

## **Interactive comment on “Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios” by Aloïs Tilloy et al.**

We thank both reviewers for their time in reading and commenting critically on our manuscript. We believe that the changes to our manuscript based on the reviewer comments will significantly improve it, and we hope this will be of use to the broader natural hazards’ community. Below we respond in detail to each reviewer, first outlining the comments, and then our reply.

Lines numbers referred to by reviewers are from our original manuscript, and line numbers referred to in our replies are from our revised manuscript.

We also attach to this reply a track change document of the changes made to date, with green indicating items that have been ‘moved’ from one section to another section.

## Authors Reply to Anonymous Referee #1

### [R1 Introduction]

The paper assesses the capability of six bivariate statistical approaches to model 60 distinct synthetic (bivariate) datasets, some of which possess asymptotic independence and others asymptotic independence. The results are used to develop a systematic framework for selecting among the competing statistical models. The framework is then demonstrated by way of two real world examples. The framework offers a novel approach for selecting among multivariate models. The manuscript is generally well written and relevant to the topic of “Advances in extreme value analysis and application to natural hazards”. In the opinion of this reviewer it is therefore worthy of publication in this special issue of Natural Hazards and Earth System Science. Nevertheless, the manuscript would benefit from a slight reorganization, additional explanation at certain points and a thorough review of grammar and spelling.

[AR to General Comments] We thank the Anonymous Reviewer 1 (R1) for this summary, and suggestion of reorganization, additional explanation and thorough review of grammar and spelling. Based on these comments, we have revised the manuscript, particularly restructuring parts of our manuscript. Below, we provide detailed replies to the reviewer 1 comments (Author Replies—AR).

### [R1 Specific comments]

[R1 Comment 1] The abstract is missing a sentence describing the link between the synthetic datasets part and the application of the framework to environmental data (i.e., the work on the synthetic datasets aids in the creation of the framework). At present, the abstract states that the benefit of a systematic modeling framework are highlighted without any introduction/description of the framework.

[AR 1] We agree with the reviewer about the fact that a linking sentence is missing in the abstract. We therefore modified the abstract to define with more clarity the framework and its purpose (L10-11) and (L12-13).

[R1 Comment 2] The manuscript would benefit from a more precise definition of what constitutes a hazard in relation to other recent literature on compound and cascading hazards. For instance, Zscheischler et al. (2018) defines compound events as “The combination of multiple drivers and/or hazards that contributes to societal or environmental impacts.”. Therefore, they may consider rain, lightening and hail as drivers and a landslide as a hazard, whereas here all four are considered as hazards.

[AR 2] We thank the reviewer for this comment as we agree on the importance of defining what constitutes a hazard. We recognize that there are different definitions. Here, the term hazard will follow the definition by UNISDR (2009), which refers to a natural hazard (hereafter referred to as a ‘hazard’) as a natural process or phenomenon that may have negative impacts on society. As hail, lightning and extreme rainfall fall in the category of “hazards”. A sentence has therefore been added to the introduction.

[R1 Comment 3] P2 L49 to P3 L83 in the introduction focuses on methodology and introduces specific subsections of the methodology section before the methodology section is introduced in the final (roadmap) paragraph at the end of the introduction. I recommend moving the text to the start of the methodology section. Perhaps, a paragraph giving a broad summary of the synthetic dataset work and the modelling framework including the link between them could be added.

[AR 3] We agree with the reviewer on this. Parts of the mentioned paragraph has therefore been moved. Paragraph P2 L52-59 has been moved to the start of the methodology section (P4 L101-111). The sentences P3 L 80-83 is now incorporated in the “roadmap” paragraph (P3 L88-190).

[R1 Comment 4] For the wildfire example, would  $\chi=0.7$  with marginals AC also be a relevant test case.

[AR 4] Thank you for this comment. It also raises the question of the selection of analogous datasets, which is currently quite subjective. To tackle this issue, we decided to add a small tool to help the selection. The tool is based on the empirical estimates of the 2 dependence measures ( $\chi$  and  $\eta$ ) and their uncertainty bounds. It therefore suggests analogous which have a dependence measure within the uncertainty bound. We believe this tool strengthen the whole methodology. Using this tool,  $\chi=0.7$  with AC marginal is not a relevant test case. We also reviewed our previous choices to be in accordance with the selection tool: we added the analogous AB and AC with  $\chi=0.3$  for wildfires and  $\chi=0.1$  AB and BB.

[R1 Comment 5] The (subjective) selection of the AND should also be discussed in the “Choices influencing the results of the simulation study” part of the conclusion.

[AR 5] Very good point. We now state the potential influence of the selection of the “AND” probability in the conclusion (P32 L780-782).

[Reviewer 1 Technical corrections]

[AR to R1 Technical corrections] We thank the Reviewer 1 for these detailed technical corrections. We have below replied to each one and taken on board the vast majority of them.

Abstract: “match” implies that synthetic data was based on specific examples of environment data “be representative of” maybe a more appropriate expression.

We totally agree, thanks, we have now made the change.

P2 L46: “Copulas” refers to a specific type of model whereas “multivariate model” is more general, perhaps replace the latter with a more specific description of the models.

We replace “multivariate models” by “Multivariate extreme models”. We did not find a more specific appellation.

P4 L94-95: The sentence “A theoretical background on extreme value theory is given in Supplement S1.1.” should appear earlier in the paragraph, after the sentence which starts “Extreme Value Theory :::”.

Change done

P4 L106: Typo. Remove “Then”.

[Change done](#)

P4 L116: Typo. Remove “.”.

[Change done](#)

P6 L166: Typo. Replace “F\_X2 (x\_2))” with “F\_X2 (x\_2)”.

[Change done](#)

P6 L170: Use “e.g.” here and elsewhere before citations where there are other relevant papers omitted due to the need for brevity.

[Change done](#).

P7 L175: “Formally, the application of a copula model can be summarized in four main steps”. It should be made clear here and elsewhere that “the application” refers to the application in this study and may vary elsewhere for instance in terms of marginal distributions etc.

[Change done](#)

P7 L181: “joint distribution” may refer to the full multivariate distribution including marginals. Consider changing to “dependence structure” or similar.

[Change done](#)

P7 L210: “extrapolate the conditional model to simulate new extreme data.” Do you do this?

[Not exactly, that sentence was misleading and not well written, it has been rephrased to improve clarity.](#)

P8 L218: Subjective term. Remove “very”.

[Change done](#)

P11 L186: Perhaps change “amount” to “number” to be more specific.

[Change done](#)

P13 L323: “Confront” seems like an unusual term to use here. Consider replacing it with “compare and contrast” or simply remove “confront or”.

[Change done](#)

Table 2: Spelling. Change “thorough” to “throughout”. P13 L334: Typo. “Random variables x”. Consider capitalizing X and Y as they denote random variables.

[Change done](#)

P14 L338: Why characterize log-normal distributors by the coefficient of variation?

[This has been the topic of a long debate between authors. The idea behind using the coefficient of variation is to \(i\) relate to previous literature working with log-normal distribution. \(ii\) characterise the](#)

log normal distributions used with one single parameter instead of two. We clarified this by adding a sentence composed of the two aforementioned points (P14 L370-371).

Table 3: Typo. Replace “dataset” with “datasets”.

[Change done](#)

P16 L379 Typo. Add space “test”.

[Change done](#)

P16 L378 & L379 & L392: Typo. Remove first names “(Arnold, Taylor and Emerson, John, 2011)”.

[Change done](#)

P16 L389: Grammar. “issue is” or “issues are”.

[Change done](#)

P16 L391-393: “The measures mentioned above are not suitable as they imply parametric distributions to be compared against observations (Stephens, 1970; Arnold, Taylor and Emerson, John, 2011). It is then not possible to compare the goodness-of-fit on the whole range of the data.” I am not clear as to the exact limitation being discussed here.

The main limitation being discussed is the need for models to be parametric (e.g., copula) to be compared with the discussed method. As we worked in this study with parametric and non(semi) parametric models, it is not possible to use the previously discussed Goodness of Fit methods. We changed the text to make it clearer.

P16 L399: A description of the reference level curve is required that it comes from the “underlying bivariate ( $X_1, X_2$ ) distribution of the data” as stated in Figure 6 should be added to the main body of text.

[Thank you, Change done](#)

P19 L461: Consider adding “in general” or similar before “outperform all the other: :”.

[Change done](#)

P19 L468-469: Grammar. Rephrase “Gumbel and Galambos copulas show very similar behaviours with respectively 68% [53–93%] and 68% [52–93%] cases with  $wd < 0.1$ .”.

[Change done](#)

P20 L473 Grammar. Change “less” to “least”.

[Change done](#)

P20 L494 I believe “Fig. 6” should be “Fig. 5”.

[Change done](#)

P20 L497 Grammar. Change “abilities of each models for the selected datasets in step (ii)” to “abilities of each model for the datasets selected in step (ii)”.

[Change done](#)

P20 L510 & P21 L519: "London Heathrow airport, UK" and "Heathrow airport (LondonUK)" be consistent with names, only need to specify it is in the UK on first mention.

[Change done](#)

P21 L515: Typo. Add spaces: "Season(Hawkes".

[Change done](#)

P22 L544-545: Repetition: Remove "Daily rainfall data from E-OBS (Cortes et al 2018) and wind gust data (maximum 3 s wind 545 velocity in a day) from the Met Office (2019)." as the information is already given in the text.

[Change done](#)

P23 L555-556: Repetition. Remove "Daily rainfall data from E-OBS (Cortes et al., 2018) and wind gust data (the 555 maximum 3 s wind velocity in a day) from the Met Office (2019)." as the information is already given in the text.

[Change done](#)

P23 L560: I believe "Figs. 9 and 6" should read "Figs. 9 and 5".

[Change done](#)

P23 L565 - 568: Consider replacing circle bullet points with numbers 1-4 so the cases correspond to the relevant rows of Table 4.

[Good idea! Change done](#)

P23 L573: Add "average" before "confidence score".

[We are unclear here if it makes sense to describe the "confidence score" as an average and would require further clarification from the reviewer.](#)

P24 L578: Grammar. Replace "model" with "models".

[Change done](#)

P24 L591: Grammar. Replace "curve" with "curves".

[Change done](#)

P25 L612 & P25 L612: "2395 km<sup>2</sup>" and "1,1778,146" be consistent with use of commas to separate numbers here and throughout the manuscript.

[Change done, Thank you](#)

P27 L643 – 644: Repetition. Remove "Daily mean temperature data from E-OBS(Cortes et al., 2018) and wildfire data from Pereira et al. (2011)." as the information is already given in the text.

[Change done](#)

P28 L653-654: Repetition. Remove “Daily mean temperature data from E-OBS (Cornes et al., 2018) and wildfire data from Pereira et al. (2011). as the information is already given in the text.

[Change done](#)

P28 L660: I believe “Figs. 13 and 6” should read “Figs. 13 and 5”.

[Change done](#)

P28 L666 - 669: Consider replacing circle bullet points with numbers 1-4 so the cases correspond to the relevant rows of Table 6.

[Change done](#)

P28 L666 - 669: Replace “B-C” with “BC”.

[Change done](#)

P30 L713: Typo. Replace “four following” with “following three”.

[Change done](#)

P31 L742-744: Grammar: Change “One of these disadvantages is the that the parametric nature of copulas lead to a lack of flexibility while going to higher dimensionality.” to “One of these disadvantages is that the parametric nature of copula leads to a lack of flexibility when going to higher dimensionality.” or similar.

[Change done](#)

P31 L747: Pair Copula Construction and vine copula are not synonymous rather vine copulas are a type of Pair Copula Construction.

[Thank you for this comment, we changed the sentence accordingly.](#)

P32 L787: Typo. “\_Results” add “.”

[Change done](#)

P32 L783: Perhaps add a sentence to make it clear that the RMSE comes from the calculation of the dependence measures for the 100 realizations of the 60 datasets.

[Change done](#)

P36 L887: Remove “pp.”

[Change done](#)

### Author's Reply to Anonymous Referee #2

[R2 Introduction] This study evaluates the efficacy of various bivariate models for a variety of correlation structures in a large set of synthetic data. While I am not a fan of synthetic data, the rigorous and comprehensive evaluation of various models justify the publication of this paper in NHESS. However, there are some issues that remain to be resolved before final publication:

[AR to General Comments] We thank the Anonymous Reviewer 2 (R1) for this summary. Below, we provide detailed replies to the Reviewer comments (Author Replies—AR).

[R2 Comment 1] 1. There is a large focus on multi-hazard analysis in this paper. However, I cannot see the sector that is impacted by the presented case studies. For a multi-hazard coastal flood analysis, for example, it is intuitive how higher water levels and larger river flows lead to a larger risk of flooding. However, who is impacted by rainfall and wind? What is the final impact that is worsened by concurrence and sub-sequence of multiple hazards? Even more confusing is the wildfire case study. Wildfire is the impact of high-temperature, hence I am not clear whether it is justified to consider the wildfire and temperature as two drivers of one impact. What is going to be that impact? How is it worsened by the combination of these two hazards? The manuscript should justify the presented multi-hazard analyses.

[AR 1] Thank you for these thoughtful comments, which we respond to here including how we have modified our manuscript in response.

- To put this into context, two authors of the manuscript work for an energy company and the original framework of carrying out this research was within problems that might be of interest for the energy sector. We therefore better highlight the potential impact to society and energy infrastructure in our revised manuscript (P3 L76-84).
- We agree with the reviewer that while compound flooding is measurable in water height (and thus the potential impacts this might have on various types of infrastructure), the potential impact of compound extreme wind and extreme rainfall is harder to quantify. In the context of energy infrastructure, the combination of extreme rainfall and extreme wind can lead, for example, to power cuts, wind destroying roof leading to greater damages, damage to energy production infrastructure (due to wind) with difficulties to repair the network arising from high surface runoff/flood due to extreme rainfall. For this point (and the next one), a paragraph is added at the end of the introduction to justify the presented multi-hazard analysis and provide better context (P3 L76-84).
- In the case of wildfire and extreme air temperature, we consider both of these hazards (UNISDR, 2009). Extreme hot temperature can lead to damage on infrastructure (e.g., rail track deformation) and cause death of people (e.g., 2003 heatwave in France). In term of impact of wildfire and extreme temperature, we argue that their interrelation has the potential to put overwhelming pressure on critical infrastructure (e.g., health service, firefighters). That being said, this article is focusing on natural hazards and their interrelations and not on the measure of a combined impact, which can differ depending on the infrastructure/society considered. We hope that these explanations clarify the aims of the



study. Again, to provide further clarification in the manuscript, we have added a paragraph at the end of the introduction to justify the presented multi-hazard analysis and provide better context (P3 L76-84).

[R2 Comment 2] 2. Wildfire burned area is related to temperature not the number of fires. A fire of size 0.1 ha can occur all year around, specially for human started fires (see Balch, J. K., et al. (2017). Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, 114(11), 2946-2951.). But more importantly, what is the impact? If it is the wildfire, is it justified to have fire as the impact and as the hazard? Is high-temperature necessarily a hazard?

[AR 2] We believe that extreme high temperatures can be considered a hazard based on sources such as UNISDR (2009). It is true that small fire occurs all year round, and our data also support that statement (see manuscript Figure 12). Burned area and number of wildfires depends on many other drivers such as wind, type of fuel and soil moisture. The aim of this study is not to decipher the mechanism leading to a wildfire but rather as an exemplar of looking at two hazards and quantifying their interrelation between two hazards. We have now made this point clearer in our manuscript. Regarding impact, both extreme high temperatures and wildfire as single hazards can potentially impact society. We also consider that extreme temperature increases the probability of wildfire occurring (AghaKouchak et al., 2018). We then try to address both of the following: (i) How much does an increase in temperature impacts wildfire likelihood; (ii) Estimating return periods for events of extreme heat and fire. A paragraph has been added at the beginning of Section 4.2 to justify our choices (P26 L452-457).

[R2 Comment 3] 3. Abstract needs some improvement. I struggled to understand how the case studies are related to synthetic data.

[AR 3] This point has also been noted by Reviewer 1 (R1) and that part of the abstract has been modified (L10-11) and (L12-13)

[R2 Comment 4] 4. An important missing element in the evaluation of models is p-value. It would be interesting to see what the p-values are and determine whether the models fail/pass to represent the data! This is actually very important. The metrics used in the paper are subjective - although valuable - and a more widely accepted metric could help the general audience relate the study to other modelling practices.

[AR 4] We are unsure that p-value is applicable in the framework of this study as we are comparing curves and not the distributions themselves. However, we agree on the fact that a robust comparison point to assess the efficacy of model might be needed. We think that comparing model fitting capabilities to an empirical curve (obtained through point counting) would give valuable insight. We therefore created empirical level curves and computed their *wd* to the reference curve following the same steps as for the 6 models. For each of the 60 synthetic datasets, models having a smaller *wd* to the reference curve than the empirical estimate are considered to be "passing" to represent the data

with more accuracy than a naïve approach. Figure 7 has therefore been updated to incorporate this new information and a new paragraph has been added in Section 3.3 (P18 L484-489).

[R2 Comment 5] 5. There are many typos in the text. I highlighted some of them in the attached manuscript, but there are more.

[AR 5] Thank you for this noting of the typos. We have taken these into account in our revised manuscript by catching many of these, and will ensure that for any final proof reading we do a careful reading for typos.

[R2 Comment 6] 6. There are some specific comments that are provided in the attached manuscript. Hope the authors find the comments useful to improve their paper

[AR 6] Thank you for these comments. We have gone through these in detail and the major ones not addressed in R2 Comments 1 to 5 are the following, along with our replies:

[R2 Specific comments (from supplement)]

L49-L51 P2 provide names not acronyms  
Change done

L 72 P3 joint extreme can be AND or OR scenario, Figure 1 shows AND scenario. please clarify  
It has been clarified

L108 P 4 Notations in equation 1&2 don't seem to be correctly presented  
Change done  
L112 P4  $u$  merges to 1 not infinity (if  $u$  is marginal)  
Change done. Thank you!

L114 P4 respectively is not needed. It is confusing actually  
Change done

L142 P5 this " $u$ " is never defined before  
It was a mistake.  $U$  has been replaced by  $z$  and  $z$  has been defined.

L170 P6 "and" not "or"  
Change done

L257 P6 Summary section not needed  
Section has been removed

L270 P6 is this the same " $u$ " used before? if not, it will be confusing  
No, but the meaning of  $u$  is specified on the following line

Figure 7 P19 while the presented metrics are nice, it would be more desirable to show the p-value for each case, to examine whether the copula/any model is representative of the data, or not. Then the distances would help select the best model

The figure has been redone and the “tiles” in which models have a lower wd than a naïve empirical approach (See AR4) is highlighted. We consider that if the model represents the data better than an empirical approach in one of the cases, then the model is suitable for that case.

L472 P20 How can you say that. How is  $wd < 0.1$  selected as the criterion for goodness-of-fit, why not 0.05 or 0.2?

We modified this part and also use the fact that the wd of the model is lower than the empirical wd as a criterion for goodness of fit (See AR4).

L488 P20 needs a comma here

In general commas are missing throughout the manuscript. before "respectively" you need a comma, etc

We have changed some of these in our revised manuscript, along with doing a more thorough proofreading.

L510 P20 comma after "here"

Change done

L599 P25 not the right reference for wildfire.

Suggestion:

Littell, J. S., McKenzie, D., Peterson, D. L., & Westerling, A. L. (2009). Climate and wildfire area burned in western US ecoprovinces, 1916–2003. *Ecological Applications*, 19(4), 1003-1021.

We added the reference, thank you.

# Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios

Aloïs Tilloy<sup>1</sup>, Bruce D Malamud<sup>1</sup>, Hugo Winter<sup>2</sup>, Amélie Joly-Laugel<sup>2</sup>

<sup>1</sup>Department of Geography, King’s College London, London WC2B 4BG, United Kingdom

<sup>2</sup>EDF Energy R&D UK Centre, Croydon CR0 2AJ, United Kingdom

\*Correspondence to: Aloïs Tilloy ([alois.tilloy@kcl.ac.uk](mailto:alois.tilloy@kcl.ac.uk))

~~Estimating risks generated by multi-hazard scenarios~~Modelling multiple hazards interrelations remains a challenge for practitioners. ~~Here we evaluate~~This article primarily focuses on the interrelations between pairs of hazards. The efficacy of ~~six distinct~~ bivariate extreme ~~modelling approaches by~~models is evaluated through their fitting ~~six distinct stochastic models capabilities~~ to 60 synthetic datasets. The properties of the synthetic datasets (marginal distributions, tail dependence structure) are chosen to match bivariate time series of environmental variables. The six models are copulas (one non-parametric, one semi-parametric, four parametric). We build 60 distinct synthetic datasets based on different parameters of log-normal margins and two different copulas. ~~We contrast~~The systematic framework developed contrasts the model strengths (model flexibility) and weaknesses (poorer fits to the data). We find that no one model fits our synthetic data for all parameters, but rather a range of models ~~are more appropriated~~depending on the characteristics of the data. To highlight the benefits of the systematic modelling framework developed, we consider the following environmental data: (i) daily precipitation and maximum wind ~~gusts~~ for 1971 to 2018 in London, UK; (ii) daily mean temperature and wildfire ~~number~~numbers for 1980 to 2005 in Porto district, Portugal. In both cases there is good agreement in the estimation of bivariate return periods between models selected from the systematic framework developed in this study. Within this framework, we have explored a way to model multi-hazard events and identify the most efficient models for a given set of synthetic data and hazard sets.

Style Definition:	Normal
Style Definition:	Style1: Font color: Black
Style Definition:	Style2: Font color: Black
Style Definition:	maintitle: Font: Calibri
Style Definition:	Subtitle: Font: Font color: Custom Color(90,90,90))
Style Definition:	Normal (Web): Font:
Style Definition:	Title: Font: (Default) Cambria
Style Definition:	Revision: Font: (Default) Calibri

## 1 Introduction

A multi-hazard approach considers more than one hazard in a given place and the interrelations between these hazards (Gill and Malamud, 2014). Multi-hazard events have the potential to cause damage to infrastructure and people that may differ greatly from the associated risks posed by a single hazard (Terzi et al., 2019). Here, we consider two main mechanisms in natural hazards interrelations defined in Tilloy et al. (2019): (i) cascade interrelations (i.e., when there is a temporal order and causality between natural hazards); (ii) compound interrelations (i.e., when several natural hazards are statistically dependent without causality).

Meteorological phenomena such as mid-latitude cyclones or thunderstorms often influence cascading and compound interrelations between natural hazards (e.g., a storm can include rain, lightning, hail, with rain and hail both potentially triggering landslides). Case examples of meteorological phenomena influencing natural hazard interrelations include the following:

A multi-hazard approach considers more than one hazard in a given place and the interrelations between these hazards (Gill and Malamud, 2014). Multi-hazard events have the potential to cause damage to infrastructure and people that may differ greatly from the associated risks posed by a single hazard (Terzi et al., 2019). Here, natural hazards (which we will also refer to as ‘hazards’) will be defined as (UNISDR, 2009) a natural process or phenomenon that may have negative impacts on society. For modelling purposes, we consider two main mechanisms in natural hazards interrelations (Tilloy et al., 2019): (i) *cascade interrelations* (i.e., when there is a temporal order and causality between natural hazards); (ii) *compound interrelations* (i.e., when several natural hazards are statistically dependent without causality).

Meteorological phenomena such as extratropical cyclones or convective storms often lead to the combination of multiple drivers and/or hazards and can therefore be related to compound events as defined by Zscheischler et al. (2018). This research concentrates on cascading and compound interrelations between natural hazards (e.g., a storm can include rain, lightning, hail, with rain and hail both potentially triggering landslides). Case examples of meteorological phenomena influencing natural hazard interrelations include the following:

- (i) In 2010, storm Xynthia hit the west coast of France. The storm itself was not particularly extreme for the season but the compound effect of extreme wind, high tides, storm surge, extreme rainfall and the fact that the soils were already saturated led to huge damage due to wind and flooding (CCR, 2019).
- (ii) In summer 2010, Russia experienced a heatwave. Low precipitation in spring 2010 led to a summer drought that contributed to the heatwave having a large magnitude (Barriopedro et al., 2011; Hauser et al., 2015; Zscheischler et al., 2018). The co-occurrence of extremely dry and hot conditions resulted in widespread wildfires, which damaged crops and caused human mortality (Barriopedro et al., 2011).
- (iii) Extreme thunderstorms occurred in the Paris region in 2001, involving lightning and extreme rainfall, with the rainfall triggering flooding, mudslides and ground collapse, with subsequent damage to railway networks (CCR, 2019).

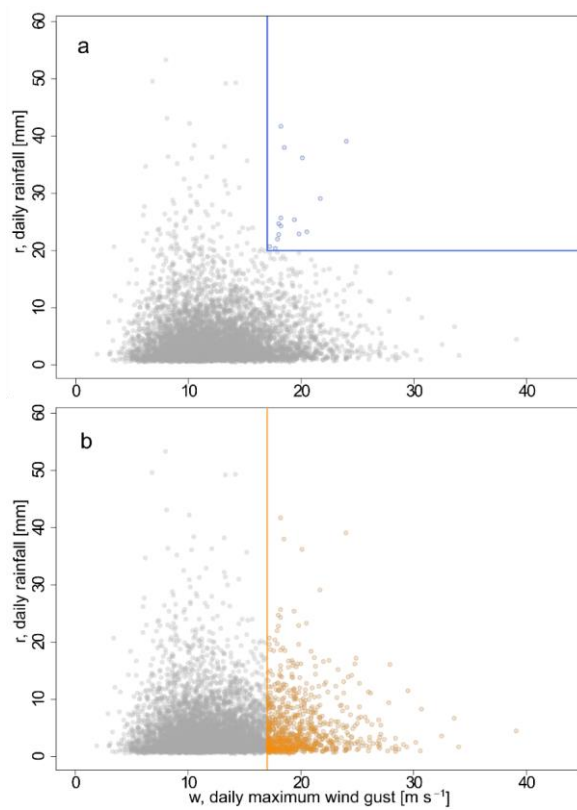
In this context, the quantification of interrelations between natural hazards can play an important role in risk mitigation and disaster risk reduction. Some of the natural hazards presented in the above examples are extreme occurrence of environmental variables (e.g. extreme temperature) which have different characteristics and statistical distributions (e.g., wind and landslides). Natural hazards are interrelated with different mechanisms (i.e. compound, cascade). For a given mechanism, interrelations also vary in strength and intensity. Additionally, as highlighted in Tilloy et al. (2019), different modelling approaches have been developed to quantify interrelations between variables. Here we focus on stochastic models that include copulas (Nelsen, 2006; Genest and Favre, 2007; Salvadori et al., 2016), and multivariate *extreme* models (Heffernan and Tawn, 2004), limiting our analysis to the bivariate

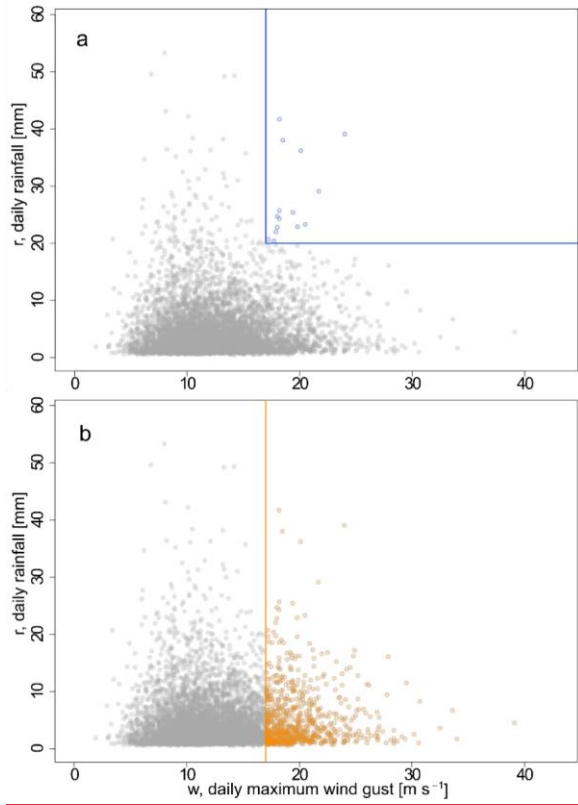
case. The potential for misinterpretation of the dependence structure of two variables clearly presents a problem when end-users try to account for hazard interrelations.

We choose six distinct bivariate models able to handle different types of tail (extreme) dependence: one non-parametric (JT-KDE), one semi-parametric (Cond-Ex) and four different parametric copulas (~~GalambosCop, Gumbelcop, FGMeop, Normaleop~~) (see **Sect. 2**). ~~Galambos, Gumbel, FGM, Normal~~ (see **Sect. 2**). ~~We are interested in modelling interrelations between hazards in the extreme domain. This implies the use of methods and concepts coming from the broad area of Extreme Value Theory (EVT). Amongst the six models compared in this study, four are directly linked to EVT (JT-KDE, Cond-Ex, GalambosCop, Gumbelcop). We investigate the applicability different bivariate models have to a wide range of natural hazard interrelations. In order to reproduce the complexity and variety of natural hazard interrelations we use 60 synthetic datasets to compare the fitting performances of the models. In these synthetics datasets we vary two main attributes of the bivariate datasets: the dependence structure and the marginal (individual) distributions. Of these 60 different synthetic datasets, 36 datasets have asymptotically dependent variables and 24 have asymptotically independent variables (see **Sect. 2.1** for a definition of these two concepts).~~ The fitting capacities of each model are compared with the estimation of *level curves*. Level curves are extensively described in **Sect. 2.3**. However, these curves correspond to probabilities that can be related to compound and cascading hazard interrelations. Compound interrelations are represented with a joint probability while cascading (sequential) interrelations are represented with conditional probabilities.

Examples of joint and conditional probabilities are given in **Fig. 1**. A joint probability is the probability of two events occurring together (~~Fig. where both variables are extreme (also called AND probability)~~ (**Fig. 1a**) and a conditional probability is the probability of an event given that another has already occurred (**Fig. 1b**). **Figure 1** illustrates the concepts of joint probability and conditional probability, with daily rainfall data from a high-resolution gridded data set of daily climate over Europe (termed ‘E-OBS’) (Cornes et al., 2018) and daily maximum wind gust data at Heathrow airport provided by the Met Office (2019). A wind gust here is defined by the Met Office (2019) as the maximum 3 s average wind speed occurring in any given period.

These datasets and the interrelation between extreme rainfall and extreme wind are discussed in **Sect. 4.1**.





**Figure 1: Illustration of joint and conditional extremes with daily rainfall  $r$  ( $\text{mm day}^{-1}$ ) and daily maximum wind gust  $w$  ( $\text{m s}^{-1}$ ) data at Heathrow airport for the period 1971–2018: (a) joint extremes (AND) of rainfall and wind gust (blue circles); (b) conditional extremes of rainfall given that wind gust is extreme (yellow circles). Daily rainfall data from E-OBS (Cornes et al., 2018) and daily maximum wind gust (3 s period) data from the Met Office (2019).**

Joint and conditional probabilities are relevant metrics for practitioners and have been studied and used in several studies in the environmental sciences (e.g., Hao et al., 2017; Zscheischler and Seneviratne, 2017). However, as the most widely used level curve is the joint probability curve, we initially focus on it. To analyse our results and compare the performances of the models we designed diagnostic tools that are presented in Sect. 3.2. (e.g., Hao et al., 2017; Zscheischler and Seneviratne, 2017). However, as the most widely used level curve is the joint probability curve, we initially focus on it.

This paper is organized as follows. We first (Sect. 2) provide a theoretical background on key concepts used in this study and present the models and methodology used. We then (Sect. 3) discuss the characteristics of our synthetic dataset and present the results of the simulation study. The diagnostic tools used to compare models are also discussed (i.e., joint return level curves and dependence measure). As a result, a map exhibiting the strength and weaknesses of our six models is presented. It aims to provide objective criteria to justify the use of one model rather than another for a given set of hazards. Two applications to pairs of natural hazards that can impact energy infrastructure are presented in Section 4. The main purpose of these data applications is to illustrate our methodology, but the natural hazard interrelations studied have the potential to negatively impact energy infrastructure. The first application looks at compound daily rainfall and wind in the United Kingdom. The combination of these two hazards can result in different and greater impacts than the addition of impacts due to extreme wind



and extreme rainfall (e.g., wind destroys roof leading to greater damages, power plants flooded with rescuers slow down by strong winds) (Martius et al., 2016). The second application studies extreme hot temperatures and wildfires in Portugal. Extreme temperatures can lead to damage on infrastructure (e.g., rail track deformation) and put pressure on the energy infrastructure by increasing the demand (Hatvani-Kovacs et al., 2016; Vogel et al., 2020), it also increases the probability of wildfires (Witte et al., 2011; Perkins, 2015) which have the potential to cause fatalities and destroy infrastructure (Tedim et al., 2018). We finish (Sect. 5) with a discussion and conclusions.

To analyse our results and compare the performances of the models we designed diagnostic tools that are presented in Sect. 3.2. The results of this study are presented as a map exhibiting the strength and weaknesses of our six models (Sect. 3.3). It aims to provide objective criteria to justify the use of one model rather than another for a given set of hazards as it is shown with two examples in Sect. 4.

This article is organized as follows. We first (Sect. 2) provide a theoretical background on key concepts used in this study and present the models and methodology used. We then (Sect. 3) discuss the characteristics of our synthetic dataset and present the results of simulation study. The diagnostic tools used to compare models are also discussed (i.e., joint return level curves and dependence measure). Two applications (Sect. 4) to pairs of natural hazard data are presented: (i) compound daily rainfall and wind in the United Kingdom and (ii) extreme hot temperatures and wildfires in Portugal. We finish (Sect. 5) with a discussion and conclusions.

## 2 Methods

As stated in the introduction, we are interested in modelling the dependence structure between variables in the tails or extreme domain. We are interested in modelling interrelations between hazards in the extreme domain. This implies the use of methods and concepts coming from the broad area of Extreme Value Theory (EVT). Amongst the six models compared in this study, four are directly linked to EVT (JT-KDE, Cond-Ex, Galambos, Gumbel). Extreme Value Theory has its roots in univariate studies (Coles, 2001) and has been extended to the multivariate framework (Pickands, 1981; Davison and Huser, 2015). A theoretical background on extreme value theory is given in Supplement S1.1. In this study, we focus on modelling the dependence between two variables. Bivariate extreme value models developed within the statistical community (Resnick, 1987; Heffernan and Tawn, 2004; Cooley et al., 2019) have recently been used for environmental application and therefore natural hazard interrelations (De Haan and De Ronde, 1998; Zheng et al., 2014; Sadegh et al., 2017). In order to reproduce the complexity and variety of natural hazard interrelations we use 60 synthetic datasets to compare the fitting performances of the models. In these synthetics datasets we vary two main attributes of the bivariate datasets: the dependence structure and the marginal (individual) distributions. Of these 60 different synthetic datasets, 36 datasets have asymptotically dependent variables and 24 have asymptotically independent variables (see Sect. 2.1 for a definition of these two concepts). Extreme Value Theory has its roots in univariate studies (Coles, 2001) and has been extended to the multivariate framework (Pickands, 1981; Davison and Huser, 2015). In this study we mainly focus on modelling the dependence between two variables. Bivariate extreme value models developed within the statistical community (Resnick, 1987; Heffernan and Tawn, 2004; Cooley et al., 2019) have recently been used for environmental application and therefore natural hazard interrelations (De Haan and De Ronde, 1998; Zheng et al., 2014; Sadegh et al., 2017). A theoretical background on extreme value theory is given in Supplement S1.1.

In this section, we first present the two types of asymptotic behaviour in bivariate extreme value statistics: asymptotic dependence and asymptotic independence and discuss different dependence measures for the estimation of the relationship

Formatted: French (France)

Field Code Changed

Formatted: English (United Kingdom)

Formatted: English (United Kingdom)

145 between two variables (Sect. 2.1). The six bivariate models are then described (Sect. 2.2). Finally, we discuss the concept of  
the return level in the bivariate framework (Sect. 2.3).

## 2.1 Bivariate extreme dependence

### 2.1.1 Asymptotic dependence and asymptotic independence

150 Let  $X_1, \dots, X_n$  be  $n$  different variables, with each variable a vector that can take on multiple values. Assume that these vectors  
are random and independent and identically distributed (i.i.d). ~~Then~~The asymptotic dependence implies that if one variable  $X_k$   
for  $k \in (1, n)$  has values  $X_k$  that are large it is possible for the other variables to take on values that are simultaneously extreme  
(Coles et al., 1999)(Coles et al., 1999). One way to characterise extremal dependence structures is to split them into those with  
asymptotic dependence and those with asymptotic independence. In the bivariate case, for  $(X_1, X_2)$  random pair with joint  
distribution  $G$ , the random variables  $X_1$  and  $X_2$  are asymptotically dependent if the following conditional probability  
155 (Heffernan, 2000)(Heffernan, 2000)

$$P(X_1 > x | X_2 > x) \rightarrow c > 0 \text{ as } x \rightarrow x^* \quad (1)$$

Where  $X_1 > x$  are those values of variable  $X_1$  that are greater than a threshold  $x$ , the probability of both  $X_1 > x$  and  $X_2 > x$  is  $c \in (0, 1]$  and  $x^*$  is the upper end point (maximum) of the common marginal distribution.

The variables  $X_1$  and  $X_2$  are asymptotically independent if (Heffernan, 2000)

$$P(X_1 > x | X_2 > x) \rightarrow 0 \text{ as } u \rightarrow \infty \quad (2)$$

160 where  $u$  is a high threshold. In practice (Davison and Huser, 2015), extremal dependence is often observed to weaken at high  
levels (i.e., as  $u \rightarrow \infty, x \rightarrow 1$ ), and it can happen that dependence between variables is observed in the body of the joint  
distribution, but that the multivariate distribution is in fact in the max-domain of attraction of independence (Davison and  
Huser, 2015).

Using models that take the assumption of asymptotic dependence (~~respectively~~-independence) in the case of asymptotically  
independent (~~respectively~~-dependent) variables can lead to a large overestimation (~~respectively~~-underestimation) of the  
165 probability of joint extreme events (Ledford, 1996; Mazas and Hamm, 2017; Cooley et al., 2019). (Ledford, 1996; Mazas and  
Hamm, 2017; Cooley et al., 2019). Multivariate extreme value and regular variation theory presented in the **Supplement S1.2**  
provides a rich theory for asymptotic dependence (De Haan and Resnick, 1977; Pickands, 1981) but are not able to distinguish  
between asymptotic independence and full independence.

### 2.1.2 Tail dependence measures

170 A popular method to analyse hazard interrelationships is to compute dependence measures (Zheng et al., 2013; Petroligakis,  
2018). Dependence measures aim to describe how two (or more) variables are correlated.

When focusing on the dependence in the tails or extreme part of distributions, linear or rank dependence measures might not  
be accurate and other coefficients appear more relevant (Hao and Singh, 2016). Dependence between variables in the joint tail  
domain has been widely studied in the statistics community (Coles and Tawn, 1991; Ledford and Tawn, 1997; Coles et al.,  
175 1999; Heffernan and Tawn, 2004; Zheng et al., 2014). As explained in **Sect. 2.1.1**, in the tails, two variables can be either  
asymptotically independent or asymptotically dependent; different diagnostics and coefficients previously developed are  
summarized in Heffernan (2000).

In this study, we use the following tail dependence measures:

Formatted Table

Formatted Table

- the extremal dependence measures  $\chi$  and  $\bar{\chi}$  introduced by Coles et al. (1999);
- the coefficient of tail dependence  $\eta$ , introduced by Ledford and Tawn (1996).

These coefficients aim to measure the extremal dependence for bivariate random variables  $(X_1, X_2)$  and assume initially that  $(X_1, X_2)$  have a common marginal distribution. Coles *et al.* (1999) defined the extremal dependence measure:

$$\chi(x) = P(X_2 > x | X_1 > x) \text{ with } \lim_{x \rightarrow x^*} \chi(x) = \chi \quad (3)$$

with  $x$  a sufficiently high threshold. A sufficiently high threshold  $x$  is a value that can be considered as extreme within a given distribution (corresponding to a high quantile); the value of the threshold depends on the marginal distribution. The extremal dependence measure  $\chi(x)$  is the probability of one variable ( $X_1$  or  $X_2$ ) being extreme given the other is extreme ( $X_2$  or  $X_1$ ). This measure  $\chi$  varies in the range  $[0,1]$ , where a value of  $\chi=0$  means that the two variables are asymptotically independent and  $\chi=1$  means that they are perfectly dependent. The extremal dependence measure  $\chi$  is only suitable for asymptotic dependence. In the case of asymptotic independence ( $\chi=0$ ), Coles *et al.* (1999) introduced the measure  $\bar{\chi}$  which falls between the range  $[-1,1]$ , 1 being asymptotic independence. Ledford and Tawn (1996) defined their coefficient of tail dependence to be able to assess the strength of dependence between two asymptotically independent variables. They show that the joint survivor function for random variables  $(Z_1, Z_2)$  with common standard Fréchet margins [can be expressed as](#) (See **Supplement S1.2**):

$$P(Z_1 > z, Z_2 > z) \sim \mathcal{L}(z)(P(Z_1 > z))^{1/\eta} \quad (4)$$

with  $\mathcal{L}(z)$  a slowly varying function while  $z \rightarrow \infty$  and  $\eta$  is the coefficient of tail dependence, lying in the range  $[0,1]$ . Different values of each coefficient and their implications are summarized in **Fig. 2**. For large  $z$ , the three tail dependence measures presented above are related in the following way (Ledford and Tawn, 2003):

with  $z$  a sufficiently high threshold in the standard Fréchet space,  $\mathcal{L}(z)$  a slowly varying function while  $z \rightarrow \infty$  and  $\eta$  is the coefficient of tail dependence, lying in the range  $[0,1]$ . Different values of each coefficient and their implications are summarized in **Fig. 2**. For large  $z$ , the three tail dependence measures presented above are related in the following way (Ledford and Tawn, 2003):

$$\bar{\chi} = 2\eta - 1 \quad (5)$$

$$\chi = \begin{cases} 0 & \text{if } \bar{\chi} = 1 \text{ and } \mathcal{L}(z) \rightarrow c > 0 \text{ as } z \rightarrow z^* \\ 1 & \text{if } \bar{\chi} < 1 \end{cases}$$

Formatted Table

Formatted Table

Formatted: Font: Not Bold

Formatted Table

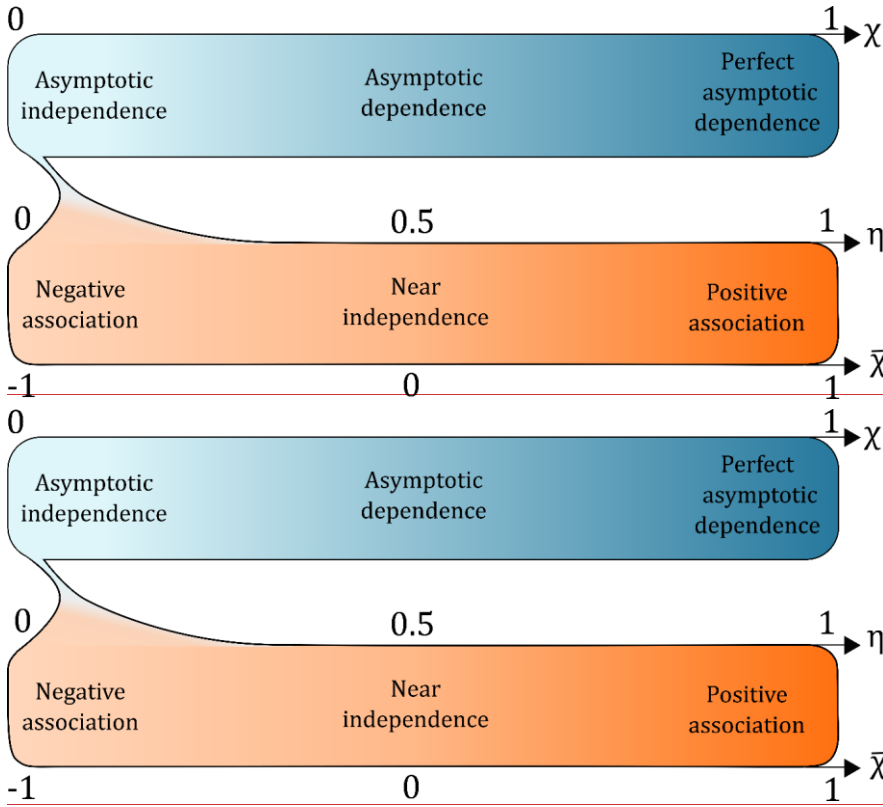


Figure 2: The three coefficients used in this study to assess the dependence between two variables at an extreme level. In the upper part of the plot (blue), the coefficient  $\chi$  varies between perfect asymptotic dependence (light blue,  $\chi = 0$ ) and asymptotic independence (dark blue,  $\chi = 1$ ). In the lower part of the plot (orange), which is in the asymptotic independence domain (in other words,  $\chi = 0$ ) the coefficients  $\bar{\chi}$  and  $\eta$  both vary between negative association (light orange,  $\bar{\chi} = -1$ ;  $\eta = 0$ ) and positive association (dark orange,  $\bar{\chi} = \eta = 1$ ).

## 2.2 Bivariate models

Dependence measures are empirical measures which estimate the strength of the correlation, or dependence between two (or more) variables. Despite the fact that these measures provide crucial information, these do not allow to model joint (or conditional) exceedance probabilities. To model joint exceedance probabilities which represent the joint occurrence of hazards (here represented by extremes of environmental variables) in time and space, the use of stochastic models is required. In this section we present the three stochastic approaches for multivariate modelling that are used in the simulation study: parametric copulas, the semi-parametric conditional extremes model and a non-parametric approach based on multivariate extreme value theory (see **Supplement S1.2**) and kernel density estimation.

### 2.2.1 Copulas

In the bivariate case, a copula is a joint distribution function which defines the dependence between two variables independently from the marginal distributions of these variables (Heffernan, 2000; Nelsen, 2006; Genest and Favre, 2007; Hao

and Singh, 2016). Let the random variables  $(X_1, X_2)$  be vectors of i.i.d. values with marginal distributions  $F_1(x_1)$  and  $F_2(x_2)$  and a joint cumulative distribution function  $F_{1,2}(x_1, x_2)$ . Any bivariate distribution function with marginal distribution functions  $F_{X1}(x_1)$  and  $F_{X2}(x_2)$  can be expressed as a copula function as follows (Sklar, 1959; Nelsen, 2006):

$$F_1(x_1, x_2) = C\{F_1(x_1), F_2(x_2)\}, \quad (6)$$

where  $C$  is the copula function. Copulas are not limited to two variables and Eq. 6 can be extended to higher dimensions. Several classes of copula with different properties are available, including Archimedean copulas, elliptical and extreme value copulas (e.g., Joe, 1997; Nelsen, 2006). Extreme value copulas have been used within various domains such as finance, insurance and hydrology because of their ability to model extremal dependence structures (Genest and Nešlehová, 2013). However, extreme value copulas are by definition asymptotically dependent as they follow the rules of multivariate extreme value theory (see Supplement S1.2). The two types of extremal dependence were presented in Sect. 2.1 and show that it is important to also consider asymptotic independence. Many copulas are asymptotically independent, including the normal copula and the Farlie-Gumbel-Morgenstern (FGM) copula (Heffernan, 2000). These two copulas will be used in the simulation analysis as asymptotically independent models (Sect. 3) (Heffernan, 2000). These two copulas will be used in the simulation analysis as asymptotically independent models (Sect. 3).

Formally, In the present study, the application of a copula model can be summarized in four main steps:

- (i) Fitting marginal distributions to the two variables and then an empirical cumulative distribution function below a threshold and generalised Pareto distribution (GPD) above this threshold.
- (ii) Transforming the variables to uniform margins. The transformed datasets no longer have information on the marginal distributions but keep the information about the joint distribution dependence structure (Nelsen, 2006).
- (iii) Fitting the copula function to the pseudo-observations by estimating the copula parameter(s) with an estimator (Genest and Favre, 2007).
- (iv) Estimating the probability of joint events with the copula function previously fitted.

## 2.2.2 Conditional extreme model

The conditional extremes model (Heffernan and Tawn, 2004; Keef et al., 2013) (Heffernan and Tawn, 2004; Keef et al., 2013) is a semi-parametric model designed to overcome several limitations of copulas and other approaches such as the joint tail methods in which all variables must become large at the same rate. The aforementioned methods can typically handle only one form of extremal dependence, either asymptotic dependence or asymptotic independence. The conditional extremes model has the ability to be more flexible with asymptotic dependence classes; it can account for asymptotic independence and asymptotic dependence (Heffernan and Tawn, 2004; Keef et al., 2013) (Heffernan and Tawn, 2004; Keef et al., 2013). It can also be used to analyse more than two i.i.d variables more easily than copula-based methods (Winter and Tawn, 2016); we restrict the theory provided here to the bivariate case. The conditional model has been used for different purposes: spatial or temporal dependence between extremes (Winter and Tawn, 2016; Winter et al., 2016) (Winter and Tawn, 2016; Winter et al., 2016), dependence between extreme hazards (Zheng et al., 2014) (Zheng et al., 2014) and even financial purposes (Hilal et al., 2011) (Hilal et al., 2011).

The conditional extremes model assesses the dependence structure between several variables conditioning on one being extreme and aims to model the conditional distribution. As in joint-tail models, the first step is to transform the marginal distributions; here the preferred marginal choice are Laplace (or Gumbel) margins (Heffernan and Tawn, 2004; Keef et al., 2013) (Heffernan and Tawn, 2004; Keef et al., 2013). Let the random variables  $(Y_1, Y_2)$  be vectors of i.i.d. values with Laplace distributions. The conditional extremes model aims to identify two normalizing functions  $a(y_i)$  and  $b(y_i)$  such that  $a$  satisfies  $\mathbb{R}_+ \rightarrow \mathbb{R}$  and  $b$  satisfies  $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ . Both are defined such that for  $y > 0$  (Winter, 2016):

Field Code Changed

$$P\left(\frac{Y_2 - a[Y_1]}{b[Y_1]} \leq z, Y_1 - u > y | Y_1 > u\right) \rightarrow \exp(-y) G(z) \quad (7)$$

as  $u \rightarrow \infty$ , where  $G(z)$  is a non-degenerate distribution function. In the case of Laplace margins the normalising functions  $a$  and  $b$  are given by (Winter, 2016):

$$a[y] = \alpha y \text{ and } b[y] = y^\beta \quad (8)$$

where  $\alpha \in [-1, 1]$  and  $\beta \in (-\infty, 1)$ . The different values of  $\alpha$  and  $\beta$  characterise different forms of tail dependence. In the case where  $\alpha = 1$  and  $\beta = 0$ , variables  $(Y_1, Y_2)$  exhibit asymptotic positive dependence and the case of asymptotic negative dependence is given when  $\alpha = -1$  and  $\beta = 0$  (Winter, 2016).

Formally, the application of the conditional extreme model can be summarized in four main steps:

- (i) Fitting marginal distributions to the two variables; an empirical cumulative distribution function below a threshold and generalised Pareto distribution (GPD) above this threshold.
- (ii) Transforming those distributions onto Laplace (or Gumbel) margins.
- (iii) Estimating the dependence parameters using non-linear regression.
- (iv) Estimating the probability of joint events ~~and extrapolate by simulating new extreme data through the conditional model to simulate new extreme data.~~

### 2.2.3 Joint tail KDE (kernel density estimation) approach

The non-parametric approach used in this paper is an adaptation of the non-parametric approach presented by Cooley et al. (2019). Moreover, the dependence measures  $\eta$  is estimated to determine whether data are asymptotically dependent or asymptotically independent. This approach is based on the 2D kernel density estimator and the multivariate extreme value framework (see **Supplement S1.2**).

The kernel density estimation (KDE) method has the advantage of being a non-parametric way to estimate the joint distribution of  $n$  variables. With KDE, we make no assumption about the underlying distribution of the margins or about the dependence structure. The KDE centres a smooth kernel at each observation. The choice of the bandwidth is ~~very~~ crucial when using this method (Duong, 2007; Hao and Singh, 2016) (Duong, 2007; Hao and Singh, 2016). This selection was done automatically in our case within the kernel survival function estimation function from the R package ks (Duong, 2007, 2015) (Duong, 2007, 2016).

The kernel density estimator is used here to estimate an empirical density distribution  $\hat{f}(X)$  and a joint survival distribution  $\hat{F}(X)$  of the bivariate dataset where  $X=(X_1, X_2)$ . The joint survival distribution corresponds to the joint exceedance probability of the two variables (See **Section 2.3**). From the joint survival distribution, it is possible to estimate level curves which are isolines corresponding to given joint probabilities of exceedance (see **Sect. 2.3**).

After estimating the joint survival distribution of the two variables with a kernel density estimator, the cumulative distributions  $\hat{F}_i(x)$  of the two random variables  $X_i$  ( $i = 1, 2, \dots$ ) are estimated empirically below a threshold and from a Generalized Pareto distribution above the threshold. The two marginal cumulative distribution functions are then transformed to Fréchet margins to allow the use of multivariate extreme value theory (Cooley et al., 2019) (Cooley et al., 2019):

$$\hat{T}_i(x) = \frac{-1}{\ln(\hat{F}_i(x))}. \quad (9)$$

Therefore,  $Z = T(X) = (T_1(X_1), T_2(X_2))$  can be assumed to be regularly varying with an index of regular variation 1 (see **Supplement S1**). An extrapolation from a base probability  $p_{\text{base}}$  (blue area in **Fig. 3**) estimated with a kernel density to an objective probability  $p_{\text{obj}}$  (purple area in **Fig. 3**) is then done on the transformed space. Thus, on the transformed scale, it is

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

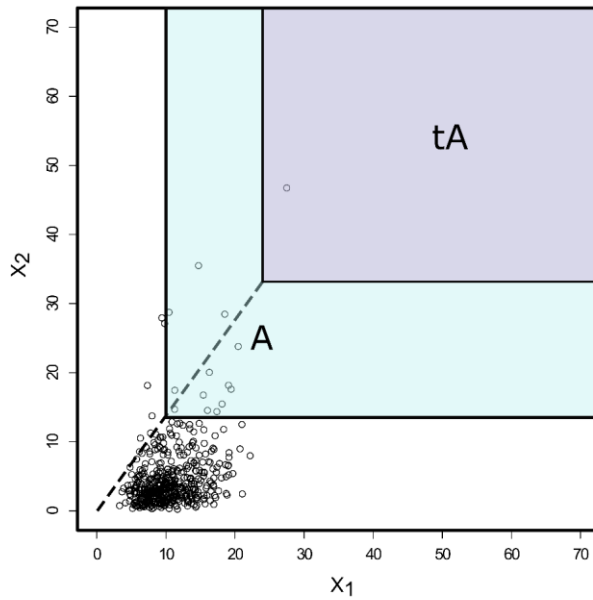
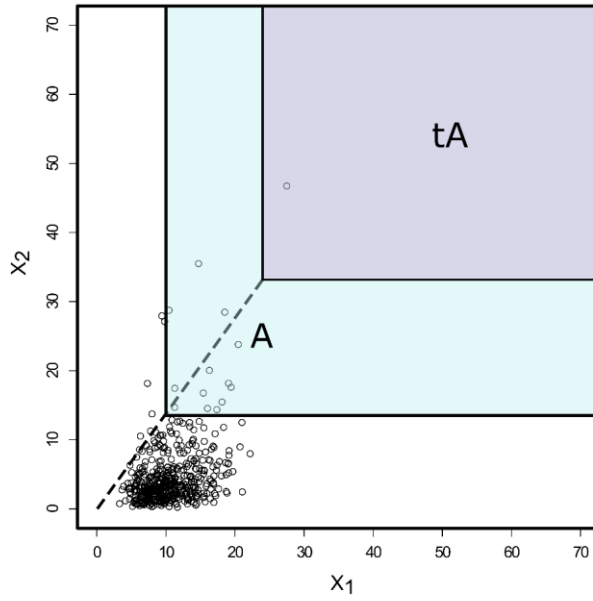
Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

possible to construct  $\hat{l}_{Z(\text{obj})} = t \hat{l}_{Z(\text{base})}$  (Cooley et al., 2019)(Cooley et al., 2019). To produce level curves on the original scale, the transformation in **Eq. 11** is reversed:  $\hat{l}_{\text{obj}} = T^{-1} \hat{l}_{Z(\text{obj})}$ . **Figure 3** gives a graphical representation of the extrapolation done within the joint tail KDE approach.



**Figure 3: Extrapolation in a regularly varying tail for a distribution in the max-domain of attraction of some multivariate extreme value distribution.** Black circles represent an asymptotically dependent bivariate dataset. In order to estimate the extreme joint probability  $P(tA)$  (where  $tA$  is an extreme set represented by the purple area), one can compute  $P(A) = \Pr\{Z \in A\}$ , (where  $A$  is a less extreme set than  $tA$ )



300 represented by the light blue area) with  $t < 1$ . More data points are available in  $A$  than  $tA$ . Then, from the regular variation framework  $\{tP(tA)\} \approx tP(A)$ . Adapted from Huser (2013)

The methodology presented above is only valid when the two variables  $X_1, X_2$  are asymptotically dependent. In the asymptotic independence case, one needs to adjust the methodology. Two asymptotically independent variables follow the properties of hidden regular variation (Resnick, 2002; Maulik and Resnick, 2005) (see Supplement S1.2.3). (Resnick, 2002; Maulik and Resnick, 2005) (see Supplement S1.2.3). Formally, the coefficient of tail dependence  $\eta$  is introduced such as (Cooley et al., 2019):

$$\hat{l}_{Z(ob)} = t^{\frac{1}{\eta}} \hat{l}_{Z(base)} \quad (10)$$

The specificity of this approach (presented below) is that it combines a non-parametric estimation of the joint density and the framework of multivariate extreme value presented in the Supplement S1.2. It can deal with both asymptotic dependence and independence. The coefficient of tail dependence estimation has an influence on the extrapolation process in the asymptotic independence case. Here we used the oneestimator presented in Winter (2016) which is derived from the joint-tail model of Ledford and Tawn (1997).

Formally, the application of the joint tail KDE model can be summarized in five main steps:

- (i) Estimating the joint cumulative distribution of the variables with a kernel density estimator.
- (ii) Fitting marginal distributions to the two variables; empirical distribution below a threshold and General Pareto Distribution (GPD) above this threshold.
- (iii) Transforming those distributions into Fréchet margins.
- (iv) Determining whether variables are asymptotically dependent or asymptotically independent by estimating the coefficients of tail dependence  $\eta$ .
- (v) Estimating the probability of joint events and extrapolate the base isoline to an objective isoline.

#### 2.2.4 Summary

In this section we have presented three modelling approaches: copulas, conditional extreme model and a non-parametric joint-tail model. All of these different approaches aim to model and reproduce the behaviour of multivariate data with different dependence structures. In the next section we discuss the concept of return levels in the bivariate framework as an output measure of interest.

### 2.3 Return levels in the bivariate framework:

Studying natural hazards as multivariate — and particularly bivariate — events is a growing practice in multiple disciplines, including the following: coastal engineering (Hawkes et al., 2002; Mazas and Hamm, 2017); climatology Hao et al., 2017, 2018; Zscheischler and Seneviratne, 2017); and hydrology (Zheng et al., 2014; Hao and Singh, 2016). There has been debate among scientists trying to define a “multivariate return period” (Serinaldi, 2015; Gouldby et al., 2017). Serinaldi (2015) defined seven different types of probabilities that can be considered as bivariate probabilities of exceedance. These can be expressed through copula notation.

Let the random variables  $(X_1, X_2)$  be vectors of i.i.d. values with marginal distributions  $F_i(x_i)$  with  $i=1,2$ ,  $C$  their copula function (Sect. 2.3.1) and  $F_{1,2}(x_1, x_2)$  Studying natural hazards as multivariate — and particularly bivariate — events is a growing practice in multiple disciplines, including the following: coastal engineering (Hawkes et al., 2002; Mazas and Hamm, 2017); climatology Hao et al., 2017, 2018; Zscheischler and Seneviratne, 2017); and hydrology (Zheng et al., 2014; Hao and Singh, 2016). There has been debate among scientists trying to define a “multivariate return period” (Serinaldi, 2015; Gouldby et al.,

Field Code Changed

Formatted: Font color: Black

2017). Serinaldi (2015) defined seven different types of probabilities that can be considered as bivariate probabilities of exceedance. These can be expressed through copula notation.

Let the random variables  $(X_1, X_2)$  be vectors of i.i.d. values with marginal distributions  $F_i(x_i)$  with  $i=1,2$ ,  $C$  their copula function (Sect. 2.3.1) and  $F_{1,2}(x_1, x_2) = C\{F_1(x_1), F_2(x_2)\} = C(u, v)$  where  $F_{1,2}$  is the bivariate distribution function of  $X_1$  and  $X_2$ ,  $u = F_1$  and  $v = F_2$  are standard uniform random variables. The seven types of probability and their equations are given in Table 1.

Table 1: Types of probabilities for bivariate  $(X, Y)$  return period estimation.  $u$  and  $v$  are extreme thresholds. From (Serinaldi, 2015).

Type of probability	Equation	Eq. #
$P_{AND}$	$P(X > u \cap Y > v) = 1 - u - v + C(u, v)$	(11)
$P_{OR}$	$P(X > u \cup Y > v) = 1 - C(u, v)$	(12)
$P_{COND1}$	$P(X > u   Y > v) = (1 - u - v + C(u, v)) / (1 - v)$	(13)
$P_{COND2}$	$P(X > u   Y \leq v) = 1 - \frac{C(u, v)}{v}$	(14)
$P_{COND3}$	$P(X > u   Y = v) = 1 - \frac{\partial C(u, v)}{\partial v}$	(15)
$P_{Kc}$	$P(C(u, v) > t) = 1 - K_c(t)$	(16)
$P_{Sg}$	$P(g(U, V)) = 1 - F_Z(z)$	(17)

The function  $K_c$  in Eq. 16 is the Kendall function and represents the distribution function of the copula (Salvadori and De Michele, 2010; Serinaldi, 2015). (Salvadori and De Michele, 2010; Serinaldi, 2015). Equation 17 refers to the ‘structure-based’ return period introduced by Volpi and Fiori (2014). Among these seven types of probabilities, we selected the “AND” and the “COND1” probabilities (see Fig. 4) as these are commonly used in the literature (Chebana and Ouada, 2011; Tencer et al., 2014; Sadegh et al., 2018) and correspond to the two types of interrelations we are interested in (i.e., compound and cascade).

Field Code Changed

Formatted: Font: Bold

Formatted Table

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

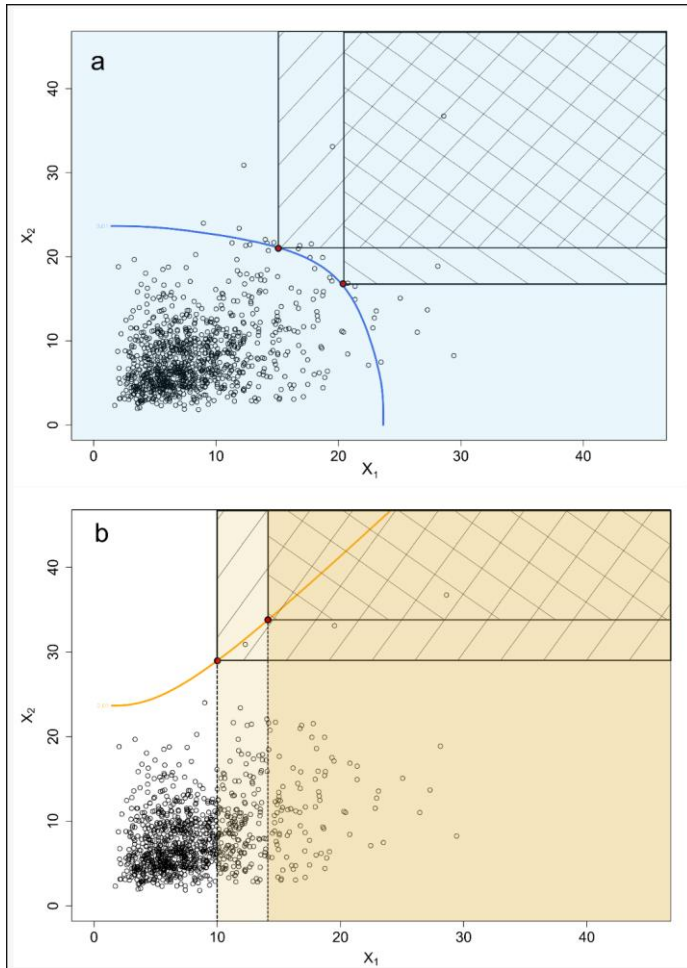
Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Field Code Changed



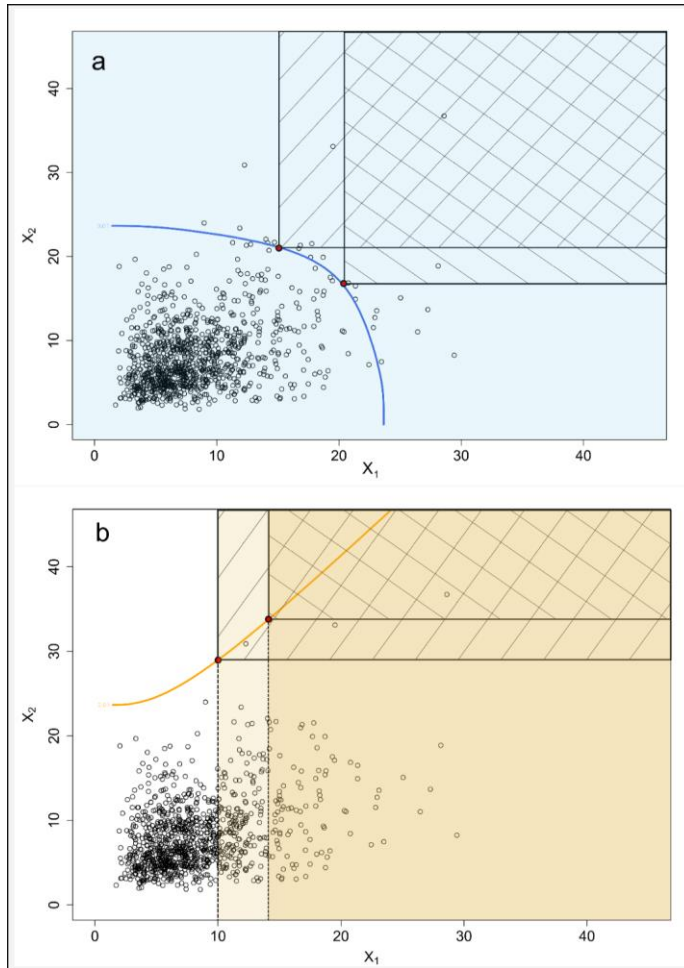


Figure 4: Graphical representation of two bivariate  $(X_1, X_2)$  probabilities of exceedance: (a)  $P_{AND}$  probability and (b)  $P_{COND}$  probability with level curves (blue in 'a' and orange in 'b') representing  $p = 0.01$  (1000 data points on a Gumbel copula with log-normal marginal distributions). Colours represent the domain on which the probabilities are computed while the areas with diagonal hatching represent the critical regions which are the regions corresponding to the given probabilities.

In 2D space, probabilities of exceedance (or quantiles) are not represented by a single value but by a curve with an infinite amount of points with the same probability of exceedance. However, as shown in Fig. 4, these probabilities are defined by: (i) the domain where these are computed and (ii) the critical region corresponding to the probability type. For the AND probability, the computation domain remains similar when moving along the curve while the critical region evolves constantly. For the COND1 probability, both computation domain and critical region evolve when moving along the curve (see Fig. 4). Bivariate probabilities of exceedance are curves. These curves have been given various names in different research papers including the following:

- *isolines* (Salvadori, 2004; De Michele et al., 2007; Salvadori et al., 2016; Sadegh et al., 2017, 2018)

Field Code Changed

- level curves (Coles, 2001; Salvadori, 2004; De Michele et al., 2007; Volpi and Fiori, 2012; Serinaldi, 2015, 2016; Bevacqua et al., 2017).

For the specific case of the AND probability, the following names have been used:

- joint exceedance curves (Hawkes et al., 2002; Hawkes, 2008; Mazas and Hamm, 2017).
- quantile curves (De Haan and De Ronde, 1998; Chebana and Ouarda, 2011).

### 3 Simulation study

- Here we are interested in comparing the abilities of six different models presented in **Sect. 2.3** to reproduce a given dependence structure. We create 60 different synthetic dataset types with varying marginal distributions and dependence structures. By doing this, we aim to produce bivariate synthetic datasets comparable to the ones studied in bivariate hazard analysis (Zheng et al., 2014; Hendry et al., 2019) (Zheng et al., 2014; Hendry et al., 2019). This will allow us to confront the six models against the synthetic datasets, as a reference for bivariate hazard interrelation analysis (See **Sect. 4**). The six models compared in this simulation study are:

- joint exceedance curves (Hawkes et al., 2002; Hawkes, 2008; Mazas and Hamm, 2017).
- quantile curves (De Haan and De Ronde, 1998; Chebana and Ouarda, 2011).

### 3 Simulation study

Here we are interested in comparing the abilities of six different models presented in **Sect. 2.3** to reproduce a given dependence structure. We create 60 different synthetic dataset types with varying marginal distributions and dependence structures. By doing this, we aim to produce bivariate synthetic datasets comparable to the ones studied in bivariate hazard analysis (Zheng et al., 2014; Hendry et al., 2019) (Zheng et al., 2014; Hendry et al., 2019). This will allow us to confront the six models against the synthetic datasets, as a reference for bivariate hazard interrelation analysis (See **Sect. 4**). The six models compared in this simulation study are:

- the conditional extremes model (Cond-Ex) (**Sect. 2.3.2**);
- the non-parametric joint-tail model (JT-KDE) (**Sect. 2.3.3**);
- the Gumbel copula (Gumcop) (**Sect. 2.3.1**);
- the normal copula (Normalcop) (**Sect. 2.3.1**);
- the Farlie-Gumbel-Morgenstern (FGMcop) copula (**Sect. 2.3.1**);
- the Galambos copula (Galamboscop) (**Sect. 2.3.1**).

Among the four copulas used here, two are asymptotically dependent (Gumbel and Galambos) and two are asymptotically independent (normal and FGM). A description of the six models is given in **Table 2**. **Table 2** synthesizes a range of information about all the six models used in this simulation study including their type (nonparametric, semiparametric, parametric), equation, parameter range (if there is a parameter) and asymptotic modelling domain. This latter information is important to interpret the result of the simulation study in **Sect. 3.3**.

395

**Table 2: Description of the six statistical models compared in this article.** The description includes the model name and acronym (used [thoroughthrougough](#) the article), type of model (parametric, semi-parametric, non-parametric), the mathematical description, the parameter range (where relevant) and the asymptotic modelling domain (AI for asymptotic independence and AD for asymptotic dependence)

Model name	Model acronym	Model type	Mathematical description	Parameter range	Asymptotic modelling domain
Joint tail KDE	JT-KDE	Non-Parametric	$P(Z \in sA^+) \approx s^{-\frac{1}{\theta}} P(Z \in A^+)$		AD
		Semi-parametric	$P(Z \in sA^+) \approx s^{-\frac{1}{\theta}} P(Z \in A^+)$	$\eta \in [0,1]$	AI
Conditional Extremes Model	Cond-Ex	Semi-Parametric	$P\left(\frac{Y_2 - a[Y_1]}{b[Y_1]} \leq z, Y_2 - u > y   Y_1 > u\right) \rightarrow \exp(-y) G(z)$ $P\left(\frac{Y_2 - a[Y_1]}{b[Y_1]} \leq z, Y_1 - u > y   Y_1 > u\right) \rightarrow \exp(-y) G(z)$ <p>for <math>y &gt; 0</math>, as <math>u \rightarrow \infty</math> where <math>G(z)</math> is a non-degenerate distribution function.</p>		AI and AD
Gumbel copula	GumCop	Parametric	$C(u,v) = \exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right\}$ $C(u,v) = \exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right\}$	$\theta \in [1,\infty)$	AD
Normal copula	NormalCop	Parametric	$C(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right) dx dy$ $C(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right) dx dy$ <p>With <math>\Phi(\cdot)</math> the standard Gaussian distribution function</p>	$\theta \in [-1,1]$	AI
FGM copula	FGMCop	Parametric	$C(u,v) = uv[1 + \theta(1-u)(1-v)]$	$\theta \in [-1,1]$	AI
Galambos copula	GalambosCop	Parametric	$C(u,v) = uv \exp\left\{-\left[(-\ln(u))^{-\theta} + (-\ln(v))^{-\theta}\right]^{-\frac{1}{\theta}}\right\}$	$\theta \in [0,\infty)$	AD

Formatted: Font: Bold

Formatted Table

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

In this section, we first describe and display the synthetic data that have been generated to conduct this study. We shall then present the measures used in this study to compare the level curves and the dependence measures estimated from the six models presented in **Table 2**. Finally, results of the simulation will be displayed and analysed.

3.1 Synthetic data

Synthetic datasets are often used to [confront or](#) compare different statistical models ([Chebana and Ouarda, 2014; Zheng et al., 2014; Cooley et al., 2019](#))([Chebana and Ouarda, 2011; Zheng et al., 2014; Cooley et al., 2019](#)). Here we generated 60 bivariate synthetic datasets representative of environmental data such as daily rainfall, daily wind gust and daily wildfire occurrences (see **Sect. 4**). The number of synthetic data points we use here have been fixed to 5000 for each dataset. For the asymptotic dependence case, 36 distinct datasets are generated from a Gumbel copula (see **Supplement S1.3.1**); for the asymptotic independence case, 24 datasets are generated from a normal copula (see **Supplement S1.3.2**). Each synthetic dataset set of parameters has been used to generate 100 realizations to produce confidence intervals.

The synthetic datasets are generated from two marginal distributions and a dependence model (i.e., copula). Both marginal distributions are log-normal; the log-normal distribution has been used (among others) for the modelling of a wide range of natural hazards, including wind, flood and rainfall ([Malamud and Turcotte, 2006; Clare et al., 2016; Loukatou et al., 2018; Nguyen Sinh et al., 2019](#)).

Random variables  $X$  with a log-normal distribution are governed by two parameters: the location parameter  $\mu$  and the shape parameter  $\sigma$  which correspond respectively to the mean and the standard deviation of  $Y$ , the variable's natural logarithm, i.e.,  $Y = \ln(X)$  (Aitchison, 1957). The parameter  $\sigma$  influences the shape of the distribution and the heaviness of the tail and the dispersion of a log-normal distribution mostly depends on the shape parameter (Koopmans et al., 1964)

Field Code Changed

We can characterize log-normal distributions with the coefficient of variation  $c_v$  which is the ratio of the standard deviation  $s$  of the log-normally distributed variable  $x$  to its nonzero mean  $\bar{x}$  (Malamud and Turcotte, 1999)(Malamud and Turcotte, 1999):

$$c_v = \frac{s}{\bar{x}} \quad (18)$$

The standard deviation  $s$  and the nonzero mean  $\bar{x}$  are both related to the two parameters  $\mu$  and  $\sigma$  of the log-normal distribution (see Table 3). The use of the coefficient of variation characterises the log-normal distribution with one single parameter instead of two. The distribution used in the simulation study, the parameters and the relationship between these parameters and the different tail dependence measures are summarised in Table 3.

**Table 3: Marginal distributions and copula used for the synthetic datasets**

Distribution	Cumulative density function	Parameters	Parameter $s$ values
Log-normal distribution	$F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$ $F(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$	$\mu, \sigma$ of $y$ , where $y = \ln(x)$ $\bar{x} = \exp(\mu + \sigma^2/2)$ $s = \sqrt{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)}$ $c_v = s/\bar{x}$	A: $c_v = 0.25$ B: $c_v = 0.53$ C: $c_v = 2.91$
Gumbel copula	$C(u, v) = \exp\left\{-\left[-\ln(u)\right]^\theta - \left[-\ln(v)\right]^\theta\right\}^{1/\theta}$ $C(u, v) = \exp\left\{-\left[-\ln(u)\right]^\theta - \left[-\ln(v)\right]^\theta\right\}^{1/\theta}$	$\theta = \log_2(2 - \chi)$	$\chi = 0.05, 0.10, 0.30, 0.50, 0.70, 0.90$
Normal copula	$C(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right) dx dy$ $C(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right) dx dy$	$\theta = 2\eta - 1$	$\eta = 0.25, 0.50, 0.75, 0.90$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution

We use three different coefficient of variations  $c_v = 0.25$  (labelled as **A** for the rest of this paper),  $0.53$  (labelled **B**) and  $2.91$  (labelled **C**) (See Table 3). The log-normal distribution **A** ( $c_v = 0.25$ ) produces a distribution close to the normal distribution. The distribution **C** ( $c_v = 2.91$ ) is a highly right-skewed distribution. The distribution **B** ( $c_v = 0.53$ ) is intermediate skewness between **A** and **B**. In the bivariate context, there are six possible combinations of these distributions: **AA**, **AB**, **AC**, **BB**, **BC**, and **CC**.

The dependence structure is represented by a Gumbel copula in the case of asymptotic dependence (AD) and a normal copula in the case of asymptotic independence (AI) as no copula can be both asymptotically independent and asymptotically dependent (Heffernan, 2000; Coles, 2001)(Heffernan, 2000; Coles, 2001). The Gumbel copula is an extreme value copula, asymptotically dependent (see Supplement Eq. S19). The Gumbel copula function only has one parameter  $\theta$  which can be related the extremal dependence measure  $\chi$ . Here, we vary  $\chi$  between  $0.05$  (very weak asymptotic dependence) and  $0.9$  (strong asymptotic dependence) (see Fig. 5). The Normal (or Gaussian) copula is asymptotically independent. Its unique parameter is related to the coefficient of tail dependence  $\eta$  (Heffernan, 2000). We vary  $\eta$  from  $\eta = 0.25$  (negative sub-asymptotic dependence) to  $\eta = 0.9$  (positive sub-asymptotic dependence) (see Fig. 5). In total, ten different dependence structures were simulated for each of the six combinations of marginal distributions. The 60 bivariate synthetic datasets used in this study are displayed in Fig. 5.

Formatted Table

Formatted: Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font: 12 pt, Font color: Black

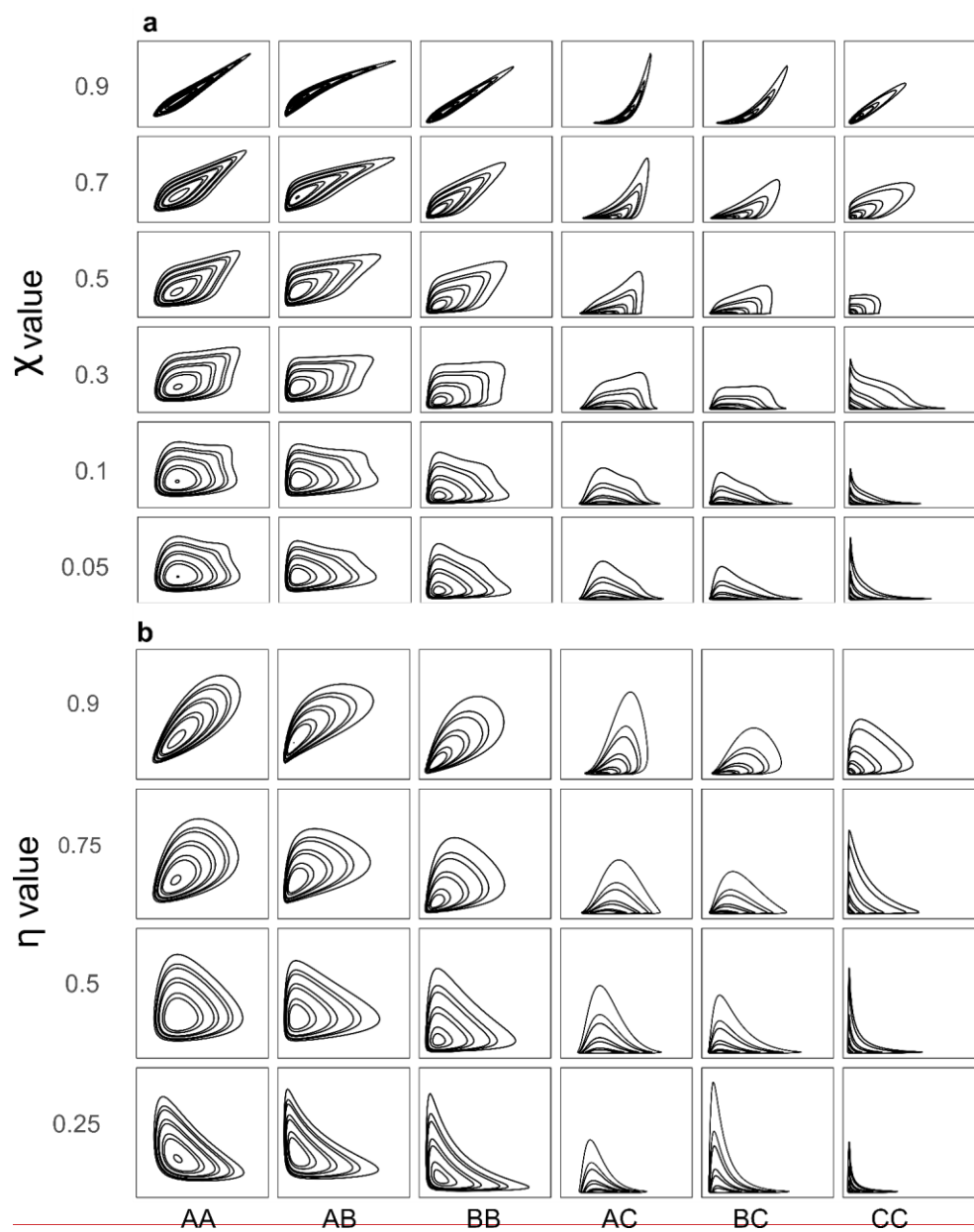
Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font: 12 pt, Font color: Black

Formatted: Font color: Black

Field Code Changed





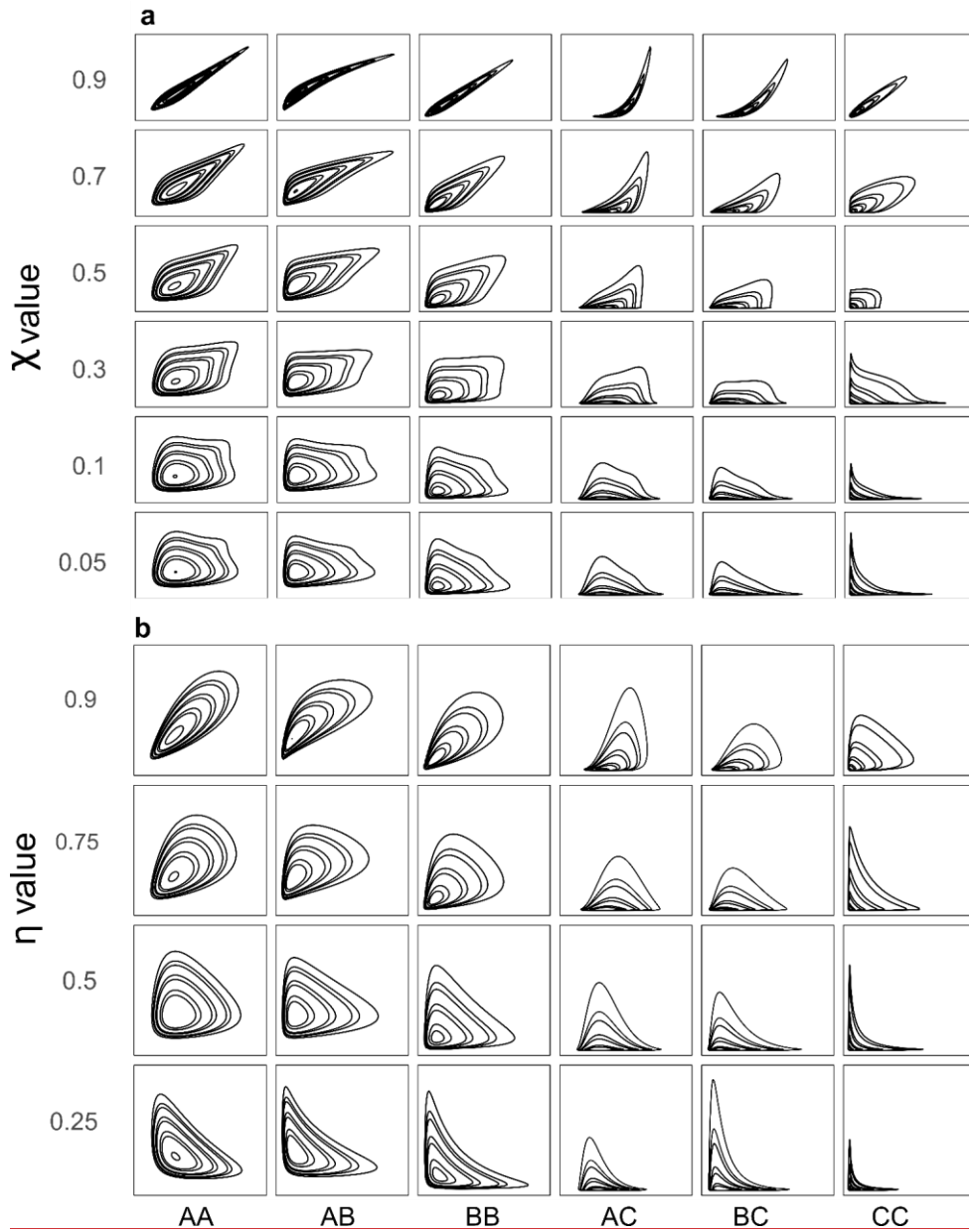


Figure 5: The 60 different synthetic bivariate datasets used in our simulation study. On the y-axis: the dependence strength (a)  $\chi$  (for asymptotic dependence) and (b)  $\eta$  (for asymptotic independence), vary from slightly negative association to heavily dependent (see also Fig. 2). On the x-axis AA to CC represent the marginal distributions that are part of the bivariate distributions (see Table 3) with A, B, C representing log-normal distributions with different coefficient of variations  $c_V$  (A:  $c_V = 0.25$ ; B:  $c_V = 0.47$ ; C:  $c_V = 0.95$ ).

To compare the fitting capabilities of the different models presented in **Sect. 2.3**, we vary several characteristics of the synthetic dataset:

- (i) *The shape of the marginal distributions.* Natural hazards can exhibit very diverse statistical properties depending not only on their type but also on the location where they occur (Sachs et al., 2012).
- (ii) *The strength of the dependence represented by the parameter of the copula function.* The type and strength of the relationship between natural hazards can vary within a broad range depending on the natural hazard studied or the location (Gill and Malamud, 2014; Martius et al., 2016). In order to take into account both AD and AI cases, the two parameters  $\chi$  and  $\eta$  (**Sect. 2.2**) are used.

Field Code Changed

### 3.2 Diagnostic tools

There are many diagnostic tools, to assess the goodness-of-fit of parametric bivariate models (Arnold and Emerson, 2011; Couasnon et al., 2018; Genest et al., 2009, 2011; Genest and Nešlehová, 2013; Sadegh et al., 2017). Amongst these, some of the most popular are the following:

- Cramer–von Mises statistic (Arnold, Taylor and Emerson, John, 2011);
- (ii) *The strength of the dependence represented by the parameter of the copula function.* The type and strength of the relationship between natural hazards can vary within a broad range depending on the natural hazard studied or the location (Gill and Malamud, 2014; Martius et al., 2016). In order to take into account both AD and AI cases, the two parameters  $\chi$  and  $\eta$  (**Sect. 2.2**) are used.

### 3.2 Diagnostic tools

There are many diagnostic tools, to assess the goodness-of-fit of parametric bivariate models (Arnold and Emerson, 2011; Couasnon et al., 2018; Genest et al., 2009, 2011; Genest and Nešlehová, 2013; Sadegh et al., 2017). Amongst these, some of the most popular are the following:

- Cramer–von Mises statistic (Arnold, Taylor and Emerson, John, 2011);
- Kolmogorov-Smirnov test (Arnold, Taylor and Emerson, John, 2011); (Arnold, Taylor and Emerson, John, 2011);
- Akaike information criterion (AIC) (Akaike, 1974);
- Bayesian information criterion (BIC) (Schwarz, 1978).

These measures have been developed in a univariate framework and then extended to the bivariate framework. Genest (2009) proposed several approaches for Cramer–von Mises and Kolmogorov-Smirnov goodness-of-fit tests for copulas. There are two issues we faced using these measures for our study:

- (i) these criteria are designed to fit on the dependence structure of the whole dataset and not the extreme dependence which can be different;
- (ii) in our study we aim to compare parametric and non-parametric models.

To tackle the first issue, goodness-of-fit test have been developed for extreme value copulas (Genest et al., 2011). The latter issues is more complicated; each modelling approach has its own fitting methodologies, and although it is now possible to compare copulas against each other (Sadegh et al., 2017; Couasnon et al., 2018), it is more difficult to compare copulas against semi-parametric or non-parametric models. The measures mentioned above are not suitable as they imply parametric distributions to be compared against observations (Stephens, 1970; Arnold, Taylor and Emerson, John, 2011). It is then not possible to compare the goodness-of-fit on the whole range of the data.

However, we are interested in fitting capabilities in the extremes. The models will then be compared on the estimation of two attributes of the synthetic data detailed below:

Field Code Changed

(i) The  $P_{AND}$  probability of exceedance (Sect. 2.3) represented by the level curve at  $p = 0.001$ .

(ii) The tail dependence measures  $\chi$  and  $\eta$  (Sect. 2.2.2).

We present here the diagnostic tools related to the level curve. To tackle the first issue, goodness-of-fit tests have been developed for extreme value copulas (Genest et al., 2011). The latter issue is more complicated; each modelling approach has its own fitting methodologies, and although it is now possible to compare copulas against each other (Sadegh et al., 2017; Couasnon et al., 2018), it is more difficult to compare copulas against semi-parametric or non-parametric models. The measures mentioned above are not suitable for the present study as they require models to be parametric to be compared against observations (Stephens, 1970; Arnold, Taylor and Emerson, John, 2011). It is then not possible to compare the goodness-of-fit of the six models used in this study all together.

However, we are interested in fitting capabilities in the extremes. The models will then be compared on the estimation of two attributes of the synthetic data detailed below:

(i) The  $P_{AND}$  probability of exceedance (Sect. 2.3) represented by the level curve at  $p = 0.001$ .

(ii) The tail dependence measures  $\chi$  and  $\eta$  (Sect. 2.2.2).

We present here the diagnostic tools related to the level curve. The tools used to compare tails dependence measures can be found in **Appendix A**. Here we chose to compare our six models with respect to their ability to reproduce a reference level curve from the underlying bivariate  $(X_1, X_2)$  distribution of the data  $l_{obj}$  ('obj' is again used to indicate objective) which corresponds to an extreme joint probability  $p = 0.001$ . For each model  $i$  a level curve  $l_{obj,i}$  is computed. Several methods and criteria have been used in the literature to compare level curves to a reference including comparing the curves with a vertical point-wise distances between the underlying curves (Chebana and Ouarda, 2011). This approach finds its limitation when level curves do not share the same  $x$ -axis coordinate ( $X_1$  axis). In **Fig. 6** is presented our procedure for computation of the goodness-of-fit indicators (described in further detail below). In **Fig. 6** the example modelled and reference curves do not reach the same coordinate on the  $X_1$  axis, making it impossible to compare these two level curves between  $X_2=0.0$  and  $X_2=0.3$ . Cooley et al. (2019) divided level curves into two parts, comparing six  $x$ -axis coordinates on one part and six  $y$ -axis on the other part, to overcome the aforementioned limitation. Here we chose to use a consistent criterion all along the curves to evaluate the distance between each modelled curve and the reference curve. The four steps we use are the following:

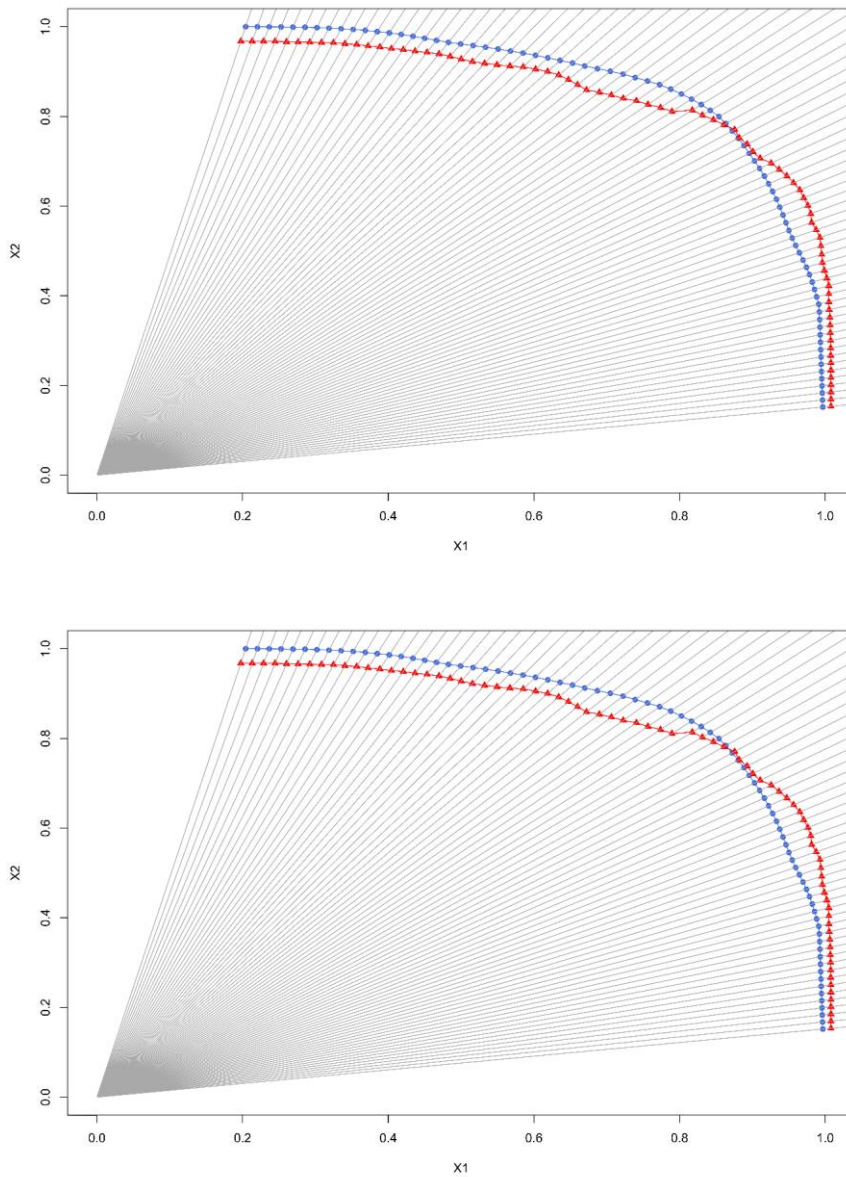
(i) Each modelled and reference level curve is normalized by dividing its coordinates by their maximum values. With that process, the curves are bounded in the  $[0,1]$  by  $[0,1]$  space. The different indicators are then computed in this normalized space.

(ii) Cartesian coordinates  $(x,y)$  of the modelled and reference level curves are transformed to polar coordinates  $(\theta, r)$ .

(iii) Each modelled and reference level curve is discretized via linear interpolation into points. Each point corresponds to an angle value (triangles and dots on the curves in **Fig. 6**).

(iv) Points from both the modelled and reference level curves with the same angle are coupled. Indicators are computed at each of the 80 couples of points (see **Fig. 6**).

The indicator designed in this study is derived from the distance between the two curves and are listed in **Table 4**:



**Figure 6: Procedure for computation of the goodness-of-fit indicators.** Two variables are given,  $X_2$  as a function of  $X_1$ . The red triangles and red curve represent the modelled level curve from a given model. The blue circles and blue curve are the reference level curve from the underlying bivariate  $(X_1, X_2)$  distribution of the data. Distance between the curves are calculated along the radius at 80  $(X_1, X_2)$  coordinates (e.g., between the blue circles and the red triangles).

We used a weighted euclidean distance ( $wd$ ) as comparison criteria. The density of level curves (described in the **Supplement S2**) allows one to weight the Euclidean distance of each of the 80 points by the local density of the curve. By weighting the Euclidean distance according to the reference bivariate distribution probability density function, we give more importance to the proper part of curve where a bivariate event is more likely to occur, rather than the naïve part (where the naïve part is defined as where the bivariate event is less likely to occur) (Chebana and Ouarda, 2011; Volpi and Fiori, 2012).

We used a weighted Euclidean distance ( $wd$ ) as comparison criteria. The density of level curves (described in the **Supplement S2**) allows one to weight the Euclidean distance of each of the 80 points by the local density of the curve. By weighting the Euclidean distance according to the reference bivariate distribution probability density function, we give more importance to the proper part of curve where a bivariate event is more likely to occur, rather than the naïve part (here the naïve part is defined as where the bivariate event is less likely to occur) (Chebana and Ouarda, 2011; Volpi and Fiori, 2012).

$$\sum_{i=1}^N w_i \left( \sqrt{(x_{mod,i} - x_{ref,i})^2 + (y_{mod,i} - y_{ref,i})^2} \right) \quad (19)$$

### 3.3 Results

Two analyses are conducted in parallel, one for asymptotic dependence (**AD**) and one for asymptotic independence (**AI**). In the case of asymptotic dependence, the Gumbel copula is used with 5000 data points. The  $\chi$  value is the one of interest under **AD**, values taken by  $\chi$  have been presented in **Sect. 3.1**. For each  $\chi$  value, we generated 100 realizations of the dataset from the same underlying bivariate distribution. The 100 realizations generated have two purposes: (i) increase the robustness of the results and (ii) create a confidence interval around the median which was set at the 95% confidence level by taking the quantiles Q2.5 and Q97.5 of the 100 realizations. To confront this approach, we generated two sets of 100 realizations which showed very small variations in the values of Q2.5, Q50 and Q97.5 without impacting our interpretation of the results. In the case of asymptotic independence, the normal copula is used. The  $\eta$  value is the one of interest in the **AI**, values taken by  $\eta$  are displayed in **Table 3**. For each  $\eta$  value, we generated 100 realizations of the dataset. The 100 realizations generated allowed us to create a 95% confidence interval using the same approach as above.

The marginal distributions do not have any impact on the dependence structure (Nelsen, 2006; Genest and Favre, 2007). The marginal distributions do not have any impact on the dependence structure (Nelsen, 2006; Genest and Favre, 2007). We show in **Appendix A** that marginal distributions also have a very small impact on the estimation of dependence measures. All the methods used in this study include a transformation of marginal distributions and the fitting of a GPD above an extreme threshold (**Sect. 2.3**). By varying the marginal distribution of the variables of our synthetic dataset we aim to capture uncertainties and errors arising from both the fitting of the marginal distributions and the dependence structure.

For both **AD** and asymptotic dependence **AD** and asymptotic independence **AI**, the objective level curve  $l_{obj}$  to be compared has been fixed at the probability  $p_{obj} = 0.001$ . For each of the 60 bivariate datasets, the six models presented are fitted to the 100 realizations. The dependence measures  $\hat{\chi}_i, \hat{\eta}_i$  as well as the level curve  $\hat{l}_{obj,i}$  are estimated for each six model, with  $i \in (1:6)$  correspond to each model. We then use the diagnostic tool and criteria presented in **Sect. 3.2** to compare the performance of the models. From the 100 realizations, 100 level curves  $\hat{l}_{obj,i}$  are generated for each model. Three curves are designed: (i) the 2.5% quantile level curve, (ii) The median level curve, (iii) the 97.5 % quantile level curve.

In an analogous way, for each of the diagnostic tools presented in **Sect. 3.2**, three values are computed: (i) the 2.5% quantile, (ii) the median, (iii) the 97.5% quantile. **Figure 7** displays the values of the  $wd$  for each model applied to each bivariate dataset. To assess more accurately whether the models manage to represent the synthetic data in the large value extremes, we compared their fitting capabilities to a naïve approach. Here, the naïve approach is an empirical level curve. For each of the 60 synthetic datasets, we compute the  $wd$  of the empirical level curves to the reference curves following the same steps as for

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font color: Black

Formatted: Font: Bold

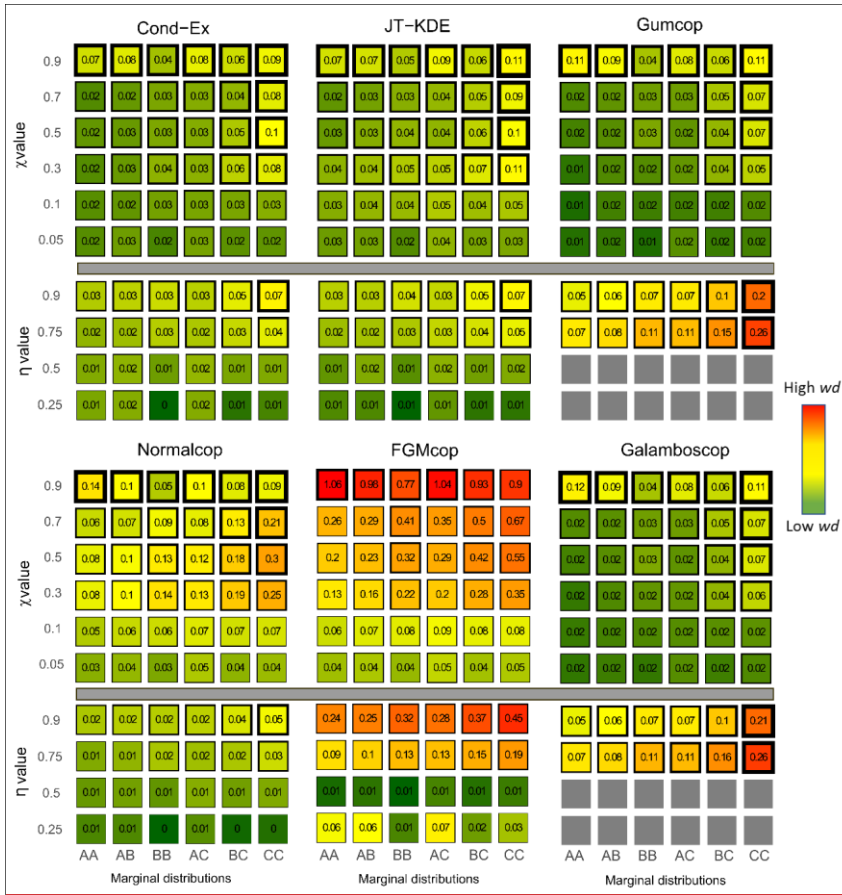
Formatted: Font: Bold

Formatted: Font: Italic

Formatted: Font: Bold

Formatted: Font: Bold

the six models. The empirical  $wd$  ( $wd_{naive}$ ) is therefore compared to the  $wd$  of each model. Models that represent the data with more accuracy than a naïve approach ( $wd < wd_{naive}$ ) are considered to be representative of the data. **Figure 7** displays the values of the  $wd$  for each model applied to each bivariate dataset and highlight the cases where models outperform a naïve approach (blue bold). Squares are coloured according to the median of the  $wd$  and thickness of the edges is proportional to the size the confidence interval (i.e., the distance between the quantiles  $Q2.5$  and  $Q97.5$ ).



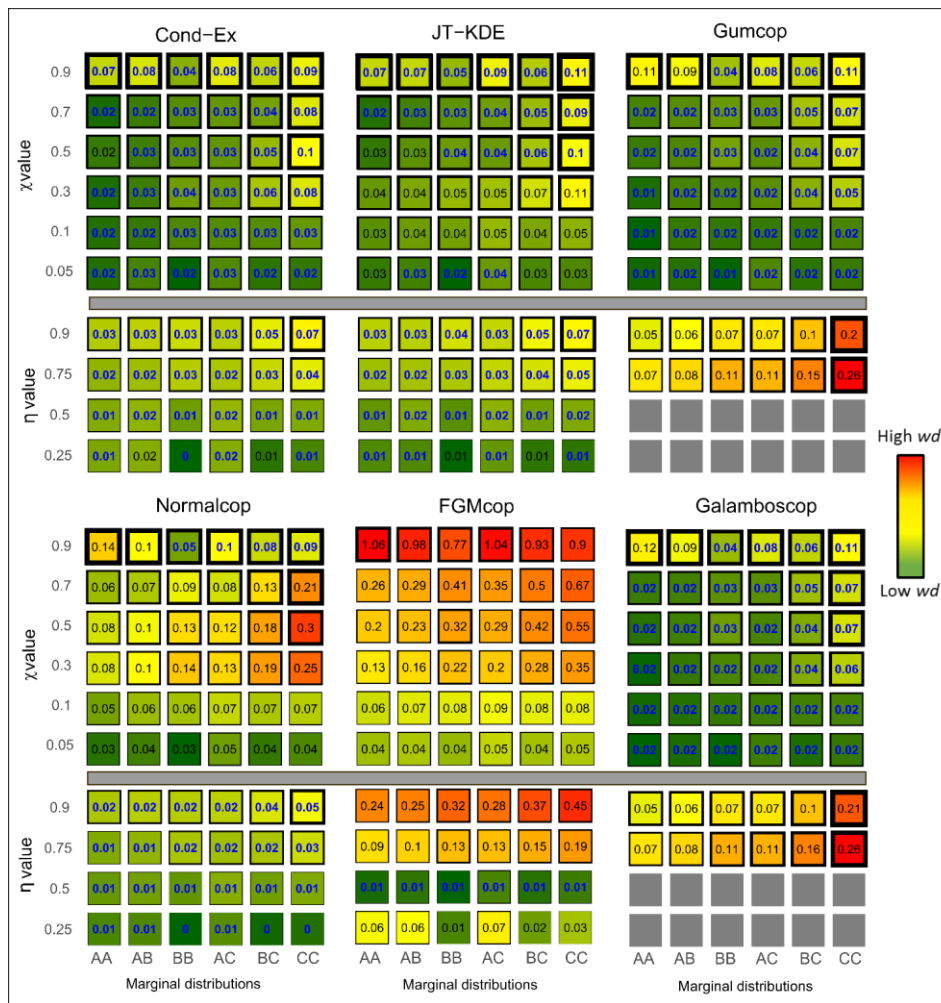


Figure 7: Weighted Normalized Euclidean Distance ( $wd$ ) to the reference curve for all 60 different synthetic datasets. Fitting capacities of each model are represented. Values in cells and colours represent the median  $wd$  from low (dark green) to high (red). Thickness of borders represent the 95% uncertainty around the median value on a logarithmic scale.

It is important here to note that we tested more **AD** (36) cases than **AI** (24) cases, 60% cases than **AI** (24-40%) cases. To assess, the flexibility of models, additionally to comparison to the naïve approach, we also consider the proportion of cases where model have a  $wd < 0.1$ . From Fig. 7, we observe the following:

- The Gumbel and normal copulas, which have been used to generate the synthetic datasets with **AD** and **AI**, generally outperform all the other models in **AD** and **AI** cases respectively.
- The conditional extremes model and the joint-tail KDE model are the most flexible **modelmodels** tested here as **they** can handle **respectively(Cond-Ex)** 98% [72–100%] and **(JT KDE)** 97% [65–100%] of the situation with a  $wd < 0.1$ ; these values reach 100% for the **AI** cases. However, the Cond-Ex model is slightly more flexible, having a representative fit to more datasets (95%) than the JT-KDE model (68%).

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold



- The normal copula, even if asymptotically independent, is the most flexible copula model with a  $wd < wd_{naive}$  in 47% of the cases, more than the number of AD datasets. The normal copula has a low  $wd (< 0.1)$  in 76% [60–90%] of the cases, and has a representative fit to the data for every AI cases and in some AD cases.
- Gumbel and Galambos copulas show very similar behaviours with respectively 68% [53–93%] and 68% [52–93%] cases with  $wd < 0.1$ . When have representative fits to only considering 57% of the AD datasets. Among the 36 asymptotic-independent AD cases, this value reaches 94%. they fail to represent only two with  $\chi=0.9$ . It is important to note that both aforementioned copulas cannot handle complete independence ( $\eta=0.5$ ) or negative dependence ( $\eta=0.25$ ).
- The FGM copula can only handle one type of extremal dependence, which is asymptotic independence (AD) with  $\eta=0.5$ . Consequently, it is the least flexible model in our results with a  $wd < 0.1$   $wd_{naive}$  in only 41% [40–43%] 10% of the cases.
- All models have poorer goodness-of-fit while shape parameters are increasing. It is particularly striking with the conditional extremes approach which exhibits high uncertainty and high  $wd$  when both margins have a standard deviation  $\sigma=1.5$ .

The Cond-Ex and JT-KDE provide similar results according to **Fig. 7** despite adopting very different approaches. Thus, their flexibility arises from their semi parametric nature. **Figure 7** also displays the uncertainty of the estimate of  $wd$ . For all models, a more accurate fit is accompanied with a reduction in uncertainties. However, both Cond-Ex and JT-KDE have on average more uncertainty around its  $wd$  despite their good fitting capabilities. On average, copulas tend to have less uncertainty due to their parametric nature.

However, the copulas are penalized by the weighting function as they usually reproduce quite well the naïve part of the curve. By considering again the percentage of situations with a criterion below 0.1, the normal copula has its performances reduced by the weighting function (–6% compared to  $d$ ). The JT-KDE model has its performance boosted by the weighting function (+7% compared to  $d$ ).

#### 4 Application to natural hazards

Results from the simulation study presented in the previous section (**Sect. 3**) can provide useful insights when modelling the interrelations between two natural hazards. In this section, we will show how results previously presented can be useful to identify the most relevant models for a given dataset according to its visual characteristics. The concordance (or discordance) of the relevant models can also increase (decrease) confidence around the results.

The methodology for model selection presented here is composed of five steps to select the most relevant models estimate joint exceedance probability level curves:

- (i) The two-tail dependence measures are estimated empirically with a 95% confidence interval. The dataset with a tail dependence measure falling in that confidence interval are suggested as analogue to the studied bivariate dataset. To select relevant combinations of marginal distribution, a scatterplot is compared visually to density plots for the 60 different datasets simulated in **Sect. 3** and displayed in **Fig. 65**.
- (ii) From the aforementioned 60 datasets, a set of one to six analogous datasets (i.e. with similar bivariate distribution) is taken.
- (iii) A confidence score is used to compare the abilities of each models for the datasets selected datasets in step (ii). For each model, the confidence score is  $\overline{wd}$  the average of the computed weighted Euclidian distance  $wd$  for all datasets selected in step (ii). By taking the average of  $wd$ , a poor fit on one analogous dataset will have a high influence on the confidence score.

- (iv) Models are ~~fitted-onfit to~~ the bivariate hazard dataset and level curves from the most relevant models are kept.
- (v) Tail dependence measures are estimated using the most relevant model with a possible new iteration of the four previous steps according to the value of the dependence measures.

To produce a confidence interval as was done in the simulation study (Sect. 3) and to visually measure the uncertainty associated to each level curve as in Sect. 3, we use a nonparametric bootstrap procedure. The function *tsboot* from the R package *boot* (Davison and Hinkley, 1997; Canty and Ripley, 2019) is used to generate 100 bootstrapped replicate datasets with the same number of observations as the original (but some are repeated). Our six models are then fitted to the original dataset and on the 100 bootstrapped replicates.

#### 4.1 Rain and wind gusts at Heathrow Airport (Asymptotic independence)

~~Here we study the interrelation between daily extreme wind gusts ( $w$ ) and extreme rainfall ( $r$ ) at London Heathrow airport, UK for the period 1 January 1971 to 31 May 2018, both introduced in Fig. 1. The relationship between wind and rainfall has been studied both globally (Martius et al., 2016) and locally (Johansson and Chen, 2003; Ming et al., 2015). These two hazards are often associated with different types of storms (Dowdy and Catto, 2017) and in particular cyclones (Ming et al., 2015; Raveh-Rubin and Wernli, 2016). In South England, these two hazards are mostly associated with mid-latitude cyclones in the winter season and thunderstorms in summer season (Hawkes, 2008; Anderson and Klugmann, 2014; Webb and Elsom, 2016; Hendry et al., 2019).~~

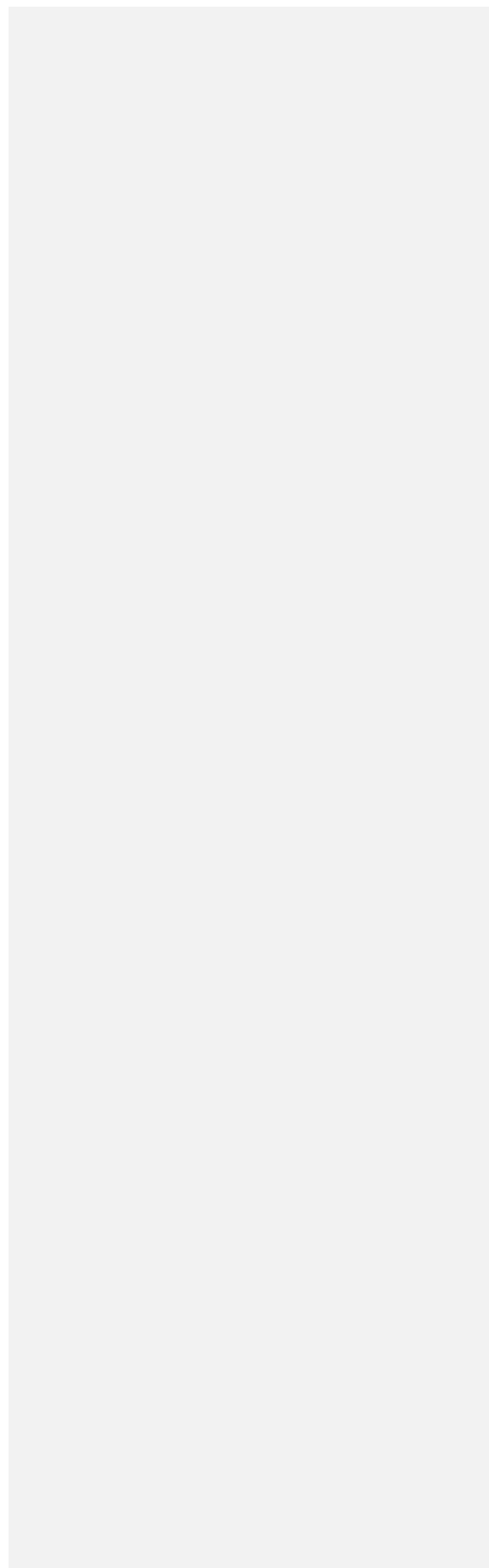
The bivariate dataset used to study the interrelation between wing gusts and rainfall at Heathrow airport is composed of the following data:

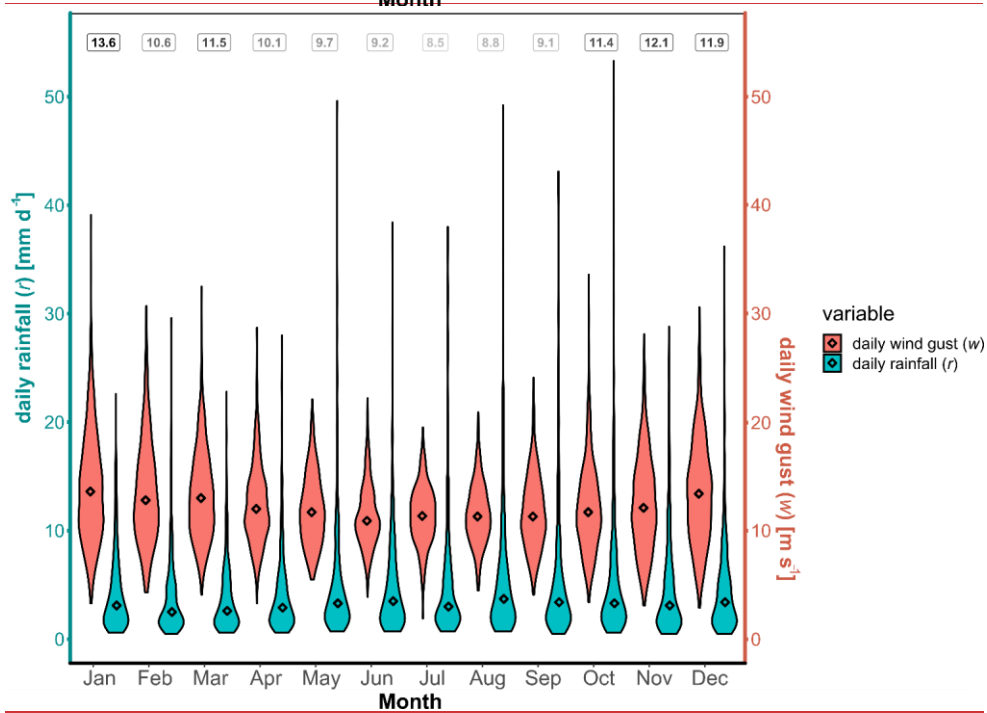
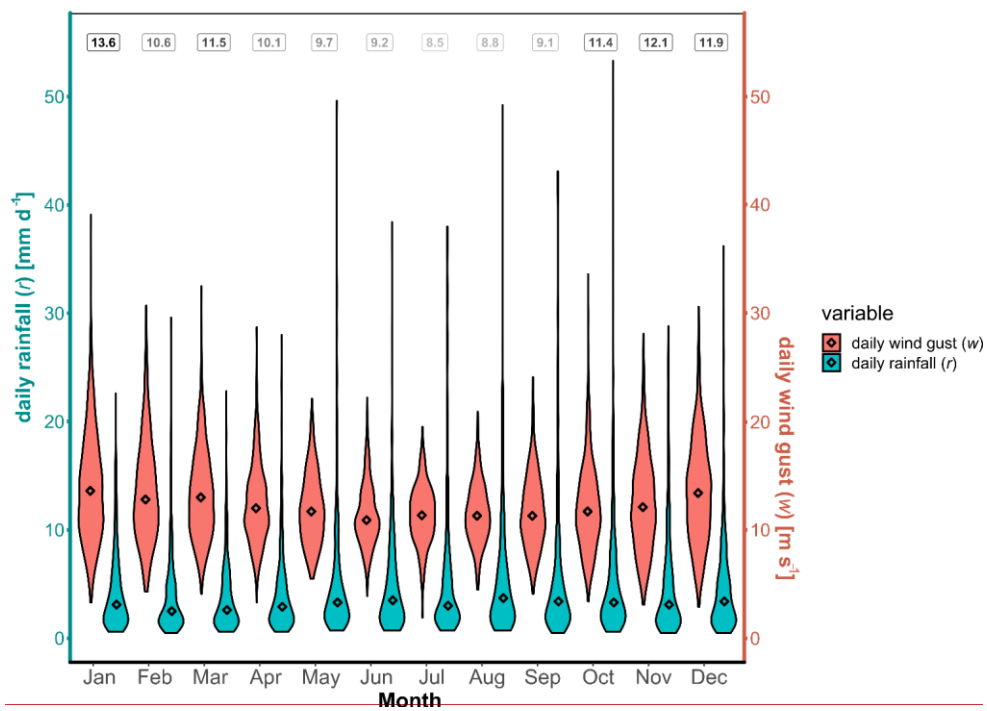
~~Here, we study the interrelation between daily extreme wind gusts ( $w$ ) and extreme rainfall ( $r$ ) at London Heathrow airport, UK for the period 1 January 1971 to 31 May 2018, both introduced in Fig. 1. The relationship between wind and rainfall has been studied both globally (Martius et al., 2016) and locally (Johansson and Chen, 2003; Ming et al., 2015). These two hazards are often associated with different types of storms (Dowdy and Catto, 2017) and in particular cyclones (Ming et al., 2015; Raveh-Rubin and Wernli, 2016). In South England, these two hazards are mostly associated with extratropical cyclones in the winter season and thunderstorms in summer season (Hawkes, 2008; Anderson and Klugmann, 2014; Webb and Elsom, 2016; Hendry et al., 2019).~~

The bivariate dataset used to study the interrelation between wing gusts and rainfall at Heathrow airport is composed of the following data:

- a) *Daily Wind Gust ( $w$ )*: daily maximum wind gust at ~~London~~ Heathrow airport (~~London~~, UK) weather station where a gust is the maximum 3 second average of wind speed. Wind gusts are short lived wind peaks in speed that can inflict great damage during a storm; ~~however~~. ~~However~~, it might not capture the overall wind intensity (Met Office, 2019). The time range of the observations is 38 years, from 1 January 1971 to 31 May 2018 of which 74 days (0.4% of the data) had no values recorded and all other values in the dataset had  $w > 0 \text{ m s}^{-1}$ . This observation data has been provided by the Met Office (2019).
- b) *Daily Rainfall ( $r$ )*: daily total precipitation in a grid cell containing ~~London~~ Heathrow airport- (~~UK~~). The data have been extracted from the E-OBS gridded database (Cornes et al., 2018) which is formed from the interpolation of observations from 18,595 meteorological stations through Europe and the Mediterranean (including Heathrow airport station). It has been shown that E-OBS has excellent correlation with other high-resolution gridded datasets even if this correlation tends to decrease for extremes (Hofstra et al., 2009)(Hofstra et al., 2009). However, by selecting a grid containing a weather station we limit uncertainties arising from interpolation. The spatial resolution in the E-OBS dataset is  $0.1^\circ \times 0.1^\circ$  and the period covered is 1950 to 2019.

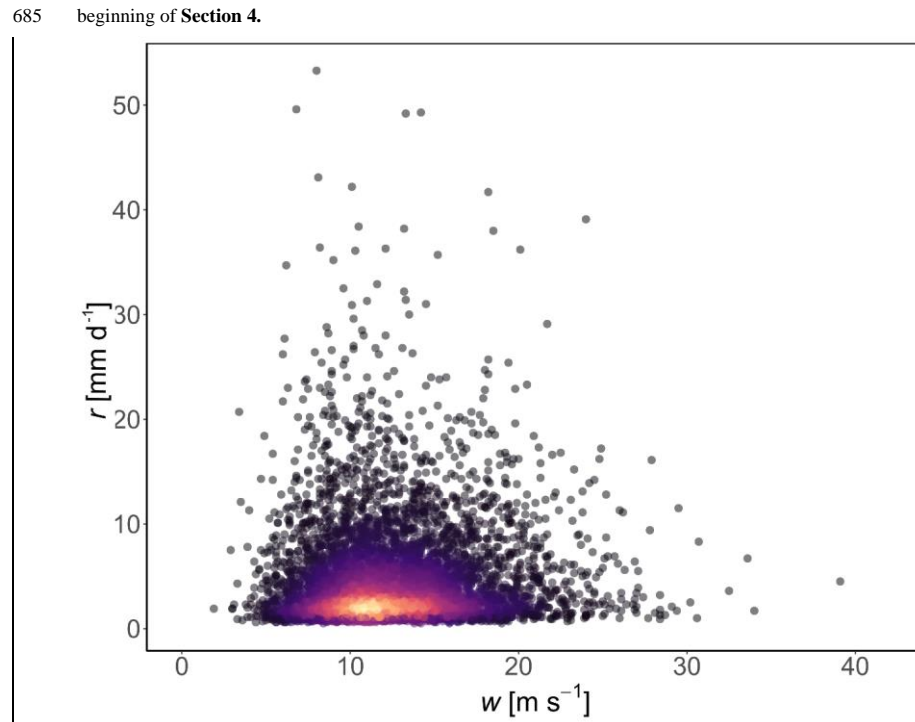
665 Data from 1 January 1971 to 31 May 2018 (38 years) in the cell containing Heathrow airport is used, with a total  
of 6074 days (35.1% of the dataset) having nonzero rainfall  $r > 0 \text{ mm d}^{-1}$ .  
From 1 January 1971 to 31 May 2018 there are a total of 17,318 days (including leap years). Our bivariate wind gust-rainfall  
dataset is composed of those values where there is both non-zero rainfall  $r > 0 \text{ mm d}^{-1}$  and windgusts  $w > 0 \text{ m s}^{-1}$  recorded,  
resulting in a total of 6044 bivariate observations (34.9% of the days in our record). An overview of both daily rainfall and  
670 daily wind gust is displayed in **Fig. 8** in the form of monthly violin plots, where the probability density of  $w$  and  $r$  at different  
values are given, smoothed by a kernel density estimator.

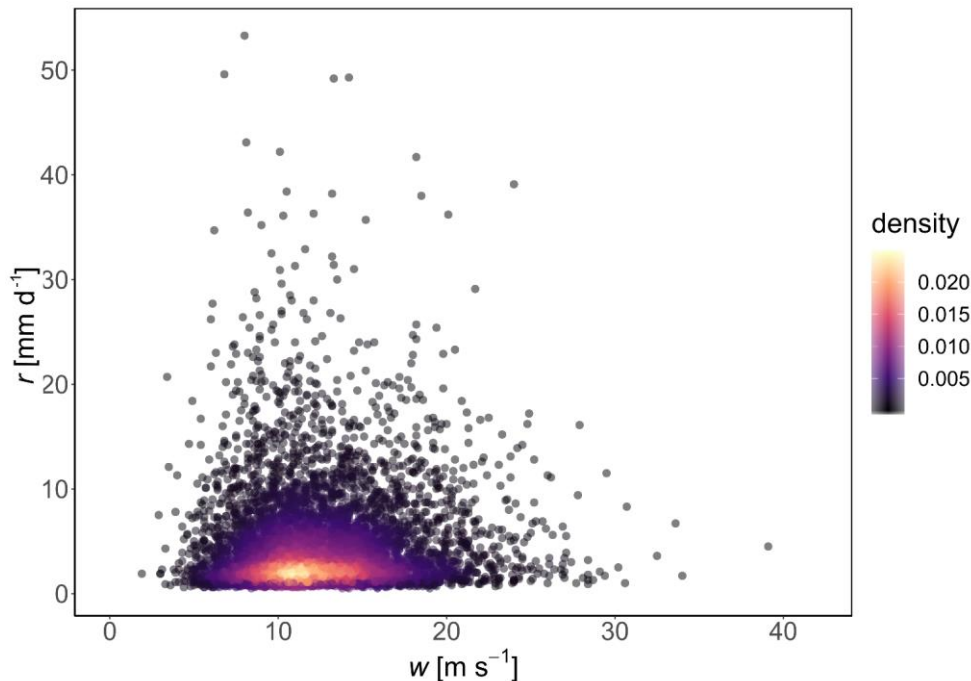




675 **Figure 8:** Violin plots of daily wind gust  $w$  (red) and daily non-zero rainfall  $r$  (blue) by month for the period 1 January 1971 to 31 May 2018 at Heathrow airport weather station, UK. Diamonds represent the median of all values for that month from 1971-2018. Numbers at the top of the graph represent the average number of days per month where there is recorded both non-zero rainfall  $r > 0 \text{ mm d}^{-1}$  and windgusts  $w > 0 \text{ m s}^{-1}$ . Daily rainfall data from E-OBS (Corney et al., 2018) and wind gust data (maximum 3 s wind velocity in a day) from the Met Office (2019).

680 From **Fig. 8** we observe a seasonality in daily wind gust speed. January is the month with the highest median (diamond symbol) and range of most values in the violin plot while July is the month with the lowest median and range of most values in the violin plot. The daily non-zero rainfall median per month varies between 2.5 mm in February and 3.5 mm in June, with the highest individual daily values occurring in October (53.3 mm d<sup>-1</sup>), May (49.6 mm d<sup>-1</sup>) and June (49.2 mm d<sup>-1</sup>). The dataset is also represented as a scatterplot in **Fig. 9**. The scatterplot will be used for the model selection methodology presented at the beginning of **Section 4**.





**Figure 9:** Days where there are recorded both daily wind gust ( $\text{m s}^{-1}$ )  $w$  and nonzero daily rainfall ( $\text{mm d}^{-1}$ )  $r > 0 \text{ mm d}^{-1}$  at Heathrow airport (London, UK) for the period 1971–2018. Daily rainfall data from E-OBS (Cornes et al., 2018) and wind gust data (the maximum 3 s wind velocity in a day) from the Met Office (2019). Colours (legend) represent the bivariate density estimated from a kernel density estimator with higher values and lighter colours representing a higher density of points at that bivariate value ( $r, w$ ).

Extreme rainfall and extreme wind have a compound interrelation according to Tilloy et al. (2019). We then estimate the joint exceedance probability curve, corresponding to a  $P_{\text{AND}}$  probability (Sect. 2.3).

We now go through the four steps presented for rainfall and wind gusts in Heathrow.

- (i) From Figs. 95 and 69, along with empirical estimates of  $\chi$  and  $\eta$ , we hypothesize that over our time range 1971–2018, daily rainfall and daily maximum wind gusts in London Heathrow Airport are asymptotically independent or weakly dependent ( $\eta = 0.5 / \chi = 0.05 / \chi = 0.1$ ) and that both marginal distributions have a small shape parameter (A-B-B-BAB, BB).
- (ii) This then gives us four analogous datasets and it is then possible to visually infer from Fig. 6 which models are the most adapted to suitable for these conditions. The four analogous datasets are the following:

1.  $\chi = 0.05$  and A-BAB
2.  $\chi = 0.05$  and B-BBB
3.  $\eta = 0.5$  and A-BAB
4.  $\eta = 0.5$  and B-BBB
5.  $\chi = 0.1$  and AB
6.  $\chi = 0.1$  and BB

Field Code Changed

Formatted: Font: Not Bold

Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 1.9 cm + Indent at: 2.54 cm

Formatted: Indent: Left: 2.54 cm, No bullets or numbering

710

715

720

725

(iii) The *confidence score* for each model is  $\overline{wd}$  the average of the weighted Euclidean distance  $wd$  from the four situations above. For the Gumbel and Galambos copulas, the cases of independence or negative dependence between variables are outside the modelling range (Sect. 2.3.1), and thus the confidence score for these models has been penalized by putting  $wd = 1.0$  for  $\eta = 0.5$  and  $\eta = 0.25$ . The conditional extremes model, normal copula and JT-KDE model all have a  $\overline{wd} = 0.02$  and is representative for all six analogous datasets. The JT-KDE model has a  $\overline{wd} = 0.03$  and is representative for 4 out of 6 analogous. The FGM and Normal copula have a confidence score of  $\overline{wd} = 0.0304$  and are only representative in AI cases. Gumbel and Galambos copulas have a confidence score of  $\overline{wd} = 0.535$  due to their penalty (Table 4).

According to these three first steps, the conditional extremes model appears to be the most suitable. However, we can identify the four most relevant models for the bivariate dataset of daily rainfall and daily wind gust at London Heathrow Airport. The conditional extreme model, the JT-KDE model, the normal copula and the FGM copula are the most relevant models for our dataset as it can be seen in Table 54.

Table 4: Euclidian weighted distance ( $wd$ ) for datasets 1 to 4 based on wind-rainfall and six models, along with confidence scores (average of the  $wd$  for datasets 1 to 4). In bold are highlighted the four models with confidence scores  $< 0.1$ .

Dataset	Cond-Ex	JT-KDE	Gumbelcop	Normalcop	FGMcop	GalambosCop
1	0.03	0.03	0.02	0.04	0.04	0.02
2	0.02	0.02	0.01	0.03	0.04	0.02
3	0.02	0.02	1.00	0.01	0.01	1.00
4	0.01	0.01	1.00	0.01	0.01	1.00
Average	0.02	0.0204	0.5102	0.0206	0.0307	0.5102
6	0.03	0.04	0.02	0.06	0.08	0.02
Average	0.02	0.03	0.35	0.04	0.04	0.35

(iv) For illustration and/or confronting our models with the data, the six models are fit to the dataset and joint exceedance level curve are produced with a joint exceedance probability set at  $p = 0.001$ , corresponding to a bivariate return period of 8 years. However, another joint exceedance probability could have been chosen. In Fig. 10 are displayed the level curves produced from the four models that were selected after steps (i) to (iii) above (Cond-Ex, JT-KDE, NormalCop and FGMCop) and presented in bold numbers in Table 4.

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Not Bold, Font color: Auto

Formatted: Font: Not Bold

Formatted: Right

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted Table

Formatted: Font color: Custom Color(0,51,204))

Formatted: Font: Not Bold, Font color: Auto

Formatted: Font: Not Bold

Formatted: Font: Not Bold

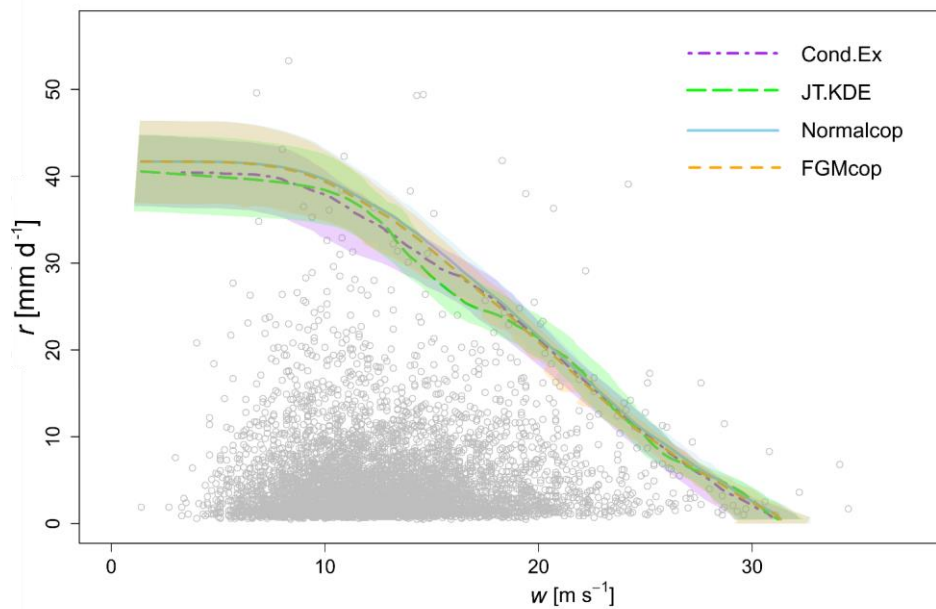
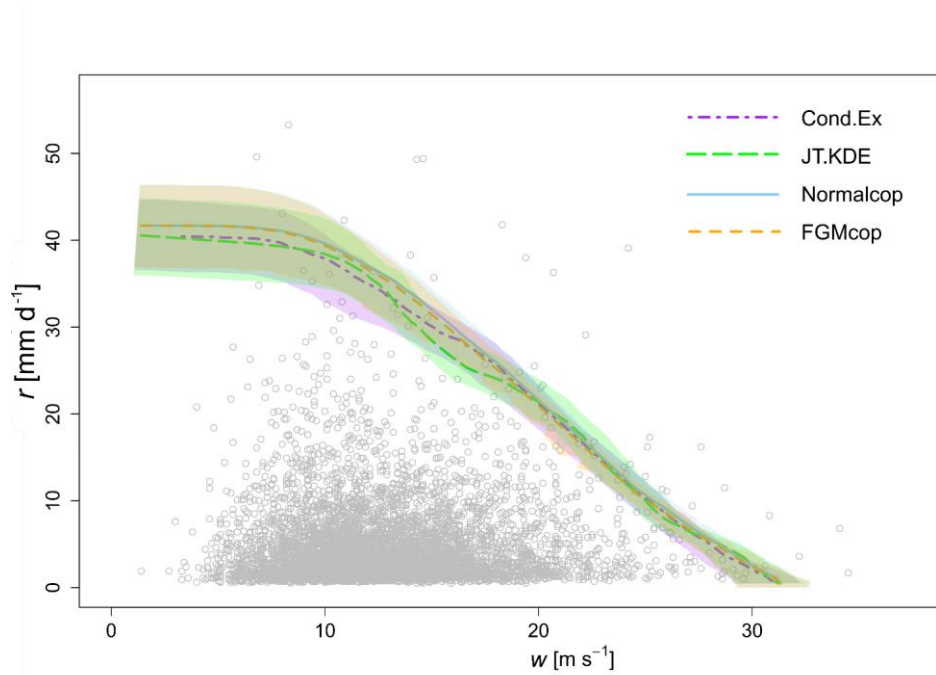
Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Not Bold

Formatted: Font: Bold, Font color: Custom Color(0,51,204))





**Figure 10:** Level curves for a  $P_{\text{and}}$  joint probability  $p = 0.001$  of daily wind gust and daily rainfall at Heathrow airport (London, UK). Level curves from the four models selected through the model selection methodology are displayed.

From **Fig. 10**, we can observe that the conditional extremes model, the FGM and the normal copula all produce very similar joint exceedance [curves](#) and that their confidence intervals overlap. **Table 5** displays the estimates (with bounds of the 95% confidence interval) of the two dependence parameters  $\chi$  and  $\eta$  from the six models. These estimates converge toward a very weak asymptotic dependence. However, the estimation of dependence parameters in near independence is highly uncertain (**Sect. 3.3.2**).

**Table 5:** Estimates of dependence parameters  $\chi$  and  $\eta$  for extreme rainfall and wind gust at Heathrow airport for the time range 1971-2018

<i>Models</i>	<i>Cond-Ex</i>	<i>JT-KDE</i>	<i>Gumcop</i>	<i>Normalcop</i>	<i>FGMcop</i>	<i>GalambosCop</i>
$\chi$	0.01 [0.00,0.02]	0.06[0.05,0.09]	0.04[0.01,0.06]	0[0,0]	0[0,0]	0.04[02,0.06]
$\eta$	0.49[0.45,0.54]	0.54[0.49,0.59]	1[1,1]	0.52[0.51,0.54]	0.5[0.5,0.5]	1[1,1]

#### 4.2 Wildfire [Daily wildfire number](#) and [extreme temperature extremes](#) in Portugal (Asymptotic dependence)

Here we present a second example of applying our models to natural hazards data, using as a case study [daily temperature and daily number of wildfires in Portugal](#). Wildfire variables such as daily number and burned area depend on many influences such as wind speed/direction/gustiness, topography, and type of fuel and soil moisture (Hinks et al., 2013). The aim of our study is not to decipher the processes leading to a wildfire but rather to provide an exemplar study examining the relationship between the two variables, daily temperature and daily number of wildfires, in a given case study area. It has been shown that dry and warm conditions increase the risk of wildfire ([AghaKouchak et al., 2018](#); [Zscheischler et al., 2018](#))([Littell et al., 2009](#); [AghaKouchak et al., 2018](#)), [Witte et al. \(2011\)](#)[Witte et al. \(2011\)](#) established a direct link between a persistent heatwave and wildfire outbreaks in Russia and Eastern Europe in 2010.

The Northern Mediterranean countries (Portugal, Spain, France, Italy and Greece) are particularly affected by summer fires (Vitolo et al., 2019). Among these, Portugal holds the highest number of wildfires per land area (Pereira et al., 2011). There are many environmental and anthropogenic factors influencing the rural fire regime in Portugal and making its territory a fire prone area. However, the majority of rural fires are recorded during hot and dry conditions in the summer (Pereira et al., 2011). Here, we used the mainland continental Portuguese Rural Fire Database, that includes 450,000 fires and covers the period 1980-2005 (Pereira et al., 2011), and includes data for all 18 districts in Portugal. This database is the largest such database in Europe in terms of total number of recorded fires in the 1980–2005 period (Pereira et al., 2011) and includes fires recorded down to a size of 0.001 ha. From the Portuguese Rural Fire database, we chose to focus on the Porto district, which was the worst affected in the period (out of the 18 Portugal districts) in term of number of wildfires with 21.6% of the total fire recorded in the dataset between 1980 and 2005. The Porto district is situated in the northern part of Portugal (see **Fig. 11**), has an area of [23952.395](#) km² and is one of the most populated districts of Portugal with an estimated population of 1,778,146 in 2018 (Instituto Nacional de Estatística Portugal, 2019).

Formatted: Font: 13 pt, Italic

Formatted: Font: 13 pt, Italic

Formatted Table

Formatted: Font: 13 pt, Italic

Formatted: Font color: Custom Color(RGB(54,95,145))

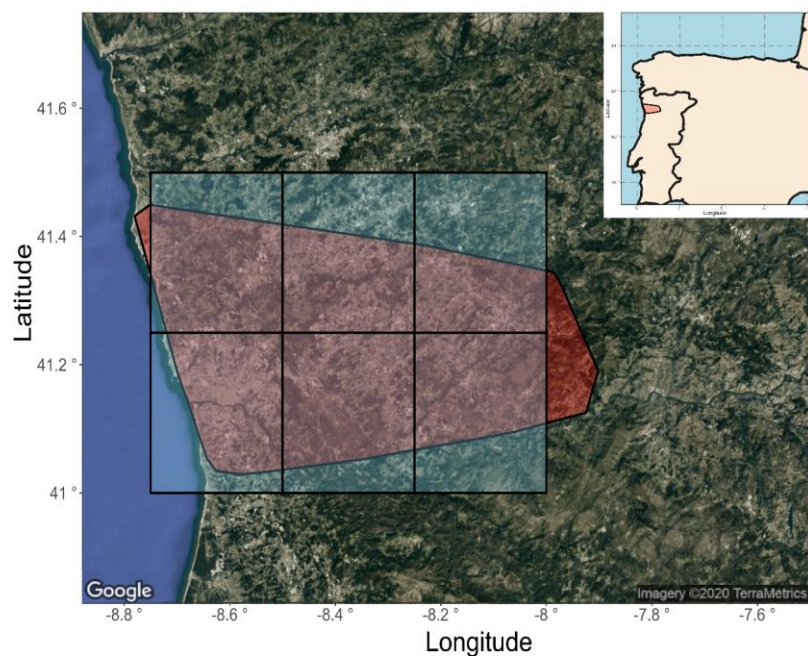
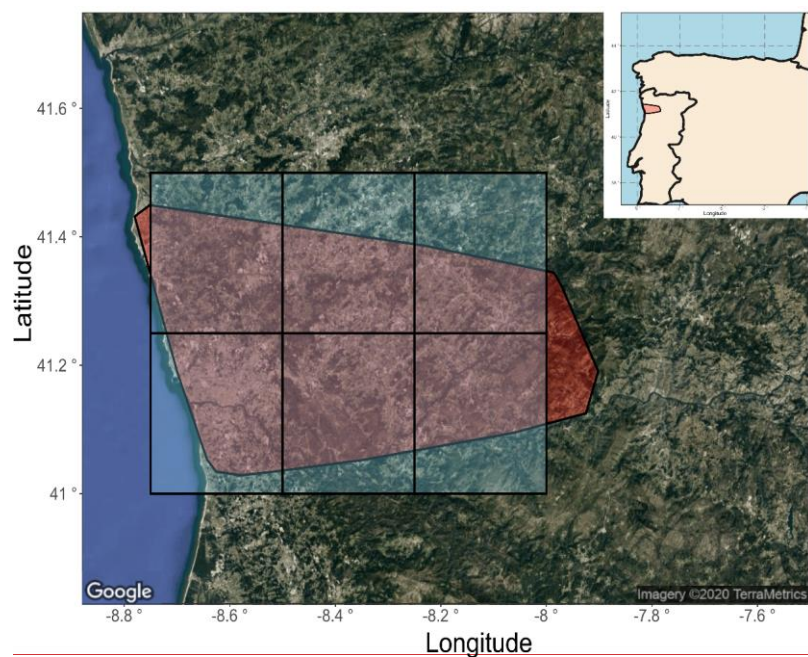


Figure 11: Portugal study area for the interrelation between extreme hot temperature and wildfire burned areas. The red area represents the Porto district in Portugal containing studied wildfire burned areas. The blue tiles represent cells from the high-resolution gridded data set of daily climates over Europe (E-OBS) (Cornes et al., 2018) containing mean daily temperature data.

Satellite image retrieved with ggmap (Kahle and Wickham, 2013).Satellite image retrieved with ggmap (Kahle and Wickham, 2013). © Google Maps (2020).

The bivariate dataset used to study the interrelation between extreme temperature and wildfire burned areas in the Porto district is composed of the following data:

- a) *Daily number of ~~wildfire~~wildfires* ( $f$ ). Daily number of wildfires for the 26-year period 1980–2005 for the Porto district were extracted from the Portuguese Rural Fire Database dataset from Pereira et al. (2011). To account for under sampling of smaller wildfires in earlier years, and as suggested by Pereira et al. (2011), we used only those fires with a burned area  $A_F \geq 0.1$  ha, resulting in 59,522 fires, an average of 6.3 fires per day (for those days with at least one fire occurrence) over the Porto district in Portugal (2395 km<sup>2</sup>).
- b) *Daily temperature data* ( $t$ ). Daily mean temperature was extracted from the E-OBS gridded dataset (Cornes et al., 2018). We approximate the area in red in **Fig. 11** (Porto district) for each day with one temperature value by taking the average of daily temperatures in each of the six 0.25° x 0.25° cells represented by blue rectangles in **Fig. 11**. This assumption reduces the confidence in return level values and adds up with other interpolation uncertainties arising from the data (Hofstra et al., 2009)(Hofstra et al., 2009). Moreover, the temperature in the six cells are strongly correlated (Pearson correlation coefficient  $\rho > 0.98$ ) and variations in temperature are mostly due to the distance to the sea and altitude (Miranda et al., 2002).

The 26 years from 1980–2005 have a total of 9496 days. Of these, a total of 3442 days (36% of the days) have both non-zero days for number of wildfires and a mean temperature value, which are used in our final bivariate dataset. An overview of both daily mean temperature and daily number of wildfires is displayed in **Fig 12** in the form of monthly violin plots.

Field Code Changed

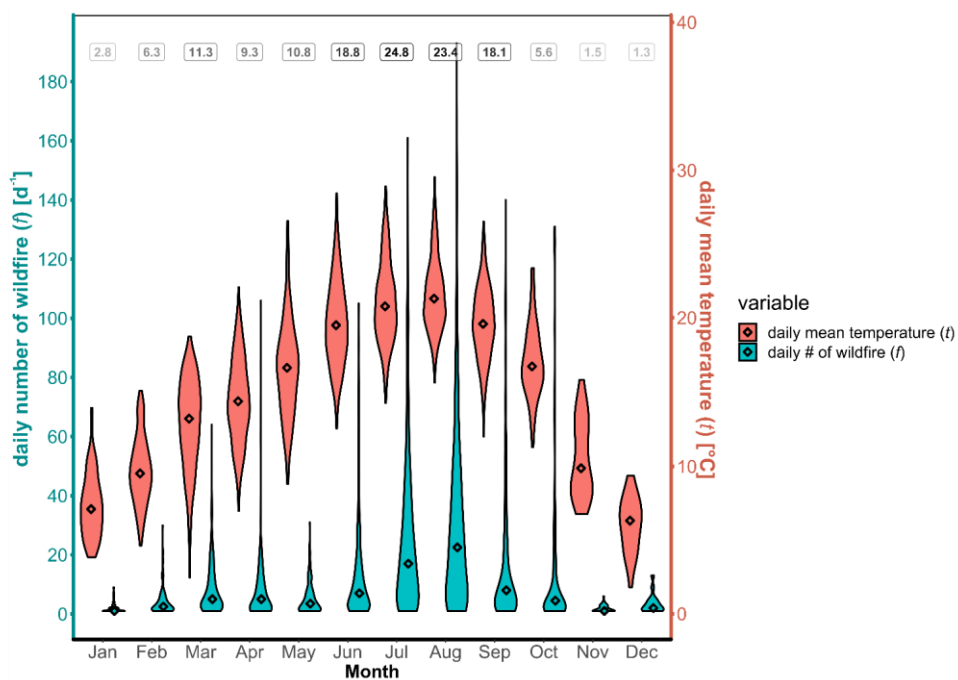
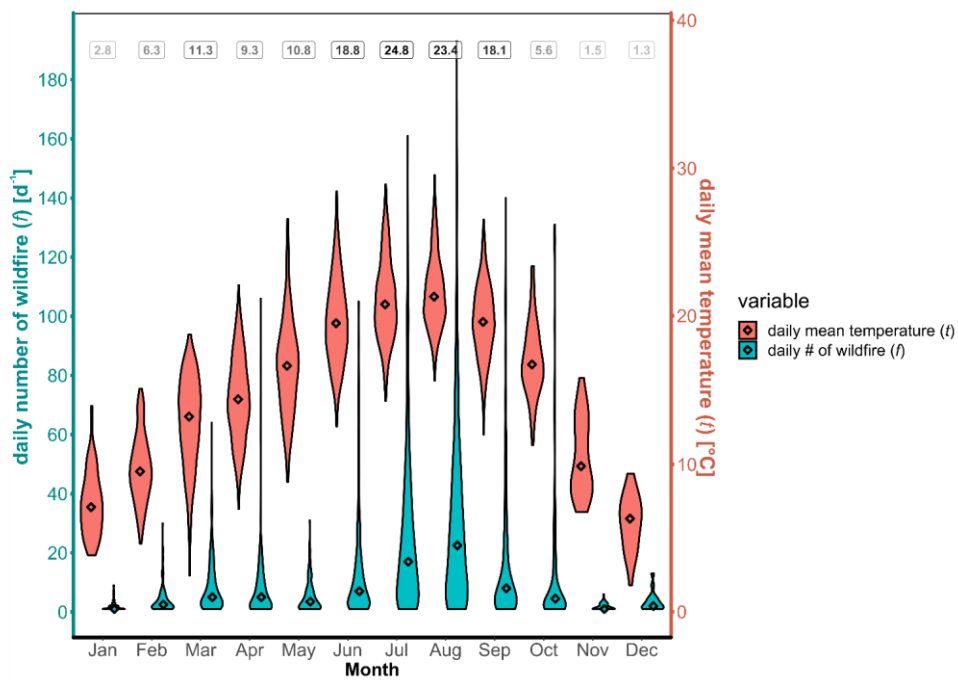
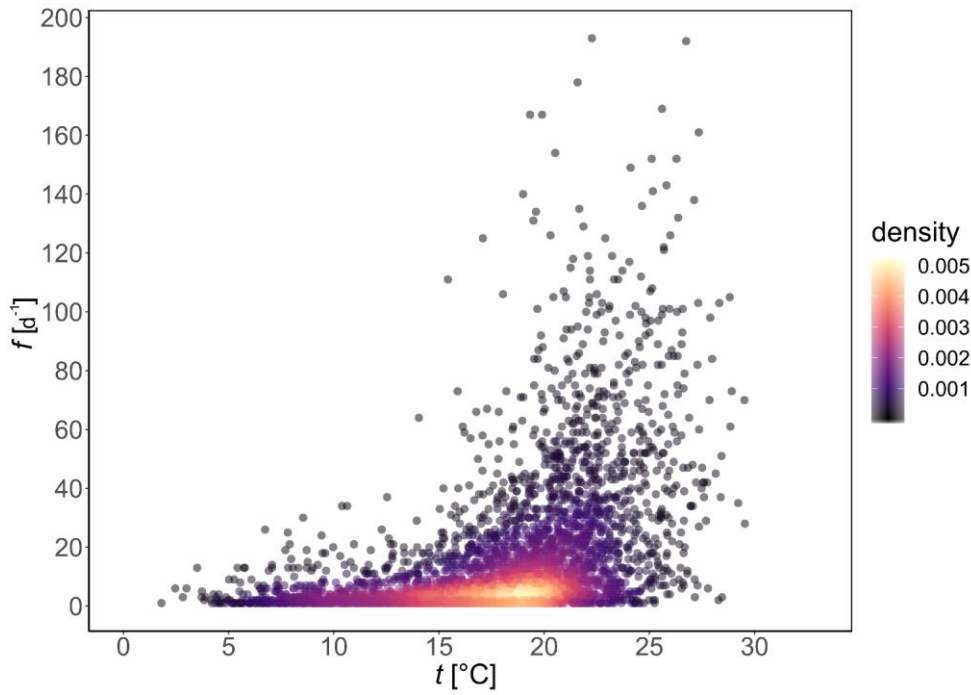


Figure 12: Violin plot of those days with both daily mean temperature (red, upper violin plots)  $t$  and daily number of wildfires (blue, lower violin plots)  $f \geq 1$  fire  $d^{-1}$ , by month for the period 1980–2005 in Porto district (Portugal). Only those wildfires with burned area  $A_F \geq 0.1$  ha are included. Diamonds for both temperature and wildfires represent the median of all values in that month over the period of record. Numbers at the top of the graph represent the average number of days per month where there are recorded both a temperature value  $t$  and at least one wildfire ( $f \geq 1$  fire  $d^{-1}$ ). Daily mean temperature data from E-OBS (Cornes et al., 2018) and wildfire data from Pereira et al. (2011).

From Fig. 12 we observe the seasonality in daily mean temperature with January the coldest month (median = 8.3°C) and August the warmest (median = 21.0°C). Daily number of wildfires (with burned area  $A_F \geq 0.1$  ha) per month varies between median of 1.0–2.5 fire  $d^{-1}$  in winter months (November to February) and 7.0–22.5 fire  $d^{-1}$  in summer months (from June to September). The dataset is also represented as a scatterplot in Fig. 13. The scatterplot will be used for the model selection methodology presented at the beginning of Section 4.



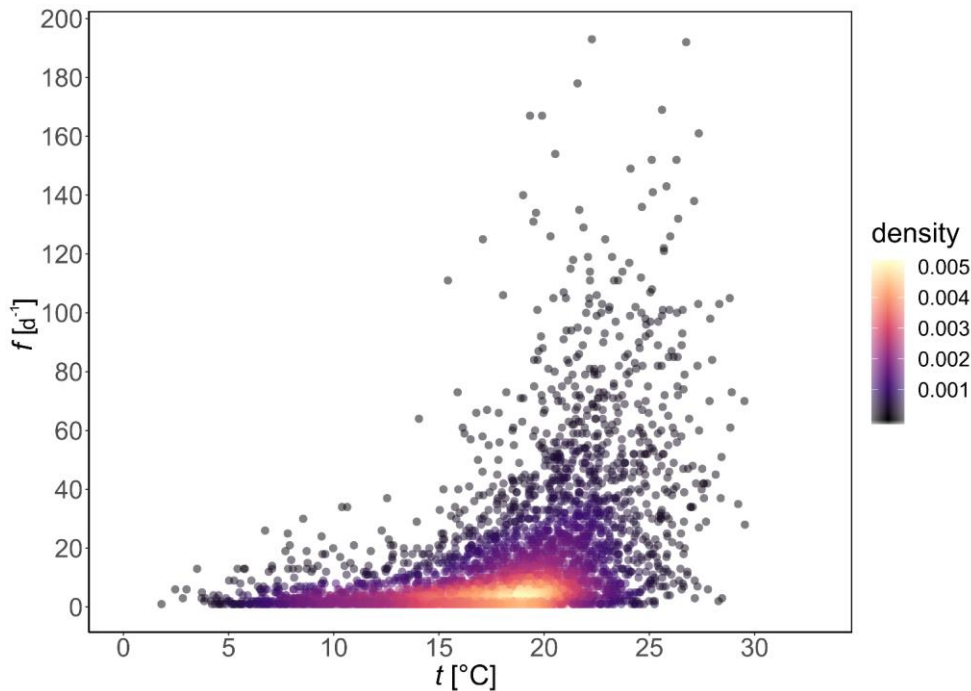


Figure 13: Scatter plot of temperature as a dependence of wildfire occurrence in Porto district, Portugal for the period 1980–2005, for those days where there are recorded both mean daily temperature ( $t$ ) and at least one fire, with  $f$  the number of wildfires in one day. Only those wildfires with burned area  $A_F \geq 0.1$  ha are included. Daily mean temperature data from E-OBS (Cornes et al., 2018) and wildfire data from Pereira et al. (2011). Colours represent the bivariate density estimated from a kernel density estimator.

As discussed in the beginning of this section, extreme (hot) temperature and wildfire are interrelated. Indeed, extreme (hot) temperature may promote the development of wildfires (Witte et al., 2011; Sutanto et al., 2020). According to Tilloy et al. (2019), this is a change condition interrelation (i.e., one hazard changes environmental parameter that moves toward a change in the likelihood of another hazard). We then estimate the conditional exceedance probability curve (Sect. 2.3).

We now go through the four steps introduced at the beginning of Sect. 4.

- (i) From Figs. 135 and 613, along with empirical estimates of  $\chi$  and  $\eta$ , we can hypothesize that over our time range, there is asymptotic dependence for the mean daily temperature and the number of wildfire per day are either asymptotically dependent ( $\chi = 0.5$ ) or asymptotically independent but with strong positive association ( $\eta - \chi = 0.93$ ) and that one marginal distributions have a slightly small shape parameter and the other one is heavily right-skewed (AC, BC).
- (ii) This then gives us four analogous datasets and it is then possible to know from Fig. 8 which models are the most adapted to these conditions. The four datasets are the following:
  1.  $\chi = 0.5$  and AC
  2.  $\chi = 0.5$  and BC
  3.  $\eta - \chi = 0.93$  and AC
  4.  $\eta - \chi = 0.93$  and B-CBC

Field Code Changed

Formatted: Font: Not Bold

Formatted: Numbered + Level: 1 + Numbering Style: 1, 2, 3, ... + Start at: 1 + Alignment: Left + Aligned at: 1.9 cm + Indent at: 2.54 cm



(iii) The confidence score for each model is the average of the  $wd$  from the four aforementioned datasets, and are presented in Table 6. Again, we have two asymptotically dependent and two asymptotically independent datasets. Based on the Table 6, the normal copula and FGM copula do not seem suitable to model the joint occurrence of wildfire and extreme temperature as these poorly fit the four datasets. The Gumbel and Galambos copula have a confidence score of respectively ( $\overline{wd} = 0.0502$ ) and the conditional extremes model ( $\overline{wd} = 0.06$  even though these 04) are asymptotically dependent representative for the four analogous datasets. The joint tail-KDE model also has a confidence score of 0.05 while the conditional extremes model has a score of 0.04.  $\overline{wd} = 0.05$  and is representative for two analogous datasets.

According to these three first steps, we can identify the most relevant model for the bivariate dataset of daily maximum temperature and daily wildfire occurrence in Porto district: the Gumbel copula, Galambos copula, the JT-KDE model and the conditional extremes model are the most relevant models for our dataset.

**Table 6: Weighted Euclidian distance ( $wd$ ) for datasets 1 to 4 based on extreme temperature-wildfire and six models, along with confidence scores (average of the  $wd$  for dataets 1 to 4). In bold are highlighted the four models with confidence scores < 0.1.**

Dataset	Cond-Ex	JT-KDE	Gumbelcop	Normalcop	FGMcop	Galamboscop
1	<b>0.03</b>	<b>0.04</b>	<b>0.02</b>	0.12	0.29	<b>0.02</b>
2	<b>0.05</b>	<b>0.06</b>	<b>0.04</b>	0.18	0.42	<b>0.04</b>
3	<b>0.03</b>	0.0305	<b>0.0702</b>	0.0213	0.262	<b>0.0702</b>
4	<b>0.0506</b>	0.0507	<b>0.0704</b>	0.1419	0.3728	<b>0.1004</b>
Average	<b>0.04</b>	<b>0.05</b>	<b>0.0503</b>	0.1215	0.3430	<b>0.0603</b>

(iv) For illustration and/or confronting of the models with the data, the six models are fit to the dataset and the joint exceedance level curves are produced with a joint exceedance probability set at  $p = 0.001$ , corresponding to a bivariate return period of approximately 8 years.

In Fig. 14 are displayed the conditional level curves produced from the four models that were selected after steps (i) to (iii) and shown in bold values in Table 6 (Cond-Ex, JT-KDE, Gumbelcop and GalambosCop).

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))

Formatted: Font: Bold, Font color: Custom Color(0,51,204))



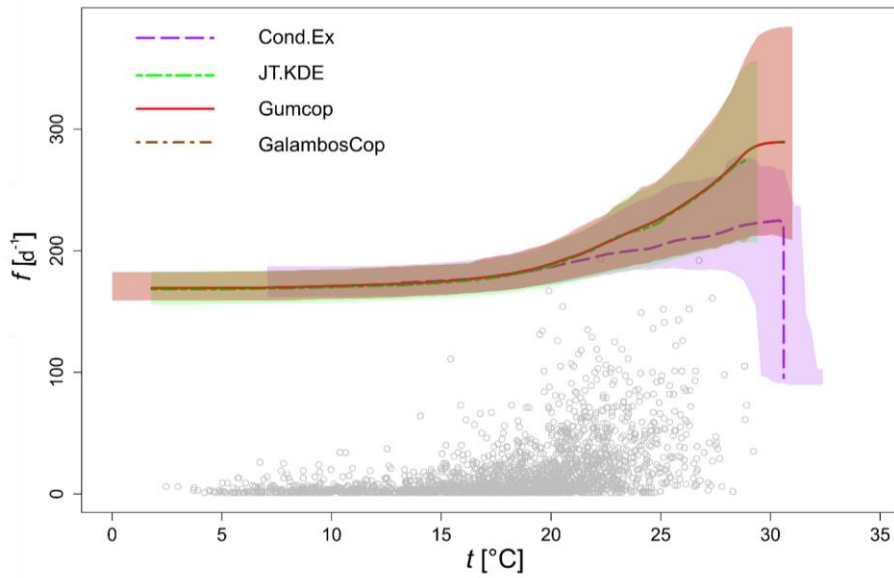
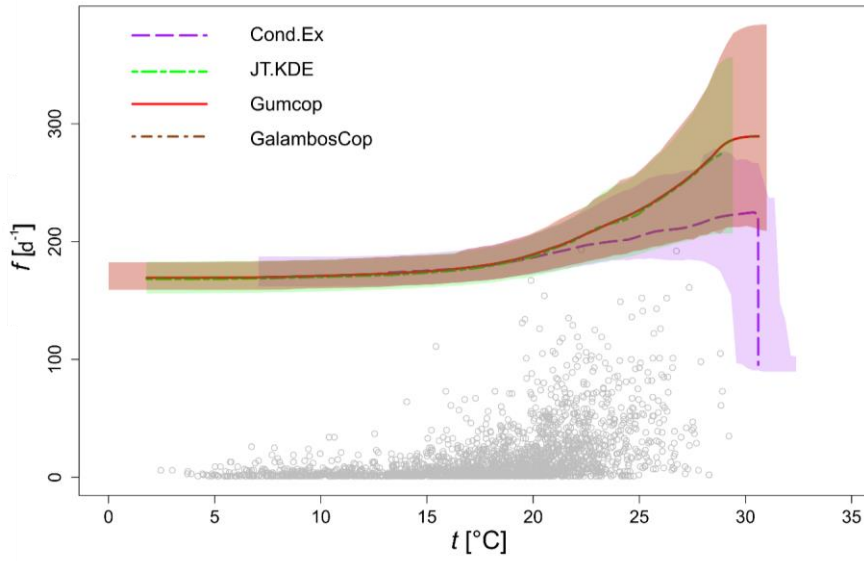


Figure 14: Level curves for a  $P_{\text{and}}$  joint probability  $p=0.001$  of daily mean temperature and daily number wildfire occurrences in Porto district, Portugal, for the period 1980-2005. Level curves from the four models selected through the model selection methodology are displayed.

From Fig. 14, we can observe that the JT-KDE and the Gumbel copula produce very similar conditional exceedance curves and that their confidence intervals **highlystrongly** overlap. However, the conditional extreme model provides a lower estimate than the other approaches the number of wildfire conditioning on the temperature being above a given threshold.

In **Table 7**, we present the estimates (with bounds of the 95% confidence interval) of the two dependence parameters  $\chi$  and  $\eta$  from the six models provide a bit more insight about the dependence structure. These estimates converge toward a moderate asymptotic dependence varying from  $\chi = 0.15$  (Cond-Ex) to  $\chi = 0.47$  (GumCop). Even if all models tend to show asymptotic dependence between the two variables, estimates of  $\eta$  are less than 1.0 for the normal copula, the JT-KDE model and the Cond-Ex with values varying between 0.67 and 0.79. This still implies a positive association between the two variables.

**Table 7: Estimates of dependence parameters  $\chi$  and  $\eta$  for mean daily temperature and daily occurrences of wildfire in Porto district for the period 1980-2005**

Models	Cond-Ex	JT-KDE	Gumcop	Normalcop	FGMcop	GalambosCop
$\chi$	0.15 [0.06, 0.20]	0.26 [0.21, 0.30]	0.47 [0.45, 0.49]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.46 [0.44, 0.49]
$\eta$	0.67 [0.59, 0.72]	0.67 [0.62, 0.71]	1.00 [1.00, 1.00]	0.79 [0.78, 0.80]	0.50 [0.50, 0.50]	1.00 [1.00, 1.00]

Formatted: Font: 13 pt, Italic

Formatted: Font: 13 pt, Italic

Formatted: Font: 13 pt, Italic

## 5 Discussion and Conclusions

Quantifying and measuring the interrelations between different natural hazards is a key element when adopting a multi-hazard approach (Gill and Malamud, 2014; Leonard et al., 2014)(Gill and Malamud, 2014; Leonard et al., 2014). In this study, we focused on statistical approaches that are often used to characterize and model interrelations between hazards. Another focus has been on modelling relationships between hazards at an extreme level. In total six statistical models with different characteristics (nature of asymptotic dependence, parametric/semi-parametric) were compared. Some of these models have already been used to study compound extremes in hydrology and climatology (Hao et al., 2018; Liu et al., 2018; Sadegh et al., 2018; Cooley et al., 2019). However, these have not been compared over a broad range of bivariate datasets and applied to the same natural hazards in the same location.

This section will discuss the ~~four~~ following three themes before a short conclusion: (a) choices influencing the results of the simulation study; (b) uncertainties at the interface between asymptotic dependence and asymptotic independence; (c) possible extensions of this approach to more than two hazards.

*Choices influencing the results of the simulation study.* This study aimed to assess the fitting ability of several bivariate models to a broad range of datasets. In order to do so, models were compared on their ability to reproduce an extreme level curve (see **Sect. 3.2.1**). The level curve corresponding to the  $P_{AND}$  probability has been selected as a comparison point because it is commonly used in the literature and is relevant for practitioners. The choice of this level curve, and its shape could influence our results. The extreme level curve probability was set at  $p = 0.001$ . The multivariate regular variation framework (Resnick, 1987)(Resnick, 1987) provides evidence supporting the fact that the dependence structure remains identical in the whole extreme domain. However, some results shown in **Sect. 3.3** might have been influenced by the value of the joint exceedance probability. In particular, it is likely that when decreasing the level curve probability (i.e., to more extreme values), the flexibility and abilities of the asymptotically independent normal copula will decrease. There are many copulas other than the four selected in this study (Nelsen, 2006; Sadegh et al., 2017)(Nelsen, 2006; Sadegh et al., 2017) that have been developed. Nevertheless, we believe the four copulas used in this study are suitable for bivariate extreme value analysis and are amongst the most widely used in the literature (Genest and Favre, 2007; Genest and Nešlehová, 2013)(Genest and Favre, 2007; Genest and Nešlehová, 2013). Another influential choice in this study has been the number of synthetic data points generated in each realization of the dataset. The number of data points and data set size is an important influence on uncertainty in natural hazard modelling and probabilistic approaches (Frau et al., 2017; Liu et al., 2018). For each simulation, we simulated  $n = 5000$  data points. Some other simulation studies took a higher number of data points (Zheng et al., 2014; Cooley et al., 2019)(Zheng et al., 2014; Cooley et al., 2019); however, ~~by~~ we replicated 100 times and produced confidence intervals, thus ensuring

Field Code Changed

consistency of our results. We also found that threshold selections, to fit the generalised Pareto distributions of the marginal distributions and to estimate the extremal dependence measures, also have an influence on our results.

*Uncertainties at the interface between asymptotic dependence and asymptotic independence.* From the results of the simulation study (Sect. 3.3) and the two case study applications (Sect. 4), one can observe that the interface between asymptotic dependence and asymptotic independence can be unclear. In Sect. 3.3, we discussed the decrease in model performance and the increase in uncertainty for low values of  $\chi$  and high values of  $\eta$ . Taking the assumption of asymptotic independence or asymptotic dependence can have a significant impact on the estimation of joint return levels. We find that extra care is required when dealing with bivariate datasets which are near independence as in Sect. 4.1.

*Possible extension of the approaches to more than two hazards.* As presented through this paper, the study of interrelations between natural hazards has primarily been done by hazard pairs (e.g., Gill and Malamud, 2014)(e.g., Gill and Malamud, 2014). Dependence measures and a variety of different models or level curves, all presented in this article, are powerful tools to assess, quantify and model interrelations between two hazards. However, in many cases, multi-hazard events include more than two hazards interacting in various ways (e.g., Gill and Malamud, 2014; Leonard et al., 2014)(e.g., Gill and Malamud, 2014; Leonard et al., 2014). The use of models presented in this article can be extended to more than two variables, sometimes with disadvantages. One of these disadvantages is ~~the~~ that the parametric nature of copulas ~~lead~~leads to a lack of flexibility ~~while~~when going to higher dimensionality (Bevacqua et al., 2017; Hao et al., 2018)(Bevacqua et al., 2017; Hao et al., 2018). The JT-KDE and Cond-Ex models are suitable for higher dimensions (Davison and Huser, 2015; Cooley et al., 2019)(Davison and Huser, 2015; Cooley et al., 2019), although, these have not been tested for high dimensional multi-hazard modelling yet (Tilloy et al., 2019)(Tilloy et al., 2019). Recent research conducted suggest pair-copula construction (PCC, also known as vine-copula) (Bedford and Cooke, 2002; Hashemi et al., 2016; Bevacqua et al., 2017, Lui et al., 2018) and non-parametric Bayesian networks (NPBN) (Hanea et al., 2015; Couasnon et al., 2018) can be used to model multi-hazard events with more than two hazards.(Hanea et al., 2015; Couasnon et al., 2018) can be used to model multi-hazard events with more than two hazards. The vine copula framework allows one to select different bivariate copulas for each ~~pairs~~pair of variables, providing a great flexibility in dependence modelling (Brechmann and Schepsmeier, 2013; Hao and Singh, 2016). Non-parametric Bayesian networks, which are associated with the structure of Bayesian network and copulas (Hanea, 2010; Hanea et al., 2010, 2015), have been used to study multiple dependences between river discharge and storm surges in the USA during a hurricane (Couasnon et al., 2018).

In conclusion, we have compared and examined the strength and weaknesses of six distinct bivariate extreme models in the context of hazard interrelations. These six models are grounded in multivariate extreme value theory and represent the diversity of approaches (e.g., non-parametric vs parametric) currently applied to hazard interrelation analysis. With this study we aimed to contribute to a better understanding on the applicability of bivariate extreme models to a wide range of natural hazard interrelations. The methodology developed in this article is aimed to be widely applicable to a variety of different hazards and different interrelations, here represented by the 60 synthetic datasets created. Abilities of each model have been assessed with two metrics: (i) dependence measure; (ii) bivariate return level (level curves). These two metrics and the different diagnostic tools developed in this study offer new intuitive ways to decipher the dependence between two variables. We recommend selecting a range of models rather than one when studying interrelations between two hazards. To highlight the benefits of the systematic framework developed, we studied the dependence between extremes (natural hazards) of the following environmental data: (i) daily precipitation accumulation and daily maximum wind gust (maximum over a period of 3 s) at Heathrow airport (UK) over the period 1971-2018; (ii) daily mean temperature and daily number of wildfires in Porto district, Portugal over the period 1980-2005. The two datasets represent different hazard interrelations: (i) compound interrelation between extreme wind and extreme rainfall and (ii) change condition interrelation where higher air temperature change condition for wildfire occurrence. In both cases, a sample of the most relevant model among the six used in this study have

been selected and fitted to the bivariate datasets. The good agreement in the estimation of bivariate return period between models corroborate the relevance of the comparison metrics we developed.

Appendix A: Comparing model abilities through tail dependence measures

A.1. Tail dependence measures estimations

Tail dependence measures  $\eta$  and  $\chi$  are estimated by each model. For copulas, these measures are related to the copula parameters. In our set of four copulas, two are asymptotically dependent (Gumbel and Galambos) with  $\eta=1$  and two are asymptotically independent (normal and FGM) with  $\chi=0$ .

For the nonparametric joint tail approach, the  $\chi$  and  $\eta$  measures are estimated following the procedure used by Winter (2016).

For the conditional model, both measures are estimated from the simulated points. Marginal distributions  $(X_1, X_2)$  are transformed to the uniform margins  $(U_1, U_2)$ . The  $\chi$  measure is estimated by calculating the probability  $P(V > u|U > u)$  (Eq. 4). The  $\eta$  measure is estimated in two steps. First we estimate  $\bar{\chi}(u)$  as (Coles et al., 1999)(Coles et al., 1999):

$$\bar{\chi}(u) = \frac{2\log P(U>u)}{\log P(U>u,V>u)} - 1 = \frac{2\log(1-u)}{\log(\chi(u)(1-u))} - 1, \tag{A1}$$

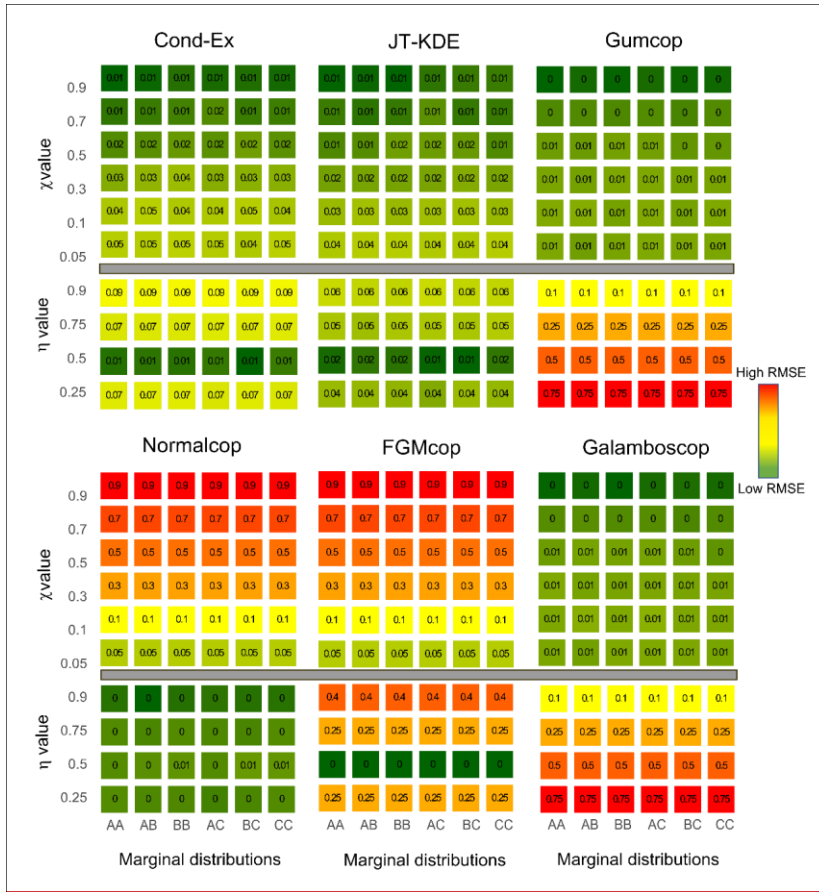
for  $0 \leq u \leq 1$  with  $u$  a sufficiently high threshold. Second, the  $\eta$  measure is estimated from  $\bar{\chi}$  (Eq. A1).

To compare the estimated dependence measure to the reference value, the root-mean-square error (RMSE), a measure of efficiency that accounts for both the bias and variance of the estimates is used, similarly to Zheng et al. (2014). Similarly to the metrics used in Sect. 3, the RMSE is calculated from 100 realizations of the 60 datasets.

A.2. Comparison of model abilities

The estimation of dependence measure is an important step in bivariate analysis (Coles et al., 1999; Heffernan, 2000; Zheng et al., 2013, 2014; Dutfoy et al., 2014). Models have also been compared on their ability to estimate the dependence measures  $\chi$  and  $\eta$ . Results arising from this comparison provide a different perspective on the abilities of each models. Figure A1 shows the RMSE of the dependence measures estimations for each of the 60 synthetic datasets.

Formatted Table



## A.2. Comparison of model abilities

The estimation of dependence measure is an important step in bivariate analysis (Coles et al., 1999; Heffernan, 2000; Zheng et al., 2013, 2014; Dutfoy et al., 2014). Models have also been compared on their ability to estimate the dependence measures  $\chi$  and  $\eta$ . Results arising from this comparison provide a different perspective on the abilities of each models. **Figure A1** shows the RMSE of the dependence measures estimations for each of the 60 synthetic datasets.

Field Code Changed

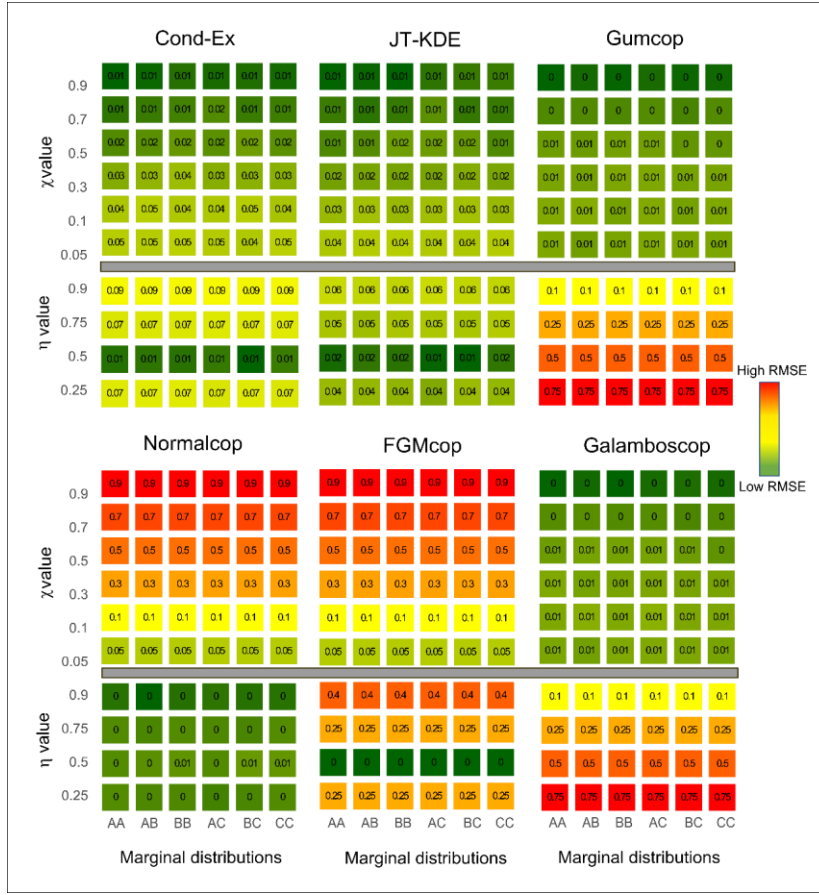


Figure A1: RMSE (root-mean-square error) of the estimated dependence measures to the reference for all 60 different datasets. Fitting capacities of each model are represented. Values in cells and colours represent the median RMSE from low (dark green) to high (red). Thickness of cell borders represent the 95% uncertainty around the median value

From Fig. A1, we observe the following:

- Marginal distributions do not have a significant impact on the accuracy of the estimation of these measures for the copulas.
- Marginal distributions have a small impact on estimation of the dependence measures for the conditional extremes model and the joint-tail model, however this impact is not as important as for the level curve estimation
- All copulas estimate very accurately the dependence measure within their operating range (AI for normal copula, near independence for FGM copula and AD for Gumbel and Galambos copula). However, only the conditional extremes model and the joint-tail model can estimate both  $\eta$  and  $\chi$ .
- The dependence measure estimator used in the joint-tail KDE approach offers slightly more accurate estimation for, in particular for  $\eta$ .

- Estimation performance of both joint tail KDE and condition extreme models decrease when approaching the interface between asymptotic dependence and asymptotic independence. The RMSE at  $\chi=0.05$  is close the 100% of the value of  $\chi$  while the RMSE  $\eta=0.9$  is at its highest for both Cond-EX and JT-KDE models. It is then hard to decipher with confidence the nature of the dependence in the asymptotic domain for low  $\chi$  values and high  $\eta$  values.



980 **Code availability**

The codes used to produce synthetic data and the 6 models used in this study are publicly available on GitLab: [https://gitlab.com/doudeg/bivariate\\_models.git](https://gitlab.com/doudeg/bivariate_models.git)

**Data availability**

985 We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (<http://www.uerra.eu>) and the data providers in the ECA&D project (<https://www.ecad.eu>). The wind gust data was obtained from the UK Met Office. The fire data is available from Pereira et al. (2011). The bivariate (wildfire, temperature) data used is available at: [https://gitlab.com/doudeg/bivariate\\_models.git](https://gitlab.com/doudeg/bivariate_models.git).

**Author contribution**

990 All authors discussed the whole plan of this article, Tilloy designed and implemented all the experiments, prepared all the data and finish the draft, including all figures in the article. Malamud, Winter and Joly-Laugel revised the article. Winter supported the methods and techniques.

**Competing interests** Malamud is on the editorial board of this journal (the article has been overseen independently by another editor of this journal).

995 **Special issue statement.** This research article is part of the special issue “Advances in extreme value analysis and application to natural hazards”.

**Acknowledgment**

Funding: The first author was supported by an EDF R&D PhD studentship.

1000 **References**

- AghaKouchak, A., Huning, L. S., Chiang, F., Sadegh, M., Vahedifard, F., Mazdiyasni, O., Moftakhari, H. and Mallakpour, I.: How do natural hazards cascade to cause disasters?, *Nature*, 561(7724), 458–460, 2018.
- Aitchison, J.: *Lognormal Distribution*, Cambridge Univ. Press., 1957.
- Akaike, H.: A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, 19(6), 716–723, 1974.
- 1005 Anderson, G. and Klugmann, D.: A European lightning density analysis using 5 years of ATDnet data, *Nat. Hazards Earth Syst. Sci.*, 14(4), 815–829, 2014.
- Arnold, T. B. and Emerson, J. W.: Nonparametric goodness-of-fit tests for discrete null distributions, *The R Journal*, 3(2), 34–39, 2011.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. and García-Herrera, R.: The hot summer of 2010: Redrawing the temperature map of Europe, *Science*, 332, 220–224, 2011.
- 1010 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(6), 2701–2723, 2017.
- Canty, A. and Ripley, B. D.: *boot: Bootstrap R (S-Plus) Functions*, 2019.
- CCR. E-risk database (Paris, France) [Online] Available at: <https://catastrophes-naturelles.ccr.fr> [Accessed 1 December 2019],

1015 2019.

Chebana, F. and Ouarda, T. B. M. J.: Multivariate quantiles in hydrological frequency analysis, *Environmetrics*, 22(1), 63–78, 2011.

Clare, M. A., Talling, P. J., Challenor, P. G. and Hunt, J. E.: Tempo and Triggering of Large Submarine Landslides: Statistical Analysis for Hazard Assessment, In: *Advances in Natural and Technological Hazards Research*, vol. 41, edited by G. Lamarche, J. Mountjoy, S. Bull, T. Hubble, S. Krastel, E. Lane, A. Micallef, L. Moscardelli, C. Mueller, I. Pecher, and S. Woelz, pp. 509–517, Springer International Publishing, Cham., 2016.

1020 Coles, S. and Tawn, J. A.: Modelling extreme multivariate events, *J. R. Stat. Soc.*, 53(2), 285–339, 1991.

Coles, S., Heffernan, J. E. and Tawn, J. A.: Dependence measure for extreme value analyses, *Extremes*, 1225(June), 41–42, 1999.

1025 Coles, S.: *An Introduction to Statistical Modelling of Extreme Values*, Springer London, London., 2001.

Cooley, D., Thibaud, E., Castillo, F. and Wehner, M. F.: A nonparametric method for producing isolines of bivariate exceedance probabilities, *Extremes*, 22, 373–390, 2019.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, *J. Geophys. Res. Atmos.*, 123(17), 9391–9409, 2018.

1030 Couasnon, A., Sebastian, A. and Morales-Nápoles, O.: A copula-based Bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: an application to the Houston ship channel, Texas, *Water*, 10(9), 1190, <https://doi.org/10.3390/w10091190>, 2018.

Davison, A. C. and Hinkley, D. V.: *Bootstrap Methods and Their Applications*, Cambridge University Press, Cambridge., 1997.

Davison, A. C. and Huser, R.: Statistics of extremes, *Annu. Rev. Stat. Its Appl.*, 2(1), 203–235, 2015.

1035 Dowdy, A. J. and Catto, J. L.: Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences, *Sci. Rep.*, 7, 1–8, 2017.

Duong, T.: ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R, *J. Stat. Softw.*, 21(7), [dx.doi.org/10.18637/jss.v021.i07](https://doi.org/10.18637/jss.v021.i07), 2007.

Duong, T.: Non-parametric smoothed estimation of multivariate cumulative distribution and survival functions, and receiver operating characteristic curves, *J. Korean Stat. Soc.*, 45(1), 33–50, 2015.

1040 Dutfey, A., Parey, S. and Roche, N.: Multivariate extreme value theory - A tutorial with applications to hydrology and meteorology, *Depend. Model.*, 2(1), 30–48, 2014.

Genest, C. and Favre, A.-C.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12(4), 347–368, 2007.

1045 Genest, C., Rémillard, B. and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, 44(2), 199–213, 2009.

Genest, C., Kojadinovic, I., Nešlehová, J. and Yan, J.: A goodness-of-fit test for bivariate extreme-value copulas, *Bernoulli*, 17(1), 253–275, 2011.

Genest, C. and Nešlehová, J.: Copula modeling for extremes, *Encycl. Environmetrics*, 1–16, 2013.

1050 Gill, J. C. and Malamud, B. D.: Reviewing and visualizing the interactions of natural hazards, *Rev. Geophys.*, 52(4), 680–722, 2014.

Gouldby, B., Wyncoll, D., Panzeri, M., Franklin, M., Hunt, T., Hames, D., Tozer, N., Hawkes, P., Dornbusch, U. and Pullen, T.: Multivariate extreme value modelling of sea conditions around the coast of England, *Proc. Inst. Civ. Eng. Marit. Eng.*, 170(1), 3–20, 2017.

1055 Gumbel, E. J.: *Statistics of extremes*, Columbia U., New York., 1958.

De Haan, L. and Resnick, S.: Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of*

- Probability, pp.97-124, 1996.
- De Haan, L. and De Ronde, J.: Sea and wind: multivariate extremes at work, *Extremes*, 1(1), 7–45, 1998.
- Hanea, A., Morales Napoles, O. and Ababei, D.: Non-parametric Bayesian networks: Improving theory and reviewing applications, *Reliab. Eng. Syst. Saf.*, 144, 265–284, 2015.
- Hao, Z. and Singh, V. P.: Review of dependence modeling in hydrology and water resources, *Prog. Phys. Geogr.*, 40(4), 549–578, 2016.
- Hao, Z., Hao, F., Singh, V. P. and Ouyang, W.: Quantitative risk assessment of the effects of drought on extreme temperature in eastern China, *J. Geophys. Res. Atmos.*, 122(17), 9050–9059, 2017.
- Hao, Z., Singh, V. and Hao, F.: Compound extremes in hydroclimatology: A review, *Water*, 10(6), 718, <https://doi.org/10.3390/w10060718>, 2018.
- [Hauser, M., Orth, R., Hatvani-Kovacs, G., Belusko, M., Pockett, J. and Boland, J. Assessment of heatwave impacts. \*Procedia Engineering\*, 169, pp.316-323, 2016](#)[Hauser, M., Orth, R. and Seneviratne, S. I.: Role of soil moisture vs . recent climate change for heat waves in western Russia, \*Geophys. Res. Lett.\*, 43, 2819–2826, 2015.](#)
- Hawkes, P. J.: Joint probability analysis for estimation of extremes, *J. Hydraul. Res.*, 46(sup2), 246–256, 2008.
- Hawkes, P. J., Gouldby, B. P., Tawn, J. A. and Owen, M. W.: The joint probability of waves and water levels in coastal engineering design, *J. Hydraul. Res.*, 40(3), 241–251, 2002.
- Heffernan, J. E.: A directory of coefficients of tail dependence, *Extremes*, 3, 279–290, 2000.
- Heffernan, J. E. and Tawn, J. A.: A conditional approach for multivariate extreme values (with discussion), *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, 66(3), 497–546, 2004.
- Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A. and Darby, S. E.: Assessing the characteristics and drivers of compound flooding events around the UK coast, *Hydrol. Earth Syst. Sci.*, 23(7), 3117–3139, 2019.
- Hilal, S., Poon, S. H. and Tawn, J.: Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX, *J. Bank. Financ.*, 35(9), 2374–2387, 2011.
- [Hincks, T.K., Malamud, B.D., Sparks, R.S.J., Wooster, M.J. and Lynham, T.J., 2013. Risk assessment and management of wildfires. \*Risk and uncertainty assessment for natural hazards\*, p.398-444.](#)
- Hofstra, N., Haylock, M., New, M. and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res. Atmos.*, 114(D21), <https://doi.org/10.1029/2009JD011799>, 2009.
- Johansson, B. and Chen, D.: The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modelling, *Int. J. Climatol.*, 23(12), 1523–1535, 2003.
- Kahle, D. and Wickham, H. ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144–161, 2013.
- Keef, C., Papastathopoulos, I. and Tawn, J. A.: Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model, *J. Multivar. Anal.*, 115, 396–404, 2013.
- Koopmans, L. H., Owen, D. B. and Rosenblatt, J. I.: Confidence intervals for the coefficient of variation for the normal and log normal distributions, *Biometrika*, 51(1–2), 25–32, 1964.
- Ledford, A. W. and Tawn, J. A.: Statistics for near independence in multivariate extreme values, *Biometrika*, 83(1), 169–187, 1996.
- Ledford, A. W. and Tawn, J. A.: Modelling dependence within joint tail regions, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 59(2), 475–499, 1997.
- Ledford, A. W. and Tawn, J. A.: Diagnostics for dependence within time series extremes, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 65(2), 521–543, 2003.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D. and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, *Wiley Interdiscip. Rev. Clim. Chang.*, 5(1), 113–128, 2014.

[Littell, J. S., McKenzie, D., Peterson, D. L., & Westerling, A. L. Climate and wildfire area burned in western US ecoprovinces, 1916–2003. \*Ecological Applications\*, 19\(4\), 1003-1021, 2009.](#)

Liu, Z., Cheng, L., Hao, Z., Li, J., Thorstensen, A. and Gao, H.: A framework for exploring joint effects of conditional factors on compound floods, *Water Resour. Res.*, 54(4), 2681–2696, 2018.

Loukatou, A., Howell, S., Johnson, P. and Duck, P.: Stochastic wind speed modelling for estimation of expected wind power output, *Appl. Energy*, 228, 1328–1340, 2018.

Malamud, B. D. and Turcotte, D. L.: Self-affine time series: I. Generation and analyses, in *Advances in Geophysics*, 40, 1–90, 1999.

Malamud, B. D. and Turcotte, D. L.: The applicability of power-law frequency statistics to floods, *J. Hydrol.*, 322(1–4), 168–180, 2006.

Martius, O., Pfahl, S. and Chevalier, C.: A global quantification of compound precipitation and wind extremes, *Geophys. Res. Lett.*, 43(14), 7709–7717, 2016.

Maulik, K. and Resnick, S.: Characterizations and examples of hidden regular variation, *Extremes*, 7(1), 31–67, 2005.

Mazas, F. and Hamm, L.: An event-based approach for extreme joint probabilities of waves and sea levels, *Coast. Eng.*, 122, 44–59, 2017.

De Michele, C., Salvadori, G., Passoni, G. and Vezzoli, R.: A multivariate model of sea storms using copulas, *Coast. Eng.*, 54(10), 734–751, 2007.

Ming, X., Xu, W., Li, Y., Du, J., Liu, B. and Shi, P.: Quantitative multi-hazard risk assessment with vulnerability surface and hazard joint return period, *Stoch. Environ. Res. Risk Assess.*, 29(1), 35–44, 2015.

Miranda, P. M. A., Coelho, F. E. S., Tomé, A. R., Valente, M. A., Carvalho, A., Pires, C., Pires, H. O., Pires, V. C. and Ramalho, C.: 20th Century Portuguese Climate and Climate Scenarios, In: *Climate Change in Portugal. Scenarios, Impacts and Adaptation Measures—SIAM Project*, Santos, F. D., Forbes, K., Moita, R. (eds), Gradiva Publishers, 28–83, 2002.

Nelsen, R.: *An Introduction to Copulas*, Springer New York, New York, NY., 2006.

Nguyen Sinh, H., Lombardo, F. T. and Letchford, C.: Multivariate simulation for assessing the joint wind and ice hazard in the United States, *J. Wind Eng. Ind. Aerodyn.*, 184, 436–444, 2019.

Pereira, M. G., Malamud, B. D., Trigo, R. M. and Alves, P. I.: The history and characteristics of the 1980-2005 Portuguese rural fire database, *Nat. Hazards Earth Syst. Sci.*, 11(12), 3343–3358, 2011.

Pickands, J.: Multivariate extreme value distribution, in *Proceedings 43rd Session of International Statistical Institute*, 859–878, 1981.

Raveh-Rubin, S. and Wernli, H.: Large-scale wind and precipitation extremes in the Mediterranean: dynamical aspects of five selected cyclone events, *Q. J. R. Meteorol. Soc.*, 142(701), 3097–3114, 2016.

Resnick, S.: Hidden regular variation, second order regular variation and asymptotic independence, *Extremes*, 303–336, 2002.

Resnick, S. I.: *Extreme Values, Regular Variation and Point Processes*, Springer New York, New York, NY., 1987.

Sachs, M. K., Yoder, M. R., Turcotte, D. L., Rundle, J. B. and Malamud, B. D.: Black swans, power laws, and dragon-kings: Earthquakes, volcanic eruptions, landslides, wildfires, floods, and SOC models, *Eur. Phys. J. Spec. Top.*, 205(1), 167–182, 2012.

Sadegh, M., Ragno, E. and AghaKouchak, A.: Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resour. Res.*, 53(6), 5166–5183, 2017.

Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasn, O., Sanders, B., Matthew, R. and AghaKouchak, A.:

Multihazard scenarios for analysis of compound extreme events, *Geophys. Res. Lett.*, 45(11), 5470–5480, 2018.

Salvadori, G.: Bivariate return periods via 2-Copulas, *Stat. Methodol.*, 1(1–2), 129–144, 2004.

Salvadori, G. and De Michele, C.: Multivariate multiparameter extreme value models and return periods: A copula approach, *Water Resour. Res.*, 46(10), <https://doi.org/10.1029/2009WR009040>, 2010.

1145 Salvadori, G., Durante, F., De Michele, C., Bernardi, M. and Petrella, L.: A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, *Water Resour. Res.*, 52(5), 3701–3721, 2016.

Schwarz, G.: Estimating the dimension of a model, *Ann. Stat.*, 6(2), 461–464, 1978.

Serinaldi, F.: Dismissing return periods!, *Stoch. Environ. Res. Risk Assess.*, 29(4), 1179–1189, 2015.

Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, *Stoch. Environ. Res. Risk Assess.*, 30(6), 1691–1704, 2016.

1150 Stephens, M. A.: Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 32(1), 115–122, 1970.

Sutanto, S. J., Vitolo, C., Di Napoli, C., D’Andrea, M. and Van Lanen, H. A. J.: Heatwaves, droughts, and fires: Exploring compound and cascading dry hazards at the pan-European scale, *Environ. Int.*, 134, 105276, 2020.

1155 Tencer, B., Weaver, A. and Zwiers, F.: Joint occurrence of daily temperature and precipitation extreme events over Canada, *J. Appl. Meteorol. Climatol.*, 53(9), 2148–2162, 2014.

Terzi, S., Torresan, S., Schneiderbauer, S., Critto, A., Zebisch, M. and Marcomini, A.: Multi-risk assessment in mountain regions: A review of modelling approaches for climate change adaptation, *J. Environ. Manage.*, 232(1), 759–771, 2019.

Tilloy, A., Malamud, B. D., Winter, H. and Joly-Laugel, A.: A review of quantification methodologies for multi-hazard interrelationships, *Earth-Science Rev.*, 196, <https://doi.org/10.1016/j.earscirev.2019.102881>, 2019.

1160 Vitolo, C., Di Napoli, C., Di Giuseppe, F., Cloke, H. L. and Pappenberger, F.: Mapping combined wildfire and heat stress hazards to improve evidence-based decision making, *Environ. Int.*, 127, 21–34, 2019.

Volpi, E. and Fiori, A.: Design event selection in bivariate hydrological frequency analysis, *Hydrol. Sci. J.*, 57(8), 1506–1515, 2012.

1165 Webb, J. D. C. and Elsom, D. M.: Severe Hailstorms in the United Kingdom and Ireland: A Climatological Survey with Recent and Historical Case Studies, In: *Extreme Weather: Forty Years of the Tornado and Storm Research Organisation (TORRO)*, Doe, R. K. (ed.), John Wiley Sons, ch. 9, pp.155-194, 2016.

Winter, H. C.: Extreme Value Modelling of Heatwaves, PhD thesis, Lancaster Univ., Available at: <https://eprints.lancs.ac.uk/id/eprint/79961/> (Accessed 28 December 2019), 2016.

1170 Winter, H. C. and Tawn, J. A.: Modelling heatwaves in central France: A case-study in extremal dependence, *J. R. Stat. Soc. Ser. C Appl. Stat.*, 65(3), 345–365, 2016.

Winter, H. C., Tawn, J. A. and Brown, S. J.: Modelling the effect of the El Niño-southern oscillation on extreme spatial temperature events over Australia, *Ann. Appl. Stat.*, 10(4), 2075–2101, 2016.

1175 Witte, J. C., Douglass, A. R., Da Silva, A., Torres, O., Levy, R. and Duncan, B. N.: NASA A-Train and Terra observations of the 2010 Russian wildfires, *Atmos. Chem. Phys.*, 11(17), 9287–9301, 2011.

Zheng, F., Westra, S. and Sisson, S. A.: Quantifying the dependence between extreme rainfall and storm surge in the coastal zone, *J. Hydrol.*, 505, 172–187, 2013.

Zheng, F., Westra, S., Leonard, M. and Sisson, S. A.: Modeling dependence between extreme rainfall and storm surge to estimate coastal flooding risk, *Water Resour. Res.*, 50(3), 2050–2071, 2014.

1180 Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, *Sci. Adv.*, 3(6), e1700263, 2017.

Zscheischler, J., Westra, S., Van Den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., Aghakouchak, A., Bresch,

D. N., Leonard, M., Wahl, T. and Zhang, X.: Future climate risk from compound events, *Nat. Clim. Chang.*, 8(6), 469–477, 2018.