Dear editor,

Thank you very much for your and the reviewers' assessment of our manuscript. I revised the manuscript according to the reviewers' comments with a particular focus on improving the discussion of different descriptors of multivariate extremes and on expanding the description of modeling approaches for multivariate extremes by including descriptions of (1) suitable univariate metrics for multivariate extremes, (2) bivariate distributions and return periods, (3) multivariate distributions, and (4) multivariate simulation approaches. I hope that you find the revised version of this manuscript suitable for publication in HESS. Thank you very much for your re-assessment.

Best regards,

Manuela Brunner

Reviewer 1

The manuscript is presented as a review paper on multivariate extremes, specifically flood, and drought. The multivariate aspect of such extremes is intended in space, in time, and in their characteristics. The topic is relevant for preparedness and risk management in the current and future climates. However, the manuscript in its current form presents some limitations.
Reply: *Thank you very much for acknowledging the relevance of the review topic and for taking the time to provide this constructive feedback, which I address point by point below.*

The introduction on the drawbacks of the univariate approach seems in contrast with the types of multivariate extremes identified. The regional and temporal extremes fall back on a univariate approach. Indeed, they are defined based on whether, e.g., flood magnitude is above a given threshold or with a given return period at one single location. When does an extreme in one location become a multivariate extreme? How many locations should be flooded? Is the regional extent of the univariate floods an indicator of whether an extreme is multivariate or not? How so? These kinds of questions are difficult to answer from the definitions of multivariate extremes provided and it makes questioning whether it is necessary to move away from the univariate approach.
Reply: *Thank you for stressing the need to clarify the link between studying multivariate extremes and univariate frequency analyses. I agree that one good strategy of studying multivariate extremes is by defining univariate metrics that describe them, e.g. spatial flood extent for spatially compounding flood events. The point I would like to make here is that analyses of hydrological extreme events should go beyond focusing on one variable only and consider extreme events from a multivariate perspective. I rewrote the introduction by removing the part about univariate frequency analysis which gave the wrong impression that this tool is inappropriate to study multivariate extremes. Instead, the new introduction stresses that multivariate extremes consider more than one variable compared to univariate extremes focusing on one variable only:*

'In July 2021, a severe and widespread flood event affected Western Germany and parts of Belgium and the Netherlands where it led to numerous fatalities and considerable damage to infrastructure (Ibebuchi et al. 2022). After such exceptional flood events, we ask: 'how frequently do such events occur?' To answer this question, one can rely on frequency

analyses which establish a link between the magnitude and frequency of events. Such analyses are often performed by focusing on one variable only, i.e. by taking a univariate perspective. In the case of the Germany flood, this would e.g. be flood peaks in one individual catchment. While such a focus on one variable enables the development of suitable preparedness and adaptation measures by providing magnitude and frequency estimates of extreme events, they have a major drawback: they neglect that extremes are often not univariate but multivariate phenomena, i.e. affect more than one variable.
To illustrate the multivariate nature of hydrologic extremes, let's again look at the 2021 flood. This flood event was not just extreme in terms of peak discharge at one location, it was also extreme in terms of the flood volume generated. Furthermore, it affected not just one catchment but multiple catchments in Germany, Belgium, and the Netherlands.
This example highlights that the multivariate nature of hydrological extremes can take multiple forms. In the case of peak discharge and volume, we are looking at an extreme event characterized by multiple variables and in the case of multiple affected locations at a regional extreme event. These different types of multivariate extremes have in common that they involve multiple interdependent variables, which requires a multivariate perspective. In this review, I first provide an overview of different types of multivariate hydrological extremes including regional extremes, consecutive extremes, extremes with multiple characteristics, and extremes transitions. In addition, I review tools, measures, and descriptors available to describe these different types of extremes. Second, I present modeling approaches available to model extremes in a multivariate framework, such as copula models and multivariate simulation approaches. Last, I discuss challenges related to multivariate hydrological extremes, including the regionalization of multivariate extremes to ungauged basins and the assessment of future changes in multivariate extreme events.'
Modification: p.1, l.14-33

In my opinion, more emphasis should be given to the descriptors of multivariate extremes, as defined by the Author, their differences, and the implication of using one descriptor rather than another. As a matter of fact, the definition of an extreme cannot be decoupled from the descriptor used. In the manuscript, they are simply listed in tables without further implications on their use.
Reply: *Thank you for highlighting the need to emphasize the descriptors of multivariate extremes. I substantially expanded the description of the different descriptors and provide an overview on what types of analyses the different descriptors can be used for:*

*'**Descriptors of regional extremes:** A diverse range of tools can be used to quantify the spatial dependence and spatial extents of floods and droughts. These tools include areal coverage, spatial extent, conditional spatial dependence, synchrony scale, length scale, probability of regional extremes, connectedness, severity-area-frequency curves, and severity-area-duration curves (Table 1). A first category of descriptors describes the spatial extent of extreme events at an event scale. This category comprises areal coverage, i.e. the percentage of a region or river basin under extreme conditions; spatial extent, i.e. the area under extreme conditions usually derived from gridded data; and conditional spatial dependence, i.e. the expected proportion of sites in the vicinity of a specific catchment that exceed their pth quantile during an event in which this catchment exceeds its pth quantile. While these descriptors focus on describing individual events, a second group of descriptors summarizes the behavior of regional extremes at a catchment scale. For example, the synchrony scale measures over which distance around a catchment, multiple rivers flood at*

the same time. A third group of metrics comprises metrics that summarizes regional relationships in extremes occurrence e.g. through a semivariogram or more specifically the length scale (i.e. the range of the semi-variogram) or the probability of regional extremes, i.e. the probability that a certain percentage of catchments within a region is jointly under extreme conditions. A fourth group of metrics includes pairwise measures such as connectedness determined either based on the number of co-occurrences at a pair of catchments or on the correlation between flood magnitudes at a pair of catchments. A last group of descriptors are frequency or duration curves, e.g. severity-area-frequency curves or severity-area-duration curves. Depending on which metric is chosen to describe regional extremes, the results of an analysis will differ. For example, change assessments may find different changes in regional extremes when looking at pairwise relationships than when focusing at the event-scale.'

Modification: p.5, l.118-132

*'**Descriptors of consecutive extremes:** The persistence and periodic features of hydrological extreme events have been documented using a range of measures including the Hurst exponent, power spectra derived using the Fourier transform, dry-to-dry transition probabilities, and others (Table 2). A very simple measure to characterize consecutive extremes is the number of consecutive events, e.g. the number of successive extreme months/years. Also related to individual events, one can compute extreme event transition probabilities, i.e. the probability of observing a subsequent extreme event given that an extreme event has occurred in the previous time unit (e.g. week/month/year). Instead of focusing on events, the temporal persistence of extremes can be summarized for entire time series of extreme events, for example by the Hurst exponent, which measures the long-term memory of a time series, or the average power spectrum, i.e. the average power over all frequencies after the Fourier transform. In addition, consecutive extreme events can be described by measures that characterize the temporal clustering behavior of extreme events including the dispersion index, which quantifies the departure of an observed process from a homogeneous Poisson process, Ripley's K, which counts the average number of extreme events in the temporal neighborhood of extreme events, and Kernel estimation, which estimates the time variation of extreme event counts as a smooth function of time. Another possibility to describe consecutive extremes is to identify flood/drought-rich and -poor periods using scan statistics. That is, unusual periods in the observations that are inconsistent with the assumption of independent and identically distributed random variables, i.e. periods encompassing very few or very many events, are identified with a moving window approach. If it is not just of interest to describe consecutive extremes but to identify their drivers, one can rely on cox regression models, which examine the dependence of the rate of occurrence of extremes on covariate processes, e.g. different types of teleconnection patterns. The choice of a specific descriptor will depend on the specific research question or application, i.e. on whether one would like to test for clustering significance, in which case Ripley's K or the dispersion index can be used, or whether one would like to identify specific periods particularly abundant in extremes occurrence, in which case scan statistics or Kernel estimation can be used, or one would like to explain temporal dependence, in which case one can rely on cox regression models.'*

Modification: p.8, l.185-203

*'**Descriptors of extremes with multiple characteristics**: The interdependencies between multiple characteristics of hydrological extreme events can be assessed using various*

dependence measures, including different correlation and tail dependence measures focusing on bivariate variable relationships (Table 3). Linear relationships can be quantified using Pearson's correlation coefficient while non-linear relationships can be described using Spearman's or Kendall's rank correlation coefficients. If the focus is not on the bulk of the distribution but on its tails, one can use the extremal dependence coefficient, which describes the probability of one variable being extreme given that the other one is extreme.'
Modification: p.11, l.243-247

*'**Descriptors of extremes transitions**: The transitions between dry and wet periods have been described using transition times and transition frequencies as summarized in Table 4. The transition time describes the time elapsing between dry and wet periods while the transition frequency describes the frequency of transitions between dry and wet periods.'*
Modification: p.12, l.271-273

Section 3 on modeling multivariate extremes is about models for assessing the frequency and magnitude of multivariate hydrologic extreme events (as summarized by the Author in lines 241-243). In this section, bivariate copula models are described way more extensively compared to other methods. However, it is unclear why such a detailed description and how copula models differ from the descriptors of hydrological extremes with multiple characteristics. As a matter of fact, copulas model the dependence between two variables, where the dependence between the variables is measured by the correlation between two variables (descriptors in Table 3). It would be useful to discuss whether bivariate copulas can be applied also to regional and temporal multivariate extremes and how. Moreover, limiting the description of multivariate models to bivariate statistical methods in a review paper on multivariate extremes is not enough. I encourage the Authors to add studies and methods for higher dimensions.
Reply: *Thank you for stressing the need to expand the discussion of multivariate models and distributions beyond the bivariate case. I introduce bivariate copula models in detail because they are a useful tool to describe return periods in a bivariate setting, which is often used because return periods are difficult to generalize to higher than two-dimensional data. However, I fully agree that it is important to also introduce multivariate distributions and models going beyond 2 dimensions because some of the extremes discussed in this review (e.g. the spatial extremes) are higher dimensional phenomena. Therefore, I substantially expanded section 3 (Modeling multivariate extremes) by including descriptions of (1) suitable univariate metrics for multivariate extremes, (2) bivariate distributions and return periods, (3) multivariate distributions, and (4) multivariate simulation approaches.*

*'**Univariate metrics for multivariate extremes:** Different approaches have been developed to quantify the frequency of multivariate extremes. The easiest work around for dealing with multivariate extremes is to describe the complex phenomena with a suitable univariate descriptor, such as describing regional floods by flood extent. Such univariate descriptors can be used in a univariate frequency analysis to determine the frequency and magnitude of events. Such a univariate frequency analysis first defines a sample of extreme events using either a block maxima/minima or a peak-over-threshold/threshold-level approach (Meylan et al. 2012). Second, it fits a suitable theoretical distribution to the sample of extreme events. In the case of block maxima, one usually works with a Generalized Extreme Value (GEV) distribution and in the case of threshold exceedances with a Generalized Pareto distribution (GPD) (Coles 2001). The goodness-of-fit of the distribution chosen is assessed using a test for*

extreme values such as the Anderson--Darling or Cramér-von-Mises test (Laio et al. 2004). Once a suitable distribution has been identified, one can use the probability distribution function to determine the probability of occurrence of a certain event or the quantile function to determine the magnitude of an event with a certain non-exceedance probability or return period (Figure 6). The relationship between the non-exceedance probability p and the corresponding return period T is expressed as follows:

T = mu/(1-p),

where mu is the mean inter-arrival time between two successive events, which is defined as one divided by the number of flood occurrences per year (Gumbel 1941, Salvadori et al. 2010, Brunner et al. 2016). Using this relationship, one can answer questions such as 'how often does an extreme event with a certain magnitude occur' or 'how big is an event with a certain return period'.'

Modification: p.13, l.279-295

*'**Bivariate distributions and return periods:** In many cases, however, univariate descriptors of multivariate extremes as described above do not exist, e.g. when we are interested in floods characterized by multiple variables such as magnitude, volume, and duration. Because multivariate definitions of return periods are difficult to establish, one often tries to break down the problem to bivariate relationships, for which bivariate distributions and return period definitions exist. […] '*

Modification: p.13, l.297-300

*'**Multivariate distributions:** Different models for multivariate extremes have been proposed in the literature, including multivariate distributions such as the logistic model (Kotz et al. 2000), conditional exceedance models (Heffernan and Tawn 2004, Neal et al. 2013, Keef et al. 2013), the multivariate skew-t distribution (Ghizzoni et al. 2010, Ghizzoni et al. 2012), hierarchical Bayesian models (Yan et al. 2015), max-stable models (Ribatet et al. 2013), the multivariate generalized Pareto distribution (Rootzen et al. 2006, Rootzen et al. 2018), and copula models such as pair-copula constructions (Graeler et al. 2014, Schulte et al. 2015, Bevacqua et al. 2017), factor copula models (Lee et al. 2017), vine copulas (Bedford and Cook 2002, Graeler et al. 2013), chi-square copulas (Bardossy et al. 2006, Quessy et al. 2016) the Fisher copula (Favre et al. 2018, Brunner et al. 2018). Classical multivariate distributions such as the logistic model, have mostly been defined for the bivariate or trivariate case because the complexity linked to the solution of multivariate problems increases strongly with the dimension (Kotz 2000).*
This dimensionality problem can be overcome by using conditional exceedance models as proposed by Heffernan and Tawn (2004), which can be applied to phenomena of any dimension, e.g. to model spatial extremes (Keef et al. 2013, Neal et al. 2013). In such a spatial extremes context, these models are defined in terms of the statistical distribution of a variable (e.g. streamflow) at a set of locations conditional on the variable exceeding a certain threshold at one of these locations. Applications are not limited to spatial extremes and could also be extended to extremes with multiple characteristics by quantifying the conditional distribution of one variable (e.g. flood peak) being extreme given that another variable (e.g. flood volume) is high (Salvadori et al. 2014).
However, in order to account for the full range of possible models, the use of conditional exceedance models requires the fitting of several models (e.g. by conditioning on each

variable once).
Multivariate distributions of higher dimension also exist both for componentwise maxima and threshold exceedances. Max-stable distributions arise from the limiting behavior of vectors of componentwise maxima (block maxima) (Segers et al. 2012, Ribatet et al. 2013) and there exist a number of parametric max-stable models, e.g. Brown-Resnick processes, the Smith model, or the Hüsler-Reiss model (Davison et al. 2012). Max-stable process models have e.g. been used to model the spatial dependence of rainfall extremes (Davison et al. 2012, Le et al. 2018).
Similarly, multivariate generalized Pareto distributions result from the limit distributions of exceedances over multivariate thresholds of different variables (Rootzen et al. 2006, Rootzen et al. 2018, Kiriliouk et al. 2018). These multivariate generalized Pareto distributions can be applied to a wider range of applications than max-stable models because they do not require the definition of pairwise extremes.
Another flexible alternative to max-stable models are multivariate copula models such as vine copulas which extend to higher than two to three dimensions (Bedford and Cook 2002, Graeler et al. 2013). Vine copulas construct high-dimensional copulas by mixing conditional bivariate copulas in a stagewise procedure, i.e. by modeling pairwise dependencies with bivariate copulas (Graeler et al. 2013).'
Modification: p.15, l.350-377

Point-by-point comments:

Line 28 and Line 280: my suggestion is to cite textbooks or the original journal papers where these concepts are first defined. For example, G. Salvadori and C. De Michele earlier works.
Reply: *Thank you for this suggestion. I included a few of the original journal papers, where the return period concept was introduced, and the paper by Salvadori and De Michele.*
Modification: p.13, l.293

Line 88: "precipitation dependence" dependence to what?
Reply: *I specified that I was referring to 'precipitation spatial dependence'.*
Modification: p.4, l.75

Figure 4: it would help to have more information on how drought is defined
Reply: *I specified that 'droughts were here defined using a variable threshold at the 15th flow percentile.'*
Modification: p.6, caption Figure 3

Lines 182 – 200: discussion about variables dependence is a bit vague. Which variables? Is it a bi-variate dependence? The example of dependence between peak and volume for hydraulic design should be elaborated further.
Reply: *Thank you very much for pointing out the need for clarification. I provide a few examples of variable pairs of interest, specify that we are talking about bivariate dependence, and use the example of peak-volume to illustrate the importance of considering bivariate/multivariate relationships in hydraulic design by adding the following sentences: 'For example, flood duration and volume or flood volume and flood peak show strong correlations (Figure 5), i.e. they show bivariate dependence. […] Such bivariate interdependence is e.g. found for drought deficit and duration or drought deficit and intensity (Figure 5a,b).[…] Such dependence, e.g. between peak discharge and flood volume, is*

important for hydraulic design because dam failure depends not only on flood peak but also volume (De Michele et al. 2005).
Modification: p.9, l.210-211

Line 321: studies in higher dimensions should be added to the manuscript
Reply: *I rewrote Section 3 (Modeling multivariate extremes) and added a subsection called 'multivariate distributions', which discusses different models for multivariate extremes including conditional exceedance models, max-stable models, the multivariate generalized Pareto distribution, and high-dimensional copula models such as vine copulas.*
Modification: p.12-16, Sections 3.1-3.3

Reviewer 2

This manuscript presents a review of some hydrological problems that can be characterized in terms of a multivariate extreme value distribution. The identified hydrological conditions that require a probabilistic estimation in terms of event magnitude and occurrence are listed and briefly discussed along with the metrics that can be generally used to determine the dependence among the variables. Further, copula for multivariate frequency analysis and continuous time serie simulation are introduced as strategies for modeling those phenomena. Due to the variety and complexity of the problems mentioned in the review paper, each of them is only hinted at, missing an in-depth discussion about several important issues. Further, many interesting works about multivariate statistical modeling are not mentioined at all; indeed, also the most recent literature on the topic is very rich. Based on these consideration, I suggest the Author to revise her work trying to improve the description of the phenomena, especially those problems that are still unsolved, and enlarge the state of the art description referring the interested readers to the most recent papers (and books) that provide well established and innovative solutions with a deeper insight into the mentioned problems.
Reply: *Thank you very much for your assessment and for highlighting that the discussion of multivariate statistical models was too superficial in the first version of the manuscript. I rewrote Section 3 (Modeling multivariate extremes) by including multivariate statistical models going beyond copula approaches, which are applicable to higher dimensional problems such as spatial extremes. Furthermore, I substantially expanded the description of metrics used to describe the four types of multivariate hydrologic extremes I focus on in this review, i.e. regional extremes, consecutive extremes, extremes with multiple characteristics, and extremes transitions.*
Modification: p.12-16, Sections 3.1-3.3