# Reply to Erwin Zehe on "Analysis and Modelling of a 9.3 kyr Palaeoflood Record: Correlations, Clustering and Cycles" by Witt *et al.* (2017)

We thank Erwin Zehe (EZ) for his thoughtful comments on our manuscript hess-2016-470 (Witt *et al.*, 2017). In this revision of our manuscript, we address the reviewer comments as given below, and believe that the result will be an improved manuscript.

**EZ MAJOR COMMENT 1**. "The presentation of the manuscript is in many parts a little too wordy. E.g. there is no need to explain the concept of correlation in such a depth in the intro, as the auto correlation within a time series is well known. Unfortunately the authors mix the presentation of the statistical methods and results, this makes the manuscript difficult to follow and to evaluate the partly very interesting findings. I suggest to rearrange the presentation in a more old school manner and to present methods, results and discussion of those in separate sections."

**Author reply**

(i) **Wordiness and detail of explanation.** We agree that this is a long paper (but we hope well organized, so people can dip in and out of certain sections), and that in places the manuscript tends towards facts/information that might be well known by certain parts of the community. We will relook at the entire manuscript, editing words out where possible, but also maintaining a level of readability and accessibility by those who do not have a background in statistical methods related to correlations, clustering, and cyclicity. We will also examine whether parts of the manuscript should be moved towards an appendix.

(ii) **Mixture of statistical methods and results. Rearrange manuscript to be methods, results, discussion**. We propose to take all statistics 'methodology' sections in Section 3, pull these out into a new section. We would then have as outline the following:

1.0 Introduction
2.0 Data
3.0 Definitions and methods for the analysis of event series
    3.1 Poisson process
    3.2 Autocorrelation
    3.3 Interevent occurrence time probability distributions
    3.4 Long range correlations using detrended fluctuation analysis and power-spectral analysis
4.0 Results of analysis of the palaeoflood time series
    4.1 Number of palaeofloods per year, decade and century, compared to a Poisson process
    4.2 Temporal correlations using autocorrelation
    4.3 Clustering using interevent occurrence times
    4.4 Quantifying long-range correlations using DFA and power-spectral analysis
    4.5 Cyclicity and fluctuations in the 9.3 kyr palaeoflood record
    4.6 Summary of clustering, correlation and cyclicity of the 9.3 kyr Piànico-Sèllere palaeoflood record
5.0 Creating a peaks over threshold model to capture correlations and clustering of the observational data
    *[Sections 5.1, 5.2, … the same as 4.1, 4.2, … previously]*
6.0 Summary

**EZ MAJOR COMMENT 2** "I wonder why the authors do not employ variograms/semivariances to estimate the spatial correlation time. This will better separate the un-correlated from the correlated variability in their records. In this context I wonder, how much of the increase in correlation times at the hifher aggregation levels may be attributed to well-known scaling effects arising form to the aggregation process itself. The range of variogram estimated from data points increased when these are aggregated to blocks (or time intervals here) while nugget increases and sill decreases – this is well known from variogram regularization. This effect could be easily determined using data from your Poisson null model."

**Author reply:**

The reviewer has suggested that we use semivariograms as a model for finite range correlations or short-range persistence (i.e., a temporal correlation time), where at a given lag or less in our time series, correlations are observed, but that at that lag or greater, correlations are not observed, or the correlations fall below a statistical threshold. The reviewer further then suggests that any increase in the correlation time (based on short-range persistence), might be due to the scaling effects from the aggregation process itself.

In our manuscript, we have used autocorrelation as a way of first exploring temporal correlations, and we found up to 200 years lags there were correlations (see also reply to Comment #3). In other words, we did not identify (up to 200 years) a temporal correlation time. We agree that we should go farther in our temporal lags, and will do so in the next version of the manuscript.

Autocorrelation is linked to the semivariogram, but includes the mean of the time series. We will also do a semivariogram analysis to see if a short-range correlation model might also be appropriate within statistical significance.

We also note that in our manuscript after our initial explorations of the time series using autocorrelation, we decided to apply a model of long-range correlations (i.e., one that does not have a temporal correlation time). Certainly, semivariograms can be used a long-range persistence model. However, instead of semivariogram analysis we applied Detrended Fluctuation Analysis (DFA) and Power-Spectral analysis (PS) for quantifying the long-range persistence strength of our palaeoflood and synthetic time series. Both semi-variogram analysis and DFA operate in the time domain, and power-spectral analysis in the frequency domain. In research we did four years ago and published as a review paper (Witt & Malamud, 2013) we examined systematically, using 17,000 synthetic benchmark series, the types of errors when evaluating long-range persistence using these three techniques. All three are appropriate measures for long-range persistence; however, the systematic and random errors (biases and standard errors) of DFA and PS analysis are significantly smaller (particularly for time series with non-Gaussian one-point probability distribution, and with very few values in the time series) compared to semi-variogram analysis, and both DFA and PS analysis are appropriate for a much broader range of long-range persistence strengths compared to semi-variogram analysis. We put our research into the context of the wider literature which has found similar results. We therefore in this paper use DFA and power-spectral analysis, due to it being much more robust than semivariograms as a long-range model. We will add to our manuscript a short explanation why we have chosen DFA and power-spectral analysis (compared to semivariograms) for examining long-range persistence.

**EZ MAJOR COMMENT 3** "I appreciate that one prefers the use of power law functions with infinite correlation length – but what is the physical meaning of this?"

**Author reply:** We believe (as this refers to correlation length) this comment refers to our presentation of the auto-correlation function applied to our palaeoflood time series (per year, per decade, per century), and the auto-correlation decaying as an inverse power-law with respect to the temporal lag. We will take "correlation length" here as meaning the lag at which values are separated and still statistically correlated. One could also consider power-spectral analysis and DFA, in terms of an equivalent correlation length for which values are still, on average, statistically correlated. Self-affine long-range memory is defined to be a time series where 'all' values in the time series are on average statistically correlated to one another. In the case of long-range memory the correlation length is defined to be infinite, due to the power-law decay of the autocorrelation function as a function of the temporal lag. The first physical significance of this 'infinite' correlation length is that 'all' values are statistical correlated with one another in the time series. A second physical significance, is that autocorrelation functions with a power-law decay indicate a self-affine geometry in the original time series examined, i.e., the considered object can be explained as a stochastic fractal. In the case of our palaeoflood series this means that the floods are organised in time as the points of a Cantor dust organised in space. We will add some sentences about this aspect to our manuscript.

**EZ MINOR COMMENT 1**: "The authors use the presence of gravel enriched strata as indicator for channelized inflow due to a flood into the lake. This seems plausible as those events might be associated with bedload transport and sedimentation into the lake. However, there is no linear relation between flow velocity and bed load transport capacity. The literature is full of empirical formulas for this. This implies that the thickness of these strata is not linearly related to pealk discharge (flow velocities). The authors might consider this as an additional argument against the use of layer thickness as proxy for flood peak discharge."
**Author reply**: We agree that strata (layer) thickness is not an indicator for flood magnitude. Note that in our detrital layers, which represent the flood events, the detrital layers themselves are mainly constituted of dolomite from the catchment, i.e. fine grained sediment with no gravel. We will add (in yellow below) a short statement to our original manuscript to clarify.

"Although detrital layers, which are mainly constituted of dolomite, indicate the occurrence (year) of an extreme precipitation event (flood), the thickness is not taken as representative of the magnitude of the event, as the thickness of a layer may depend on the distance between the source of the detritus and the studied section, the amount of detritus available at the time of the extreme precipitation and weaker vs. stronger rainfall during the extreme precipitation events. In addition, over the 9336 yr period we investigate, the source of given storms might result in different thickness signatures. We also note that although we can identify when a flood is thought to have occurred, this does not preclude the possibility that other floods occurred but were not identified as a detrital layer. …"

**EZ MINOR COMMENT 2**: "Why not showing a long time series of annual flood maxima, they also exhibit clustering of extremes and long time periods – this example is maybe closer to the object of desire?"
**Author reply:** If we understand what is being suggested, is that we replace parts C and D of Figure 1 in our manuscript (Atlantic Storm Maximum Wind speed) with a similar figure related to floods. We probably would not use an annual flood maxima time series, as one year might have a flood that occurs that is less than three 'independent' floods in the next year. Rather, we would use previous techniques that we have worked with to do partial duration flood series, and then choose the largest 100 for the period considered. We agree this could make the Figure 1 more relevant to our paper.

**EZ MINOR COMMENT 3**: "Could you also infer on clustering of intensities above a threshold using indicator auto correlations?"
**Author reply:** We understand this question to ask what would happen if we applied the autocorrelation function to our palaeoflood time series (per year, per decade, per century) to all values above a given threshold, in terms of learning more about the clustering (vs. correlation) behaviour of the time series. The values above the threshold could either (time series 1) retain their value, and all values below the threshold set to '0', or (time series 2) all values above the threshold set to a given value (e.g., '1') (and all values below the threshold '0').

Currently we apply autocorrelation analysis to the number of floods per year, per decade, and per century, but without a threshold. The problem we have with our palaeoflood time series, is that rather a time series that is 'just' stochastic, there is also the presence of a long-term cyclicity. In case of weakly-stationary data such as a fractional noise (no underlining long term cycle), we could use the decay rate of the autocorrelation function for quantifying correlations (and also examine values above the threshold to infer clustering behaviour).

However, because of the presence of a long-term cyclicity, the autocorrelation function will be strongly impacted, and we therefore use Detrended Fluctuation Analysis (DFA) and power-spectral analysis, which are more appropriate for fractional noises with cyclicity. For example, DFA is detrending segments of the aggregated time series.

In our manuscript, to examine clustering, we use the one point probability distribution (and in particular the shape parameter of the Weibull distribution) of the time intervals between floods as an indicator for clustering. For our specific model, small shape parameters are equivalent to a large strength of long-range correlations and vice versa (see Figure 12 a,b and c) and thus strong clusters appear in case of strong long-range correlations.

We will add in a sentence to our manuscript to refer to the fact why we do not examine autocorrelation as an indicator of clustering.

**EZ MINOR COMMENT 4**: "I appreciate the authors effort to provide a reproducible paper, but the explanation of the Poisson process is too detailed, this is text book knowledge. The parameter lambda is the mean value of floods / time interval, which depends on the aggregation level. As far as I know the Poisson process produces cluster data, if lambda is larger than 1?"
**Author reply:**
(a) We will shorten the introduction of Poisson processes in our introduction, but would still like enough detail that a non-expert will have enough to follow the flow of our manuscript.
(b) The Poisson process as introduced in our original manuscript page 8, does not lead to temporal clustering of the realisation (in time) of a Poisson process, for any lambda > 0 (Cox and Lewis, 1978). We also note that for lambda > 1 the one-point probability distribution of events per time unit is one humped and has a defined mode.

**EZ MINOR COMMENT 5**: "Page 9 line 5: In case the flood do always occur in the first three years of the decade the average inter event period is not 1 year but, 2/3 *1 + 1/3*7= 3 years?"
**Author reply**: We thank the reviewer for noting this error. We will change
"For example, if in a given decade, we have three varves each with one flood, Eq. (2) will be different if the three floods are the first three years of the decade (i.e., mean $\Delta = 1$ yr) vs. a flood in years 1, 5, and 9 of the decade (i.e., mean $\Delta = 4$ yr)."
to read
"For example, if in a given decade, we have three varves each with one flood, Eq. (2) will be different if the three years with floods are in years 1, 2, 3 ($\Delta_1 = \Delta_2 = 1$ yr, $\Delta_3 \geq 8$ yr because the 4$^{th}$ flood will be in a subsequent decade) vs. one flood each in years 1, 5, and 9 of the decade (i.e., $\Delta_1 = \Delta_2 = 4$ yr, $\Delta_3 \geq 2$ yr)."

## References Cited

Cox, D.R., and Lewis, P.A.W.: *The Statistical Analysis of Series of Events*, Chapman and Hall, London, 1978.
Witt, A., and Malamud, B.D.: Quantification of long-range persistence in geophysical time series: Conventional and benchmark-based improvement techniques. *Surveys in Geophysics* 34 541–651, 2013.