Journal: Natural Hazards and Earth System Sciences (NHESS)
Title: A hazard model of subfreezing temperatures in the United Kingdom using vine copulas
Author(s): Symeon Koumoutsaris
MS No.: nhess-2017-389
MS Type: Research article

General comment

The author assesses the return period of extreme cold winter across UK, based on the Air Freezing Index (AFI). The first part of the paper is based on the estimation of AFI return periods for single spatial grid boxes; in the second part the author makes use of pair-copula constructions for estimating the return period of the average AFI (or weighted AFI based on local population in grid boxes) over larger regions (i.e. UK, South England, North England & Northern Ireland, and Scotland). The latter is useful, e.g., for insurance portfolio loss analysis which requires hazard calculation across large regions. The author concludes saying that according to the model, the extreme winter 1962/63 has a return period of ~89 years, although this result may be affected by uncertainties due to the shortness of the data used for the model calibration, and to the background climate change trend. The work looks conceptually interesting to me, although the structure of the presentation, the figure descriptions, and the explanation of procedure details, would need improvements and more attention. Above all, I have the following comment.

Main Comment

The large (non-estimated) uncertainties due to the shortness of the data can strongly compromise the interpretation of the results. Although Vine Copula is a flexible mathematical tool for modelling multivariate probability density functions (pdfs), in general the sample size used for fitting multivariate pdfs should be large enough. The sample size needed for a reasonable fit of the pdf increases very fast (~exponentially) with the dimension of the pdf. In this study 51 observations are used to fit a 170-dimensional pdf (I hope I have not misunderstood...). I believe that this is a very small size for such a fit. Moreover, eventual serial correlations in the time series of the 170 variables would even corresponds to a reduction of the actual available data sample size (Serinaldi et al., 2015). In the case of Vine Copula, the limitation associated with the small sample size occur mostly when fitting conditional pair-copulas. The author says that he is using many independent copula: if this is a reasonable choice then it corresponds to somehow virtually reduce the dimension of the pdf. The structure of the vine is not fully specified, however we can estimate that around 2417 pair-copulas are conditional and non-independent***, which is a very large number. Serinaldi et al. (2015) and Bevacqua et al. (2017) did show that even in much better conditions (much higher ratio: sample size / vine dimension), the uncertainties of the pdf are very large, with propagation of the uncertainties to the estimated quantity of interest (return periods). Personally, I am surprised about how small is the ratio sample size / vine dimension of this work, and I am not aware about studies where a similar ratio is used. For example: (1) Brechmann and Schepsmeier (2013) (cited by the author in the paper) use 396 observations for modelling a 6-dimensional vine; (2) Hobæk Haff et al. (2015) use 40 years of daily data for modelling a 64-dimensional vine; (3) Dißmann et al. (2013) used 2337

observations for fitting a 16-dimensional vine; (4) Brechmann et al. (2012) used 1107 observation to model a 19-dimensional vine.
Here, in my opinion the small sample size can represent a strong limitation as I expect huge uncertainties associated with the results. Although the author says that there are uncertainties about the results, I think that this issue should be explicitly discussed referring to the ratio sample size / vine dimension. According to me, the outcome of this discussion should be considered as crucial for deciding whether the paper should be published. This discussion should refer to literature where similar ratio (sample size / vine dimension) were involved in the analysis (literature where they show that the approach is reasonable). If these papers are not available, I think it should be shown if the approach is reasonable, for example via quantifying the impact of the short sample size on the uncertainties of the results (via bootstrap, similarly to what has been done by, e.g., Serinaldi et al. (2015) and Bevacqua et al. (2017).

*** For a D-vine: there are totally n(n-1)/2=170(169)/2=14365 pair copulas in the Vine (169 are non-conditional copulas). As the 82% of copulas are independent, about 2586 copulas (18% of 14365) are non-independent. The non-conditional copulas are 169, therefore at least about 2586-169=2417 copulas are conditional.

References:
Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, Stoch. Env. Res. Risk A., 30, 1691, https://doi.org/10.1007/s00477-015-1124-3, 2015.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701-2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.

Brechmann, E. C. and Schepsmeier, U.: Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine, Journal of Statistical Software, doi:10.18637/jss.v052.i03, https://www.jstatsoft.org/article/view/v052i03, 2013.

Hobæk Haff, I., A. Frigessi, and D. Maraun (2015), How well do regional climate models simulate the spatial dependence of precipitation? An application of pair-copula constructions, J. Geophys. Res. Atmos., 120, 2624–2646, doi:10.1002/2014JD022748.

Dißmann, J., Brechmann, E., Czado, C., and Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns, Computational Statistics & Data Analysis, 59, 52 – 69, doi:https://doi.org/10.1016/j.csda.2012.08.010, http://www.sciencedirect.com/science/article/pii/S0167947312003131, 2013.

Brechmann EC, Czado C, Aas K (2012). “Truncated Regular Vines in High Dimensions with Applications to Financial Data.” Canadian Journal of Statistics, 40(1), 68–85.

Specific comments

Structure of the paper

I think that a more typical structure would improve the manuscript, e.g. Introduction, Data, Methods, Results, Discussion, Conclusion. At the moment, the structure is not as a usual reader would expect.

About the vine (P9)

1) An equation with an example of a Vine (e.g. in 4 dimension) would be helpful for the reader. In particular this should be shown in combination with the uniform variables used for the vine fit (i.e. the "marginal variables" coming from the GEV).
2) The structure of the used vine is not clear. A table with the percentage of family types used in each tree would be appreciated by the reader.
3) There is not enough information about the procedure used for the fitting of the vine, e.g. what criteria was used for the selection of the RVM structure, what criteria was used to fit the pair copulas, or how you assigned independence to some of the par-copulas. There are references to the R-package, however this is not enough, also considering that in the package different approaches for fit can be used.

Methodology

The part where return periods of averaged AFI are computed based on the variables modelled by the copula is very similar to the so called *structural approach* used in the following references (which I suggest to cite to show similar applications to the reader):

1) Salvadori, G., Durante, F., Tomasicchio, G., and D'alessandro, F.: Practical guidelines for the multivariate assessment of the structural risk in coastal and off-shore engineering, Coast. Eng., 95, 77–83, https://doi.org/10.1016/j.coastaleng.2014.09.007, 2015.
2) Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701-2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.
3) Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, Stoch. Env. Res. Risk A., 30, 1691, https://doi.org/10.1007/s00477-015-1124-3, 2015.

P1 l1: *the third coldest winter ever recorded.* Where and according to what criteria?

P3 l5 *It is based on rigorously quality checked station data interpolated to a regular grid using inverse-distance weighting, as desribed in Perry et al. (2009).*

It should be mentioned here or later that therefore the dependencies catched by the copulas may be partially due to the interpolation itself.

P3 l10 *Nevertheless, local temperature may be subtly different in certain micro-climates, such*

as upland and urban regions.

I would mention that however the resolution 5km x 5km may not always be realistic, depending on the number of stations which were available for the creation of the data set.

P3 l29 *98.3°C.*
Based on line 17, I expected negative values for the AFI. Could you mention that you take the absolute values of the temperature? Also, it would be appreciated if you would show the equation of the AFI.

P3 l32 *After 1962/63, a long run of mild winters followed until late 1978 and early 1979 (Figure 2).*
Is this in Figure 2 the AFI averaged over UK?
Please, use °C in the y label of Fig. 2.

P5 l4 *An additional term was included, the probability of no hazard (P0), in order to account for the cells mainly on the south England coast that have years with no negative temperatures at all.*

1) **Does this mean that for some cells the GEV is fitted on very few data?** Please give information about this, and on the goodness of the fit for these cells.
2) Please, specify how P0 is estimated, e.g. N_occurence/N_years.

FIG 3

1) I assume that the "*historical AFI GEV fit (black circles)*" is the empirical estimate. If yes, is this computed as written in P6 L5? Please, specify this.
2) Could you specify the estimated parameters, or also only making clear to the reader whether the difference is due to the selection of different family type (Gumbel, Frechet, and Weibull distributions)?

P5 l20 *As an example, the GEV fit for a single cell over London is shown in Figure 3. The grey line represents the GEV fit without any weighting applied, while the black curve is estimated using the TWMLE method with an improved fit towards the tail of the distribution (i.e. the more extreme events).*
I would rather say that you get a curve that is nearer to the empirical estimate.

P6 l2 *Other urban regions (e.g. Manchester or Midlands area) do not stand out as much as a result of the low grid resolution.*
Can this also be due to the original data format? For example there may be not enough stations around some urban areas.

P9 l8 *At the first level, 49% of the selected bivariate copulas are found to be Gumbel which implies greater dependence at larger AFI values.*
You refer to the tail dependence, I assume. Make it more clear, please. Greater with respect to what?

P9 l13 *The RVM is used to simulate 10,000 years of winter-seasons in the UK. This amount of realisations should be long enough in order to estimate with enough confidence the 200 year RP hazard, which is commonly associated with capital and regulatory requirements.*

The 10,000 years time series should be long enough to neglect uncertainties associated with the Monte Carlo simulations (which is the method used for extracting the return period associated with the fitted parametric pdf) (Serinaldi et al. (2015) and Bevacqua et al. (2017)). One should ensure if the sample is “long enough” via repeating the (10,000 years) simulations several times and checking if the there are differences in the estimated return period (if there are no differences, the 10,000 years sample is long enough). Performing a long enough simulations allows one to get a convergence to the true return period that one would get analytically from the fitted pdf (given the complexity of the problem it is impracticable to get an analytical derivation of the RP).
Performing a long simulation does not solve the issue about the model uncertainties (uncertainties existing about the pdf), which is there because the pdf is calibrated on a finite - very short - sample. I suggest to discuss this in a way to make difference between these different type of uncertainties.

P9 l 27 *The exceedance probability (EP) curve of wAFI is shown in Figure 7, both for the historical and the stochastic data.*
So far you talked about RP. Personally, I think that it would be better to keep the same terminology instead of introducing EP, or at least use also RP here.

P9 l27 *The uncertainty intervals in the historical data are computed as the 5th and 95th quantile of the probability density function (Folland and Anderson, 2002).*
I suggest to use: ”The uncertainty intervals in the return period (estimated empirically?) of the historical data are computed via the 5th and 95th quantile of the probability density function”

P10 l2 *low tail dependence.* Gaussian and Frank copula have zero tail dependence, not “low”. It may be helpful to better introduce the tail dependence in a sentence where you talk about it for the first time.

P10 l2 *On the other hand, the low impact of the other copula familes is due to the fact that the extreme hazard values are mainly driven by the large dependencies between nearby cells, especially at the first tree levels.*
Could you please argue this better?

P10 l16 *However, recent studies suggest that cold weather in the UK is likely to be less severe, to occur less frequently, and to last for a shorter period of time than was historically the case due to anthropogenic induced climate change (on Climate Change, 2017).*
I would already mention here that there is debate about this (as you then specify in the next paragraph).

P11 l7 *As shown in Figure 8, South England is in general warmer than the North England and Northern Ireland region, partially driven by the urban micro-climate effect of the London area. The 1962/63 winter was less extreme in this region (wAFI of 139° C) with an estimated return period of 1 in 79 years. On the other hand, Scotland is usually significantly colder than the rest of UK, reaching for example AFI values of 100 ◦ C almost 2 times more often.*

Please, make more clear in the text (and in the figure captions) when you talk about AFI, wAFI, averaged non-weighted AFI (and in which area is computed the average (UK, or sub-regions)). Also, when introducing eq. (3), I suggest to anticipate that you are going to use the wAFI both on UK and subregional scale.
Figure captions. Please improve the Figure captions with more information. For example in Fig 2 what is the NAOI (North Atlantic Oscillation Index)?

Technical corrections
P3 l6 *desribed.* Described

P4 l1 *that winter*. You may use “winter 1978/79”.

Fig 4 and 5. Could you please use the same scale range, i.e. 0-400°C

P10 l2 *familes.* Families