Dear Editor,
Dear reviewers,
I would like first to thank the reviewers for their constructive and very helpful comments. I am replying below to your comments in red.

With best regards, Symeon Koumoutsaris

## REVIEW 1

## Main Comment

This is a fair comment. Indeed the sample size is small relative to the dimension of the problem. In order to estimate the associated uncertainties, I followed a bootstrap approach, as suggested by the reviewer. In particular, similarly to Bevacqua et al. (2017a):

- I simulate $B=1000$ samples with the same length as the observed data (i.e. 51 years)
- On each of the $B=1000$ samples, I fit a RVM, whose structure is the same as the fitted with the observed data, while the pair-copula families are re-selected for each sample.
- For each of these $B=1000$ RVMs, simulate 10 K years of correlated samples and estimate the corresponding wAFI return period levels for the entire UK, and for England S, England N \& N. Ireland, and Scotland shown in Figures $7 \& 8$ of the submitted manuscript.
- Estimate the uncertainties in the return levels by identifying the $95 \%$ confidence interval (i.e. the range $2.5-97.5 \%$ ) of the $B=1000$ return level curves.

For simplicity, let's call this uncertainty as the "RVM uncertainty".
In addition to this uncertainty and in conjunction with comment P9 I13 of the reviewer, I perform another round of re-simulations, in order to estimate the uncertainty associated with the 10K years of the Monte Carlo simulation, as follows:

- I use the "default" RVM to simulate 10K years of correlated samples x 1000 times.
- For each of these 1000 simulations, I estimate the corresponding wAFI return period levels for the entire UK and for the three sub regions, as above.
- Estimate the uncertainties in the return levels by identifying the $95 \%$ confidence interval (i.e. the range 2.5-97.5 \%) of the 1000 return level curves.
- I also take into advantage of these 1000 re-simulations in order to estimate more precisely the return period curves using the median of them.

For simplicity, let's call this uncertainty as the "10K uncertainty".

Figure 1 below shows the wAFI RP curves for the whole UK and the three sub-regions (i.e. similar to the figures 7 and 8 of the submitted manuscript), including the $95 \%$ confidence intervals (CI) of the RVM (orange dashed lines) and of the 10K (black dashed lines) as computed above.

It is clear from Figure 1 that the two uncertainty ranges are very close to each other. This means that the uncertainty in the RVM model is mainly driven by the randomness due to the 10K year sampling (the 10K uncertainty). In other words, the RVM fitting introduces only a minor additional uncertainty in comparison to the uncertainty due to the Monte Carlo sampling. This happens because a) $64 \%$ of the fitted copulas are selected to be independent and b) due to the nature of the hazard, the dependencies between neighboring cells are strong and these are driven almost entirely within the first few trees of the RVM. Both (a) and (b) above, lead to a reduction in the dimensions of the pdf, as the reviewer also mentions.

As a further confirmation, Figure 2 shows the resulting RP curves from models truncated above the first three levels of the RVM tree (using only one 10K year simulation though). Here, the truncated model above level 3 is very close to the default one with no truncation, which suggests that only the first few levels drive the dependencies between the cells.

However, uncertainty is indeed important: as shown in Table 1, the 1962/63 winter RP estimate for the entire UK ranges from 82 to 122 years. Notice as well that the more precise estimate of the RP of the 1962/63 winter derived from all the 1000 simulations of the 10K years (as described above) is 97 years ( $\sim$ equivalent of 10 M years of simulations) instead of 89 years in the submitted manuscript based only on one 10 K years simulation. However, notice that the uncertainty is much smaller in comparison to the uncertainty in the empirical curve (i.e. directly from the observed data, grey lines in Figure 1). Because of that, I believe that the model is useful but indeed its results should be interpreted being aware of these uncertainties.

It is also important to note that uncertainties at different regions can be larger and this can be particularly important for example when calculating monetary losses of insurance portfolios where risks may be concentrated in certain areas.

I agree therefore with both reviewers and I will make sure to convey more clearly the importance of those uncertainties in the revised manuscript.

Table 1: Return periods (in years) for the 1962/63 event for regions in the UK. The 95\% confidence intervals are also shown for the "10K" and "RVM" uncertainties (see text).

|  | 1962/63 RP (in years) |  |  |
| :--- | :---: | :---: | :---: |
| Region | Median | 95\% CI RVM | $95 \%$ CI 10K |
| UK | 97 | $81-120$ | $82-122$ |
| England S | 111 | $91-138$ | $92-140$ |
|  <br> N. Ireland | 84 | $70-101$ | $71-103$ |
| Scotland | 55 | $49-65$ | $49-67$ |



Figure 1: Weighted AFI RP curves for (a) the whole UK, (b) South England, (c), North England \& Northern Ireland, and (d) Scotland. Empirical estimates are shown in grey, including their confidence intervals (grey dashed lines). The black solid line respresents the stochastic curve. Confidence intervals due to the 10K years Monte Carlo simulation and due to the RVM fits are shown as the black and orange dashed lines, respectively.

Figure 2: Weighted AFI RP curves for the whole UK. The empirical EP is shown in grey. Results are
 shown using RVM with truncation above level 1 (red line), above level 2 (blue line), and above level 3 (pink line) or no truncation (default, black line).

## Specific comments

## Structure of the paper

I agree and I will re-structure the manuscript as suggested.

## 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).

Agreed.
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.

I rewrote this paragraph in order to make it clearer. Table 4 at the end of this document also contains the percentage of family types used in each of the tree level.
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.

I add the following paragraph in order to give more information about the fitting choices:
"The copula family types for each selected pair in the first tree are determined by using the Akaike information criterion (see Brechmann and Schepsmeier, (2013)). For computational reasons, the twoparameter Archimedean copulas are excluded from this analysis (which however has only a negligible impact in the results, see Figure 5 A1 of the Appendix). The copula parameters are estimated sequentially (using maximum likelihood estimation) starting from the top tree until the last tree, as described in Czado et al. (2013). This approach only involves estimation of bivariate copulas and has been chosen since it is computationally much simpler than joint maximum likelihood estimation of all parameters at once."

## Methodology

P1 I1: the third coldest winter ever recorded. Where and according to what criteria?

I will update the abstract to make this clearer. This is according to the Central England Temperature (CET) record, the oldest continuously running temperature dataset in the world (Manley, 1974), only two other winters (1683/84 and 1739/40) have been colder than 1962/63 in the last 350 years (also mentioned in the submitted manuscript).

P3 I5 It is based on rigorously quality checked station data interpolated to a regular grid using inversedistance weighting, as described 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. Agreed

P3 I10 Nevertheless, local temperature may be subtly different in certain micro-climates, such as upland and urban regions. I would mention that however the resolution $5 \mathrm{~km} \times 5 \mathrm{~km}$ may not always be realistic, depending on the number of stations which were available for the creation of the data set.

I agree. However, notice that I re-grid the data to a lower resolution of $50 \times 50 \mathrm{~km} 2$. The station network contains 540 stations with an average density of $21 \times 21 \mathrm{~km} 2$, with also more stations near urban areas (see Figures 1 and 2 in Perry and Hollis, 2005).

P3 $12998.3^{\circ} \mathrm{C}$. 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.

That's correct. The exact equation is

$$
A F I=\sum_{\text {day }=1 / 6 / \text { Year }}^{31 / 5 /(\text { Year }+1)}\left|T_{\text {day }}\right| \quad \text { if } T_{\text {day }}<0^{\circ} \mathrm{C}
$$

P3 I32 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 ${ }^{\circ} \mathrm{C}$ in the y label of Fig. 2.

Yes, that's has been corrected.
P5 I4 An additional term was included, the probability of no hazard (PO), 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.

In order to improve the fits at those cells, I applied a geographical smoothing of the GEV parameters as well; I had erroneously missed to discuss this part from the submitted text. I will update the text to reflect in detail this methodology.

More precisely, along with the TWMLE method described in the text, I applied a second modification in order to geographically smooth the GEV parameters. The smoothing is incorporated into the fitting process by minimizing the local (ranked) log-likelihood. More precisely, the log-likelihood at each grid cell $i$ is calculated using all grid points but weighted by their distance $d_{i j}$ :
$\log L_{i}=\sum_{j=1}^{170}\left(w_{i j} \cdot \log L_{j}\right)$, where $w_{i j}=\frac{1}{\sqrt{2 \pi}} e^{-\frac{d_{i j}^{2}}{2 L^{2}}}$
where $L$ here is the length scale or smoothing parameter and $\log L_{j}$ is the ranked log-likelihood for cell j . Because the historical gridded data are already geographically smoothed, I decided to use a small length scale parameter $L$ of 15 km (in comparison to the 50 km grid size).

In general, the increase of the sample size at each grid point allows for a more precise estimation of the parameters, especially for the shape parameter which is highly influential in estimating the hazard levels and high return periods. This methodology also permits the estimation of the parameters in cells with no data.

The parameters at each cell are shown in Table 3 (at the end of this document) and will also be included in the final form of the manuscript. I' also attaching a pdf which contains the GEV fits for all the 170 cells of the domain (gevPlots.pdf).
2) Please, specify how PO is estimated, e.g. N_occurence/N_years. FIG 3

P0 = N_occurrence / (Nyears + 1)

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.

Yes, that's is correct. I will make this clear in the caption.
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)?

The difference between the two fitted curves in Figure 3 of the submitted manuscript is the different methodology: MLE vs. the selected TWMLE which gives more weight to the higher order observations as described in the text.

P5 I20 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.

I agree with the reviewer and I will update the text.

P6 12 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.

Figure 3 shows the AFI values for the (empirically estimated) 50 years return period based on the historical data at their original resolution ( $5 \mathrm{~km} \times 5 \mathrm{~km}$ ). Apart from the London area, other cities (e.g. Liverpool, Manchester) stand out with lower values than their surroundings. It is therefore mainly the re-gridding process to $50 x 50 \mathrm{~km}$ (necessary for computing reasons) that masks the cities in this study's results.


Figure 3: High resolution maps of AFI values (in ${ }^{\circ} \mathrm{C}$ ) for return period of 50 years.

P9 I8 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?

I meant greater with respect to the low AFI values, so indeed I refer to the tail dependence. I will rephrase it to make it more clear.

P9 I13 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.

I agree with this comment. Please refer to my answer at the main comment where I discuss both uncertainties.

P9 I 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.

I agree with the reviewer and I will update the text to reflect that.
P9 127 The uncertainty intervals in the historical data are computed as the 5th and $95^{\text {th }}$ 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" Agreed.

P10 I2 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. Agreed. I will introduce the tail dependence definition and functions also in conjunction with the reply of the following question.

P10 I2 On the other hand, the low impact of the other copula families 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?

Unfortunately I found an error in Figure 1 and 2 of the Appendix in the submitted manuscript: the curve based on the Clayton RVM was computed erroneously. It is corrected as shown in Figure 4a below. So, Gaussian, Frank and Clayton copulas now show differences in comparison to the default RVM. This is to be expected since all three of those copulas do not show right tail dependence in the limits. In
particular, away from the extremes, normal shows greater right tail concentration than Frank and Clayton copulas (see Figure 4b).


Figure 4a: Sensitivity tests for the weighted AFI RP over the entire U.K. based on RVM fitted using: all but the two-parameter
Archimedean copulas (black line), only one Copula family each time, i.e. Gaussian (blue line line), Student's t (pink), Clayton (light blue), Gumbel (orange), Frank (grey), and Joe (green).


Figure 4b: Lower and upper tail dependence plot for the Gauss, Gumbel, Clayton, Joe, and Frank copulas, assuming a Kendall tau value of 0.5.

Concerning the argument that the large dependencies drive the extreme hazard values, I believe that this is to be expected: the co-occurrence of extreme AFI values in many cells (which lead to large regionally averaged AFI) is mainly driven by large dependencies among the cells.

In order to demonstrate the large influence of the larger dependencies in the model, I manually adjust the copula parameters in order to test the sensitivity of the very large ( $\tau \geq 0.5$ ) vs. the lower ( $\tau<0.5$ ) dependencies. For simplicity select a test RVM using only Gumbel (or independent) copula families (RVM04), which however matches relatively well the results from the default as shown in Figure 5 (and also described in the manuscript). I manually adjust RVM04 by reducing the Gumbel alpha parameters (at all levels) such as the new corresponding Kendall's tau ( $\tau$ ) value would be very small (equal to 0.01). This is simple for the Gumbel copula since the Gumbel copula parameter $\alpha$ is given by: $\alpha=1$ / ( $1-\tau$ ). I adjust separately the copulas with very large dependencies (with $\tau \geq 0.5$ or $\alpha \geq 2$ ) and those with lower ( $\tau$
< 0.5 or $\alpha<2$ ) dependencies and the resulting RP curves are shown with the blue and orange lines, respectively.


Figure 5: RP curves of weighted for the whole UK. The empirical EP is shown in grey. The black line shows the curve based on the default RVM while the red line is based on a RVM fitted only with Gumbel copula families. The orange and blue lines represent the curves after manually adjusting the selected Gumbel parameters as described in the text.

P10 I16 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).

## Agreed

P11 I7 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 $139^{\circ} \mathrm{C}$ ) 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 $100^{\circ} \mathrm{C}$ 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.

Agreed
Figure captions. Please improve the Figure captions with more information. For example in Fig 2 what is the NAOI (North Atlantic Oscillation Index)?

Agreed

## Technical corrections

P3 I6 desribed. Described

P4 I1 that winter . You may use "winter 1978/79".

Fig 4 and 5. Could you please use the same scale range, i.e. $0-400^{\circ} \mathrm{C}$

P10 12 familes. Families

Agreed.

## REVIEW 2

## GENERAL COMMENTS

The paper is an interesting one, and outlines an original multivariate investigation concerning subfreezing temperatures. The comments posted by Referee 1 already provide an excellent, detailed review, with which I (almost) fully agree. Below, please find further notes: my objections should be read as constructive advices. Some relevant bibliography is reported at the end of this review.

1. I noticed that there is some confusion between the notions of probability distribution function and probability density function (e.g., Page 10, Lines 5-7: "The uncertainty intervals in the historical data are computed as the 5th and 95th quantile of the probability density function (Folland and Anderson, 2002)"). The probability distribution function is the integral of the probability density function (if it exists). The quantiles are the inverses of the probability distribution function (a nondecreasing one), not of the density function (which may not even be monotone). The Author must check the paper and fix all the points where such a confusion arises, otherwise the paper is not correct from a probabilistic point of view. Agreed.
2. I was puzzled by the comment of Referee 1 concerning the sample size, and I ask the Author to clarify the issue: here, 170 variables are at play, each observed 51 times. To the best of my understanding, the idea revolving around Vine copulas is that any multivariate density can be decomposed into a (suitable, maybe not unique) product of univariate marginal densities and bivariate copula densities: in turn, only univariate and bivariate fits should be needed, isn't it? Thus, apparently, the fitting problem may not be so severe. Clearly, trying and fitting the upper tail of a GEV law using only 51 observations may be difficult (although the TWMLE escamotage is used), but it may not be impossible. Similarly, trying and fitting a bivariate copula using only 51 pairs may not be advisable, but it is not uncommon in practical applications. Overall, should my interpretation be correct, the game played by the Author may not be a "Mission Impossible", rather an "Uncertain Mission". . . Thus, I kindly ask the Author to clearly explain the situation, and to provide estimates of the uncertainties as explained below.

This question is related to the main comment of the first reviewer. Indeed there are indeed 170 variables each observed 51 times. Please see my answer to the main comment of the first reviewer at the beginning of this document.
3. I definitely agree with the comment of Referee 1 concerning the procedure to estimate the uncertainties (Page 9, Lines 23-ff.). As a rule of thumb, 1000 independent repetitions of the 10,000years Monte Carlo simulations are usually suggested in literature, in order to provide "reasonable(?)" estimates of the confidence intervals of interest (clearly, it may be adjusted depending on the computational burden).

I have re-simulated 1000 independent repetitions of the 10K years Monte Carlo simulations as suggested by the reviewer. I used these repetitions to construct confidence intervals: please see my answer to the main comment of the first reviewer at the beginning of this document.
4. My main "perplexity" concerns a methodological issue. In this work, I can see the Mathematics/ Statistics, but I do not see the Physics, which, instead, should be the starting point. To be clear, and to the best of my knowledge, the procedure used to construct the 170-dimensional copula finds its justification in an aggregation/clustering algorithm based solely on statistical considerations (Page 9, Lines 13-14: "The method follows an automatic strategy of jointly searching for an appropriate R-Vine tree structure"). If I remember it correctly, the algorithm is based on the Kendall and/or on the Kendall Distribution Function K, and/or, in general, on the strength of the statistical association between the variables at play. While interesting and meaningful from a mathematical point of view, such a procedure may eventually (statistically) associate grid cells having little, or negligible, physical link (for instance, could this be the case of the grid cells corresponding to Edinburgh and London, quite far apart from a spatial and a climatic point of view?) In other words, important information like, e.g., the latitude (corresponding to different climatic regions) may not be considered/used by the statistical procedure adopted for constructing the overall copula. The Author is kindly asked to discuss the issue, and to provide suitable justifications. Is it possible to modify the construction of the 170-dimensional copula in order to take into account the physics of the phenomenon?

I believe that the physics are captured via the large scale circulation, which is the driver of winter temperatures in UK, and which is modeled through the multivariate joint distribution. I don't agree with the reviewer with respect to the association of distant locations: winter temperatures are usually coherent across large part of UK as they are primarily associated with large-scale atmospheric circulation (Scaife and Knight, 2008). Notice that AFI assess the freezing temperatures at an annual temporal resolution, i.e. during the entire winter season and, thus, is less associated with small scale local phenomena which can cause differences between locations. This can be seen in Figure 6, where the 51-year long observed AFI values over London correlate significantly with all the remaining UK cells with linear correlation coefficients above 0.5 . The vine copula model's role is to capture these large scale temperature associations across the UK.

Moreover, the dominant large scale mode of variability in the Euro-Atlantic region is the North Atlantic Oscillation, which I tried to include (not in the copula but) in the GEV fits (see section 3.1, page 12, lines 11-25); however, including it was not improving the model fits, possibly due to the quite noisy character of the phenomenon and the relatively short historical record used in this study.

Recently, Bevacqua et al (2017a, 2017b) has developed an R package called CDVineCopulaConditional (a reference suggested by the first reviewer in fact) which offers the possibility to include meteorological predictors for the contributing variables; including for example NAOI would be something that I would like to include in the future.


Figure 6: Linear correlation map between the 51-year long AFI values in a cell over London (blue dot) and all other cells of the model domain.
5. The Author has modeled the historical data, but, should the climate be really changing, then (at least from an Insurance point of view) the Author should account for it in his model, e.g. by introducing (in the long term simulations) suitable temporal patterns in the GEV/copula parameters according to available projections of the future climate (like, e.g., in IPCC scenarios). A comment is required on this issue.

From an Insurance point of view, the focus is mainly in the next year or sometimes a bit longer (but up to 4 years) depending on the (re) insurance contracts. Therefore, the long term changes in the climate based on future climate projections are not much of interest.

However, what is important is that the model here has been constructed under the assumption of a stationary climate, i.e. under the assumption that the climate has not changed (significantly) during the last 51 years. Although not mentioned in the original manuscript, I had tested this assumption in a similar method as for NAO (see section 3.1): I incorporated a GLM into the statistical distribution parameter estimates of the GEV fits, by defining the location parameter as a function of the year: $\mu=\beta_{0}$ $+\beta_{1} \times$ year. However, this was not improving the model fits (the beta parameters were not statistically different from zero) and thus it has not been included it to the final model.
6. In Section 3.1 "Results and discussion", the Author mentions the actual debate about climate changes (already commented by Referee 1). I would suggest to take a look at a recent paper by Vezzoli et al. (2017), where the traditional validation criteria of climate models are discussed, and an advanced/ thorough distributional perspective is outlined: it may partially explain why several crucial hypotheses are "still largely under debate" (as claimed by the Author and Referee 1), and may partially account for the general inability to draw up clear settlements.

Thank you I take on board this comment.

## SPECIFIC COMMENTS.

Page(s) 2, Line(s) 23-ff.

For the benefit of unskilled readers and practitioners, here the Author should provide general references involving seminal books, papers, and guidelines concerning copulas, like writing: "For a theoretical introduction to copulas, see Nelsen (2006); Joe (2014); Durante and Sempi (2015); for a practical/engineering approach and guidelines, see Genest and Favre (2007); Salvadori and De Michele (2007); Salvadori et al. (2007, 2014, 2015)". Instead, citations concerning Vine copulas, being more specific and related to the modeling outlined in this work, may be postponed later. Page(s) 9, Line(s) 2022.

I agree with the reviewer and I will update the manuscript as suggested.

Author. "Goodness-of-fit is performed for the final selected R-Vine Model (RVM) based on the RVineGofTest algorithm of the same R package (Schepsmeier, 2013). The Cramer von Mises test, which compares the empirical copula with the RVM, has a value of 0.019 and a p.value $=1$, which indicates that the fitted RVM cannot be rejected at a 5\% significance level."

Referee. I am puzzled by such a large p-Value: in my opinion, it may entail a large probability of Type II error, i.e. accepting a False Null Assumption (this a typical performance of Cramer-von- Mises and similar tests, when the sample size is insufficient). The Author is kindly asked to discuss the issue, and to provide suitable justifications.

Unfortunately, the gof tests for Vine copulas show poor behavior in small sample sizes and also at higher dimensions, as is the case for this work (Schepsmeier, 2013). I should also mention that different gof methods implemented in the VineCopula R package (via the function RVineGofTest) can show very different values for the CvM statistic as shown in the table 2 below. I have also tried to include a range in the GoF test values by resampling the historical AFI years (sampling 51 years with replacement $\times 10$ times). Due to the long runtime, it has not been possible to sample more times; the min/max ranges are shown in parenthesis for the ECP2 and ECP methods in the table below. It is even harder to estimate the power of these tests. Thus, my main justification would be the performance of the model in comparison to the empirical estimates: as shown in Figures 7 and 8 of the submitted manuscript, the simulated RPs follow reasonable well the empirical (historical) estimates not only at the entire UK but also at smaller regions.

Table 2: GOF valus for the CvM statistic based on different methods implemented in the VineCopula R package

| GoF method | CvM | p.val |
| :---: | :---: | :---: |
| ECP2 | $0.019(0.019-0.15)$ | $1(0.95-1)$ |
| ECP | $2.3(2.1-4.3)$ | $0.73(0.61-0.73)$ |
| Breymann | 1.8 | 0.185 |
| Berg2 | 0.19 | 1 |
| Berg | 1.11 | 1 |
| White | not enough memory |  |
| IR | not enough memory |  |

## References

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, 2017a.

Bevacqua E., CDVineCopulaConditional: Sampling from Conditional C- and D-Vine Copulas. R package version 0.1.0, 2017b

Bowman, A. W. and Azzalini A., Applied smoothing techniques for data analysis, Oxford University Press, London, 1997

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.

Czado, E., Brechmann E. C., and Gruber, I.: Selection of Vine Copulas. In: P. Jarkowski, F. Durante, and W. Haerdle, (editors). Copulae Mathematical and Quantitative Finance, 16, 775-787, 2013.

Perry, M. and Hollis, D., The generation of monthly gridded datasets for a range of climatic variables over the UK. Int. J. Climatol., 25: 1041-1054. doi:10.1002/joc.1161, 2015.

Scaife, A. A. and Knight, J. R.: Ensemble simulations of the cold European winter of 2005-2006, Quarterly Journal of the Royal Meteorological Society, 134, 1647-1659, doi:10.1002/qj.312, http://dx.doi.org/10.1002/qj.312, 2008.

Schepsmeier, U.: Estimating standard errors and efficient goodness-of- t tests for regular vine copula models, Ph.D. thesis, Fakultat fur Mathematik Technische Universitat Munchen, http://mediatum.ub.tum.de/doc/1175739/document.pdf, 2013.



