Journal: Natural Hazards and Earth System Sciences (NHESS)
Second revised version of: "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

Dear reviewer and editor,

I would like to thank you both for your thoughtful comments. Please find below my answers to your comments (in red) and a copy and a copy of the manuscript where all my changes are annotated.

With best regards,
Symeon Koumoutsaris

1. Reviewer comments

Main comments about the non-stationary model.
- I have concerns regarding using the CO2 as a predictor for the local UK warming. While global average warming is correlated with CO2 concentration, the global average warming can differ from the average local warming and even more from the local (potential) warming of the extreme cold events (being the latter the focus of the study). Therefore, I think that using a predictor of global warming (CO2) as a predictor for changes of the local cold extreme temperature in the UK is problematic. Especially when the author extrapolates information about the future climate based on changes in CO2. The author himself highlights that under global warming (increasing CO2) the Arctic amplification has been occurring, and this might lead to non-linear responses of the local UK climate. In particular, it is under debate whether very cold winters might be experienced in the UK in the future as indirect result of the increasing average global temperature. Thus, it is clear that potentially relevant physical mechanisms are not included within the selected predictor (CO2). This, might lead to a mismatch between the predicted future climate and the real future climate. Thus, personally, I would not use the CO2 as a predictor at all, and especially I would not present results for the future. An alternative would be to strongly highlight the limitations or - if possible - justify the use of this predictor, and in any case to have a dedicated discussion on that. I believe that the decision is up to the editor.

A dedicated discussion discussing the caveats but justifying the choice of the predictors is added in section 3.2.2.

I believe that the choice of CO2 (and NAO) as predictor is justified for the following reasons:

- **Global CO_2 concentration and the subsequent global warming is an important driver of temperature change in UK.** The global increase in temperature during the last century is mirrored in the UK climate: average annual UK temperatures over land and the surrounding seas have increased in line with global observations, with a trend towards milder winters and hotter summers in recent decades (UKCCRA, 2017). In terms of extreme cold events, Massey et al. (2012) and Christidis and Stott (2012) find that human influence has significantly reduced the probability of such a severe winter in UK. This study also finds a negative correlation is found between the average UK AFI (mAFI) and ΔF_{CO2} forcing (ρ = -0.17, pval= 0.08). Global climate modelling studies also suggest that increased greenhouse concentrations will move UK climate towards warmer, wetter winters and hotter, drier summers. The projected warming is estimated to be larger for the high emissions scenario compared to the medium or low cases, particularly during the second half of the 21st century (UKCCRA, 2017).

- **Covariates such as global mean temperature, CO_2 concentration, and indexes of natural variability (ENSO, NAO etc) have been employed in many studies (Edwards and Challenor, 2013).** In particular, global average temperature is often considered as a covariate in the nonstationary GEV since it is the most robust indicator of human-induced climate change with the strongest attribution of causes (Mondal et al., 2016). Notice that using either global average temperature or CO_2 concertation (or radiative forcing) is equivalent as these are highly correlated: The Pearson correlation coefficient between annual global average temperature and global average CO_2 concentration is 0.93. For example, Hauser et al. (2016) analyzed extreme daily temperatures in western Russia using global mean temperature and time as covariates in the GEV location parameter. Grinsted et al. (2013) used non-stationary generalized extreme value analysis and found that global average surface temperature is a better predictor of Atlantic cyclone activity than local grid cell temperatures. Mestre et al. (2009) found that extremes of maximum temperatures in future climate show a strong rise in the location GEV parameter, which is largely driven by the increasing CO_2 concentration. Hurricane wind speeds have been modelled with respect to global mean temperature, NAO, El Niño Southern Oscillation (ENSO) and other indexes by Jagger and Elsner (2006). Risser and Wehner (2017) analyzed observed extreme precipitation by using two time-dependent covariates, total atmospheric CO_2 concentration and the ENSO Index. Several studies focusing on extreme precipitation have used global warming (Oldenborgh et al. 2016; Agilan and Umamahesh, 2017) or CO_2 concentration (Risser and Wehner 2017; van Oldenborgh et al. 2017) as covariates of their GEV parameters.
- **Both CO_2 and NAO are accurately measured.** Although a model that relies on global mean surface temperature may not have as strong correlations as the casual link is more indirect, it has the advantage that it does not rely on subtle regional patterns that are difficult for models to capture.
- **They provide a reasonable way to isolate the human and natural influences on extreme temperatures.** For the same reason as well, Risser and Wehner (2017) used total atmospheric CO_2 concentration and the ENSO Index to analyse observed extreme precipitation. While it would be possible to use a more locally defined metric of future change (such as the change in the mean UK temperature for example), this would include more unforced naturally occurring internal variability of the climate system, making it difficult to identify the changes that are driven by anthropogenic CO_2 emissions.
- **Finally, using a covariate such as the change in CO_2 forcing avoids the difficulty with determining the start of the trend and also results can be easily rescaled to different time period or emission scenario which is helpful for mitigation strategies.**

Nevertheless, there are some caveats on the choice of predictors which I discuss this in the revised document. At the same time, I agree with the reviewer extrapolating far in the future is particularly problematic, since it assumes that the trends will remain the same in the future. For this reason, I adjust the future climate scenario to a closer year in the future (year 2030). As suggested by the reviewer further down below I also use the RCP emission scenario to estimate the radiative forcing.

- Also, the non-stationary model does not consider potential the non-stationarity in the dependencies described by the joint pdf, and this would need to be highlighted better (at the moment this is only discussed at the end of the results for NAO only, and not for CO2, while this would need to be said in the methods too).

I have added the following phrase in the methods section:

Notice, however, that the effect of NAO/ CO_2 on the residual hazard dependency structure (for example the fact that the AFI between two locations might be more or less correlated as a result of changing NAO/ CO_2 values) is not taken into account here. Recently, a methodology that offers the possibility to include such meteorological predictors in a vine copula model has been developed by Bevacqua (2017a, 2017b) and is something to be addressed in a future study.

Also, according to my understanding, the dependencies described by the copula in this model are not typical spatial dependencies between the locations as in the stationary model, and this is not discussed anywhere. (The marginal pdfs include predictors through linear models. Thus, to my understanding, the uniform variables (obtained from the marginals) modelled by the copula are not the usual ranks associated with the AFI in each location. It seems to me that the uniform variables are rather the ranks of the residual of the linear models. However, this in the parenthesis is only my interpretation, while I think that the reader should not interpret this, rather it would be better if the author provides an explanation.) I think that a discussion on this is needed, such that it is made clear what the model is actually modelling (also considering whether this might lead to limitations).

I have added the following paragraph in the methods section:

In the case of the stationary model, the vine copula is employed to model the entire spatial dependence of the AFI in the UK. On the other hand, the spatial AFI structure in the case of the non-stationary model is modelled in two ways: a) by quantifying the dependence on NAO/ CO_2 in each location, treating each location as conditionally independent, then inducing spatial dependence through the variation of NAO/CO_2 and b) by fitting the RVM model to all the residual dependencies associated with the AFI between the cells; these account for dependencies between cells resulting from other large-scale circulation patterns, and regional climate variability (e.g. due to effects of local orography, land-sea contrast, and small scale atmospheric features such as convective cells).

Minor comments:
P1 l9, rephrase done
P1 l10, “such an event”? done
P2 l4, I would delete “entire”. As also the author says later, the NAO affects mostly the weather around the Atlantic basin. The AO, on the other hand, influences the weather over the entire Northern Hemisphere. done
P3 l18, this sentence about the wind is confusing me and seems, to me, out of context. done
P5 l13, at the end: “period” -> “analysed period”? done
P6 l1 I would start with “Based on the AFI, the winter ...” done
P10 l5, it should be made clear in the text that also the copula is stationary. As mentioned above I have added a phrase in section 3.3.1 (at the copula section).
P11 l11 Please, consider to specify the percentages for mu, P0, and in total separately if you think that this would lead to some interesting additional considerations.

Here below you can find a table with the percentages split.

	P0	**μ**
NAOI	72%	55%
ΔF	39%	21%

As described in the text, NAO is found to affect more cells in total in comparison to anthropogenic climate change, and this is also seen both for P0 and μ. The reason for that I believe is the same (the hidden climate change trend in the limited observational period) as is already mentioned in the manuscript so I prefer to not add this in the revised manuscript.

.

Paragraph P15 l4
- Should "RVM" be "pdf"? Corrected
- Sentence "Performing...", rephrase done
- "long enough to neglect the Monte Carlo uncertainty" Corrected
(P15 l15 Thought: it could have been possible to employ a more standard RCP scenario for future, defined based on the DeltaF, rather than on the change in CO2.) Done
P16 l14
"Instead of 100K", please make a clear link to what is said at the beginning of the section, where you say that 100K would be necessary to reduce the Monte Carlo uncertainty. Probably you could simply move the Paragraph on P15 l4 here. I make a clear link:
Due to computational constraints, confidence intervals are computed only for the stationary model. In addition, the simulation length has been reduced to 10K years (instead of 100K), which implies that part of the calculated uncertainty is due to Monte Carlo sampling variability.

P16 l14
I would start the sentence with "In order to investigate…" and then say that you separate the uncertainties. Rephrased as follows: "In order to investigate further the sources of this uncertainty, the uncertainty associated with the RVM only is separated from the uncertainty of the full model, i.e. of the joint pdf, by calculating confidence intervals with the same approach as described above, but using the same marginal pdfs in each bootstrap repetition."
Section 4.1,
- Please, make clear that you are referring to the mean return period when you say "stochastic set"
- I guess that you are simulating from the vine and then transforming the simulated uniform variables to the "real" via the marginal models containing the predictors/covariates. Here you fix the CO2 predictors to certain values, but what value is given to the NAO predictors? Please, make this clearer. Also, please make this procedure clear in the method section.
NAO is simulated using a Gaussian distribution. I make this more clear in the methods section (3.4): "Each year of the three stochastic sets above is associated with a random NAOI value that has been simulated assuming a Gaussian distribution, fitted to the historical NAOI dataset (see Figure 6). The influence of NAO on each one of these sets can thus be discerned by selecting only the simulated years with negative or positive NAOI values."
Fig10 The grey uncertainties are not well visible on my printed version. Please, check if this is only a problem of mine. I have darkened the shaded areas.
P20 l11 "originates", I would write "appears to originate" Corrected
Table 4 missing the unit: "years" Corrected
P21 l1 see the main comment on CO2 above Replied above
P21 l12 "return period of 1 in 39 years", is it not enough to write "return period of 39 years"? Also in other parts of the paper? Corrected
P21 l12 How is it a positive and negative phase defined? What is the value of NAO that is used to sample the data? In the figure it is written > or < 1, but do not you use a unique value, e.g., 1 and -1? Please, make this clear. I add the following sentence:
"Fig. 11 shows the RP curve of current climate wAFI, alongside with RP curves computed solely from simulated years with NAOI values greater than 1 (i.e. representing the positive NAO phase) or years with NAOI values lower 5 than 1 (i.e. representing the negative NAO phase)."

P23 l3. The predictor influence on the dependence is not considered. Indeed it is specified here only for the NAO, but not for the CO2. As mentioned above I have added a phrase in the methods section. I've also included CO2.
P24 l9 the occurrence has increased? The return period has increased. The same in the abstract. Corrected
P24 l16 "such extreme events", not clear which extreme events. Corrected
Best regards.

2. Editor comments

Please consider the following main issues a) to d) to be properly clarified:
a) There is an apparent inconsistency in your manuscript
Note the following. At lines 9-10 of the abstract and lines 9-10 of page 24 in the Conclusions you write that "The model suggests that the occurrence of such extreme cold events have increased approximately two times during the course of the 20th century as a result of anthropogenic climate change". Instead, at lines 24-25 you write that "the non-stationary model suggests that under current climate conditions, such an extreme event, is approximately 2 times less likely to occur than in the 1960s". Table 4 also suggests an increased (doubled) return time of the "1962/63 winter freeze event". Please correct this inconsistency or clarify. This is a typo – it's corrected
b) In the "Methods" section, defend your choice of using the change in radiative forcing from CO2 in the nonstationary model (first main comment of the reviewer) Please see above
c) In the methods section, add that the non-stationary model does not consider the potential non-stationarity in the dependencies described by the joint pdf (first part of the second comment by the reviewer) Done
d) It is not clear which variables the nonstationary model is actually using. Please, make this clear in the methods section (second part of the second main comment of the reviewer) Done
Moreover, please consider all minor points raised by the reviewer Done

Finally consider also the following comments:
i. I agree with the reviewer that at line 9 of the abstract ""such extreme cold events" is not immediately clear and the sentence should be rephrased (I presume it refers to events as anomalous as the "1962/63 winter freeze event") I have corrected the sentence: "the occurrence of extreme cold events such as the 1962/63 winter".
ii. Line 8-9 of the conclusion "According to the model, such a cold winter is estimated to occur once every approximately 400 years under current climate conditions in the UK." I see this in table 4 (please, in the text, specify that you refer to the South UK). However, this result contradicts the empirical evidence. You write at lines 20-22, page 20 that "only two other winters (1683/84 and 1739/40) have been colder than 1962/63 in the last 350 years", suggesting a return period in the range of 110-120 years. Please clarify this inconsistency.

The return period of 400 years corresponds to the current climate, i.e. corresponding to a present day (2018) concentration of CO2 (400 ppm). I delete the sentence in the conclusion to avoid confusion.

A direct comparison with the historical Central England Temperature (CET) record, can made in comparison to the stationary model. As shown in Table 4, the stationary stochastic set seems to overestimate the return period of this event (at least in comparison to the South UK).

I mention this in section 4.2.2 and I make it more clear "According to the latter, only two other winters (1683/84 and 1739/40) have been colder than 1962/63 in the last 350 years, suggesting a return period in the range of 110-120 years, as well. The stationary model overestimates this winters' return period which is estimated to 205 years across all the UK. Especially in the South of the UK the model suggests that this event has been particularly unusual. In the Northern part of UK on the other hand, the model suggests a lower return period of 106 years, closer to the empirical estimate."

iii. The definition of AFI in expression (2) at page 5 is not clear. Specify the meaning of the subscript "i" and specify appropriately the upper and lower limit of the sum Both corrected.

iv. Consider whether moving the first part of section 3.3 (describing copulas) to an appendix could improve readability of your manuscript and help the readers to focus on the novelty of your study I agree and I moved this section in the Appendix.

References

Bevacqua, E.: CDVineCopulaConditional: Sampling from Conditional C- and D-Vine Copulas. R package version 0.1.0, R package version 0.1.0, https://cran.r-project.org/web/packages/CDVineCopulaConditional/index.html, 2017.

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), Hydrology and Earth System Sciences, 21, 2701–2723, doi:10.5194/hess-21-2701-2017, https://www.hydrol-earth-syst-sci.net/21/2701/2017/, 2017a.

Christidis, N. and Stott, P. A.: Lengthened odds of the cold UK winter of 2010/2011 attributable to human influence [in "Explaining Extreme Events of 2011 from a Climate Perspective"], Bulletin of the American Meteorological Society, 93, 1060–1062, doi:10.1175/BAMS-D-12-00021.1, https://doi.org/10.1175/BAMS-D-12-00021.1, 2012.

Edwards, T. and Challenor, P.: Risk and uncertainty in hydrometeorological hazards, pp. 100–150, Cambridge University Press, United Kingdom, 2013.

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Projected Atlantic hurricane surge threat from rising temperatures, Proceedings of the National Academy of Sciences, 110, 5369–5373, doi:10.1073/pnas.1209980110, https://www.pnas.org/content/110/14/5369, 2013.

Hauser, M., Orth, R., and Seneviratne, S. I.: Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia, Geophysical Research Letters, 43, 2819–2826, doi:10.1002/2016GL068036, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL068036, 2016.

Jagger, T. H. and Elsner, J. B.: Climatology Models for Extreme HurricaneWinds near the United States, Journal of Climate, 19, 3220–3236, doi:10.1175/JCLI3913.1, https://doi.org/10.1175/JCLI3913.1, 2006.

Massey, N., Aina, T., Rye, C., Otto, F., Wilson, S., Jones, R., and Allen, M.: Have the odds of warm November temperatures and of cold December temperatures in Central England changed. Bulletin of the American Meteorological Society 93, 1057-1059, Bulletin of the American Meteorological Society, 93, 1057–1059, 2012.

Mestre, O., Denvil, S., and Salas, D.: Trend fitting of gcm temperature extremes, in: Geophysical Research Abstracts, http://meetingorganizer.copernicus.org/EGU2009/EGU2009-10262.pdf, 2009.

Mondal, A. and Mujumdar, P. P.: Detection of Change in Flood Return Levels under Global Warming, Journal of Hydrologic Engineering, 21, 04016 021, doi:10.1061/(ASCE)HE.1943-5584.0001326, 2016.

van Oldenborgh, G. J., Philip, S., Aalbers, E., Vautard, R., Otto, F., Haustein, K., Habets, F., Singh, R., and Cullen, H.: Rapid attribution of the May/June 2016 flood-inducing precipitation in France and Germany to climate change, Hydrology and Earth System Sciences Discussions, 2016, 1–23, doi:10.5194/hess-2016-308, https://www.hydrol-earth-syst-sci-discuss.net/hess-2016-308/, 2016.

van Oldenborgh, G. J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G., and Cullen, H.: Attribution of extreme rainfall from Hurricane Harvey, August 2017, Environmental Research Letters, 12, 124 009, doi:10.1088/1748-9326/aa9ef2, https://doi.org/10.1088%2F1748-9326%2Faa9ef2, 2017.

Risser, M. D. and Wehner, M. F.: Attributable Human-Induced Changes in the Likelihood and Magnitude of the Observed Extreme Precipitation during Hurricane Harvey, Geophysical Research Letters, 44, 12,457–12,464, doi:10.1002/2017GL075888, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL075888, 2017.

UKCCRA: UK Climate Change Risk Assessment 2017, Tech. rep., Department for Environment, Food & Rural Affairs, https://www.gov.uk/government/publications/uk-climate-change-risk-assessment-2017, 2007.

A hazard model of subfreezing temperatures in the United Kingdom using vine copulas

Symeon Koumoutsaris

Guy Carpenter, Tower Place, London, EC3R 5BU, UK

Correspondence to: Symeon Koumoutsaris (symeon.koumoutsaris@guycarp.com)

Abstract. Extreme cold weather events, such as the ~~winters~~ winter of 1962/63, the third coldest winter ever recorded to the Central England Temperature record, or more recently the winter of 2010/11, have significant consequences for the society and economy. This paper assesses the probability of such extreme cold weather across the United Kingdom~~. For that, a~~ , as part of a probabilistic catastrophe model for insured losses caused by the bursting of pipes. A statistical model is developed in order to model the extremes of the Air Freezing Index (AFI), which is a common measure of magnitude and duration of freezing temperatures. A novel approach in the modelling of the spatial dependence of the hazard has been followed which takes advantage of the vine copula methodology. The method allows to model complex dependencies especially between the tails of the AFI distributions which is important to assess the extreme behaviour of such events. The influence of North Atlantic Oscillation ~~(NAO)~~ and of anthropogenic climate change on the frequency of UK cold winters has also been taken into account. ~~The modelsuggests that~~ According to the model, the occurrence of ~~such~~ extreme cold events~~have increased~~ , such as the 1962/63 winter, have decreased approximately two times during the course of the 20th century as a result of anthropogenic climate change. Furthermore, the model predicts that such an event ~~will become quite~~ is expected to become more uncommon, about ~~10 times less frequently, under a 2xCO$_2$ climate scenario. The frequency of extreme~~ twice less frequent, by the year 2030. Extreme cold spells in UK has been found to be heavily modulated by NAO, as well. A cold event is estimated to occur ≈3-4 times more likely during its negative than its positive phase. However, ~~the estimated uncertainty~~ considerable uncertainty exists in these results~~is quite large and comes from the relatively~~ , owing mainly to the short record length ~~. Moreover, possible spurious trends in the historical data add considerable uncertainty to these estimates, as well~~and the large interannual variability of AFI.

1 Introduction

Extended periods of extreme cold weather can cause severe disruptions in human societies; on human health, by exacerbating previous medical conditions or due to reduction of food supply which can lead to famine and disease; agriculture, by devastating crops particularly if the freeze occurs early or late in the growing season; on infrastructure, e.g. severe disruptions in the transport system, burst of residential or system water pipes (**?**). All these consequences lead to important economic losses.

Of particular interest for the insurance industry are the economical losses that originate as a result of bursting of pipes due to freeze events. Water pipes burst because the water inside them expands as it gets close to freezing which causes an increase in pressure inside the pipe. Whether a pipe will break or not, depends on the water temperature (and consequently on the air temperature), the freezing duration, the pipe diameter and composition, the wind chill effect (due to wind and air leakage on water pipes), and the presence of insulation (**??**).

Insurance losses from burst pipes have a significant impact on the UK insurance industry. They amount to more than £900 million in the last 10 years, representing around 10% of the total insured losses, mainly due to flood and windstorm, in the United Kingdom (UK) during the same period (**?**). Particular years can be very damaging, such as, for example, the winter of 2010/2011 where losses from burst pipes have exceeded £300 million in UK making it the peril with the largest losses that year (**?**). Moreover, much more extreme cold winters have actually occurred in the UK in the last 100 years, such as the winters of 1946/47 and 1962/63. It is crucial for the insurance business to be able to anticipate the likelihood of occurrence of similar and even more extreme events so that they can adequately prepare for their financial impact (**?**). In fact, the capital requirements in (re)insurance is estimated in a 1 in 200 year return period (RP) loss basis, which is usually much larger than the available historical records.

Probabilistic catastrophe modelling is generally agreed to be the most appropriate method to analyze such problems. The main goal of catastrophe models is to estimate the full spectrum of probability of loss for a specific insurance portfolio (i.e. comprised by several residential, auto, commercial or industrial risks). This requires the ability to extrapolate the possible losses at each risk to high return periods ~~(RP)~~ which is usually achieved by simulating synthetic events that are likely to happen in the near future (typically a year). More importantly, it requires to consider also how all risks relate to each other and their potential synergy to create catastrophic losses. Such spatial dependence between risks can result from various sources, for example due to the spatial structure of the hazard (e.g. the footprint in a windstorm or the catchment area in a flood event) or due to similar building vulnerabilities between risks in the same geographical area (e.g. due to common building practices) (**?**).

Modelling the spatial dependence of the hazard is usually achieved by taking advantage of certain characteristic properties of the hazard footprint, like for example the track path and the radius of maximum wind for windstorms or the elevation in the case of floods. In the case of temperature, however, such a property cannot be easily defined; an alternative solution is to use multivariate copula models. Based on Sklar's theorem (**?**), the joint distribution of all risk sources can be fully specified by the separate marginal distributions of the variables and by their copula, which defines the dependence structure between the variables.

However, one important difficulty is the limited choice of adequate copulas for more than two dimensions. For example, standard multivariate copula models such as the elliptical and Archimedean copulas do not allow for different dependency models between pairs of variables. Vine copulas provide a flexible solution to this problem based on a pairwise decomposition of a multivariate model into bivariate copulas. This approach is very flexible, as the bivariate copulas can be selected independently for each pair, from a wide range of parametric families, which enables modelling of a wide range of complex dependencies (**??**).

In this paper, the vine copula methodology is used in a novel application to develop a catastrophe model on insurance losses due to pipe bursts resulting from freeze events in the United Kingdom. The focus here is on the hazard component (Sect. **??**) which is modeled using the Air Freezing Index (AFI), an index which takes account both the magnitude and duration of air temperature below freezing, calculated from reanalysis data from the last 110 years. The statistical models utilized to extrapolate to longer return periods are described in Sect. **??**. The model also accounts for two major drivers of climate variability in UK that are incorporated as predictors:

- the North Atlantic Oscillation (NAO), a leading pattern of weather and climate variability over the ~~entire Northern Hemisphere~~ Northern Hemisphere mid-latitudes, which accounts for more than half of the year-to-year variability in winter surface temperature over UK.
- Anthropogenic climate change and its direct effects in the temperature profile in the UK.

Stochastic winter-seasons are simulated taking into account the correlation of the hazard between all pair-cells with the help of regular vine copulas (Sect. **??**). The resulting return periods of extreme cold winters in UK, including the underlying uncertainties, are discussed in Sect. **??**. Concluding remarks are found in Sect. **??**.

2 Data

2.1 Temperature data sets

The hazard component of the catastrophe model is based on the European Centre for Medium-Range Weather Forecasts (ECMWF) twentieth century reanalysis (ERA-20C) covering the entire twentieth century from 1900 to 2010 (**?**). Reanalyses are data-assimilating weather models which are widely used as proxies for the true state of the atmosphere in the recent past. Even though centennial reanalyses, such as ERA-20C, represent the most convenient data sets for assessing the long-term historical climate, biases and uncertainties inherent in both raw observations and models mean that they should be used with caution.

~~For consistency throughout the period, the observational input in ERA-20C is limited to surface pressure and surface marine winds only, which may however lead to some reduction in accuracy (?). For~~ For example, important differences in the 2-meter temperature have been found between ERA-20C and other centennial reanalysis data sets, especially during the first half of the twentieth century as a result of the spare observational network in those early years (**??**). Furthermore, studies have suggested that long-term changes in the Arctic Oscillation, mean sea level pressure, and wintertime storminess seen in ERA-20C, may be spurious as a result of the assimilation of increasing numbers of observations (**???**).

ERA-20C product provides daily 3-hour forecast (i.e. eight forecast steps starting at 06:00UTC each day) of minimum and maximum temperature at 2 meters. These are used to compute daily minimum and maximum values at every grid cell for the entire period. The daily average temperatures are then computed as 0.5(T_{max}-T_{min}) and the data are re-gridded to a 1°x1°resolution, which corresponds to a total of 67 cells over land.

The coarse horizontal resolution is expected to have relatively small influence in most cases given that winter climate anomalies are often coherent across large parts of the UK as they are primarily associated with large-scale atmospheric circulation patterns (**?**). Nevertheless, local temperature may be subtly different in certain micro-climates, such as upland and urban regions. In particular over urban regions, which are most important from an insurance perspective, lower resolution may lead to temperatures that are biased towards lower values, leading though to a conservative view on the severity of extreme freeze events. In upland regions, on the other hand, extreme cold temperatures are most probably underestimated, although it is reasonable to expect that their damaging effects are somewhat mitigated from increased protection levels. For example, water pipes in properties located in mountainous regions are usually better protected against cold spells.

For comparison purposes, the observed daily average temperature gridded data set developed from the UK Met Office is also used (**?**). This data set is based on temperature data retrieved from 540 stations across UK with an average station density of 21 x 21 km^2 (**??**). It covers the entire UK, but for a much shorter period of 51 years (1960-2011). The original 5km x 5km resolution is re-gridded using bi-linear interpolation to 1°x1°in order to match the ERA-20C grid.

2.2 North Atlantic Oscillation Index

The NAO refers to a redistribution of atmospheric mass between the Arctic and the subtropical Atlantic ~~,~~ and swings from one phase to another producing large changes in weather, and in particular in surface air temperature, over the Atlantic and the adjacent continents (**?**). It is described by the NAO index (NAOI), a measure of the mean atmospheric pressure gradient between the Azores High and the Iceland Low. A positive NAOI is associated with depression systems taking a more northerly route across the Atlantic, which causes UK weather to be milder, while a negative NAOI is associated with depression systems taking a more southerly route, as a result of which UK weather tends to be colder and drier (**?**). In this study, the winter (December thru March) station-based index of the NAO from **?** is used, which is based on the difference of normalized sea level pressure between Lisbon, Portugal and Stykkisholmur/Reykjavik, Iceland (~~Figure~~ Fig. **??**b).

2.3 Anthropogenic forcing

Increases in concentration of greenhouse gases, such as carbon dioxide (CO_2), are accompanied by increased radiative forcing, i.e. the difference between the incoming radiation from the sun and the outgoing radiation emitted from the Earth. This forcing arises from the ability of the gases to absorb long wave radiation, thus reducing the amount of heat energy being lost to space, and increasing the warming of the earth's surface. Here we use the change in radiative forcing from CO_2 as a predictor for climate change. It is calculated using the simplified expression (**?**):

$$\Delta F_{CO_2} = 5.35 ln\left(\frac{C_i}{C_{1990}}\right) \tag{1}$$

where ΔF_{CO_2} is the radiative forcing change (in W m^{-2}), C_i is the concentration of atmospheric CO_2 at year i, and C_{1900} is the reference 'pre-industrial' CO_2 concentration at year 1900. Consequently, a doubling of CO_2 corresponds to a change in the radiative forcing of 3.7 W m^{-2}. Historical observations of global mean CO_2 concentrations (in parts per million or ppm) are ~~based on~~ taken from **?**. The temporal ~~increase~~ change in the CO_2 radiative forcing during the 20th century is shown in ~~Figure~~ Fig. **??**c. Projections of future CO_2 emissions are based on the Representative Concentration Pathway (RCP) scenarios adopted by the Intergovernmental Panel on Climate Change (IPCC) for its fifth Assessment Report (AR5) (**?**).

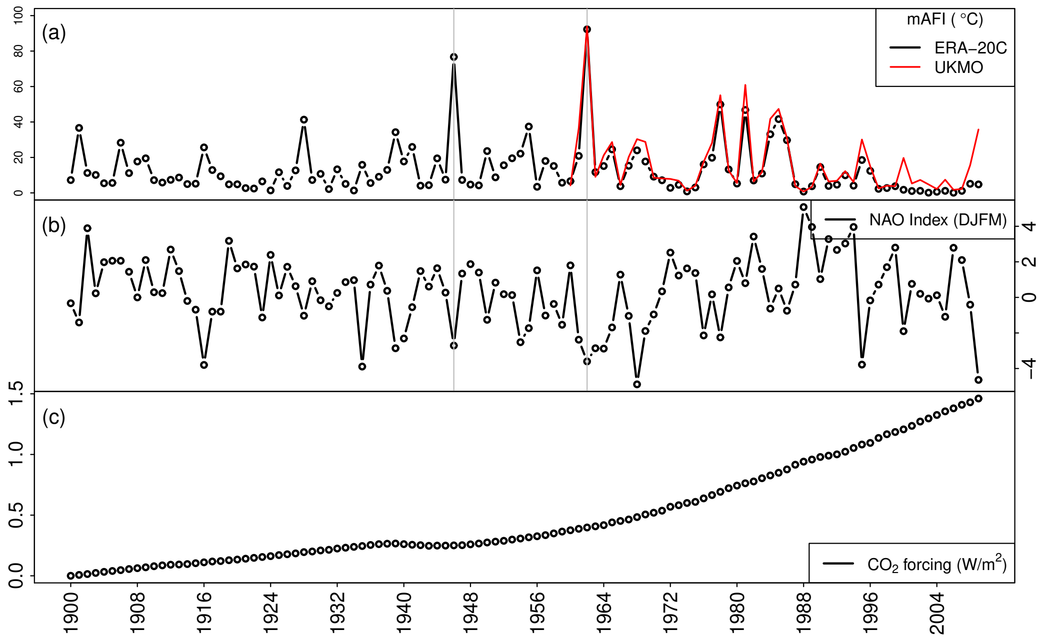

Figure 1. Interannual variation of (a) average AFI over UK (mAFI), (b) the North Atlantic Oscillation Index (NAOI), and (c) CO_2 forcing during the study period.

3 Methods

3.1 Air-Freezing Index and historical events

The daily temperature data are used to compute the AFI at each grid cell, as the sum of the absolute average daily temperatures of all days with below 0°C temperatures during the freezing period (Eq. (**??**)). The freezing period in this study is defined from first of June of year ~~y~~ to end of May of the following year~~$y+1$~~, in order to include the entire winter season. Because AFI accounts both for the magnitude and duration of the freezing period, it is commonly used for determining the freezing severity of the winter season (**??**).

$$AFI_i = \begin{cases} \sum_{d=1}^{N} |T_d|, & \text{if } T_d < 0°\text{C} \\ 0, & \text{if } T_d \geq 0°\text{C for all } d \end{cases} \tag{2}$$

where, AFI_i is the AFI at cell i, N is the number of days in a winter season, T_d the daily average temperature for a day d.

Maps of AFI values from ERA-20C for the severe winters of 1946/47, 1962/63, and 2009/10 are shown in ~~Figure~~ Fig. **??**. The winter of 1946/47 (i.e. season starting from 1st June 1946 to 31st of May 1947) was a harsh European winter noted for its impact in the United Kingdom. It was notable for a succession of snowstorms from late January until mid-March, mainly

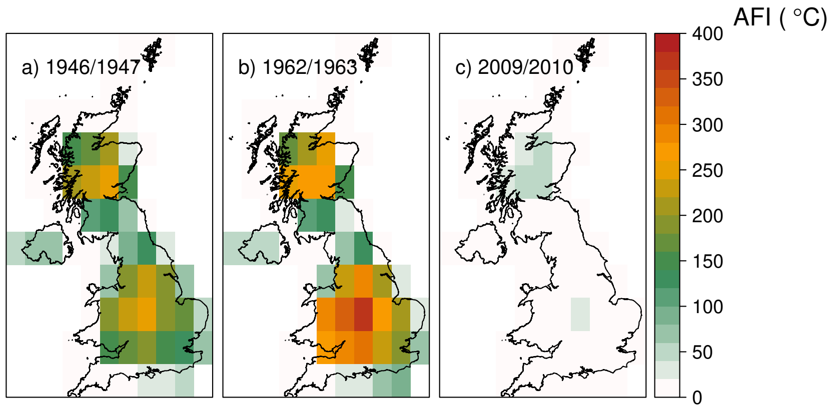

Figure 2. Map of AFI values (in °C) for the the winter-seasons of a) 1946/47, b) 1962/63, and c) 2009/10.

associated with easterly airstreams (**?**). The mean AFI value (mAFI) in the entire UK (i.e. average of AFI values across all gridcells) mounted up to 75.6°C, the second largest value during the analysed period.

~~The~~ Based on the AFI, the 1962/1963 winter season was the most severe winter in the 20th century and one of the coldest on the record in the United Kingdom (**?**). The "Big Freeze of 1962/63", as it is also known, began on the 26 of December 1962 with heavy snowfall and went on for nearly three months until March 1963. The cause of the cold conditions has been the development of a large "blocking" anticyclone over Scandinavia and north-western Russia. Easterly winds on the southern edge of this system transported cold continental air westwards, displacing the more usual mild westerly influence from the Atlantic Ocean on the British Isles. Over the Christmas period, the Scandinavian High collapsed, but a new one formed near Iceland, bringing Northerly winds. The mAFI in the entire UK (i.e. average of AFI values across all gridcells) mounted up to 90.9°C, which represents six standard deviations larger than the average of the entire 110-year period (14.0°C). The event affected more the Southern part of the country as shown in Fig. **??**.

After 1962/63, a long run of mild winters followed until late 1978 and early 1979. However, temperatures in 1978/79 were not as low and the cold weather was interrupted frequently by brief periods of thaw (**?**). The mAFI value of winter 1978/79 reached 49.2°C. The 1980s stands out as a decade with several cold spells in UK, with mAFI values above 30°C for the winters 1981/82, 1984/85, and 1985/86 (46.1, 32.6, and 41.0 °C, respectively).

For the last 10 years of our study period (from 2000 to 2010), mAFI seem to be underestimated in the re-analysis data set (Fig. **??**a). In particular, the winter of 2009/2010, which is well known to have brought frigid temperatures to the UK (**????**), has a mAFI value of only 4.7 °C (~~Figure~~ Fig. **??**c) which is much lower than the long-term average (13 °C) and over ten times lower than mAFI value according to the UKMO dataset (59.1 °C). No clear reason is known for this bias, but it might be related to possible spurious long-term trends in the atmospheric circulation (**?**).

As shown in ~~Figure~~ Fig. **??**a, the two most severe winters in the century (1946/47 and 1962/63) were associated with a negative NAO phase (**??**). As mentioned earlier, the NAO has a profound effect on winter climate variability around the Atlantic basin, accounting more than half of the year-to-year variability in winter surface temperature over UK (**??**). Not surprising, the ERA-20C mAFI over the entire UK is found to be significantly anti-correlated (ρ = -0.49, pval=6.510^{-8}) with NAOI. A negative correlation is found between mAFI and ΔF_{CO_2} forcing, but it is much less significant (ρ = -0.17, pval= 0.08). Both NAO and climate change effects are included in the statistical model as predictors in order to account for their relation to cold winter spells in UK as discussed in the following section.

3.2 Extreme value analysis

3.2.1 Stationary model

Since the historical data only extends for 110 years and our interest lies in very rare events (such as 1 in 200 years), it is necessary to extrapolate by fitting an extreme value distribution. The Generalized Extreme Value (GEV) family of distributions has been chosen, which includes the Gumbel, the Fréchet, and Weibull distributions. 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. The probability therefore that the AFI value (X) inside a cell j is lower or equal than a certain value (x) takes the form:

$$F(x) = P(X \leq x) = P0 + (1-P0)exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\} \tag{3}$$

where μ, σ, and ξ represent the location, scale, and shape parameters of the distribution, respectively. $F(x)$ is defined when $1+\xi\frac{x-\mu}{\sigma} > 0$, $\mu \in \Re$, $\sigma > 0$, and $\xi \in \Re$. Its derivative, the GEV probability density function $f(x)$ is given by:

$$f(x) = f(x) = \begin{cases} P0, & \text{if } x = 0 \\ (1-P0)\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, & \text{if } x > 0 \end{cases} \tag{4}$$

There are various methods of parameter estimation for fitting the GEV distribution, such as least squares estimation, maximum likelihood estimation (MLE), probability weighted moments, and others. Traditional parameter estimation techniques give equal weight to every observation in the data set. However, the focus in catastrophe modeling is mainly on the extreme outcomes and, thus, it is preferable to give more weight to the long return periods. The Tail-Weighted Maximum Likelihood Estimation (TWMLE) method developed by **?** is employed here in order to estimate the GEV parameters. This method introduces ranking depended weights ($w_{(r)}$) in the maximum likelihood. The weights are defined for each cell based on the historical winter-season AFI values, i.e. the lowest historical AFI value in the cell (rank r=1 out of n observations) has the lowest weight, while the largest historical AFI value (rank r=n) has the largest weight, as follows:

$$w_{(r)} = AFI_{(r)}/\sum_{r=1}^{n} AFI_{(r)} \tag{5}$$

Along with the TWMLE method described above, a second modification has been implemented 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:

$$LogL_i = \sum_{j=1}^{170} (k_{ij} * LogL_j) \tag{6}$$

where $k_{ij} = \frac{1}{\sqrt{2\pi}} e^{-\frac{d_{ij}^2}{2L^2}}$, d_{ij} is the distance between cell i and j, L is the smoothing parameter, $LogL_j$ is the ranked log-likelihood for cell j.

The smoothing increases the sample size at each grid point, which thus leads to a more precise estimation of the parameters, especially for the shape parameter which is highly influential in estimating the hazard levels ~~and~~ at high return periods. Because the data grid resolution is already coarse, a small length scale parameter L of 20 km has been used (in comparison to the grid size).

Finally, in order to avoid an over-estimate of the positive value of the shape parameter due to the small sample size (**?**), a modification of the maximum likelihood estimator using a penalty function is also applied for fitting the GEV. The penalty function penalizes estimates of ξ that are close to, or greater than 1, following **?**.

Estimates of P0 for each grid cell are obtained by fitting a logistic regression model with intercept only (Eq. (**??**)). As before, the fitting is performed against all grid cells, weighted by their distance d_{ij}, and a length scale of 20 Km has been used. The model is extended in the non-stationary model to include covariates as described in Sect. **??**.

$$ln\left(\frac{P0}{1-P0}\right) = b_0 \tag{7}$$

As an example, the GEV fit for a single cell over London is shown in Fig. **??**. The curve fitted as described above (black line) is closer to the empirical estimates (black circles, computed as described in Sect. **??**) in comparison with the GEV fit with no weighting applied (grey line). As shown in ~~table~~ Table **??**, for both fits the shape parameter is positive (i.e. both fits correspond to the Fréchet distribution), but for the approach followed here (TWMLE + geographical smoothing), the shape parameter is smaller leading to a shorter tail and a curve that is nearer to the empirical estimate.

Maps of the fitted parameters are shown in Fig. **??**. The probability of non-negative temperatures during a season (P0) is, as expected, larger around the coast which has milder and less variable climate due to the water influence. This also explains the lower mean (location parameter) and larger spread (scale parameter) in the AFI distributions ~~around the coast~~ at the coastal regions in comparison to inland. The shape parameter, which affects the skew of the distribution, shows larger values in the southern part of the UK in comparison to the north, suggesting a less rapid increase in the maximum AFI estimates.

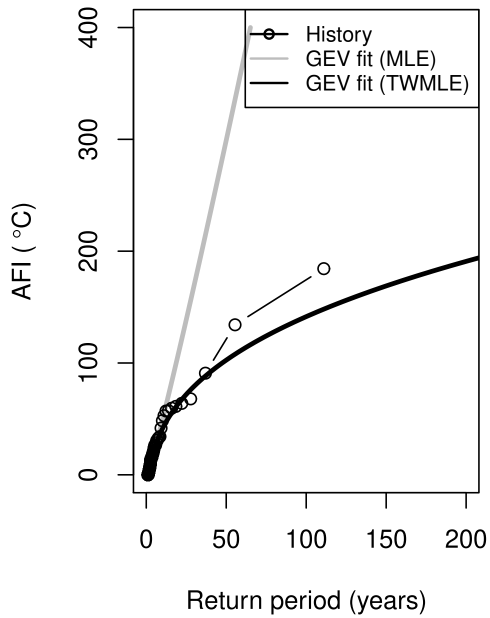

Figure 3. AFI return period curves for a single cell over London: empirical fit (black circles), GEV fitted with MLE (grey line), and GEV fitted with TWMLE and geographical smoothing (black line).

Table 1. Model parameters for a single cell over London.

method	b_0	b_1	b_2	μ	μ_0	μ_1	σ	ξ
MLE (no predictors)	-1.77	0	0	4.05	0	0	5.61	1.08
TWMLE + geographical smoothing (no predictors)	-1.77	0	0	4.87	0	0	12.67	0.35
TWMLE + geographical smoothing (with predictors)	-3.74	0.36	2.62	-2.27	-5.87	3.07	15.32	0.25

3.2.2 Non Stationary model

In stationary models, the distribution parameter space is assumed to be constant for the period under consideration. However, such assumption is not valid in the presence of atmospheric circulation patterns or anthropogenic changes. Regression approaches are often used to assess the influence of climatic factors on hazards and covariates such as global mean temperature, CO_2 concentration, and indexes of natural variability (such as NAOI) have been employed by several studies (?). In this study, a generalized linear model (GLM) is introduced into the statistical distribution parameter estimates in order to improve the non-

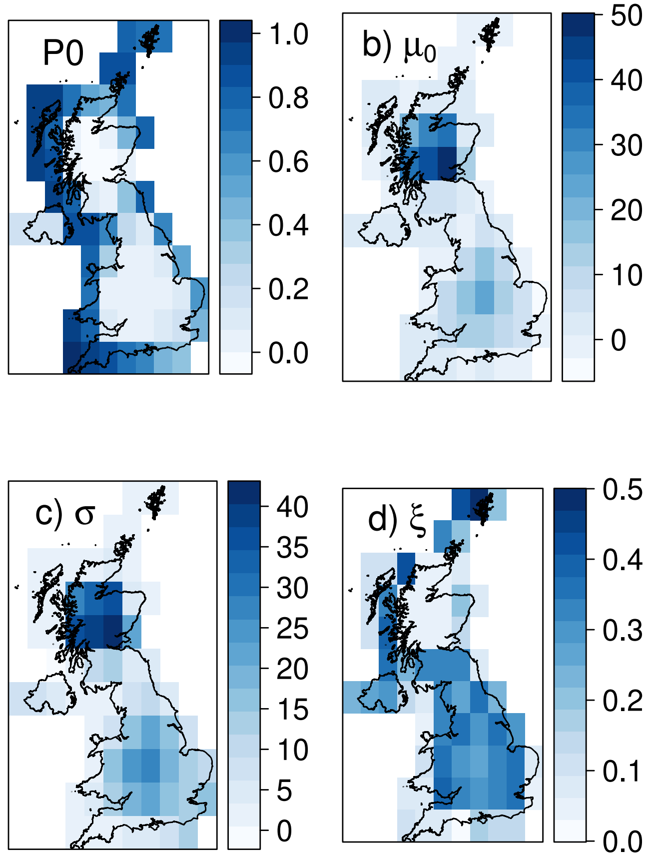

Figure 4. Maps showing the spatial distribution of the model fitted parameters: a) P0 calculated as $e^{b_0}/(e^{b_0}+1)$, b) location μ, c) scale σ, and d) shape ξ.

stationarity representation of the model. The NAOI and the global CO_2 radiative forcing are chosen as covariates. There are some important caveats to this choice. First, other natural factors apart from NAO are not accounted for and hidden co-varying effects might also be present. Also, while CO_2 radiative forcing is linearly related to the equilibrium surface temperature, the relationship to transient surface temperatures further depends on the efficacy of ocean heat uptake (**?**). Both can lead to non-linear responses of the local UK climate, especially when extrapolating far in the future.

Despite the caveats, CO_2 radiative forcing and NAO have some important advantages. First both are accurately measured. Although a model that relies on global mean surface temperature may not have as strong correlations as the casual link is more indirect, it has the advantage that it does not rely on subtle regional patterns that are difficult to capture. They also provide a reasonable way to isolate the human and natural influences on extreme temperatures (see for example **?**). While it would be possible to use a more locally defined metric (such as the change in the mean UK temperature for example), this would include more unforced naturally occurring internal variability of the climate system, making it difficult to identify the changes that are driven by anthropogenic CO_2 emissions. Finally, using a covariate such as the change in CO_2 forcing avoids the difficulty with determining the start of the trend and also results can be easily rescaled to different time period or emission scenario which is helpful for mitigation strategies.

The influence of NAO and of global warming is examined by exploring improvements to the distribution fits, after incorporating linear covariates on the distribution parameters~~.~~ , as follows:

- $ln\left(\frac{P0}{1-P0}\right) = b_0 + b_1 NAOI + b_2 \Delta F_{CO_2}$
- $\mu = \mu_0 + \mu_1 NAOI + \mu_2 \Delta F_{CO_2}$

where (b_0, μ_0) are the stationary model parameter estimates and (b_1, μ_1), (b_2, μ_2) are linear transformations of the covariates NAOI and ΔF_{CO_2} with respect to time, respectively.

~~In this study, only~~ Only non-stationarity with respect to P0 and the location parameter, μ, is discussed, since modeling temporal changes in σ and ξ reliably requires long-term observations in order to be estimated accurately (**??**). ~~A~~ In addition, a simple linear model is selected, as this is usually preferred when searching for trends in the occurrence of extreme events (**?**). Finally, even though some climate modeling studies predict changes in the nature of NAO variability in an increasing CO_2 climate (**??**), the model does not include any interaction-terms, as they have been found to be non-significant.

As before, the parameters of each cell are estimated taking also into account its neighboring cells weighted by their distance. The most pertinent model is selected, for each cell, using the χ^2 test, based on the change in deviance, between the null, one or two predictor model. If the significance value is less than 0.01, the model is estimated to have a significant improvement over the reduced model. A separate test is performed for the P0 and the GEV model. As an example, in the case of the London cell, the model with two predictors for both P0 and the location parameter has been chosen (table **??**).

The spatial distribution of the parameters of the final model is shown in Fig. **??**. Increasing NAOI or ΔF_{CO_2} are consistent with a warming trend, leading to positive values of the P0 parameters (indicating increases in the number of years with no negative temperatures) and to negative values in the location parameters (indicating lower means in the AFI distributions). The NAO is found to affect more cells in total (90%) in comparison to anthropogenic climate change (51%). Notice however that due to the internal variability of the NAO, any signal from a climate change trend can be hidden in the limited observational period.

3.3 ~~Copulas and vine copulas~~Vine copula model

The stochastic behaviour of the hazard (i.e. AFI) at each cell is fully described by its corresponding GEV probability distribution, as described in Sect. **??**. However, insurance portfolio loss analysis requires the calculation of the combined stochastic behaviour of the hazard across all the model domain (i.e. all cells). This is described by the joint distribution of the hazard which, according to Sklar's theorem, can be fully specified by the separate marginal GEV distributions and by their ~~(d-dimensional)~~ copula, which models the hazard dependence between the cells.

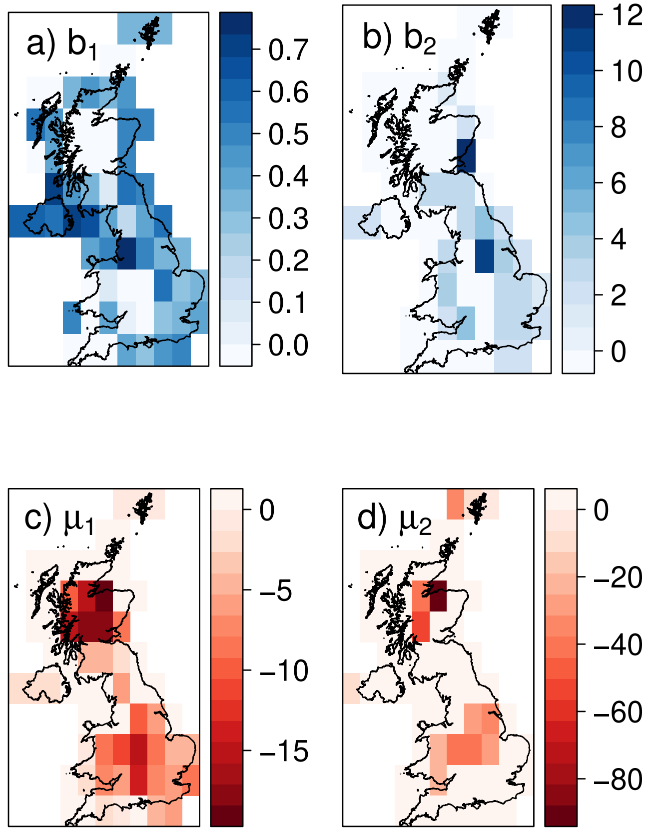

Figure 5. Maps showing the spatial distribution of the non-stationary model parameters: a) b_1, b) b_2, c) μ_1, and d) μ_2. Zero values indicate linear trends not significant at the 0.01 level.

~~More precisely, consider a vector of $X = (X_1, ..., X_d)$ of random variables with a joint probability density function (pdf), $f(x_1, ..., x_d)$. Sklar's theorem (?) states that any multivariate continuous distribution function $F(x_1, ..., x_d)$ with marginals $F_1(x_1), ..., F_d(x_d)$ can be written as:~~

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$$

~~for some appropriate d-dimensional copula C, which is uniquely determined on $0,1^d$.~~

~~The probability density function (pdf) of X, $f(x_1, ..., x_d)$, can be found by taking the partial derivatives with respect to X:~~

$$f(x_1, ..., x_d) = c(u_1, ..., u_d) \prod_{i=1}^{d} f_i(x_i)$$

~~where $c(u_1, ..., u_d)$ is the copula density, given by~~

$$c(u_1, ..., u_d) = \frac{\vartheta^d C(u_1, ..., u_d)}{\vartheta u_1 ... u_d}$$

~~Expression **??** is important in terms of modelling because it permits to define a multivariate density as the product of marginal pdfs and a copula density function that captures the dependence between the random variables (**?**). For a theoretical introduction to copulas, see **????**; for a practical/engineering approach and guidelines, see **????**~~

~~To quantify the dependence between variables, different measures have been defined, addressing different aspects of dependence. A common measure of overall dependence is the Kendall rank correlation coefficient, commonly referred to as Kendall's τ coefficient (**?**). However, dependence of rare events cannot be measured by overall correlations: even if two variables are completely uncorrelated, there can be a significant probability of a concurrent extreme event in the two, i.e., they can still be tail dependent. Tail dependence describes the amount of dependence in the lower tail or upper tail of a bivariate distribution. For its mathematical definition see **?**.~~

~~One important complication is that identifying the appropriate d-dimensional copula is not an easy task. In high dimensions, the choice of adequate families is rather limited (**?**). Standard multivariate copulas, either do not allow for tail dependence (i.e. multivariate Gaussian) or have only a single parameter to control tail dependence of all pairs of variables (Student-t and Archimedean multivariate copulas). This is particularly problematic for catastrophe modeling applications, where a flexible modeling of tails is vital to assess reliably the extreme behaviour of natural events.~~

Vine copulas provide a flexible solution to this problem based on a pairwise decomposition of a multivariate model into bivariate (conditional and unconditional) copulas, where each pair-copula can be chosen independently from the others. ~~In particular, asymmetries and tail dependence can be taken into account as well as (conditional) independence to build more parsimonious models. Vines thus combine the advantages of multivariate copula modeling, that is separation of marginal and dependence modeling, and the flexibility of bivariate copulas (**?**).~~ A brief introduction on the vine copula methodology can be found in Appendix **??**.

~~As an example, in a 4-dimensional case, the joint pdf can be decomposed as a product of 6 pair-copulas (3 uncoditional and 3 conditional) and 4 marginal pdfs as shown in Eq. (**??**):~~

$$\begin{aligned}
f(x_1, x_2, x_3, x_4) = & f(x_1) f(x_2) f(x_3) f(x_4) \\
& \times c_{12}(F_1(x_1), F_2(x_2)) \\
& \times c_{23}(F_2(x_2), F_3(x_3)) \\
& \times c_{34}(F_3(x_3), F_4(x_4)) \\
& \times c_{13|2}(F_{1|2}(x_1 \mid x_2)), F_{3|2}(x_3 \mid x_2))) \\
& \times c_{24|3}(F_{2|3}(x_2 \mid x_3)), F_{4|3}(x_4 \mid x_3))) \\
& \times c_{14|23}(F_{1|23}(x_1 \mid x_2, x_3)), F_{4|23}(x_4 \mid x_2, x_3)))
\end{aligned}$$

~~The above decomposition is not unique and **?** introduced a graphical structure called regular vine (R-Vine) structure to represent this decomposition with a set of nested trees. The dependence structure with three trees for the 4-dimensional example above is shown in Figure **??**. More details on vine copulas can be found in **????**.~~

~~Example of 4-dimensional R-Vine trees corresponding to the decomposition shown in Eq. (**??**).~~

3.3.1 ~~Selection of the Regular Vine Model (RVM)~~

In this study, the joint multivariate hazard distribution of AFI across all the model domain (67 cells) is decomposed as a product of marginal and pair-copula pdfs~~(in a similar way as shown for the 4-d case above). F(x) and f(x) represent the marginal GEV distributions here as defined by Eq. (??) and (??).~~. The pair-copulas are fitted using the R (https://www.r-project.org/) package VineCopula (**??**). The method follows an automatic strategy of jointly searching for an appropriate R-Vine tree structure, its pair-copula families, and estimating their parameters developed by **?**. This algorithm selects the tree structure by maximizing the empirical Kendall's τ values, based on the premise that variable pairs with high dependence should contribute significantly to the model fit and should be included in the first trees.

The copula family types for each selected pair in the first tree are determined by using the Akaike information criterion (**?**). For computational reasons, the two-parameter Archimedean copulas are excluded from this analysis, which however has only a negligible impact in the results (not shown). The copula parameters are estimated sequentially (using maximum likelihood estimation) starting from the top tree until the last tree, as described in **?**. This approach only involves estimation of bivariate copulas and has been chosen since it is computationally much less demanding than joint maximum likelihood estimation of all parameters at once.

The percentage of family types used for the first few trees of the selected RVM is shown in Table **??**. The large majority of the pairs in all trees are estimated to be independent (59%), but these pairs occur mainly at the higher trees, since the most important dependencies are captured in the first trees (**??**). Large dependencies, with Kendall's tau coefficients greater than 0.90, are found as expected between neighboring cells, but remain important across the whole model domain due to the nature of the hazard: AFI assess the freezing temperatures during the entire winter and, thus, is less associated with small scale local phenomena that can cause important spatial variation. At the first tree, 52% of the selected bivariate copulas are found to belong to the t-Student Copula and 35% to the Gumbel family, which exhibit positive dependence in the tails. Gumbel in particular has a greater dependence in the positive tail than in the negative and thus implies greater dependence at larger AFI values than at lower ones. From the third tree and onwards, the percentage of independent families is always larger than 40%.

The small sample size used (110 years of data) in conjunction with the high dimensions of the modelled pdf (67) is of concern in this study since this can lead to large uncertainties in the resulting pdf, which can also propagate in the estimated return periods. The impact of the short sample size on the uncertainties in the results is quantified using a bootstrap technique, as described in Sect. ~~**??**~~**??**.

Goodness-of-fit (GOF) is calculated using the Cramer von Mises test, which compares the final selected RVM with the empirical copula. The RVineGofTest algorithm of the same R package implements different methods to compute the test, which however perform usually poorly in cases of small sample sizes and at higher dimensions as is the case for this work

Table 2. Percentage of family types used for the first five trees of the R-Vine Model.

Tree	Indep	Gaussian	Student t	Clayton	Gumbel	Frank	Joe	180°Clayton	180°Gumbel	180°Joe	90°Clayton	90°Gumbel	90°Joe	270°Clayton	270°Gumbel	270°Joe
1	0	3.0	51.5	0	34.8	1.5	1.5	1.5	0	6.1	0	0	0	0	0	0
2	9.2	4.6	36.9	3.1	6.2	16.9	1.5	3.1	1.5	0	1.5	4.6	7.7	0	0	3.1
3	25	1.6	31.2	4.7	1.6	7.8	1.6	0	1.6	9.4	6.2	3.1	1.6	1.6	0	3.1
4	27	6.3	28.6	4.8	1.6	9.5	4.8	3.2	1.6	4.8	1.6	1.6	1.6	0	1.6	1.6
5	27.4	8.1	24.2	4.8	1.6	9.7	1.6	6.5	6.5	1.6	0	1.6	1.6	1.6	0	3.2
All	59.2	3.9	9.4	2.5	2.1	8.9	1.4	2.5	1.0	1.4	1.6	0.6	1.3	1.8	0.8	1.4

Table 3. Goodness-of-fit values for the Cramer von Mises statistic based on the empirical copula process (ECP) and based on the combination of probability integral transform and empirical copula process (ECP2) as implemented in the VineCopula R package.

Method	CvM	p.val
ECP	9.1	0.7
ECP2	0.009	1

(**?**). Nevertheless, ~~table~~ Table **??** shows the GOF results for two of these methods. The p.value is found to be larger than 0.05, which is an indication that the fitted RVM cannot be rejected at a 5% significance level. However, given also the quite large p.values, a Type II error cannot be excluded. Nevertheless, the suitability of the model, in comparison to the empirical data, is further discussed in the the results section as well.

3.3.1 ~~Stochastic simulation and uncertainty estimation via parametric bootstrap~~

~~The RVM is~~ In the case of the stationary model, the vine copula is employed to model the entire spatial dependence of the AFI in the UK. On the other hand, the spatial AFI structure in the case of the non-stationary model is modelled in two ways: a) by quantifying the dependence on NAO/CO_2 in each location, treating each location as conditionally independent, then inducing spatial dependence through the variation of NAO/CO_2 and b) by fitting the RVM model to all the residual dependencies associated with the AFI between the cells; these account for dependencies between cells resulting from other large-scale circulation patterns and also regional climate variability (e.g. due to effects of local orography, land-sea contrast, and small scale atmospheric features such as convective cells). Notice, however, that the effect of NAO/CO_2 on the residual hazard dependency structure is not taken into account here. Recently, a methodology that offers the possibility to include such meteorological predictors in a vine copula model has been developed by **??** and is something to be addressed in a future study.

3.4 Stochastic simulation and uncertainty estimation via parametric bootstrap

The pdf is used to simulate 100K years of winter-seasons in the UK. For each year, the simulated AFI values at each grid cell depend on the other cells based on the fitted RVM. ~~Performing long enough simulations is necessary in order to obtain~~

~~converged numerically~~ Long simulations are needed to obtain numerically converged results, i.e. to convergence to the "true" return period. Our focus here is the 200 year RP, which is commonly associated with capital and regulatory requirements. By repeating the simulation several times, it has been assessed that 100K years of winter seasons is long enough ~~and~~ to neglect the Monte Carlo ~~simulation error is negligible.~~

uncertainty. The stationary model is used to generate a stochastic set which corresponds to the current hazard experience. The non-stationary model permits us to create additional stochastic sets that represent different climate conditions. In ~~the case of NAO and following the Shapiro–Wilk test for normality, a 100-year long NAOI has been simulated assuming a Gaussian distribution (see Figure ??). In~~ order to assess the influence of climate change in UK cold spells, three separate stochastic sets, of 100K years each, have been created as follows:

- Pre-industrial climate (ΔF_{CO_2}=0 Wm^{-2}), corresponding to pre-industrial (1900) concentration of CO_2 (296 ppm).
- Current climate (ΔF_{CO_2}=1.6 Wm^{-2}), corresponding to a present day (2018) concentration of CO_2 (400 ppm).
- Future ~~double-CO_2~~ climate (ΔF_{CO_2} = ~~3.7~~ 2.0 Wm^{-2}), corresponding to ~~2 x~~ the future year 2030 concentration of CO_2 ~~concentration since pre-industrial times (592 ppm)~~(435 ppm), according to the RCP4.5 emission scenario.

The choice of year 2030 assures a relative close time distance which is more relevant for the insurance industry (**?**). At the same time, extrapolating far in the future is particularly problematic, since it relies on the assumption of the stability of these linear relationships, even though they may be significantly altered by changing boundary conditions. The change in the CO_2 radiative forcing is calculated here based on the RPC4.5 scenario (2 W m^{-2}), but similar values are projected for 2030 by the other RCP scenarios as well (**?**). Each year of the three stochastic sets above is associated with a random NAOI value that has been simulated assuming a Gaussian distribution, fitted to the historical NAOI dataset (see Fig. **??**). The influence of NAO on each one of these sets can thus be discerned by selecting only the simulated years with negative or positive NAOI values.

The small sample size used in this study (110 years of data) together with the high dimensions of the modelled pdf (67) can lead to large uncertainties in the estimated return periods. Following **?**, the model uncertainty is assessed using a parametric bootstrap approach, where a large number of models are created using as basis, instead of observations, randomly simulated data from the selected RVM. In particular, confidence intervals are constructed as follows:

- A simulation with the same length as the observed data (i.e. 110 years) is repeated for B = 500 times.
- For each of these B = 500 samples, a new full model is fitted (including new GEV and logistic regression model parameters at each cell and new RVM structure, pair-copula families and parameters) following the methodology described in Sect. ~~**??** and **??**~~**??**.
- For each of the resulting B = 500 RVMs, a simulation of 10K years of winter-seasons is performed. The uniform variables are then transformed using the (new) inverse marginal pdfs and the corresponding return period levels are estimated.
- The uncertainty in the return levels is estimated by identifying the 95% confidence interval (i.e. the range 2.5–97.5 %) from these 500 return level curves.

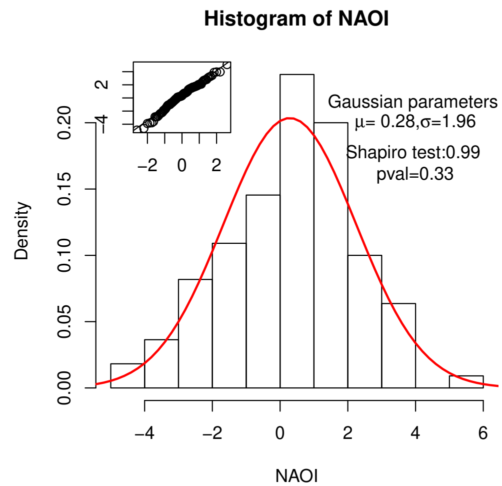

Figure 6. Histogram of the NAOI and the pdf of the fitted Gaussian distribution (red line).

Due to computational constraints, confidence intervals are computed only for the stationary model~~and~~. In addition, the simulation length has been reduced to 10K years (instead of 100K), which implies that part of the calculated uncertainty is due to Monte Carlo sampling variability. In order to ~~separate the uncertainty~~ investigate further the sources of this uncertainty, the uncertainty associated with the RVM only is separated from the uncertainty of the full model, i.e. of the joint pdf, ~~confidence intervals have been also calculated~~ by calculating confidence intervals with the same approach as described above, but using the same marginal pdfs in each bootstrap repetition.

4 Results and discussion

4.1 Return period maps

The obtained stochastic sets (see Sect. ~~??~~??) are used to create return period maps for the different climatic conditions. The top panels of Fig. **??** represent the AFI values that occur once every 10, 25, and 50 years based on the stationary model. The empirical return periods are also plotted for comparison (bottom panels). These are calculated for each cell as 1/(1-P), where P represents the cumulative probabilities of the ranked values and is calculated based on the Weibull formula P=i/(n+1) (**?**). The spatial pattern is consistent between the empirical and stochastic sets, showing largest AFI values in high elevation areas, as expected. However, the empirical values are in general somewhat larger than the stochastic set. This difference is driven by the exceptional 1962/63 event which is estimated empirically at 1 in 110 years but is predicted to be less frequent according to the GEV fits. The probability of such an event happening today is discussed in detail in Sect. **??**.

~~(Top panels) Maps of stochastic AFI values (in C) for return periods of a) 1 in 10, b) 1 in 25, and c) 1 in 50 years. (Bottom panels) Maps of the corresponding empirical AFI values.~~

Return period maps at higher return periods (100, 200, and 500 years) for the pre-industrial, current, and ~~2xCO_2~~ future climate stochastic sets are shown in Fig. ~~??~~??. UK in the beginning of the 20th century has been experiencing much colder winters than today. ~~In a 2xCO_2~~ By 2030, the future climate change scenario, ~~negative temperatures become very rare (larger than 1 in 500 years) in large part of the UK~~ extreme cold events with AFI larger than 100 °C become quite rare (above 100 years RP) everywhere except at mountainous regions. ~~It is also interesting to note that at~~ At high return periods and across all scenarios, the model predicts larger AFI values ~~are predicted~~ for the southern part of UK in comparison to the north. The extreme AFI values in the south are driven by the exceptional 1962/63 winter which has been more severe in the South than the North (see Fig. **??**b). Excluding this winter from the analysis results in almost much lower AFI values in most of the region (not shown).

4.2 Regional return period AFI curves

The vine copula methodology permits the estimation of the hazard return periods over aggregated regions in the UK. Since our focus is mainly on inhabited areas, for each simulation year (y) and for each region, the "weighted AFI" (wAFI) is computed, where the AFI value at each cell j is weighted by the corresponding number of residential properties (n_j), as shown in Eq. (**??**). The weighted AFI thus places more weight on the hazard over large populated urban areas than agricultural or mountainous areas. The number of residential properties in the UK is taken from the PERILS Industry Exposure Database (https://www.perils.org/), which contains up-to-date high quality insurance market data at Cresta level ("Catastrophe Risk Evaluation and Standardizing Target Accumulations", https://www.cresta.org/) based on data directly collected from insurance companies writing property business in the UK. Return period wAFI curves for both the empirical and the stochastic data are shown in Fig. **??**. Analogous return period plot based on mAFI, i.e. without weighting, can be found in the Appendix (Fig. **??**).

$$wAFI_y = \frac{\sum AFI_{j,y} \cdot n_j}{\sum n_j} \tag{8}$$

4.2.1 Model uncertainty

The stationary model is utilized to analyze the uncertainty in the model results and investigate its sources. Fig. **??** shows the empirical and the stochastic return period curves of wAFI for the entire UK, together with their associated uncertainties. The empirical return periods calculation is described in Sect. **??**, while their uncertainty intervals are computed from the 2.5^{th} and 97.5^{th} quantile of the beta probability distribution function (**?**). The stochastic curve and confidence intervals are computed as described in Sect. ~~??~~??. The uncertainty in the model is found to be large, only marginally lower, then the empirical estimates and is associated to the short historical record length. Most of the uncertainty (around 90% for RPs greater than 50 years) ~~originates~~ appears to originate from the uncertainty in the GEV distribution parameters, with the remaining 10% to be due to the RVM model (dark shaded area in Fig. **??**). Extreme-value theory is considered as a state-of-the-art procedure to find values for return periods that amply exceed the record length and has been utilized in this study. However, a common difficulty with extremes is that, by definition, data is rare and as a result, the shorter the record length, the more inaccurate is the estimation of

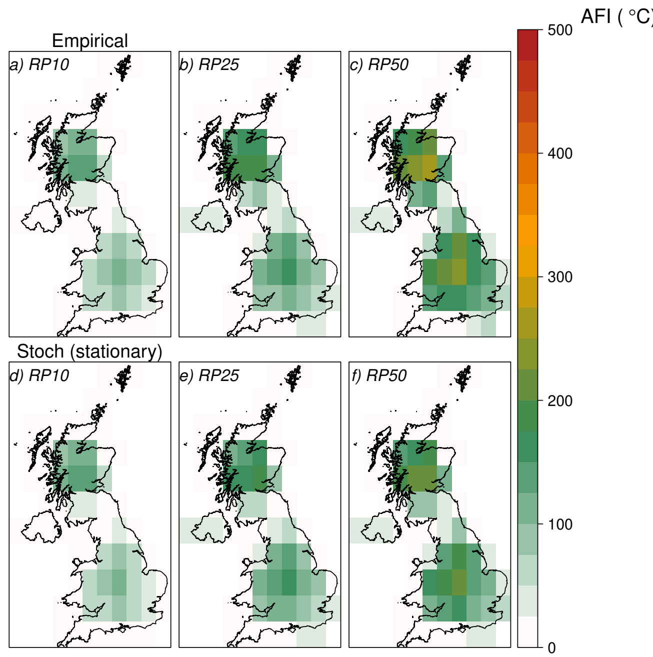

Figure 7. (Top panels) Maps of stochastic AFI values (in °C) for return periods of ~~1 in 100, 1 in 200, and 1 in 500 years for pre-industrial (top panels~~a) 10, ~~current (middle panels~~b) 25, and ~~2xCO_2~~ c) 50 years. (~~bottom~~ Bottom panels) ~~climate~~Maps of the corresponding empirical AFI values.

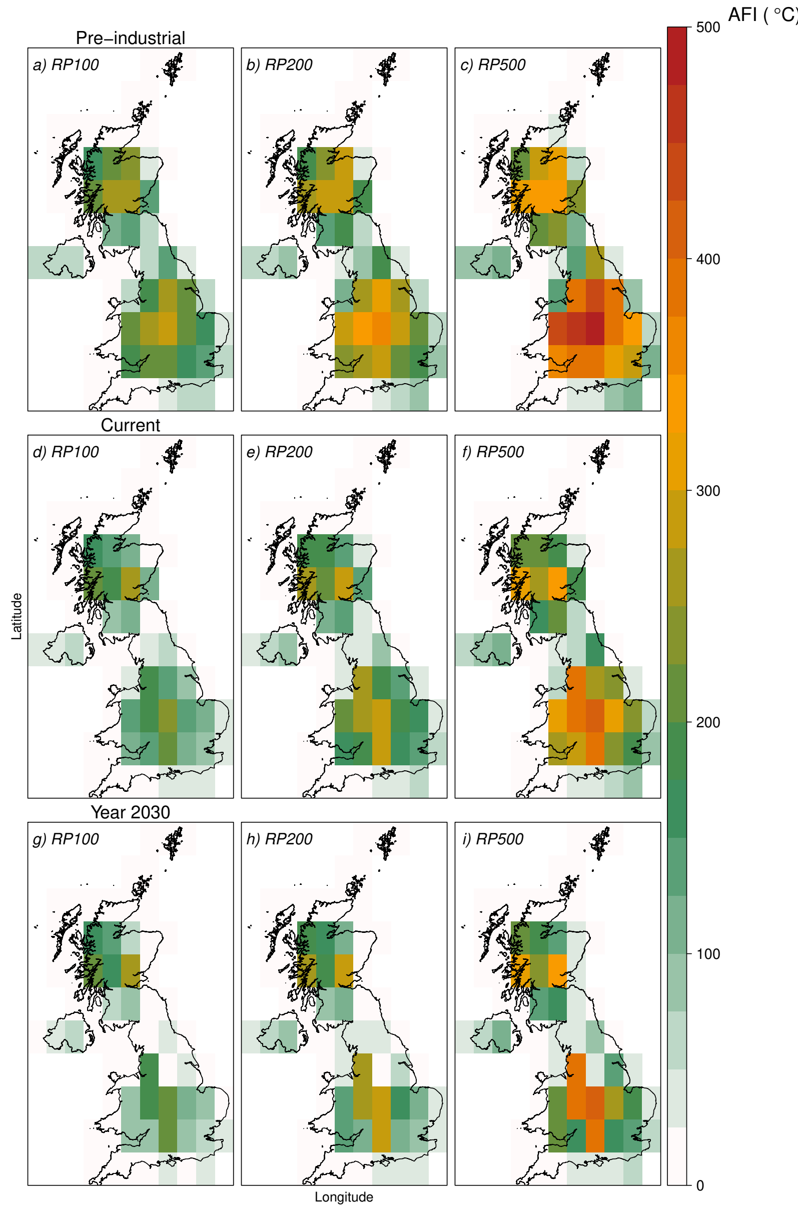

Figure 8. Maps of stochastic AFI values (in °C) for return periods of 100, 200, and 500 years for pre-industrial (top panels), current (middle panels), and future (bottom panels) climate.

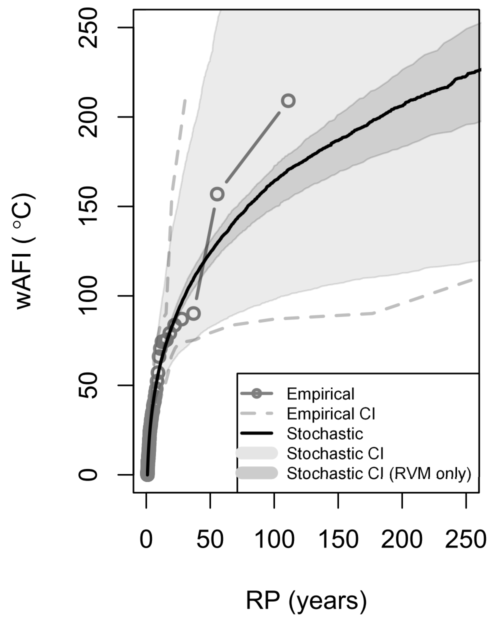

Figure 9. Return period curves of wAFI (in °C) based on the historical data (grey) and the stochastic model (black). The 95% confidence intervals are shown as dashed grey lines for the historical data and as a shaded grey area for the stochastic model. The dark shaded area represents the stochastic uncertainty due to the RVM model alone.

the GEV parameters. The results presented in the following sections should therefore be interpreted being aware of the existing uncertainties.

4.2.2 The 1962/63 winter return period and climate change influence

Return period curves for the stochastic sets under pre-industrial, current and future climate conditions are shown in Fig. ??. The 1962/63 winter, with a wAFI of 209 °C, has been the coldest in the reanalysis data in the UK and, thus, it is estimated empirically as a 1 in 110 years event (i.e. the length of the data set). This corresponds well with the Central England Temperature (CET) record, the oldest continuously running temperature data set in the world (?). According to the latter, only two other winters (1683/84 and 1739/40) have been colder than 1962/63 in the last 350 years, suggesting a return period in the range of 110-120 years, as well. The stationary model ~~suggests a longer return period of~~ overestimates this winters' return period which is estimated to 205 years ~~for~~ across all the UK ~~. In particular~~ (table ??). Especially in the South of the UK the model suggests that this event has been particularly unusual. In the Northern part of UK on the other hand, the model suggests a lower return period of 106 years, closer to the empirical estimate.

Table 4. Return period estimates (in years) for the 1962/63 winter freeze event, based on wAFI.

Method	All UK	South UK (<55°N)	North UK (>55°N)
Empirical	110	110	110
Stationary stochastic set	205	213	106
Non-stationary stochastic sets			
pre-industrial	204	209	102
1960s	216	219	112
current	433	442	222
~~2xCO_2~~ 2030s	~~4284~~ 788	~~4140~~ 789	~~5394~~ 400

The non-stationary model suggests that under current climate conditions, such an extreme event, is approximately ~~2~~ two times less likely to occur than in the 1960s~~(table ??)~~. This agrees with **?** who used climate model simulations to demonstrate that cold December temperatures in the UK are now half as likely as they were in the 1960s. **?** also indicate that human influence has reduced the probability of such a severe winter in UK by at least 20% and possibly by as much as 4 times, with a best estimate that the probability has been halved. On the other hand, some recent studies have argued that warming in the Arctic could favor the occurrence of cold winter extremes, and might have been also responsible for the unusually cold winters in the UK of 2009/10 and 2010/11 (**??**). This hypothesis though is still largely under debate, see for example **?** and **?**.

~~Under a 2xCO_2 climate, a~~ By the year 2030, an event of the same severity as 1962/63, is predicted to become almost ~~10 times more~~ twice less infrequent, having a return period of ~~around 1 in 4000~~ 788 years. Fig. **??**a shows an important reduction in the probability of occurrence of cold extreme events across the whole distribution as a result of the increase of anthropogenic CO_2 concentrations. Larger reductions are found for the most extreme events as well, which is probably related to the large increase of the probability of no negative temperatures (P0) for several cells especially around the coast (see Fig. **??**). Similar results are found in both the northern and the southern part of UK, as well (Fig. **??**b).

4.2.3 NAO influence

The profound effect of NAO on the winter surface temperature over UK has been reported by several studies (**???**). In conjunction with those, the model predicts a negative (positive) NAO phase increases (decreases) substantially the probability of a cold event in the UK~~(~~. Fig. **??** ~~). In fact, on~~ shows the RP curve of current climate wAFI, alongside with RP curves computed solely from simulated years with NAOI values greater than 1 (i.e. representing the positive NAO phase) or years with NAOI values lower than 1 (i.e. representing the negative NAO phase). On average, extreme cold winters are estimated to occur approximately 3 to 4 times more likely during the negative than the positive phase. As an example, an event with wAFI of 100 °C has a return period of ~~1 in~~ 39 years, assuming a negative phase, and 1 in 133 years, assuming a positive phase. Because of its intrinsic chaotic behaviour, NAO is difficult (if even possible) to be predicted (**?**). Nevertheless, numerical seasonal forecast

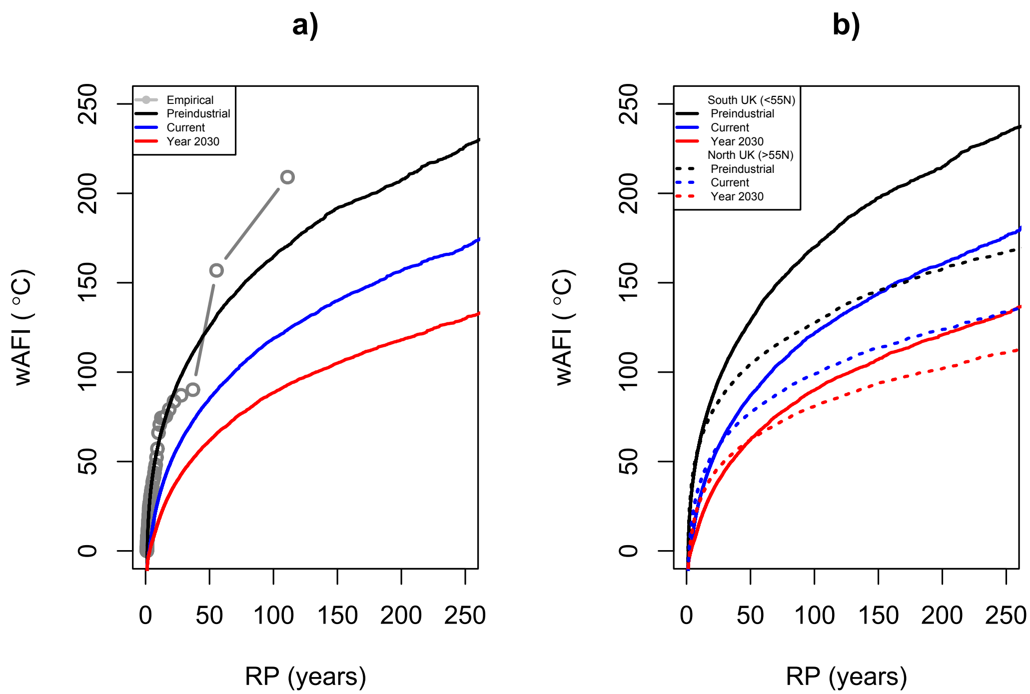

Figure 10. a) Return period curves of wAFI (in °C) based on the historical data (grey) and the stochastic model for three different climate conditions: pre-industrial (black), current (blue), and ~~2xCO_2~~ future (red).

b) Same as (a) but separated between South UK (full lines) and North UK (dashed lines). Only stochastic sets are shown.

systems are currently rapidly improving and have even shown some success in the past (**??**). Incorporating such information in models could be very useful from the catastrophe risk management perspective.

~~It is important to note that~~

As already mentioned, the effect of NAO or CO_2 radiative forcing in the hazard dependency structure has not been taken into account here ~~. Recently, a methodology that offers the possibility to include such meteorological predictors in a vine copula model has been developed by **??**~~ and is something to be addressed in a future study. ~~Finally, another~~ Another point that requires further consideration is the mechanisms that control and affect the NAO and its temporal evolution and in particular how the NAO responds to external CO_2 forcing ~~(**?**, e.g.)~~(**?**).

~~Return period curves of wAFI (in C) based on the current climate stochastic model and assuming a variable NAOI as described in the text (black line). Return period curves based on negative (lower than -1) and positive (larger than 1) values of NAOI are shown with green and orange lines, respectively.~~

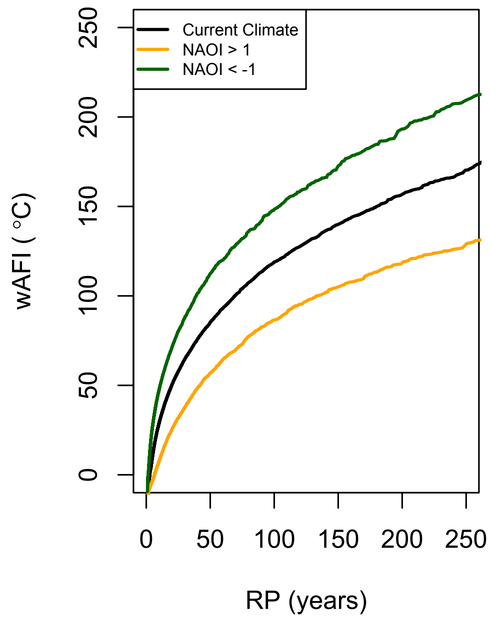

Figure 11. Return period curves of wAFI (in °C) based on the current climate stochastic model and assuming a variable NAOI as described in the text (black line). Return period curves based on negative (lower than -1) and positive (larger than 1) values of NAOI are shown with green and orange lines, respectively.

5 Conclusions

This paper presents a probabilistic model of extreme cold winters in the United Kingdom. The hazard is modeled using the Air Freezing Index, an index which takes account both the magnitude and the duration of air temperature below freezing and is calculated from the ERA-20C reanalysis temperature data covering the period from 1900 to 2010. Extreme value theory has been applied in order to estimate the probability of extreme cold winters spatially across the UK. More importantly, the spatial dependence between regions in the UK has been assessed through a novel approach which takes advantage of the vine copula methodology. This approach allows the modeling of concurrent high AFI values across the country which is necessary in order to assess reliably the extreme behaviour of ~~such~~ freeze events.

Recognizing the non-stationary nature of climate extremes, the model also incorporates NAO and climate change effects as predictors. Stochastic sets of 100K years representing different climate conditions (i.e. pre-industrial, current, or future climate and positive or negative NAO) have been generated and the return periods of extreme cold winters in UK, such as the "Big Freeze of 1962/63" have been estimated. According to the model, ~~such a cold winter is estimated to occur once every~~

~~approximately 400 years under current climate conditions in the UK. The~~ the occurrence of such an event is calculated to have ~~increased~~ decreased approximately two times during the course of the 20th century as a result of anthropogenic climate change. ~~Moreover, the model predicts that such an event will become quite uncommon and occur even more rarely, about 10 times less frequently , under 2xCO~~The model further predicts that by 2030s, extreme cold winters will become even more uncommon and will occur about twice less frequently under the influence of increasing CO_2 ~~climate conditions~~concentrations. The frequency of extreme cold spells in UK has been found to be heavily modulated by NAO, as well. A cold event is estimated to occur $\approx$3-4 times more likely during the negative than the positive phase.

However, considerable uncertainty exists in these estimates which should be interpreted with caution. The 110-year reanalysis record used in this study ~~has been~~ is estimated to be short ~~in order to estimate with enough confidence the frequencies of such extreme events~~and the level of uncertainty in extremal estimates with long return periods is high. Additional uncertainty may also be introduced by possible spurious trends in the reanalysis data set. A longer record of temperature data would be necessary in order to reduce the uncertainty and high quality long-term reanalysis products ~~including ensemble approaches~~ with multiple ensemble members could help towards this direction.

Appendix A: Copulas and vine copulas

According to Sklar's theorem, the joint multivariate distribution of a set of d random vectors can be fully specified by the separate marginal distributions and by their (d-dimensional) copula, which defines the dependence structure between them. More precisely, consider a vector of $X = (X_1,...,X_d)$ of random variables with a joint probability density function (pdf), $f(x_1,...,x_d)$. Sklar's theorem (**?**) states that any multivariate continuous distribution function $F(x_1,...,x_d)$ with marginals $F_1(x_1),...,F_d(x_d)$ can be written as:

$$F(x_1,...,x_d) = C(F_1(x_1),...,F_d(x_d)) \tag{A1}$$

for some appropriate d-dimensional copula C, which is uniquely determined on $[0,1]^d$.

The probability density function (pdf) of X, $f(x_1,...,x_d)$, can be found by taking the partial derivatives with respect to X:

$$f(x_1,...,x_d) = c(u_1,...,u_d)\prod_{i=1}^{d} f_i(x_i) \tag{A2}$$

where $c(u_1,...,u_d)$ is the copula density, given by

$$c(u_1,...,u_d) = \frac{\vartheta^d C(u_1,...,u_d)}{\vartheta u_1 ... u_d} \tag{A3}$$

Expression **??** is important in terms of modelling because it permits to define a multivariate density as the product of marginal pdfs and a copula density function that captures the dependence between the random variables (**?**). For a theoretical introduction to copulas, see **????**; for a practical/engineering approach and guidelines, see **????**

To quantify the dependence between variables, different measures have been defined, addressing different aspects of dependence. A common measure of overall dependence is the Kendall rank correlation coefficient, commonly referred to as Kendall's τ coefficient (**?**). However, dependence of rare events cannot be measured by overall correlations: even if two variables are completely uncorrelated, there can be a significant probability of a concurrent extreme event in the two, i.e., they can still be tail dependent. Tail dependence describes the amount of dependence in the lower tail or upper tail of a bivariate distribution. For its mathematical definition see **?**.

One important complication is that identifying the appropriate d-dimensional copula is not an easy task. In high dimensions, the choice of adequate families is rather limited (**?**). Standard multivariate copulas, either do not allow for tail dependence (i.e. multivariate Gaussian) or have only a single parameter to control tail dependence of all pairs of variables (Student-t and Archimedean multivariate copulas). This is particularly problematic for catastrophe modeling applications, where a flexible modeling of tails is vital to assess reliably the extreme behaviour of natural events.

Vine copulas provide a flexible solution to this problem based on a pairwise decomposition of a multivariate model into bivariate (conditional and unconditional) copulas, where each pair-copula can be chosen independently from the others. In particular, asymmetries and tail dependence can be taken into account as well as (conditional) independence to build more parsimonious models. Vines thus combine the advantages of multivariate copula modeling, that is separation of marginal and dependence modeling, and the flexibility of bivariate copulas (**?**).

As an example, in a 4-dimensional case, the joint pdf can be decomposed as a product of 6 pair-copulas (3 uncoditional and 3 conditional) and 4 marginal pdfs as shown in Eq. (**??**):

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = {} & f(x_1)f(x_2)f(x_3)f(x_4) \\ & \times c_{12}(F_1(x_1),F_2(x_2)) \\ & \times c_{23}(F_2(x_2),F_3(x_3)) \\ & \times c_{34}(F_3(x_3),F_4(x_4)) \\ & \times c_{13|2}(F_{1|2}(x_1 \mid x_2)),F_{3|2}(x_3 \mid x_2))) \\ & \times c_{24|3}(F_{2|3}(x_2 \mid x_3)),F_{4|3}(x_4 \mid x_3))) \\ & \times c_{14|23}(F_{1|23}(x_1 \mid x_2,x_3)),F_{4|23}(x_4 \mid x_2,x_3))) \end{aligned} \tag{A4}$$

The above decomposition is not unique and **?** introduced a graphical structure called regular vine (R-Vine) structure to represent this decomposition with a set of nested trees. The dependence structure with three trees for the 4-dimensional example above is shown in Fig. **??**. More details on vine copulas can be found in **????**.

A1

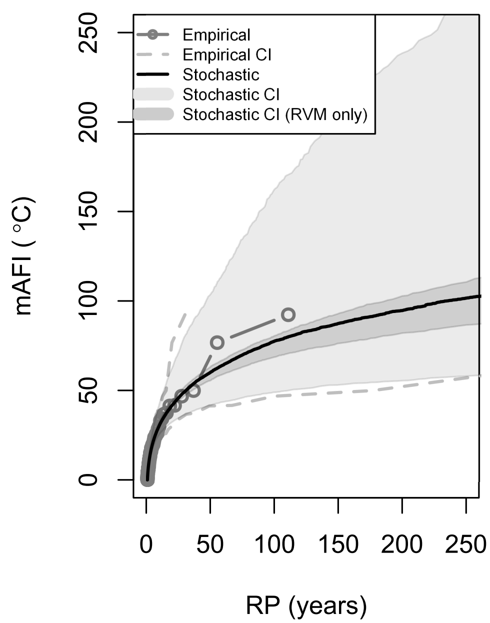

Figure A1. Similar to Fig. **??** but for mAFI (without weighting, in °C).

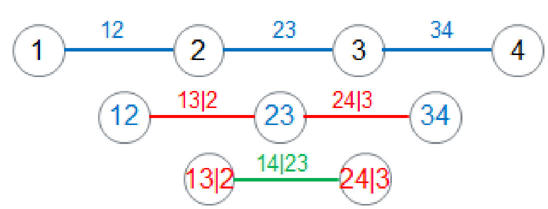

Figure A2. Example of 4-dimensional R-Vine trees corresponding to the decomposition shown in Eq. (**??**).

Disclaimer. TEXT

Acknowledgements. TEXT