Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-117-AC1, 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "Does the AO index have predictive power regarding extreme cold temperatures in Europe?" *by* Tamás Bódai and Torben Schmith

Tamás Bódai and Torben Schmith

bodai@pusan.ac.kr

Received and published: 3 June 2020

We would like to thank the referee for their time and effort with reading our manuscript and for providing feedback. We quote the reviewer's points and comment on them below (starting with an indentation).

The study is focused on modeling cold extremes by using EVT tools and covariates such as AO. The objective is to propose an approach to be used for long-term weather forecasting, although the authors tested and used just observational data. From a methodological point of view, there is nothing new except for the idea of measuring the predictive power by using the 'sharpness' criterion introduced by Gneiting in 2007.

Thus, results are somehow expected.

In our opinion, the latter is a mistaken conclusion. Employing well-established methods does not mean that the results can be expected. If anything, the methods need to be employed and the work done because we cannot foresee the results of it. As an example, a Runge-Kutta integrator scheme is most common to integrate ODE systems. Yet, when the solution is chaotic, it cannot be anticipated without actually solving the system numerically. Or, engineering consultancy most often does not need to come up with new methods to make an analysis. Yet, firms are paying very substantial money for the analysis, simply because its results cannot be foreseen.

My main concerns are on the applied tests and on the readability of the manuscript. As for the former one, I suggest using EDF tests,

We would like to note, to start with, that in our original analysis we ended up with using the Lilliefors test, because the Matlab help file for the Kolmogorov-Smirnov test, which is an EDF-based test, commented that "The result is not accurate if CDF is estimated from the data. To test x against the normal, lognormal, extreme value, Weibull, or exponential distribution without specifying distribution parameters, use lillietest instead."

https://ch.mathworks.com/help/stats/kstest.html?searchHighlight=kstest&s_tid=doc_srchtitle#namevaluepairarguments.

We did not find a similar caution about the Anderson-Darling test (AD), another EDF-based test, and, so, following the Referee's suggestion, we have evaluated the p-values in the same fashion as in the case of the original two tests used. We used the Matlab function 'adtest':

https://ch.mathworks.com/help/stats/adtest.html

The results for the four models can be seen in the diagrams attached. It turns out that the AD is the most lenient out of the three tests pursued up to now. In particular, for models M1, M2, M4, hardly any gridpoint sees the rejection of the GEVD form when using AD. (Only in the case of M3 do we see many gridpoints of rejection, which is similar to the results with the original two tests, not shown in the paper.) The sensible approach is a conservative one: if one of two tests rejects the GEVD, then we cannot go with that model even if the other one could not reject it. Therefore, aposteriori, the AD does not seem to be the one that we should rely on.

while for the latter one I think the manuscript needs to be fully restructured. Results, methodological details, technical discussion, etc. are all mixed through the entire manuscript. Just to make an example, the introduction should give more emphasis on the objectives of the study, what has been already done by others, etc. rather than focusing immediately on technical details.

We think that the paper has an adequate structure, which is not really out of line with the mainstream. In our opinion, it is not very constructive to try to precisely separate passages of text relating to methodology, results, literature review, etc. Methodological, technical descriptions and comments appear to us to be appropriate to make in the Introduction or the section conveying the main results.

We intended to write a short paper, and we believe that we did set out our main aims in the Introduction and also gave there a backdrop of it including references to relevant past work. We do think that the technicality of mentioning nonstationary extreme value statistics in the Introduction is necessary in order to be able to say what our work is really concerned with.

Nevertheless, we would be willing to streamline the Introduction and perhaps implement some sensible restructuring.

Some Specific Comments

20 Well, this is true if there is convergence

Convergence is taking place asymptotically. One of the authors worked together with

a senior colleague who kept speaking about “convergence setting in” or “convergence taking place”. But in fact, it never does. Therefore, a GEVD model estimated from data that does not actually conform precisely to that model (even if a statistical test cannot reject it – a point made also by the referee)

http://centaur.reading.ac.uk/73386/

https://www.cambridge.org/core/books/nonlinear-and-stochastic-climate-dynamics/extreme-value-analysis-in-dynamical-systems-two-case-studies/CE02FCFC8BF95D735ECA2298D1CA8E89

will take a range of values to perform well. That is, while extreme value theory is meant to facilitate an extrapolation based on universality, such an extrapolation has its practical limits. In fact, the original pioneering work by Fisher and Tippett (1928), which we cited at this point of the text, is concerned with the problem of convergence, and therefore maybe we don’t need to prompt a caveat explicitly. This topic is nowadays called Penultimate Extreme value theory (P-EVT).

25 I do not see why the parameters should change from one winter to the other, especially the shape. Maybe some more words on this would help readers to understand what the authors’ view is.

This is a very useful point. Thank you. On lines 93-97 we wrote that “different Poincaré sectioning surfaces associated with different fixed values of the co-variate should yield sections of the attractor whose geometries are different, yet, their fractal dimensions (Hall and Davies, 1995) and so the shape parameter (Holland et al., 2012; Lucarini et al., 2014; Bódai, 2017) should be the same. Nevertheless, a considerable variation of finite-size estimates is possible, even if the pre-asymptotic non-GEV characteristics do not show up (Bódai, 2017; Gálfi et al., 2017)”. Given that in the formula above line 25 Z_t is meant to be a monthly maximum, i.e., the block size is finite, we could not possibly mean the asymptotic limit. In terms of P-EVT, the shape parameter does not need to be the same for all values of the covariate – even if it actually is in the asymptotic limit, as we explained. Nevertheless, as we write on l 74 and 212, the shape parameter probably doesn't vary much with the co-variate, and, so, a lot of data would be needed to possibly reject the hypothesis of a constant shape parameter. We would leave $\xi(t)$ (instead of writing ξ denoting a constant) in the formula in question, and add that we do not restrict generality there, and, perhaps, we could refer forward to Sec. 2.1 where we do in fact settle with the constant ξ model.

60 There is no need of the first paragraph

Ok.

65 It is not that evident from Fig. 1

This is actually true. Thank you very much for pointing this problem out. We realise now that we had been primed by our preliminary figures (not shown in the paper) based on the DJF-mean AO index and corresponding mean daily minimum temperatures; one for a location in Northern Europe and one for Kiev. We attach now a diagram like that in our Fig. 1a, but with one new data line for the DJF-mean AO vs the corresponding mean daily temperature minimum. (The data points are linked for visualisation reasons only; they do not indicate chronology.) We can see that there is a correlation also for positive values of the AO index, unlike we wrote in the paper. Therefore, we propose that we replace Fig. 1a with this new figure, and modify the text on l 67-71 as follows.

"This seems to apply also to the location of Kiev, as indicated by the solid thick data line in the diagram in Fig. 1 (a). In the same diagram, a scatter plot of the daily minimum temperature in Kiev vs the monthly mean AO index (AO) provides some view also on intraseasonal variability including extremes. Our observation is that the characterisation of the AO-dependence of extremes – unlike that of the seasonal means – will have to include some nonlinearity. In particular, the overall dependence of e.g. the location parameter $\mu(AO)$ of a "nonstationary GEV distribution" is seemingly nonmonotonic."

70 Terms of the equations should be explained.

We are not sure which terms are not clear. Just before these equations, we are motivating the adoption of a model quadratic in the variable AO. That leaves all other symbols to stand for constants. We have not excused ourselves that we recycle the symbols, μ, σ, ξ appearing in the form of the GEVD in the Introduction (as we do not recycle), and, therefore, the meaning of these symbols should be clear. The location parameter μ was explicitly mentioned just above on l 70; and the scale parameter σ, key for our analysis, was discussed in the Introduction. Furthermore, in the revision, we would make the forward reference in the Introduction about the shape parameter ξ, according to the previous point made by the referee.

85 Again, the GEV is correct as soon as there is convergence, that is actually very difficult to test as requires very large samples

We agree about these two issues with convergence. That is, firstly, with finite block size the distribution of block maxima is nearly never exactly a GEVD. It is a GEVD only if the parent distribution itself is a GEVD – thanks to the max-stable property of the GEVD. Please note that this is precisely what we meant by the bracketed text on l 83. Secondly, the finite-size non-GEVD character might require a lot of data to detect. This was one of the key points of

https://www.cambridge.org/core/books/nonlinear-and-stochastic-climate-dynamics/extreme-value-analysis-in-dynamical-systems-two-case-studies/CE02FCFC8BF95D735ECA2298D1CA8E89

Indeed, we might know from theory that the data cannot conform to the GEVD, whether it is because of the finite block size or because of the fractality of the probability measure, yet we might not be able to detect it if there is not enough data and the irregularity is not so strong.

Not independent from an earlier point of the referee, we would revise the text here, extending our sentence on l95-97, as follows.

"Nevertheless, considerable variation of finite-size estimates is possible, even if the pre-asymptotic non-GEV characteristics do not show up (Bódai, 2017; Gálfi et al., 2017), which dictates a generically non-constant shape parameter, as expressed in Eq. (1) [equation number to be introduced]. Yet, we adopt a constant shape parameter model (1) [equation number will change], as it is known that the estimation of the shape parameter of a stationary process is rather sensitive to data shortage (Friederichs and Thorarinsdóttir, 2012)."

The last clause in the above is a copy from l 212-213. We are willing to take care of the details of a good restructuring and editing in the revision.

109 This is a questionable assumption if we say that external forcing and not natural variability is causing the long-term trend (or a combination of the two).

Please excuse us, we are unsure what you mean. We cannot see why the external forcing scenario (l 106-107) should jeopardize our assumption more than the decadal-scale internal variability (l 114-). Indeed, this is just an assumption, something that we have not proven. But we have two important points. One is that the assumption is rather intuitive. Just because some values of the co-variate occur more frequently in different time periods (under different "climates"), the same value might imply very similar (conditional) distributions of the observable of interest in these different periods. For example, maybe some warm Decembers will occur more frequently in the future, but even in the past the probability that the monthly maximum temperature will exceed a threshold, given the same particular value for the monthly mean, might be largely unchanged. The second point is about robustness, that you address yourself by your next point.

Furthermore, I do not see how you can be confident with the robustness. I think it is fairer to just say there is an assumption that may be violated.

We wrote on l 111-113 that if our assumption – regarding the seasonal nonstationarity – was inaccurate, that would imply poorer predictive power in terms of a larger

scale parameter. That is, we will not see a false gain of predictive power when our assumption is inaccurate, but rather the opposite. We see it as a good conservative approach.

We should make it explicit indeed that this is generic, applying also to the long-term nonstationarity. We propose to append the paragraph on l 119 by the following sentence. "Furthermore, the conservative nature of the estimate applies generically, including the climate-change-type nonstationarity."

155 I suggest to used EDF-tests instead of these two as the focus is on GEV family.

We will make a note of the results of the Anderson-Darling test as per the above, without including figures. Again, we find that it is too lenient compared to the original two tests. Otherwise, we see no fundamental reason why EDF-tests should better suit the GEVD form. We note that the Matlab implementation of e.g. the Lilliefors test

https://www.mathworks.com/help/stats/lillietest.html

takes the Gumbel distribution as one of those that can be selected by the user.

160 I suggest to rephrase this paragraph.

We will make an attempt to do this. Perhaps it is too concise as it is now.

Figure 2 is not very informative Figures 7 and 8 Maybe as SOM?

Fig. 2a shows the large variability in winter relative to the other seasons, and that it looks as though the "outgrowth from a harmonic function is towards negative temperatures". Furthermore, both Fig. 2a and b convey information about the amount of data available. It is a key theme of our paper that data scarcity needs to be handled somehow.

We suppose that "SOM" refers to a supplementary material. Ours is a short paper and it would feel not appropriate for us to have a supplementary material with just two figures. Furthermore, we think it is an important enough point that the models do not apply well to all locations. The referee himself/herself regarded the perceived issue with the statistical tests as one of the two of their main concerns.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-117, 2020.

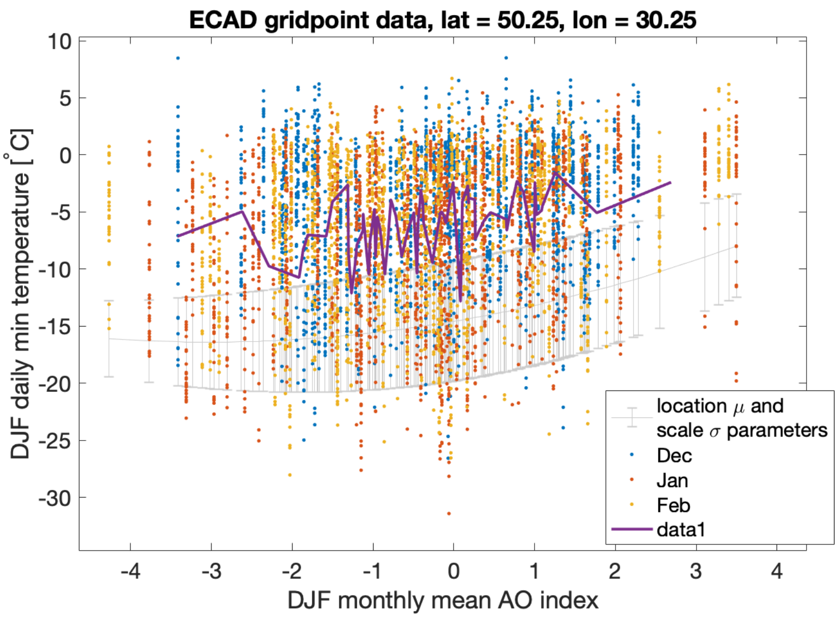

Fig. 1.

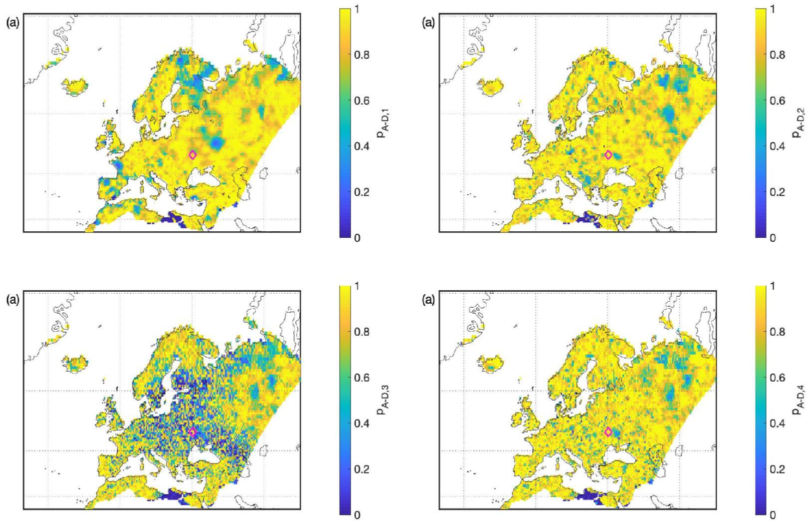

Fig. 2.