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, 2, 3, 4]

[1] School of Social Sciences, Nanyang Technological University, Singapore

[2] LKC Medicine, Nanyang Technological University, Singapore

[3] National Institute of Education, Singapore

[4] Centre for Research and Development in Learning, Nanyang Technological University, Singapore

Author Note

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: 2587

$_{33}$ **The Structure of Chaos: An Empirical Comparison of Fractal Physiology**

$_{34}$ **Complexity Indices using NeuroKit2**

$_{35}$ **Introduction**

$_{36}$ Complexity is an umbrella term for concepts derived from information theory, chaos

$_{37}$ theory, and fractal mathematics, used to quantify unpredictability, entropy, and/or

$_{38}$ randomness. Using these methods to characterize physiological signals (a subfield

$_{39}$ commonly referred to as "fractal physiology," Bassingthwaighte et al., 2013) has shown

$_{40}$ promising results in the assessment and diagnostic of the state and health of living systems

$_{41}$ Goetz (2007).

$_{42}$ There has been an exponential increase in the number of complexity indices in the

$_{43}$ past few decades (A. C. Yang & Tsai, 2013). Although these new procedures are usually

$_{44}$ mathematically well-defined and theoretically promising, limited empirical evidence is

$_{45}$ available to understand their similarities and differences (Lau et al., 2021; A. C. Yang &

$_{46}$ Tsai, 2013). Moreover, some of these methods are resource-intensive and require long

$_{47}$ computation times. This complicates their application with techniques that utilise high

$_{48}$ sampling-rates (e.g., M/EEG) and makes them impractical to implement in real-time

$_{49}$ settings - such as brain-computer interfaces (Manis et al., 2018; "Refined Composite

$_{50}$ Multiscale Dispersion Entropy and Its Application to Biomedical Signals," 2017). As such,

$_{51}$ having empirical data about the computation time of various complexity indices would

$_{52}$ prove useful, for instance to objectively guide their selection, especially in contexts where

$_{53}$ time or computational resourcse are limited.

$_{54}$ Additionally, the lack of a comprehensive open-source and user-friendly software for

$_{55}$ computing various complexity indices likely contributes to the limited availability of

$_{56}$ empirical comparison (Flood & Grimm, 2021). Indeed, most complexity indices are only

$_{57}$ described mathematically in journal articles, with reusable code seldom made available,

$_{58}$ therefore limiting their further application and validation (Flood & Grimm, 2021; A. C.

59   Yang & Tsai, 2013). To address this gap, we added a comprehensive set of

60   complexity-related features to *NeuroKit2*, a Python package for physiological signal

61   processing (Makowski et al., 2021). This submodule aims at enabling users to compute a

62   vast amount of complexity indices. The code is designed to be as fast as possible, while

63   still written in pure Python (though with the helpˆof dependencies such as Numpy or

64   Pandas, Harris et al., 2020; McKinney et al., 2010) to maximize the re-usability,

65   transparency, and correctness.

66   Leveraging this tool, the goal of this study is to empirically compare a large number

67   of complexity indices, inspect how they relate to one another, and derive recommendations

68   for indices selection. More specifically, we will quantify the complexity using 128 indices of

69   various types of signals with varying degrees of noise, using *NeuroKit2*. We will then

70   project the results on a latent space through factor analysis, and review the various indices

71   that we find the most relevant and interesting in regards to their representation of the

72   latent dimensions. This analysis will be complemented by hierarchical clustering.

73   **Methods**

74   The script to generate the data can be found at

75   **github.com/neuropsychology/NeuroKit/studies/complexity_benchmark**

76   We started by generating 5 types of signals, one random-walk, two oscillatory signals

77   made (with one made of harmonic frequencies that results in a self-repeating - fractal-like -

78   signal), and two complex signals derived from Lorenz systems (with parameters

79   ($\sigma = 10, \beta = 2.5, \rho = 28$); and ($\sigma = 20, \beta = 2, \rho = 30$), respectively). Each of this signal was

80   iteratively generated at 5 different lengths. The resulting vectors were standardized and

81   each were added 5 types of $(1/f)^{\beta}$ noise (namely violet $\beta = -2$, blue $\beta = -1$, white $\beta = 0$,

82   pink $\beta = 1$, and brown $\beta = 2$ noise). Each noise type was added at 48 different intensities

83   (linearly ranging from 0.1 to 4). Examples of generated signals are presented in **Figure 1**.

84   The combination of these parameters resulted in a total of 6000 signal iterations. For

85 each of them, we computed 128 complexity indices, and additional basic metrics such as

86 the standard deviation (*SD*), the *length* of the signal and its dominant *frequency*. We also

87 included a *random* number to make sure that our our dimensionality analyses accurately

88 discriminate this unrelated feature. The parameters used (such as the time-delay τ or the

89 embedding dimension) are documented in the data generation script. For a complete

90 description of the various indices included, please refer to NeuroKit's documentation at

91 **https://neuropsychology.github.io/NeuroKit**.

92 **Results**

93   The data analysis script, the data and the code for the figures is fully available at

94 **github.com/neuropsychology/NeuroKit/studies/complexity_benchmark**. The

95 analysis was performed in R using the *easystats* collection of packages (Lüdecke et al.,

96 2021; Lüdecke et al., 2020; Makowski et al., 2020/2022, 2020).

97   **Computation Time.**   Firstly, one should note that the computation times

98 presented in **Figure 2** are relative and do not correspond to real times, as these would

99 highly depend on the machine used. Instead, the goal is here to convey some intuition on

100 the differences between different classes of indices (using the same machine and the same

101 language of implementation, i.e., Python). It is possible that computational advances or

102 improvements in the code efficiency might change some of these values, but we believe that

103 the "big picture" should remain fairly stable, as it is to a large extend driven by the

104 inherent nature of the algorithms under consideration.

105   Despite the relative shortness of the signals considered (a few thousand points at

106 most), the fully-parallelized data generation script took 24h to run on a 48-cores machine.

107 After summarizing and sorting the indices by computation time, the most striking feature is

108 the order of magnitude of difference between the fastest and slowest indices. Additionally,

109 some indices are particularly sensitive to the signal length, a property which combined with

110 computational cost led to indices being 100,000 times slower to compute than others.

111   In particular, multiscale indices were among the slowest to compute due to their

112 iterative nature (a given index is computed multiple times on coarse-grained subseries of

113 the signal). Indices related to Recurrence Quantification Analysis (RQA) were also

114 relatively slow and did not scale well with signal length.

115   For the subsequent analyses, we removed statistically redundant indices, such as

116 *PowEn* - identical to *SD*, *CREn (100)* - identical to *CREn (10)*, and *FuzzyRCMSEn* -

117 identical to *RCMSEn.*

118   **Correlation.**   The Pearson correlation analysis revealed that complexity indices,

119 despite their multitude and their conceptual specificities, do indeed share similarities. They

120 form two major clusters that are easily observable (the blue and the red groups in **Figure

121 2**). However, these two anti-correlated groups are mostly revealing of the fact that some

122 indices, by design, index the "predictability", whereas others, the "randomness", and thus

123 are negatively related to one-another. In order to extract finer groupings, further analyses

124 procedures are applied below.

125   **Factor Analysis.**   The agreement procedure for the optimal number of factors

126 suggested that the 125 indices can be mapped on a multidimensional space of 14

127 orthogonal latent factors, that we extracted using a *varimax* rotation. We then took

128 interest in the loading profile of each index, and in particular the latent dimension that it

129 maximally relates to (see **Figure 3**).

130   This first factor is the closest to the largest amount of indices, and is positively

131 loaded by indices that are sensitive to the deviation of consecutive differences (e.g., *ShanEn*

132 *- D*, *NLDFD*, *PFD - D*). In line with this, this factor was negatively loaded by indices

133 related to Detrended Fluctuation Analysis (DFA), which tends to index the presence of

134 long-term correlations. As such, this latent factor might encapsulate the predominance of

135 short-term vs. long-term unpredictability. The second factor was strongly loaded by signal

136 *length* and *SD*, and thus might not capture features of complexity *per se.* Indices with the

137 most relation to it were indices known to be sensitive to signal length, such as *ApEn.* The

138 third factor included multiscale indices, such as *MSWPEn.* The fourth factor was loaded

139 by permutation entropy indices, such as *WPEn.* The fifth and the sixth factors were loaded

140 by indices grouped by the signal symbolization method used (by a tolerance level r, or by

141 the number of bins for the fifth and the sixth factors, respectively). The seventh factor was

142 loaded positively by the amount of noise, and negatively by multifractal indices such as

143 *MFDFA - Increment*, suggesting a sensitivity to regularity. Finally, as a manipulation

144 check for our factorization method, the random vector did not load unto any factors.

145   **Hierarchical Clustering and Connectivity Network.**   For illustration

146 purposes, we represented the correlation matrix as a connectivity graph (see **Figure 4**).

147 We then ran a hierarchical clustering (with a Ward D2 distance) to provide additional

148 information or confirmation about the groups discussed above. This allowed us to

149 fine-grain our recommendations of complimentary complexity indices (see **Figure 5**).

150   **Indices Selection.**   The selection of a subset of indices was based on the following

151 considerations: 1) high loadings on one predominant latent dimension, with additional

152 attention to the pattern of secondary loadings. For instance, an index with a positive

153 factor 1 loading and a negative factor 2 loading could complement another index with a

154 similar factor 1 loading, but a positive factor 2 loading. This was helped by 2) the

155 hierarchical clustering dendrogram, with which we attempted to indices from each

156 (meaningful) higher order clusters. Items related to clusters that we know were related to

157 noise, length or other artifacts were omitted. 3) A preference for indices with relatively

158 shorter computation times. This yielded a selection of 12 indices. Next, we computed the

159 cumulative variance explained of this selection in respect to the entirety of indices, and

160 derived the optimal order to maximize the variance explained (see **Figure 6**). The 12

161 included indices, representing 91.01% of the variance of the whole dataset, were:

162 - *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 (J. 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.,

190   1999).

191   • *MFDFA (Increment)*: The cumulative function of the squared increments of the

192   generalized Hurst's exponents between consecutive moment orders (Faini et al., 2021).

193   Finally, we visualized the expected value of our selection of indices for different types

194   of signals under different conditions of noise (see **Figure 7**). This revealed that two

195   indices, namely *ShanEn (D)* and *NLDFD*, are primarily driven by the noise intensity

196   (which is expected, as they capture the variability of successive differences). The other

197   indices appear to be able to discriminate between the various types of signals (when the

198   signal is not dominated by noise).

199   **Discussion**

200   As the span and application of complexity science grows, a systematic approach to

201   compare their "performance" becomes necessary to reinforcethe clarity and structure of the

202   field. The term *performance* is here to be understood in a relative sense, as any such

203   endeavor faces the "hard problem" of complexity science: various objective properties of

204   signals (e.g., short-term vs. long-term variability, auto-correlation, information,

205   randomness, Namdari & Li, 2019; Xiong et al., 2017) participate in forming together

206   over-arching concepts such as "complex" and "chaotic". Indices that are sensitive to some

207   of these objective properties are thus conceptually linked through these over-arching

208   framework. However, it remains unclear how these high-level concepts transfer back, in a

209   top-down fashion, into a combination of lower-level features. As such, it is conceptually

210   complicated to benchmark complexity measures against "objectively" complex

211   vs. non-complex signals. In other words, we know that different objective signal

212   characteristics can contribute to the "complexity" of a signal, but there is not a one-to-one

213   correspondence between the latter and the former.

214   To circumvent the aforementioned consideration, we adopted a paradigm where we

215 generated different types of signals to which we systematically added distinct types - and

216 amount - of perturbations. It is to note that we did not seek at measuring how complexity

217 indices can discriminate between these signal types, nor did we attempt at mimicking

218 real-life signals or scenarios. The goal was instead to generate enough variability to reliably

219 map the relationships between the indices.

220   Our results empirical confirm the plurality of underlying components of complexity

221 (although it is here defined somewhat circularly as what is measured by complexity

222 indices), and more importantly show that complexity indices vary in their sensitivity to

223 various orthogonal latent dimensions. However, the limited possibilities of interpretation of

224 these dimensions is a limitation of the present investigation, and future studies are needed

225 to investigate and discuss them in greater depth (for instance, by modulating specific

226 properties of signals and measuring their impact on these latent dimensions).

227   Taking into account the increasing role of complexity science as a field and the sheer

228 number of complexity indices already published, our study aimed at empirically map the

229 relationship between various indices and provide useful information to guide future

230 researchers in their selection. Indices that were highlighted as encapsulating information

231 about different underlying dimensions at a relatively low computational cost include

232 *ShanEn (D)*, *MSWPEn*, *CWPEn*, *FuzzyMSEn*, *AttEn*, *NLDFD*, *Hjorth*, *MFDFA (Width)*,

233 *MFDFA (Max)*, *MFDFA (Mean)*, *SVDEn*, *MFDFA (Increment)*. These indices might be

234 complimentary in offering a comprehensive profile of the complexity of a time series.

235 Moving forward, future studies are needed to validate, analyze and interpret the nature of

236 the dominant sensitivities of indices groups, so that studies results can be more easily

237 interpreted and integrated into new research 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.

Flood, M. W., & Grimm, B. (2021). EntropyHub: An open-source toolkit for entropic time series analysis. *PLOS ONE*, *16*(11), e0259448. https://doi.org/10.1371/journal.pone.0259448

Goetz, S. J. (2007). Spatial Dynamics, Networks and Modelling ? Edited by Aura Reggiani and Peter Nijkamp. *Papers in Regional Science*, *86*(3), 523–524. https://doi.org/10.1111/j.1435-5957.2007.00128.x

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

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

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

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

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

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

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

278 Lüdecke, D., Patil, I., Ben-Shachar, M. S., Wiernik, B. M., Waggoner, P., & Makowski, D.
279   (2021). see: An R package for visualizing statistical models. *Journal of Open Source*
280   *Software*, *6*(64), 3393. https://doi.org/10.21105/joss.03393

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

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

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

291 Manis, G., Aktaruzzaman, M., & Sassi, R. (2018). Low Computational Cost for Sample

292   Entropy. *Entropy, 20*(1), 61. https://doi.org/10.3390/e20010061

293 McKinney, W.others. (2010). Data structures for statistical computing in python.

294   *Proceedings of the 9th Python in Science Conference, 445*, 51–56.

295 Namdari, A., & Li, Z. (Steven). (2019). A review of entropy measures for uncertainty

296   quantification of stochastic processes. *Advances in Mechanical Engineering, 11*(6),

297   168781401985735. https://doi.org/10.1177/1687814019857350

298 Petrosian, A. (1995). Kolmogorov complexity of finite sequences and recognition of

299   different preictal EEG patterns. *Proceedings Eighth IEEE Symposium on*

300   *Computer-Based Medical Systems*, 212–217.

301 Refined composite multiscale dispersion entropy and its application to biomedical signals.

302   (2017). *IEEE Transactions on Biomedical Engineering, 64*(12), 2872–2879.

303   https://doi.org/10.1109/tbme.2017.2679136

304 Roberts, S. J., Penny, W., & Rezek, I. (1999). Temporal and spatial complexity measures

305   for electroencephalogram based brain-computer interfacing. *Medical & Biological*

306   *Engineering & Computing, 37*(1), 93–98.

307 Unakafov, A. M., & Keller, K. (2014). Conditional entropy of ordinal patterns. *Physica D:*

308   *Nonlinear Phenomena, 269*, 94–102.

309 Xiong, W., Faes, L., & Ivanov, P. Ch. (2017). Entropy measures, entropy estimators, and

310   their performance in quantifying complex dynamics: Effects of artifacts,

311   nonstationarity, and long-range correlations. *Physical Review E, 95*(6).

312   https://doi.org/10.1103/physreve.95.062114

313 Yang, A. C., & Tsai, S.-J. (2013). Is mental illness complex? From behavior to brain.

314   *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 45*, 253–257.

315   https://doi.org/10.1016/j.pnpbp.2012.09.015

316 Yang, J., Choudhary, G. I., Rahardja, S., & Franti, P. (2020). Classification of interbeat

317   interval time-series using attention entropy. *IEEE Transactions on Affective Computing.*
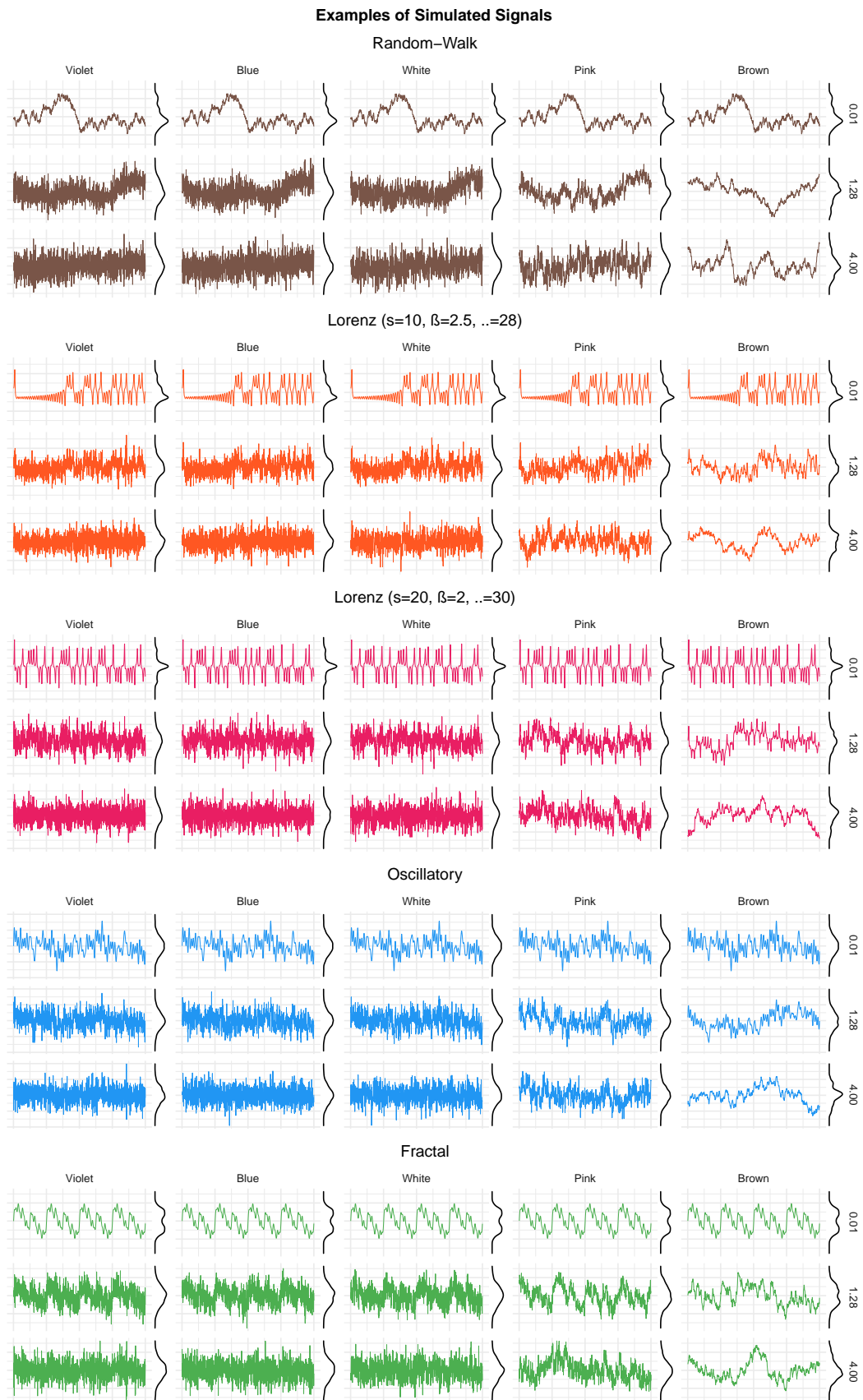
Figure 1. Different types of simulated signals, to which was added 5 types of noise (violet, blue, white, pink, and brown) with different intensities. For each signal type, the first row shows the signal with a minimal amount of noise, and the last with a maximal amount of noise. We can see
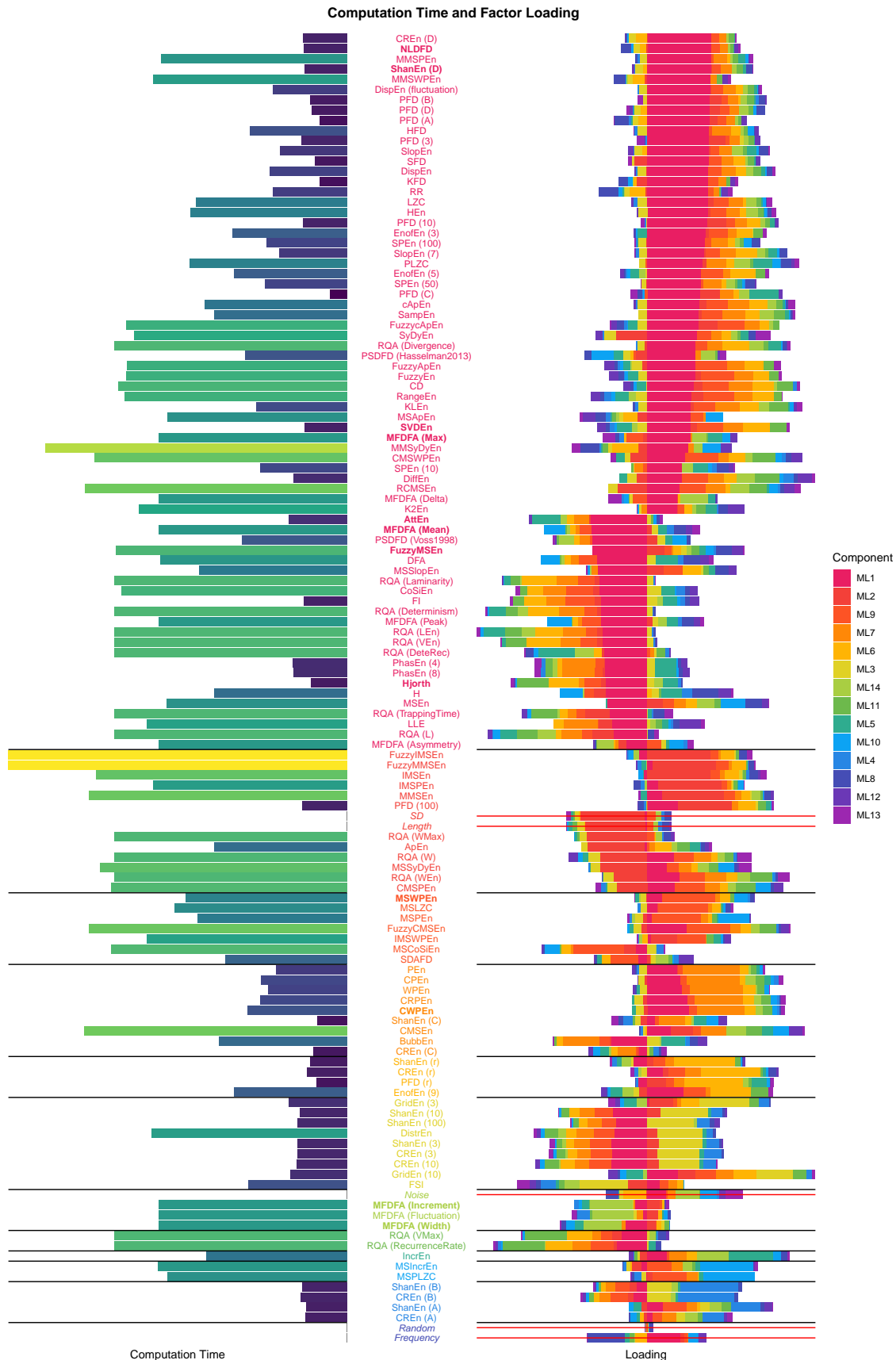
Figure 2. Correlation matrix of complexity indices.

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

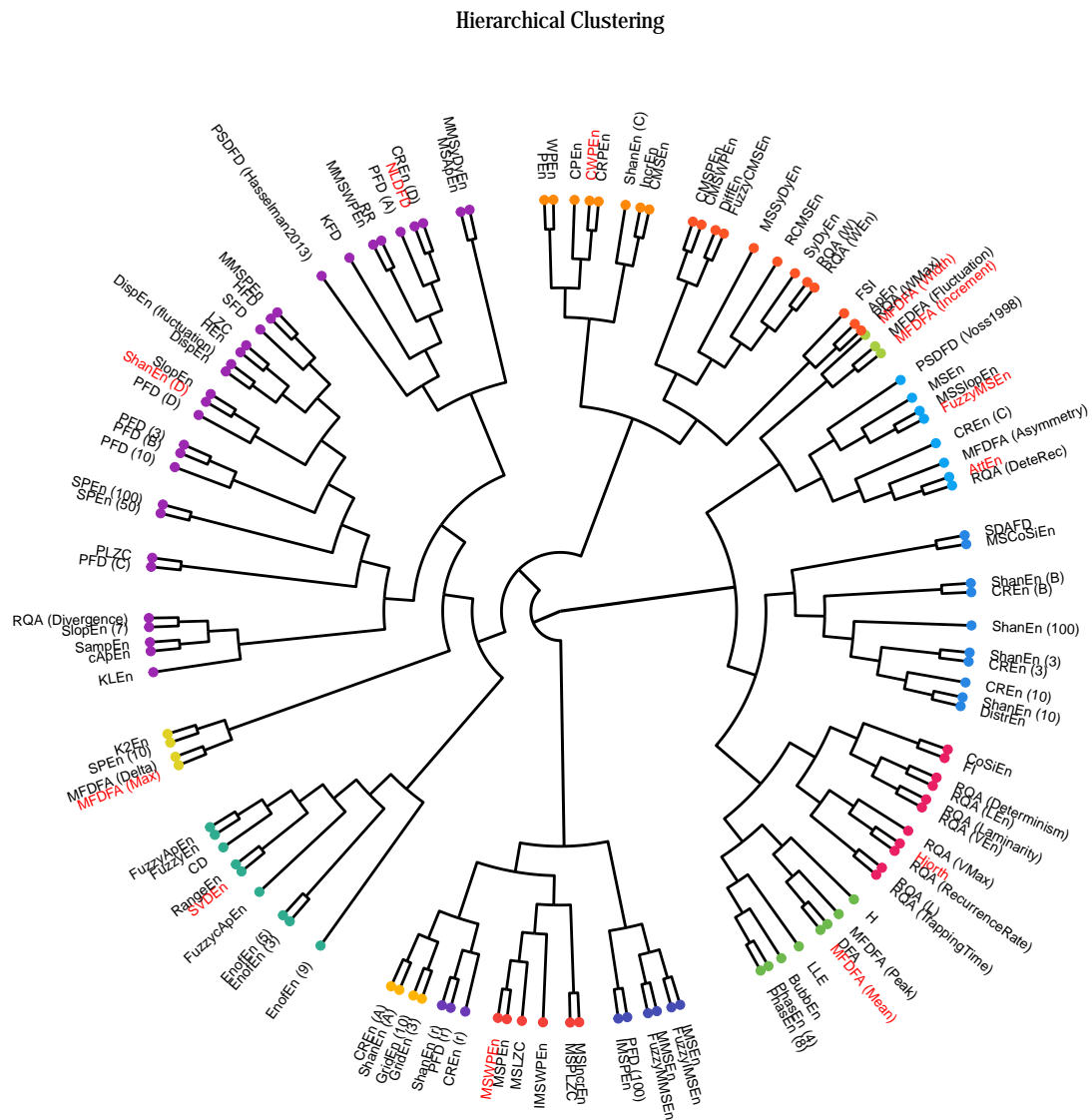**Figure 4.** Correlation network of the complexity indices. Only the links where |r| > 0.6 are displayed.

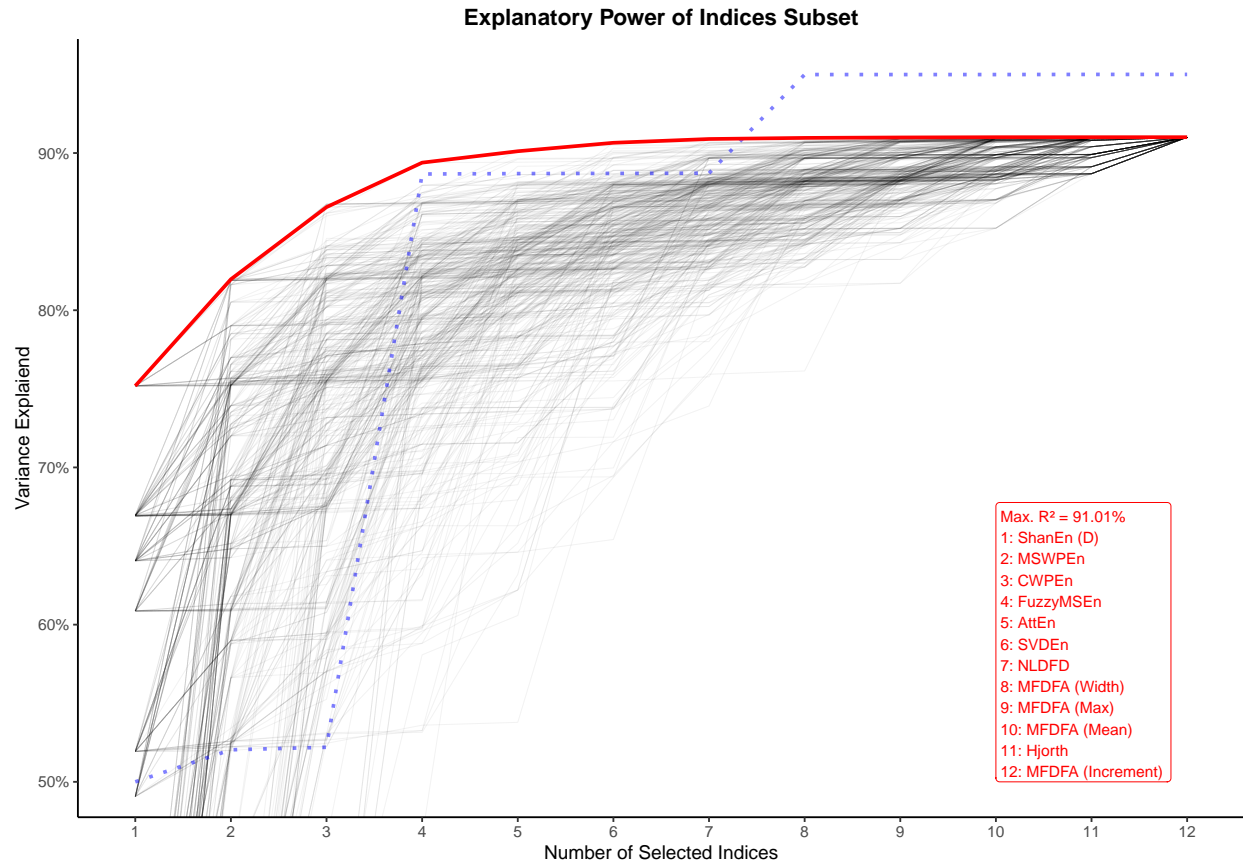**Figure 5.** Dendrogram representing the hierarchical clustering of the complexity indices.

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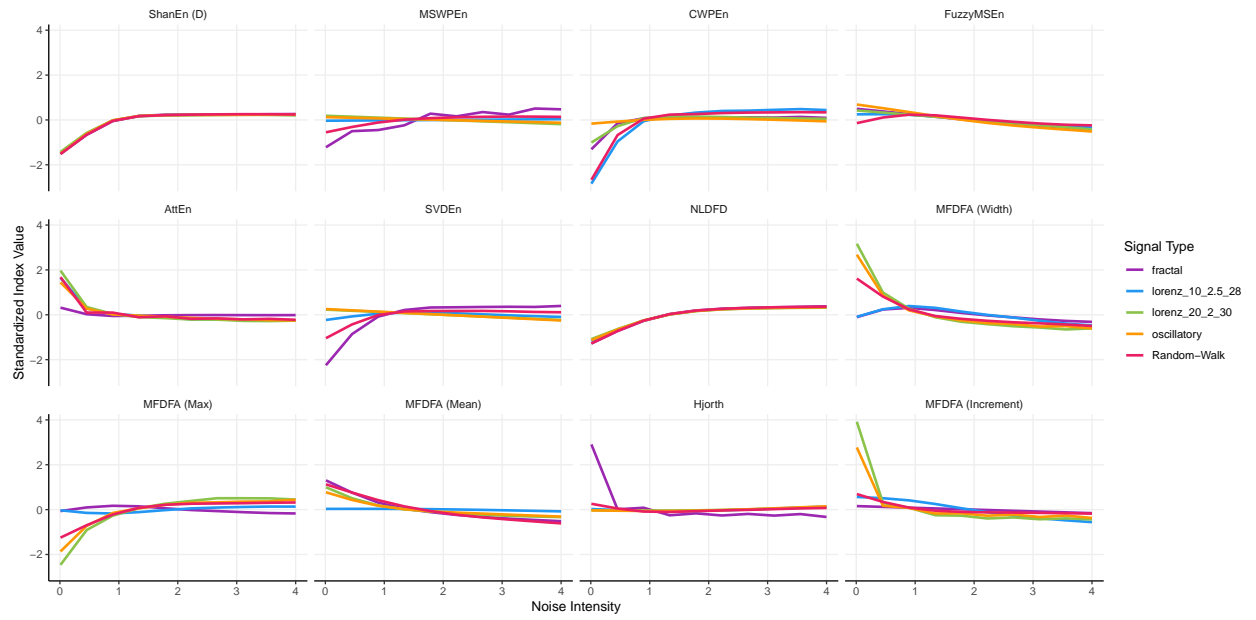
Figure 7. Visualization of the expected value of a selection of indices depending on the signal type and of the amount of noise.