Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2018-102-RC1, 2018

Natural Hazards and Earth System Sciences Discussions

Open Access

NHESSD

Interactive comment on "On the relevance of extremal dependence for spatial statistical modelling of natural hazards" *by* Laura C. Dawkins and David B. Stephenson

Anonymous Referee #1

Received and published: 9 May 2018

General comments

The aim of this paper is to examine the extremal dependence of windstorm footprint data for pairs of sites across Europe - to assess whether there a non-zero probability that there will be an unusually large wind speed at one site given that such an event has occurred at a second site? Understanding extremal dependence is vital for accurate prediction of joint loss functions for the sites in question. Joint losses are themselves key for natural disasters which affect multiple locations across a large region, in which case the combined loss over this region is of most importance. Traditionally, statistical models for extremal dependence make one of two assumptions: either the data

Printer-friendly version

Discussion paper

are asymptotically dependent (eg. max-stable models) or they are asymptotically independent. If the wrong class of extremal dependence is assumed and a model is subsequently fitted then the model predictions cannot be trusted; for example, if it is incorrectly assumed that the largest values at the two sites occur concurrently with non-zero probability then the joint loss at the two sites may be over-estimated.

In this paper, a variety of model-based estimates of the extremal dependence for wind speeds are compared to the equivalent empirical estimate and a copula-based method based on the Gaussian copula if found to give the best fit, although this data-based justification is illustrated by only a small number of sites (London-Amsterdam and London-Madrid). A justification based on a physical model for turbulent wind-speeds is also postulated. Finally the impact of the choice of extremal dependence class on the estimate of a particular definition of the conditional loss function is illustrated.

The paper is mostly well-written, although there are a number of grammatical/typographical errors which should be addressed (see below). The logical flow is clear, and results are well labelled and clearly discussed in the text.

Specific comments

— Comparison of Gumbel, Gaussian and power law copulas to the empirical estimate is shown only for a very few pairs of sites. At the end of Section 3.3 it is suggested that the results found at these sites are representative of results for other pairs of locations. How many other pairs of locations were tested? How confident are you that asymptotic independence is the dominant dependence structure across all pairs of sites? Have you considered ways in which you could formally test this over all pairs of sites?

— A thought on the justification for asymptotic independence based on the model for wind gust speeds. It is assumed in the physical model that each of the Cartesian components of wind gusts speeds $U(s)$ and $V(s)$ follows an independent Gaussian spatial process. By properties of the multivariate normal distribution, each the vector of each component $\mathbf{U} = (U(s_1), U(S_2))$ and $\mathbf{V} = (V(s_1), V(S_2))$ at any two sites s_1 and s_2 follows a bivariate normal distribution and consequently the components of each of $\mathbf{U}$ and $\mathbf{V}$ are asymptotically independent. However the Gaussian process assumption is just another modelling assumption, so the question is how accurate is it, ie. how strong is the evidence in favour of the Gaussian assumption? If actually the speed components followed some other process that had asymptotically dependence bivariate margins then the conclusion would be very different. On a related point, it is not entirely obvious where the equation for χ_{max} and the expression for $\Pr[u_1^2 > t, u_2^2 > t]$ comes from. Although these expressions are correct, Would it be possible to put a derivation in an appendix?

— The second half of Section 4 (p21 onwards) needs some re-working to clarify the points that are being made. For example, it is not clear why we might be interested in the first and second moments derived after line 326; also these are not al moments but expectation (first moment), conditional expectation (conditional first moment) and variance (function of second and first moment). Why does the expected loss not depend on the extremal dependence between the two sites (line 327)? Could you clarify what is being illustrated (line 331) in this final part of the section? Although the Gumbel/Gaussian discrepancy is clear in Figure 7, it might be informative to also look at spatial plots of the *differences* between the empirical and model-based estimates of each of $\chi(u)$ and $\bar{\chi}(u)$ to see if there is any spatial clustering in these differences, ie. do the models represent the empirical behaviour better for some regions/distances/directions than others? Finally, how does Figure 8 change with the choice of p, and does your model enable you to look at the sizes of losses at the two sites rather than just the probability of a loss jointly occurring at each site?

Technical corrections

- Throughout apposite/apposing should be opposite/opposing.
- Lines 71-76: these sentences are not entirely clear. On line 73 please clarify what the ‘parametric representations’ are representations of. May also be clearer to split the sentence on lines 73–76 into two sentences, the first to discuss what will be done (ie. two copula dependence models fitted) and the second to explain why there two copulas were chosen.
- line 94: please could you clarify exactly which ‘statistical property’ is meant here. Something like ‘the extremal dependence class estimated from the data’.
- line 135: based *on*
- line 137/138: doesn’t quite get across the message that sites separated by different distances/directions will have different levels of dependence. Also, why is this likely to be the case? Have you looked at dependence as a function of distance/direction?
- line 165: RHS sign in the inequality should be reversed.
- line 167: could append this paragraph to the previous one.
- line 171: no need to state ‘*empirical* exploration’ as it is made clear later in the sentence that the estimate is an empirical one.
- line 177-180: think this sentence can be removed as it doesn’t quite fit here and is better covered in Section 3.3
- line 182: clarify that rarity of extreme events in historical data is not specific to this particular data set.

- line 186: not sure that ‘model the asymptotic limit’ is quite right, maybe ‘predict’ instead of ‘model’. The model can only reflect the (sub-asymptotic) evidence in the data and the only extra information used in obtaining an estimate of the asymptotic limit is the assumption that the model fitted to sub-asymptotic data can be extrapolated to make predictions on higher (asymptotic) levels.
- Notation: switching from X and Y to Z_1 and Z_2 mixes two different ways of distinguishing sites (different letters *v.* subscripts). Could change (X, Y) to (X_1, X_2) or (Y_1, Y_2), *or* change (Z_1, Z_2) to (Z_X, Z_Y).
- line 196: ‘asymptotic’ $\rightarrow$ ‘extremal’, as asymptotic dependence is a particular class of extremal dependence.
- line 198: expression for $\bar{\chi}(0)$ has an excess bracket.
- line 201: ‘model’ missing after Ledford and Tawn (1996)
- lines 218-220: this sentence would be clearer split into two. First describing the models for (X, Y) and then describing how model-based predictions of $\chi(p)$ and $\bar{\chi}(p)$ are obtained from the models and are compared to the empirical estimate in Figure 3.
- Figure 6 caption ms^{-1}.