Authors' response to Referee #1

For clarity, authors' responses are inserted as blue text.

The study performs a comparative evaluation of statistical methods available in the literature using as benchmark test the quantification of fractal scaling in water quality time series with irregular sampling. While lacking technical novelty, the manuscript is written in a sober, careful manner, aptly guiding the reader throughout the key arguments, methodologies and procedures, and presenting the results in a clear and concise manner.

Response: While we thank the reader for the comments about the clarity of our presentation, and for the helpful comments the reviewer has made, we disagree with the comment about the lack of technical novelty, which is repeated elsewhere below. We respond to it there.

This being said, I raise four fundamental concerns:

1) The paper is a pure statistical exercise. It would thus benefit from a physical interpretation of the methodological structure and results: namely, discussing possible physical mechanisms responsible for the statistical signatures detected in the analysis, along with their physical consistency. For instance, whether a trend is physically sustainable and realistic in terms of system energetics, what physical mechanisms sustain the power laws detected in the data, and what physically entails the fractal behaviour. Fractals and scaling are well understood in the physical sciences but the HESS readers will be happy to learn this in the hydrological context. In doing so, the authors will be able to strengthen their arguments and diffuse concerns about whether there is any realism underlying the signatures detected in the analysis.

Response: While we appreciate the value of a physical interpretation, we will not alter the manuscript in response to this comment for the following reasons:

I. We disagree that this is a 'statistical exercise' (which we take to imply that it provides no actual illumination). As its title reflects, the manuscript focuses on the comparison of various statistical methods for quantifying fractal scaling. These methods are actually used in the literature, and we believe that it is of value to know if they actually work or not.
II. Sections 1.1 and 1.2 in the Introduction provide background information with some reference to the physical origins, and providing many citations to literature that provides the insights the reviewer is looking for. In fact, significant sections of the foundational literature on fractals and scaling were developed within the hydrology community and later picked up more widely (the Hurst exponent is an example).
III. The physical interpretations of fractal time series are many and varied, depending on the context. Those interpretations are already available in the literature, and our purpose is neither to add to them, nor to review them. Instead our purpose is to conduct benchmark tests to determine whether some widely used techniques for inferring fractal scaling are reliable or not.

2) Monte Carlo simulations can be structured and tuned for essentially any purpose and to yield any outcome, relying on wise choices made in the methodological setup and the generating system, based on the researchers' understanding or conception of its behaviour. If the methodological setup is entirely data-based, i.e. learn from some statistic or machine learning procedure derived from dataset records, there will always be a degree of case-specific empiricism that is not straightforwardly generalisable, unless there is a fundamental principle beneath. This again links to concern 1. Therefore, it is important to thoroughly provide a solid background to all the assumptions supporting the choices made in the methodological setup and operation.

Response: We believe we have provided sufficient reasoning for the assumptions adopted, but welcome comments on any specific area that is unclear or has been omitted.

3) The paper does not introduce any methodological novelty. In fact, there is a vast literature on statistics of irregularly sampled series (also known as unevenly spaced time series). Therefore, I strongly encourage the authors to look into the literature outside of hydrology, e.g. in astrophysics, neurosciences, paleoclimatology, where they will find a rich diversity of sophisticated and long-proven methods that already tackle the same problems. In doing so, the authors will necessarily tone down the false claims about novelty in new methods and frameworks, when in reality the only novelty is the application of existing methods to hydrological case studies. The key merit of the paper is essentially the comparative evaluation of well known statistical methods and their application to the hydrological sciences, namely relevant water quality issues. As such, this is a purely applied paper and should be clearly presented as such. This brings me to the fourth concern.

Response: We agree with the reviewer's comment that "the paper is essentially the comparative evaluation of well known statistical methods and their application to the hydrological sciences, namely relevant water quality issues." This is precisely stated in Section 1.3 where we define the scope of the work -- see the 2nd last paragraph in that section. But we stress that we never claimed that our work is about developing "new methods". Rather, it is stated clearly in several locations of the paper that this work is about the evaluation of existing statistical methods.

In addition, we disagree with the reviewer on the point that there are "a rich diversity of sophisticated and long-proven methods that already tackle the same problems". Many existing methods do not apply to irregularly sampled data and hence can not be used. Others have been widely used, but have not been rigorously tested. The Lomb-Scargle spectral method is well established, but has known weaknesses (as discussed in the paper and elsewhere, see Montanari et al. 1999). If the reviewer is aware of other works that solve the problem addressed by our work, we again encourage the reviewer to provide citations.

Regarding the novelty (or contribution) of this work, we are not aware of any other papers that perform a similar comparative analysis of these methods, let alone one that is tailored to the needs

of the hydrology and earth science community. The reviewer provides no evidence or citations to back up the claim that our study is not novel. If the reviewer is aware of any other such studies we would encourage the reviewer to provide them. One of us has been working on this problem for over 20 years, and in that time has seen no published work that is similar to ours.

Studies that review, compare and critically evaluate available methods are valuable contributions to the scientific literature. They are, in our opinion, useful checks on a proliferation of divergent methods that threatens to generate (at best) incomparable and (at worst) inaccurate observations of physical phenomena.

Our contribution in this regard is explicitly summarized in the last paragraph of the paper, which is copied below: "*Overall, these results provide new contributions in terms of better understanding and quantification of the proposed methods' performances for estimating the strength of fractal scaling in irregularly sampled water-quality data. In addition, the work has provided an innovative and general approach for modeling sampling irregularity in water-quality records. Moreover, this work has proposed and demonstrated a generalizable framework for data simulation (with gaps) and β estimation, which can be readily applied toward the evaluation of other methods that are not covered in this work. More generally, the findings and approaches may also be broadly applicable to irregularly sampled data in other scientific disciplines. Last but not least, we note that accurate quantification of fractal scaling in irregular water-quality time series remains an unresolved challenge for the hydrologic community and for many other disciplines that must grapple with irregular sampling.*"

4) There are no novel hydrological insights in the paper. While the statistical messages are useful (albeit not technically novel), it would be essential to bring out a substantial advance in the understanding of the hydrological and earth systems. After all, HESS is not merely a journal of applied statistics but rather one in which there should be something to be learnt in the functioning of the hydrological system.

Response: We disagree with the reviewer that this work does not provide contributions to HESS in terms of understanding of the hydrological and earth systems. As noted in Section 1.3 "Motivations and Objectives of this Work," the quantification of fractal scaling has important implications for detecting trends in water quality time series, but there is a large gap with respect to what methods are appropriate (or applicable) for quantifying fractal scaling in irregularly sampled water quality time series. By dealing with this issue, this work is highly relevant to the hydrological community.

References Cited

Montanari, A., M. S. Taqqu and V. Teverovsky, 1999. Estimating long-range dependence in the presence of periodicity: An empirical study. *Mathematical and Computer Modelling* 29:217-228, DOI: 10.1016/S0895-7177(99)00104-1.