining procedure (Fadlallah et al., 2013).
- *CWPEn*: The Conditional Weighted Permutation Entropy is based on the difference of weighted entropy between that obtained at an embedding dimension m and that

obtained at $m + 1$ (Unakafov & Keller, 2014).

- *FuzzyMSEn*: This index corresponds to the multiscale Fuzzy Sample Entropy (Ishikawa & Mieno, 1979). This algorithm is computationally expensive to run.
- *AttEn*: The Attention Entropy is based on the frequency distribution of the intervals between the local maxima and minima of the time series (Yang et al., 2020).
- *NLDFD*: The Fractal dimension via Normalized Length Density (NLD) corresponds to the average absolute consecutive differences of the standardized signal (Kalauzi et al., 2009).
- *Hjorth*: Hjorth's Complexity is defined as the ratio of the mobility of the first derivative of the signal to the mean frequency of the signal (Hjorth, 1970).
- *MFDFA (Width)*: The width of the multifractal singularity spectrum (Kantelhardt et al., 2002) obtained via Detrended Fluctuation Analysis (DFA).
- *MFDFA (Max)* : The value of singularity spectrum D corresponding to the maximum value of singularity exponent H.
- *MFDFA (Mean)* : The mean of the maximum and minimum values of singularity exponent H.
- *SVDEn*: Singular Value Decomposition (SVD) Entropy quantifies the amount of eigenvectors needed for an adequate representation of the signal (Roberts et al., 1999).
- *MFDFA (Increment)*: The cumulative function of the squared increments of the generalized Hurst's exponents between consecutive moment orders (Faini et al., 2021).

Finally, we visualized the expected value of our selection of indices for different types of signals under different conditions of noise (see **Figure 7**). This revealed that two indices, namely *ShanEn (D)* and *NLDFD*, are primarily driven by the noise intensity (which is expected, as they capture the variability of successive differences). The other indices appear to be able to discriminate between the various types of signals (when the signal is not dominated by noise).

Discussion

As complexity science grows in size and application, a systematic approach to compare their "performance" becomes necessary to increase the clarity and structure of the field. The word *performance* is here to be understood in a relative sense, as any such endeavor faces the "hard problem" of complexity science. The fact that indices are sensitive to specific objective properties of a signal that we consider part of over-arching concepts such as "complex" and "chaotic", though it is unclear how these high-level concepts transfer back, in a top-down fashion, into a combination of lower-level features, such as short-term vs. long-term variability, auto-correlation, information, randomness, and so on. As such, it is conceptually complicated to benchmark complexity measures against "objectively" complex vs. non-complex signals. In other words, we know that different characteristics can contribute to the "complexity" of a signal, but there is not a one-to-one correspondence between the latter and the former.

This explains the choice of the paradigm used in the present study, in which we generated different types of signals to which we systematically added different types and amount of perturbations. However, we did not seek at measuring how complexity indices can discriminate between these features or systems, nor did we attempt at mimicking real-life signals or scenarios. The goal was instead to generate enough variability to reliably map the relationships between the indices.

The plurality of underlying components of empirical complexity (what is measured by complexity indices) seems to be confirmed by our results, showing that complexity indices vary in their sensitivity to various orthogonal latent dimensions. One of the limitation of the current study has to do with the limited possibilities of interpretation of these underlying dimensions, and future studies are needed to investigate and discuss them in greater depth.

Indices that were highlighted as encapsulating information about different underlying dimensions at a relatively low computational cost include *ShanEn (D)*, *MSWPEn*, *CWPEn*, *FuzzyMSEn*, *AttEn*, *NLDFD*, *Hjorth*, *MFDFA (Width)*, *MFDFA (Max)*, *MFDFA (Mean)*, *SVDEn*, *MFDFA (Increment)*. These indices might be complimentary in offering a comprehensive profile of the complexity of a time series. Future studies are needed to analyze the nature of the dominant sensitivities of different groups of indices, so that results can be more easily interpreted and integrated into new studies and novel theories.

References

Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (2013). *Fractal physiology*. Springer.

Ehlers, C. L. (1995). Chaos and complexity: Can it help us to understand mood and behavior? *Archives of General Psychiatry*, *52*(11), 960–964.

Esteller, R., Vachtsevanos, G., Echauz, J., & Litt, B. (2001). A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *48*(2), 177–183.

Fadlallah, B., Chen, B., Keil, A., & Príncipe, J. (2013). Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Physical Review E*, *87*(2), 022911.

Faini, A., Parati, G., & Castiglioni, P. (2021). Multiscale assessment of the degree of multifractality for physiological time series. *Philosophical Transactions of the Royal Society A*, *379*(2212), 20200254.

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, *29*(3), 306–310.

Ishikawa, A., & Mieno, H. (1979). The fuzzy entropy concept and its application. *Fuzzy Sets and Systems*, *2*(2), 113–123.

Kalauzi, A., Bojić, T., & Rakić, L. (2009). Extracting complexity waveforms from one-dimensional signals. *Nonlinear Biomedical Physics*, *3*(1), 1–11.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and Its Applications*, *316*(1-4), 87–114.

Lau, Z. J., Pham, T., Annabel, S., & Makowski, D. (2021). *Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations*.

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Lüdecke, D., Patil, I., Ben-Shachar, M. S., Wiernik, B. M., Waggoner, P., & Makowski, D. (2021). see: An R package for visualizing statistical models. *Journal of Open Source Software*, *6*(64), 3393. https://doi.org/10.21105/joss.03393

Makowski, D., Ben-Shachar, M., Patil, I., & Lüdecke, D. (2020). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, *5*(51), 2306. https://doi.org/10.21105/joss.02306

Makowski, D., Lüdecke, D., Ben-Shachar, M. S., & Patil, I. (2022). *modelbased: Estimation of model-based predictions, contrasts and means* (Version 0.7.2.1) [Computer software]. https://CRAN.R-project.org/package=modelbased (Original work published 2020)

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Petrosian, A. (1995). Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns. *Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems*, 212–217.

Roberts, S. J., Penny, W., & Rezek, I. (1999). Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. *Medical & Biological Engineering & Computing*, *37*(1), 93–98.

Unakafov, A. M., & Keller, K. (2014). Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena*, *269*, 94–102.

Yang, J., Choudhary, G. I., Rahardja, S., & Franti, P. (2020). Classification of interbeat interval time-series using attention entropy. *IEEE Transactions on Affective Computing*.

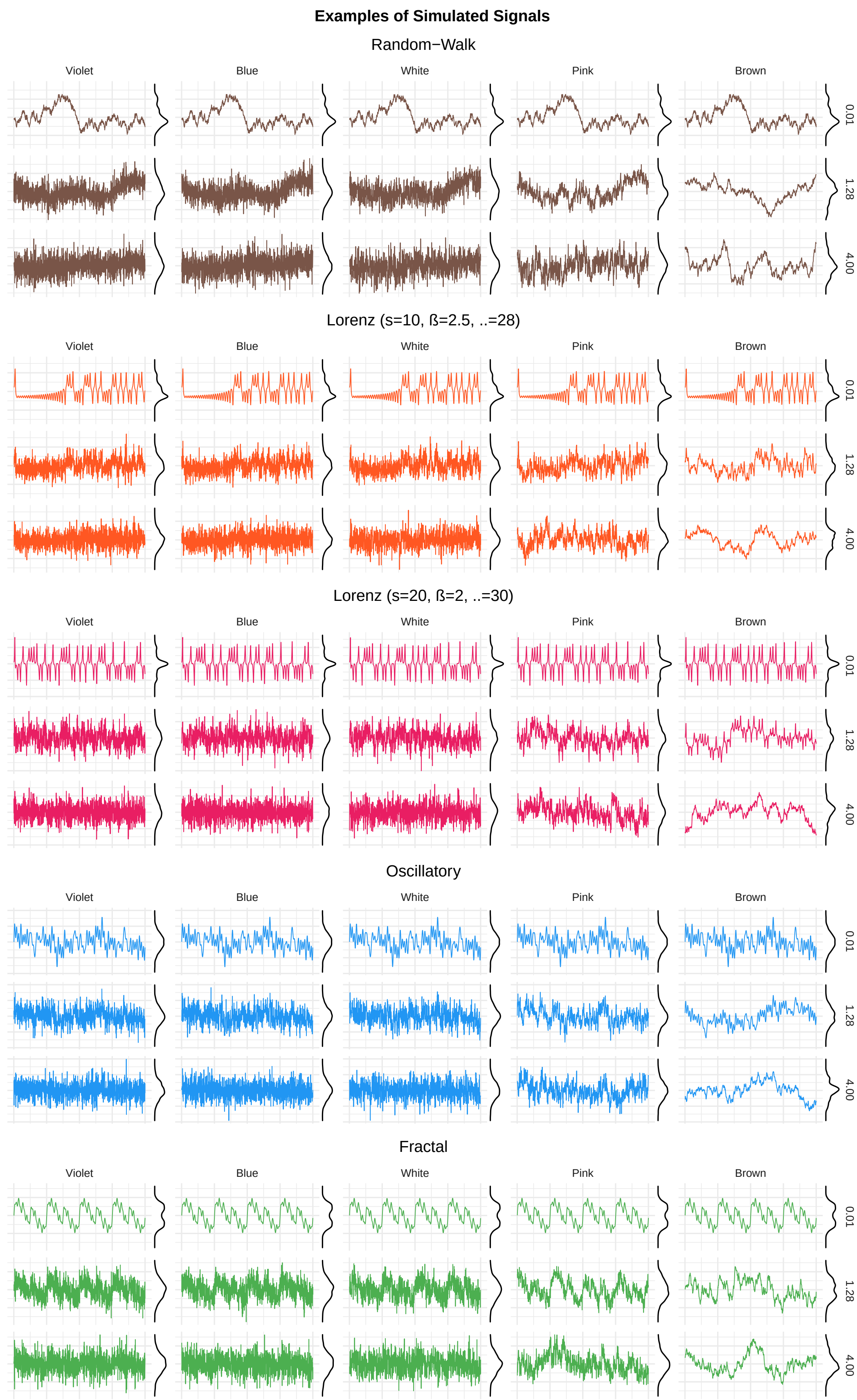

Figure 1. Different types of simulated signals, to which was added 5 types of noise (violet, blue, white, pink, and brown) with different intensities. For each signal type, the first row shows the signal with a minimal amount of noise, and the last with a maximal amount of noise. We can see

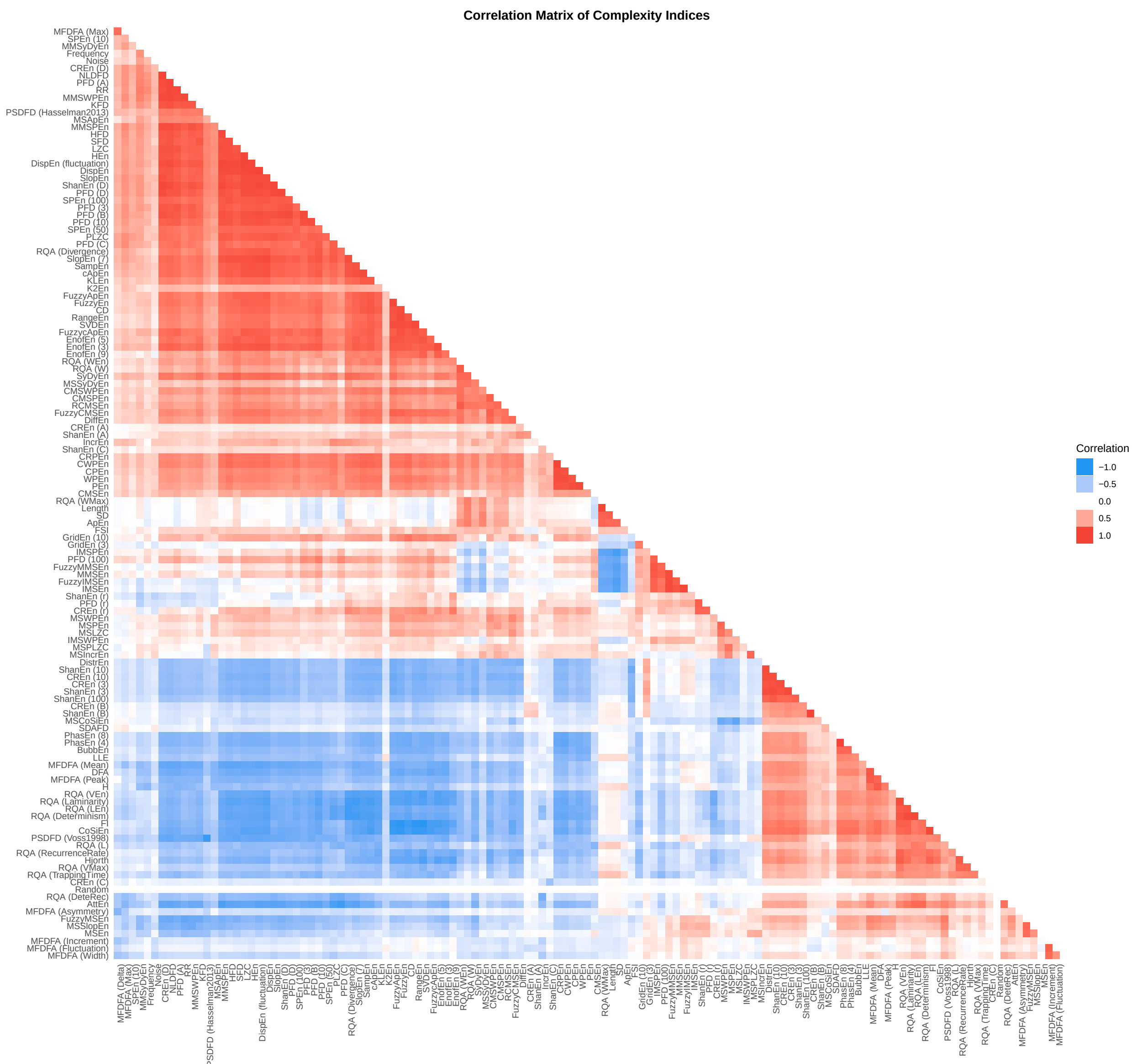

Figure 2. Correlation matrix of complexity indices.

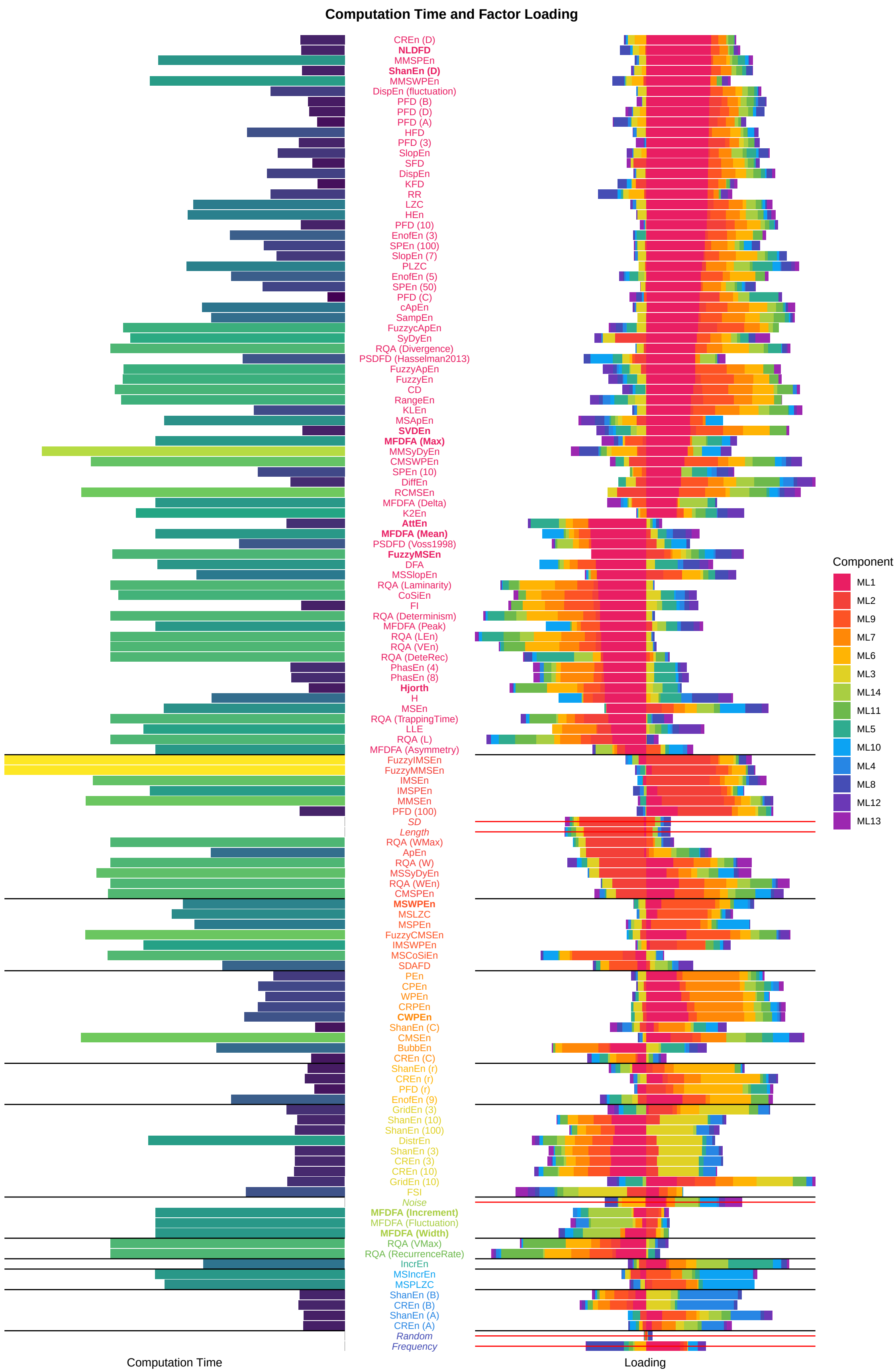

Figure 3. Factor loadings and computation times of the complexity indices, colored by the factor they represent the most.

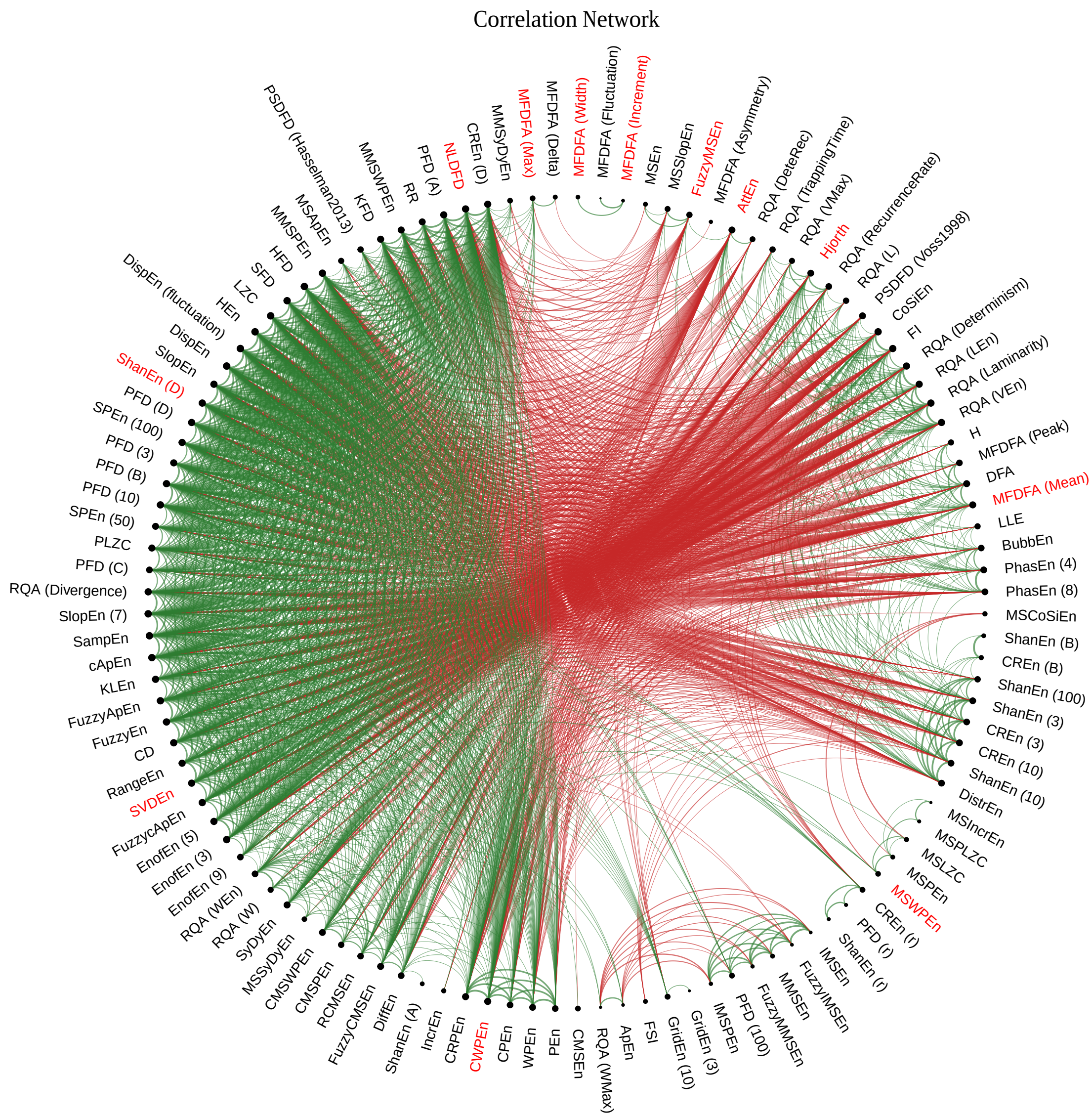

Figure 4. Correlation network of the complexity indices. Only the links where |r| > 0.6 are displayed.

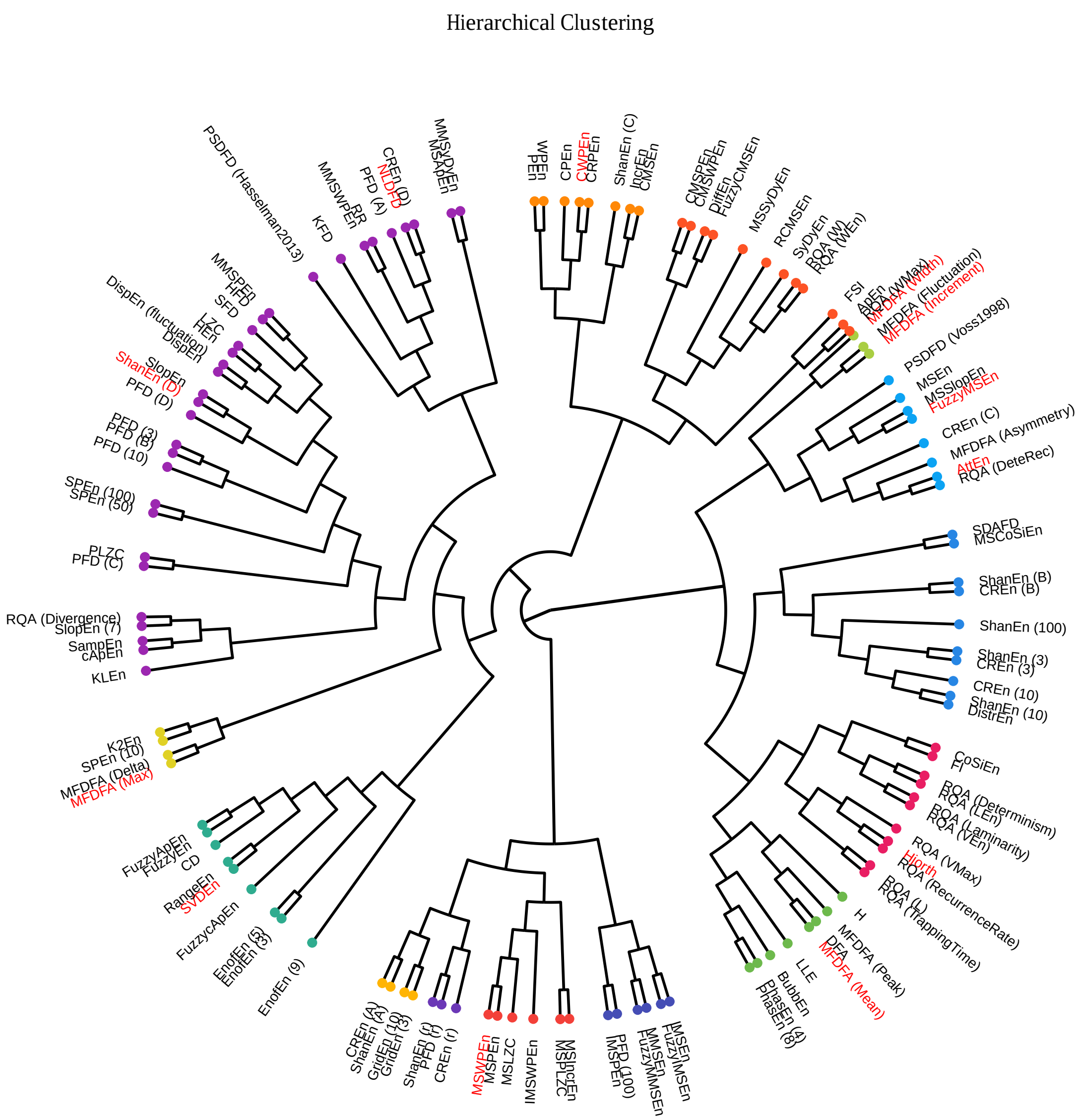

Figure 5. Dendrogram representing the hierarchical clustering of the complexity indices.

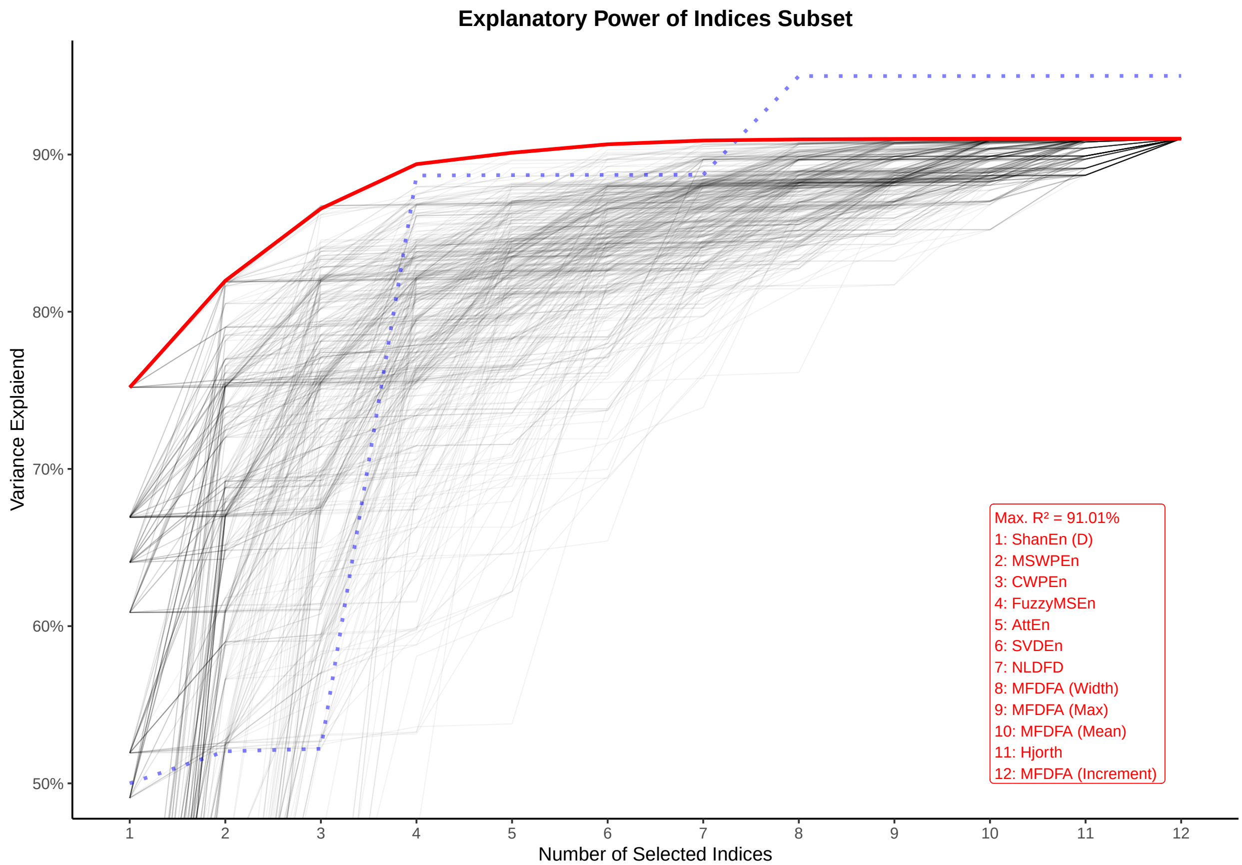

Figure 6. Variance of the whole dataset of indices explained by the subselection. Each line represents a random number of selected variables. The red line represents the optimal order (i.e., the relative importance) that maximizes the variance explained. The dotted blue line represents the cumulative relative average computation time of the selected indices, and shows that FuzzyMSEn and MFDFA indices are the most costly algorithms.

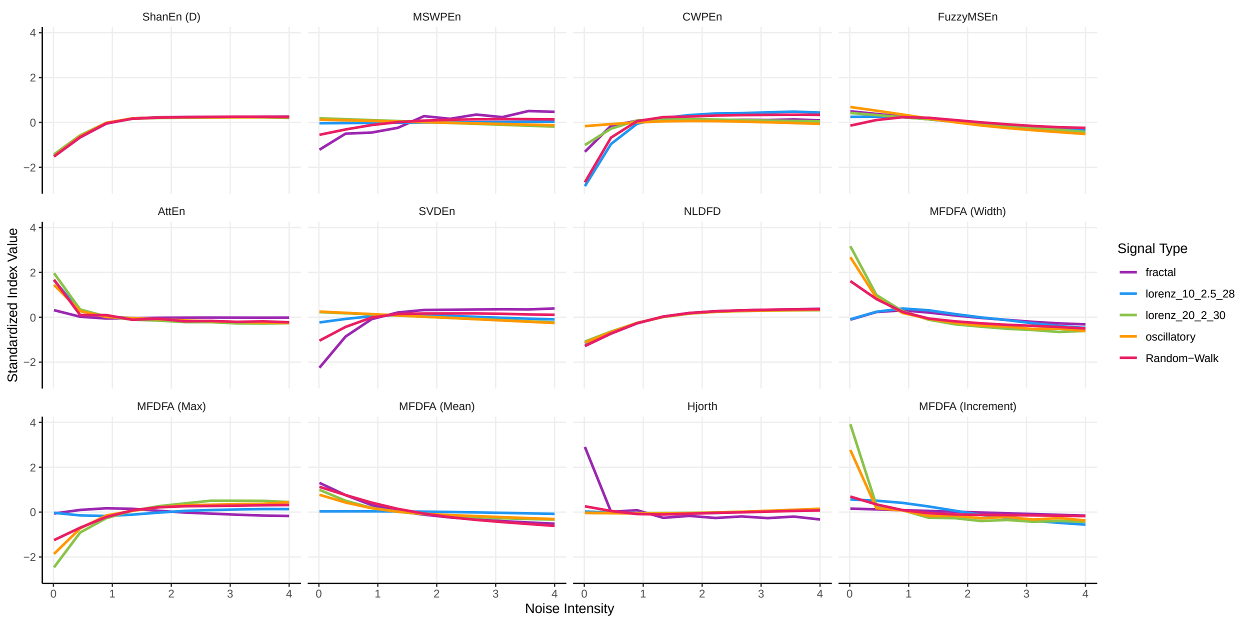

Figure 7. Visualization of the expected value of a selection of indices depending on the signal type and of the amount of noise.