REVISION

We thank the editor and the reviewers for giving us the opportunity of revising and substantially improving our manuscript.

We have addressed all the comments below. In addition to that, this revision was the opportunity to make the following improvements:

1. We slightly changed the list of complexity indices to include a few new ones (e.g., mean information gain - MIG), and removed some results that, we realized, were likely not generalizable: these included algorithms that, for instance, discretized consecutive changes based on whether these exceeded a given threshold (e.g., higher or lower than the signal’s SD), because the amplitude of consecutive change is highly sensitive to sampling rate.
2. We updated **Figure 2** to be more informative: because the median computation time is already included in Figure 4, we updated Figure 2 to include the variations as a function of signal length.
3. We updated **Figure 4** to include a measure of “archetypicity” which indicates the diversity of the loading profile of each index. This is informative to discriminate between indices that uniquely capture some of the latent dimensions suggested by our model, and those that load on various factors.

# Reviewer #1

We thank the reviewer for providing these thorough comments that helped us clarify and address several important points in the manuscript and the abstract.

This article should make clear that the list of techniques used is not fixed but should be completed. The short list in the abstract is misleading. In my opinion it should be indicated that it is a selection of the authors but for me the risk is that less scientific researchers use only the few indicators instead of the whole list of parameters.

We have added further clarification to the abstract to highlight that the total pool of 112 indices used was decided based on what is implemented in the software (and is thus non-exhaustive) (l. 26-27):

*“Using the NeuroKit2 Python software, we computed a list of 112 (predominantly used) complexity indices on signals varying in their characteristics (noise, length and frequency spectrum).”*

We also rephrased the part about the short selection of 12 indices, emphasizing that it was one possibility among others and that the information for alternative subsets is provided. This abstract’s part now reads as follows (l. 30-36):

*“Based on these considerations, we propose that a selection of 12 indices, together representing 85.97% of the total variance of all indices, might offer a parsimonious and complimentary choice in regards to the quantification of the complexity of time series. […] Elements of consideration for alternative subsets are discussed, and data, analysis scripts and code for the figures are open-source.”*

Note also that there are also recent articles, to mention only a few, such as those on the notions of symmetry (linking in recurrences -RQA and entropies (palindrome, symmetropy, symmetrical quantification analysis)) that would also deserve to be mentioned.

We recognize that in the attempt to keep our article concise and to the point, we did not include an extensive literature review on other non-included complexity approaches, such as the recently developed symmetry-based approach. We have mentioned this non-exhaustivity (l. 81-84):

*“It should be noted that, even though this is one of the largest empirical comparison of complexity measures to date to our knowledge, the list of indices used is by no means exhaustive, with new indices constantly being developed, such as for instance symmetropy (Girault & Menigot, 2022).”*

To improve the paper, the abstract should be corrected because the abstract is a review, so the results are missing, please correct.

We have improved the abstract ton include (and temper) the main results (see above).

The size of the signals is missing (unless I am mistaken).

We have corrected this by adding in the information as follows (l. 92-93):

*“Each of these signals was iteratively generated at 6 different lengths (ranging from 500 to 3000 by 500 samples).”*

There is either a problem in the numbering of the figures in the "factor analysis" part or the description is not clear and should be corrected

We note that there was indeed a mistake in the reference to the figure in the “factor analysis” section. This has since been fixed accordingly.

# Reviewer #2

We thank the reviewer for the constructive comments as well as the suggestions made to include other complexity methods that had prompted us to delve deeper into the extensive literature of complexity quantification and other available open-sourced software for running such comparative analyses.

The authors may like to cite and discuss some work by Fulcher et al. on highly comparative time series analysis (HCTSA), particularly their 2022 paper as it directly speaks to this cause.

We agree that Fulcher et al.’s work on developing the HCTSA toolbox should indeed be cited seeing its relevance to the objective of our manuscript, which is to provide researchers a guide to selecting complexity indices most pertinent to their research. However, we note also that there lies distinctions in the use of this software from that of NeuroKit, which we used to derive our analyses. We have made clarifications in the text as follows (l. 64-67):

“*Even when available and open-source, the code implementations of complexity measures are typically found scattered across different packages or scripts, or embedded within a larger goal-directed framework (e.g., HCTSA, a time-series comparison tool, Fulcher & Jones, 2017).”*

The authors never account for multifractal nonlinearity or tMF which quantifies multifractality due to nonlinearity obtained by comparison of the original spectrum with that of surrogates (Kelty-Stephen et al., 's work has been using this index as a predictor for a decade now). While I do not expect the authors to redo their modeling, this variable is something they may keep in mind for a follow-up study.

We thank the reviewer for this suggestion. In fact, we actually did manage to implement this index into NeuroKit2 but could not include it in our analysis as it is magnitudes of order slower than the other indices (since tMF recomputes MFDFA on *n*th surrogate versions of the time series). That being said, it definitely is something that should be specifically investigated in follow-up studies (for instance, one focused on multifractality assessment).