Nonlinear Processes
in Geophysics
Discussions

Open Access

EGU

NPGD

Interactive
comment

Interactive comment on "Recurrence analysis of extreme event like data" *by* Abhirup Banerjee et al.

Carsten Brandt (Referee)

carsten.brandt@uni-bremen.de

Received and published: 22 November 2020

General comments: The authors present a paper with a distance measure and its modification (the latter developed by the authors) to compare windows out of non-equally spaced time series of extreme events using recurrence quantification analysis (RQA). The paper is in general well written and fulfills from my point of view standards on scientific publications. I have several general and technical comments that should be solved/ reacted to before final publication.

Specific comments: - I must confess (and this is almost a general comment) that it was at least in first reading difficult for me to understand the edit distance (or its modification). Maybe it is good to make the basic concept clearer: in the paragraph starting at line 75, I propose to write after the first sentence that 1) the aim is to determine a distance measure between each segment and 2) that the transformation is a means to

an end, that is, the aim is creating a cost of the transformation as the distance measure. The aim is not the transformation itself. For me, this clarification/underlining is essential.

- Then, I could not find the term "edit distance" in the original paper from Victor and Purpura 1997. I think it would be good to make a remark for readers wanting to dive into the origins. - Line 254: did you really use eq. (4)? That is the Heaviside function, and you claim in chapter 3 that you use the logistic function. - Paragraph from line 296 on: I don't understand this completely. You apply a threshold on the data of daily discharge? You determine the percentile per year? A few events - you mean 3 to 4 (you wrote 3/4 = three quarters)? - Line 303 recurrence threshold: this parameter is quite crucial in RQA. You set it to 8% of the distance distribution. Is there any specific reason for this value (in the Poisson events you used 10%), was there a reasoning or trials to get to this value? Then please include these in the paper. If it was "just chosen", then write this as well please. Also important: I am missing a discussion of the difference of your modified edit distance and the original one on the Mississippi flood events, i.e., comparing figure 12 and 13 with figure 14. You mention that you computed standard edit distance and figure 14 in line 304, but I am missing a discussion about the outcome of this, or the comparison with your modified distance. A final remark: what do you think about modifying the title or amending it by "...using a modified distance measure"? This is from my point of view the main and important contribution you make. But I leave this up to you, the authors.

Technical corrections (minor contents or grammar): Line 60/61: I suggest writing either "A flood time series shows..." or "Flood time series show..." Line 67: To be strict, "RP" is the abbreviation of "recurrence plot"; "recurrence plot analysis" (RPA) is also sometimes used, but "recurrence quantification analysis" (RQA) is much more common. Line 82: "where alpha and beta... at times ta(alpha)" - isn't there "and tb(beta" missing? The definition of determinism DET has been done with a minimum line length of 2, as one can see in eq. (8). Please consider to write down "with a minimum line

length of 2" or similar to make this clear. Other line length are also used in the literature for RQA. Figure 6 caption "Deltat is the gap between two events in kept up to the maximum value 6." - I do not understand this sentence. Line 221: The symbol OL for the overlapping range might be confusing, especially in an equation in which on the right hand side (LdeltaT) L and deltaT are multiplied (I was wondering in the first place whether O shall be multiplied by L on the left hand side). Consider using an index O_L (sorry for not being able to display it here). Line 222: I do not understand the last sentence: the number of shared data points is L, is it not? This can be expressed of course as OL/deltaT. omega times deltaT minus OL makes no sense anyway, because the first summand has unit time squared and the second summand has unit time. Line 223: I suggest removing the "and". Line 234: "...different signal"s - "s" is missing Line 306 a dot after "Fig" is missing Line 306 and 3ß7: you wrote Fig. 13d twice. I guess the first shall be a different one?

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2020-41, 2020.