

## Response to reviewer comments

Dear Editor and Reviewers,

Please find attached the revised version of the Manuscript "*On causality of extreme events*", which I resubmit for consideration to PeerJ.

I would like to thank all Referees for their insightful comments and suggestions, and for the effort they have clearly devoted to review this work. They allowed indeed clarifying several aspects of the Manuscript that were not correctly described, and improve its overall value.

Below I report the remarks and comments received, and indicate how I have addressed them in the revised version of the Manuscript. In the sake of clarity, changes in the Manuscript have been marked in red.

Additionally, I'd like to add a note about the following comment from the Editor:

I agree with suggestions such as these "in particular the authors must discuss the sensitivity of their method to the computation of the probabilities. For example how the choice of the binning size or other approach to compute the probabilities affect the results."

As clearly explained to the answer to the second Referee, the method does not involve any binning process, and the process to compute the probabilities associated to extreme events is almost trivial. The only two free variables,  $\tau_b$  and  $\tau_c$ , are defined in the text, and the method for setting their values discussed. I'm therefore afraid that I did not fully understand this comment. In case this was due to a poor explanation of the algorithm, I've included a new Fig. 2, with a simple graphical representation. If nevertheless these answers are not enough, I'll be glad to discuss this topic further.

### Reviewer #1

The introduction does a good job of summarizing the discussion on causality, but overall I found it very disorganized. It introduces many different concepts and ideas very quickly and seems to be a bit unstructured. I think this can be resolved by generating more structured paragraphs. Have a clear topic sentence that defines a detailed paragraph. Having 2 or 3 sentences in a paragraph potentially leaves off key information for readers that are not as well versed in this literature. For instance, when bringing up the genetic examples, it would be nice to re-link how that example fits in the content of the paragraph.

I agree with the Referee that the introduction was somehow complex, as many different concepts were mixed together. I've completely rephrased it, by progressively introducing the main ideas of the paper: how standard metrics work, how correlation can be used to avoid the problem of the time evolution, why correlation is not causality, *etc.* I hope the text is now clearer and more accessible.

Avoid putting text in brackets when it often can be left in the sentence and separated by commas. For example: Line 30. "but not necessarily" does not need to be in brackets

Corrected.

Figure 1 should also have clear indications to separate each proportion of it. Perhaps, label figures with Roman numeral to make it easier to distinguish than "left bottom, right bottom, *etc.*".

I have included Roman numerals to reference the different panels of Fig. 1, both in the figure itself, and in the main text.

It would be interesting to have seen the data tested on a real data set rather than synthetic, although I understand the limitations behind it. I look forward to seeing it used in additional contexts.

I agree with the Referee that the most interesting point is the application of the proposed metrics to real problems. This was the rationale behind the introduction of the application to EEG data – see Fig. 8. Additionally, we are working on applying this methodology to different real-world problems, and we just had a contribution accepted at an important air transport conference. Nevertheless, I also consider of utmost importance to lay the foundations of the method from a mathematical point of view, and to demonstrate with synthetic data its usefulness – as opposite to real-world data, which can always raise doubts about how they were collected, pre-processed, *etc.*, and how these aspects may affect the obtained results.

## Reviewer #2

The novel approach should be tested against other methods in solving the same problem. From the comparison of the performance of other methods compare to their, the authors can derived constructive and tangible conclusions.

I thank the Referee for this useful suggestion. I have included some new results, in which the proposed metric is compared with standard ones – see, for instance, the new Fig. 4 and Fig. 7. Indeed these results nicely support the main message of the paper: that the proposed metric is more effective than the Granger Causality or the Transfer Entropy, both in detecting causalities in static vectors, and in discriminating true from confounding relationships.

While we find the methodology presented in the paper to be well-suited for static data, the authors do not refer enough to the other methods such as data mining and machine learning, Granger causality, cointegration, or transfer entropy which are simply quickly mentioned in the introduction. I believe that the methodology proposed might benefit from the vast literature in detecting coupling in time series using these techniques. This work need to be enriched with salient literatures providing more sound evidence to the authors' comments. In particular, each approach must be detailed with appropriate citations and the authors must provides the advantages and limitations of each of them.

This issue has been tackled in several points of the manuscript, also as the result of the suggestions of the other Referees. First of all, the introduction has been completely restructured, to highlight how standard metrics (e.g. Granger causality or Transfer Entropy) work. Second, at the end of Section II, the proposed metric is compared with some similar proposals that can be found in the Literature. Finally, Figs. 4 and 7 now include an analysis of the results obtained by the proposed metric, compared with what yielded by other algorithms.

I'd also like to add a note about data mining, which has been cited by the Referee. While it is true that data mining can be used to infer relations between vectors of data, it suffers from a problem similar to correlation: the presence of a relation does not imply the existence of a causality, as the former may be due to the presence of a confounding effect. In any case, a reference to data mining has been added in the introduction.

In particular the authors must discuss the sensitivity of their method to the computation of the probabilities. For example how the choice of the binning size or other approach to compute the probabilities affect the results.

I don't fully understand this comment. On the one hand, the method does not require computing any binning – note that the word “binning” does not even appear in the text. The identification of the extreme values does not involve any binning, contrary to how it is done in the symbolic transfer entropy; but just a comparison between the individual values and the thresholds. In order to clarify this point, I've introduced a new Fig. 2, in which this process is graphically depicted. Also note that the problem of selecting the best threshold is already discussed in the text – lines 106 and following.

Additionally, the computation of the probabilities is a very simple process: the probability of encountering extreme events is just the number of extreme values, divided by the total number of available values – lines 76-78. Therefore, there is no free parameter that may alter the results obtained.

Finally, it may be possible that the Referee is referring to Fig. 3, and the distributions there used. In this case, the vectors are reconstructed by drawing random numbers with the corresponding distributions. No further processing is performed, and values are not binned.

I apologize if I may have missed the main idea behind this comment. If this is indeed the case, and these answers are not satisfactory, I'll be glad to discuss this topic further.

In the figures, the numbers are formatted according to the European style which might confusing with regard to the American style of decimal notations. This issue should be fixed with respect to the broad audience of the journal.

Corrected!

Also, we found a typo in line 55 where p1 should read p2 instead.

Corrected, thank you.

## Reviewer #3

However, as a general comment, given the results that the metric underperforms on linear datasets (fig. 2), and requires considerable samples (fig. 3), the abstract can be revised to include such caveats.

The abstract has been amended as requested. The limitations of the proposed metric have further been discussed in the Conclusions section.

Also, the meaning of cross-section and longitudinal may not be standard, so I suggest using different terminology or phrase.

Both terms are commonly used in data analysis, and especially in the biomedical field. For the sake of completeness, I have included an improved definition for both of them in the text – both in the introduction, and in Section III.

It looks like the cross-sectional observations are similar to an ensemble of multiple short processes (not necessarily arranged in time). Could the author compare his work with [a] which, though it uses transfer entropy, is able to resolve information transfer through an ensemble of independent reputations.

[a] G. Gómez-Herrero, W. Wu, K. Rütanen, M. C. Soriano, G. Pipa, and R. Vicente, Assessing coupling dynamics from an ensemble of time series, *Entropy* 17, 1958 (2015)

I thank the Referee for suggesting this paper, which I was not aware of. In spite of the authors' attempt to eliminate the need of long time series, the method proposed in Ref. [a] is still not applicable to completely static vectors of data, as the one found in genetics. I have introduced a discussion of this difference in Section 2.

Temporal information if available should only aid in inferring causality. From reading the paper and going through the examples (fig. 6), it feels as if the paper claims that even if temporal information is available, computing it using the new metric will give better results. why?

I agree with the Referee that the availability of temporal information should only be positive for the analysis. Nevertheless, nowhere in the manuscript I suggest that the cross-sectional analysis yields better results than the longitudinal one. If

one inspects Fig. 7 (Left), previously Fig. 5, it can be appreciated that for large couplings the longitudinal analysis is performing better, probably because the cross-sectional one is confounded by the presence of a strong correlation. Note that I've added a paragraph in the manuscript to highlight this behaviour.

On the other hand, Fig. 8 does not suggest this conclusion. The fact that the cross-sectional analysis detects slightly more links than the longitudinal one does not mean that the former is better – as we have no mean to confirm that these causalities have a real biological basis. The important issue is that both of them yield similar sets of links, and that therefore they are detecting the same phenomenon. Although this is out of the scope of this contribution, there are plenty of papers that have been published on this topic, *i.e.* whether links in functional networks are significant or not from a biological point of view, and how to filter irrelevant links in brain functional networks.

This also relates to the thresholds for selecting extreme events. In the longitudinal case, it appears that the approach is no different from reference [15]. It would be nice to see a comparison, in words, or through examples (I believe the authors for that paper have also provided a MATLAB code for executing the same)

While the method proposed in Ref. 15 presents several similarities, the selection of the events of interest is different. Ref. 15 considers those events that are local maxima, independently of their amplitude; this implies that some values, which are very small local maxima, would not be considered in the proposed metric. In this sense, the method here presented considers only those events that are “not normal” or “not expected”. I've included a paragraph at the end of Section II to clarify this issue.

The problem of data size is critical here, especially for thresholds that may be high, or there is too much noise in the data. Kindly discuss that aspect in your conclusion

I agree with the Referee: the conclusions section was too optimist, and did not correctly highlight the limitations of the proposed method. The following paragraph has been added to amend this:

In spite of the advantages that the proposed metric presents, and that have been described throughout the text, two limitations have to be highlighted. First, the reduced sensitivity of the metric to linear causality relationships, and in the analysis of data without long tail distributions, *i.e.* without clear extreme events - see Fig. 2 for further details. Second, the need of large quantities of data to reach statistically significant results (Fig. 5).

Minor fixes: p1 in line 55 should be p2;  
Fig. 2 caption should say 10,000 realizations;

Both errors have been corrected.

arrows not showing in Fig. 6 for the directed links;

Arrows were present, but were too small (specially when the paper was printed). I have corrected this.

kindly define 'weight' of a node

Corrected.