Received: 1 June 2016 | Revised: 22 September 2016 | Accepted: 28 September 2016

DOI 10.1002/env.2428

WILEY

RESEARCH ARTICLE

Directional multivariate extremes in environmental phenomena

Raúl Torres[1] | Carlo De Michele[2] | Henry Laniado[3,4] | Rosa E. Lillo[5]

[1]Department of Statistics, Universidad Carlos III de Madrid, Leganés, 28911, Spain
[2]Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, 20133, Italy
[3]School of Mines, Universidad Nacional de Colombia, Medellín, 50034, Colombia
[4]Department of Mathematical Sciences, Universidad EAFIT, Medellín, 7023, Colombia
[5]Department of Statistics, Institute of Financial Big Data, Universidad Carlos III de Madrid, Getafe, 28903, Spain

Correspondence
Raúl Torres, Department of Statistics, Universidad Carlos III de Madrid, Leganés 28911, Spain.
Email: ratorres@est-econ.uc3m.es

Funding information
Spanish Ministry of Economy and Competitiveness, Grant/Award Number: ECO2015-66593-P; Departamento Administrativo de Ciencia y Tecnología, COLCIENCIAS, Grant/Award Number: 656, 2014, Es Tiempo de Volver

Several environmental phenomena can be described by different correlated variables that must be considered jointly in order to be more representative of the nature of these phenomena. For such events, identification of extremes is inappropriate if it is based on marginal analysis. Extremes have usually been linked to the notion of quantile, which is an important tool to analyze risk in the univariate setting. We propose to identify multivariate extremes and analyze environmental phenomena in terms of the directional multivariate quantile, which allows us to analyze the data considering all the variables implied in the phenomena, as well as look at the data in interesting directions that can better describe an environmental catastrophe. Because there are many references in the literature that propose extremes detection based on copula models, we also generalize the copula method by introducing the directional approach. Advantages and disadvantages of the nonparametric proposal that we introduce and the copula methods are provided in the paper. We show with simulated and real data sets how by considering the first principal component direction we can improve the visualization of extremes. Finally, two cases of study are analyzed: a synthetic case of flood risk at a dam (a three-variable case) and a real case study of sea storms (a five-variable case).

KEYWORDS
copulas, flood risk, multivariate directional quantiles, multivariate extremes, principal component analysis, sea storms

1 | INTRODUCTION

Serious economic and social consequences are generally associated with extreme environmental events such as floods, storms, and droughts (Chebana & Ouarda, 2011b), which are usually defined in terms of several correlated variables. For instance, rainfall is characterized by storm intensity and duration (e.g., De Michele & Salvadori, 2003; Salvadori & De Michele, 2004); air quality is described in terms of levels of ozone and nitrogen dioxide (e.g., Chebana & Ouarda, 2011a; Heffernan & Tawn, 2004); floods are modeled by their peak, volume, and duration (e.g., Chebana & Ouarda, 2011b; De Michele et al., 2005; Grimaldi & Serinaldi, 2006; Shiau, 2003); droughts are modeled by volume, duration, and magnitude (e.g., De Michele et al., 2013; Kim et al., 2003; Salvadori & De Michele, 2015); and sea storms are represented by wave height, peak, direction, and duration (e.g., De Michele et al., 2007). Consequently, extremes detection cannot be made on the basis of a univariate analysis.

There are references in the literature that tackle multivariate extreme detection. Some studies use copulas since the work by De Michele and Salvadori (2003), for example Salvadori (2004), Salvadori et al. (2011, 2013, 2016), whom also define multivariate versions of the return period[1], and Grimaldi and Serinaldi (2013). Another alternative is given by Serfling (2002); Chebana and Ouarda (2011a) through depth functions. However, both alternatives have drawbacks when they have to be implemented in high-dimensional scenarios. Copulas, when used in parametric form, are difficult to be applied in large dimensions for the increasing number of parameters to estimate, and depth functions are problematic due to the lack of computational implementation in most of the cases. Therefore, the first contribution of the paper is to introduce a method to detect extremes based on a nonparametric procedure suitable for high-dimensional analysis.

[1]For further information of the return period we refer to Salvadori et al. (2007)

On the other hand, extremes have been traditionally analyzed in one dimension by considering only the probabilities of exceeding quantiles related to either the distribution function or the survival function. In other words, observations are considered extreme if they are associated to lower values or upper values of the variable, which is equivalent to looking at the data in one of the two possible directions $\{-1, +1\}$. Some extensions of quantiles to the bivariate case have been proposed in Fernández-Ponce and Suárez-Llorens (2002); Shiau (2003); Salvadori (2004); Zhang and Singh (2006); Embrechts and Puccetti (2006) and to the n-dimensional setting in Gupta and Manohar (2005); Salvadori et al. (2011); Fraiman and Pateiro-López (2012); Cousin and Di Bernardino (2013); Di Bernardino et al. (2015). The generalizations of the quantile notion in all the previous references consider, as in the univariate case, the directions associated with the distribution function or the survival function.

But why not look at the data with different perspectives and take advantage of the inherent complexity of the n-dimensional setting the data lives in? There exist infinite directions to look at the data from a reference point that could help with the accuracy of the analysis and the interpretation of the results. Attempts have been made considering alternative directions, for instance, Laniado et al. (2012) and Torres et al. (2015) developed a financial application where the risk of losses is analyzed considering the direction of the investment weight composition in a portfolio; Cascos and Molchanov (2007); Hallin et al. (2010) and Kong and Mizera (2012) have applied a directional setting to define quantile trimmings. Hence, a second contribution of this paper is to outline a general approach to detect directional multivariate extremes, which can be useful in other statistical areas apart from environmental sciences.

The definition of directional multivariate extremes is based on the directional multivariate quantile introduced in Laniado et al. (2010) and Torres et al. (2015), where the free parameter of direction included can be chosen considering external information such as anthropogenic forces generating today's the environmental global change (see Hegerl et al., 2004). Specifically, we propose to use principal component analysis (PCA) in the environmental framework because the visualization of the extremes improves with respect to the use of the classical directions, as is shown in two cases of study, PCA is only a suitable method to select a direction of analysis. However, if prior information is available about the physical phenomenon, other directions can be more appropriated.

Firstly, we use the probabilistic model proposed in Salvadori et al. (2011) for the analysis of floods incoming to the Ceppo Morelli dam in Italy to perform a Monte Carlo study for a time window of 1,000 years. Our approach improves previous results by the reduction of the ratio of false positives (regular observations that are classified as extremes). Secondly, we perform a study of sea storms considering variables such as wave height, storm duration, storm magnitude, storm direction, and inter-arrival time that provides information about the period of calm between two successive storms. The study shows relevant differences with the work by De Michele et al. (2007) such as the computational feasibility of the method in the five-dimensional setting and also the visualization of the extremes with cross-sectional plots, where it is shown how the classical theory identifies an excessive number of observations as extremes.

The third contribution of the paper is to introduce the directional approach in the copula method. We obtain results that establish the equivalence between the directional approach and the copula-based methods. It is also shown with the simulations across the document how using a mixture of both settings (directional and copula approach), we can describe better a multivariate system.

The structure of the paper is the following: Section 2 introduces the notion of directional multivariate extremes and the nonparametric procedure to carry out the identification in practice. Section 3 presents a summary of the classical methodology based on copulas, and theoretical results linking copulas and the notion of directions. In Sections 4 and 5, we motivate the use of principal components (PCA) to get an interesting direction of analysis in real case studies. We also present in Section 6 some examples of the pros and cons for the extreme identification using our directional nonparametric procedure or the extended copula method. Finally, Section 7 presents some conclusions.

2 | METHODOLOGY

In this section, we present the procedure to identify directional multivariate extremes based on the directional setting proposed in Torres et al. (2015) (first contribution), a nonparametric algorithm for practical implementation (second contribution), and the motivation of the first PCA direction as a proposal of direction to be considered.

2.1 | Directional multivariate extreme value analysis

The directional multivariate setting is defined in terms of the *oriented orthant* introduced in Laniado et al. (2012).

An oriented orthant in $\mathbb{R}^n$ with vertex $\mathbf{x}$ in direction $\mathbf{u}$ is defined by,

$$\mathfrak{C}_{\mathbf{x}}^{R_{\mathbf{u}}} = \{\mathbf{z} \in \mathbb{R}^n : R_{\mathbf{u}}(\mathbf{z} - \mathbf{x}) \geqslant 0\}. \tag{1}$$

where $\mathbf{u} \in \{\mathbf{z} \in \mathbb{R}^n : ||\mathbf{z}|| = 1\}$ and $R_{\mathbf{u}}$ is an orthogonal matrix such that $R_{\mathbf{u}}\mathbf{u} = \mathbf{e}$, with $\mathbf{e} = \frac{\sqrt{n}}{n}[1, \dots, 1]'$.

Note that an oriented orthant is a translation and a rotation of the nonnegative Euclidean orthant toward a new vertex in the point $\mathbf{x}$ and a new direction $\mathbf{u}$. As is explained in Torres et al. (2015), $R_{\mathbf{u}}$ is not unique for $n \geqslant 3$. Then, in order to guarantee uniqueness in the orthogonal transformation, the QR-oriented orthant was defined in (Torres et al., 2015) as follows, let $\mathbf{u}$ be a unit vector with non-null components and let $M_{\mathbf{u}}$ and $M_{\mathbf{e}}$ be matrices defined as,

$$\begin{aligned} M_{\mathbf{u}} &= [\mathbf{u}, \quad sgn(u_2)\mathbf{e}_2, \quad \cdots, \quad sgn(u_n)\mathbf{e}_n] \quad \text{and} \\ M_{\mathbf{e}} &= [\mathbf{e}, \quad \mathbf{e}_2, \quad \cdots, \quad \mathbf{e}_n], \end{aligned} \tag{2}$$

where u_i, $i = 1, \ldots, n$ is the i-th component of $\mathbf{u}$, $sgn(\cdot)$ is the scalar sign function and $\mathbf{e}_i$ is the vector with all its components equal to zero except the i-th component equal to one. Note that the hypothesis of $u_i \neq 0$, $i = 1, \ldots, n$ guarantees that $M_{\mathbf{u}}$ always is a matrix of rank n. Now, we consider the QR decomposition of $M_{\mathbf{u}}$ and $M_{\mathbf{e}}$ (see, e.g., Horn and Johnson, 2013, Chapter 2),

$$M_{\mathbf{u}} = Q_{\mathbf{u}}T_{\mathbf{u}} \quad \text{and} \quad M_{\mathbf{e}} = Q_{\mathbf{e}}T_{\mathbf{e}},$$

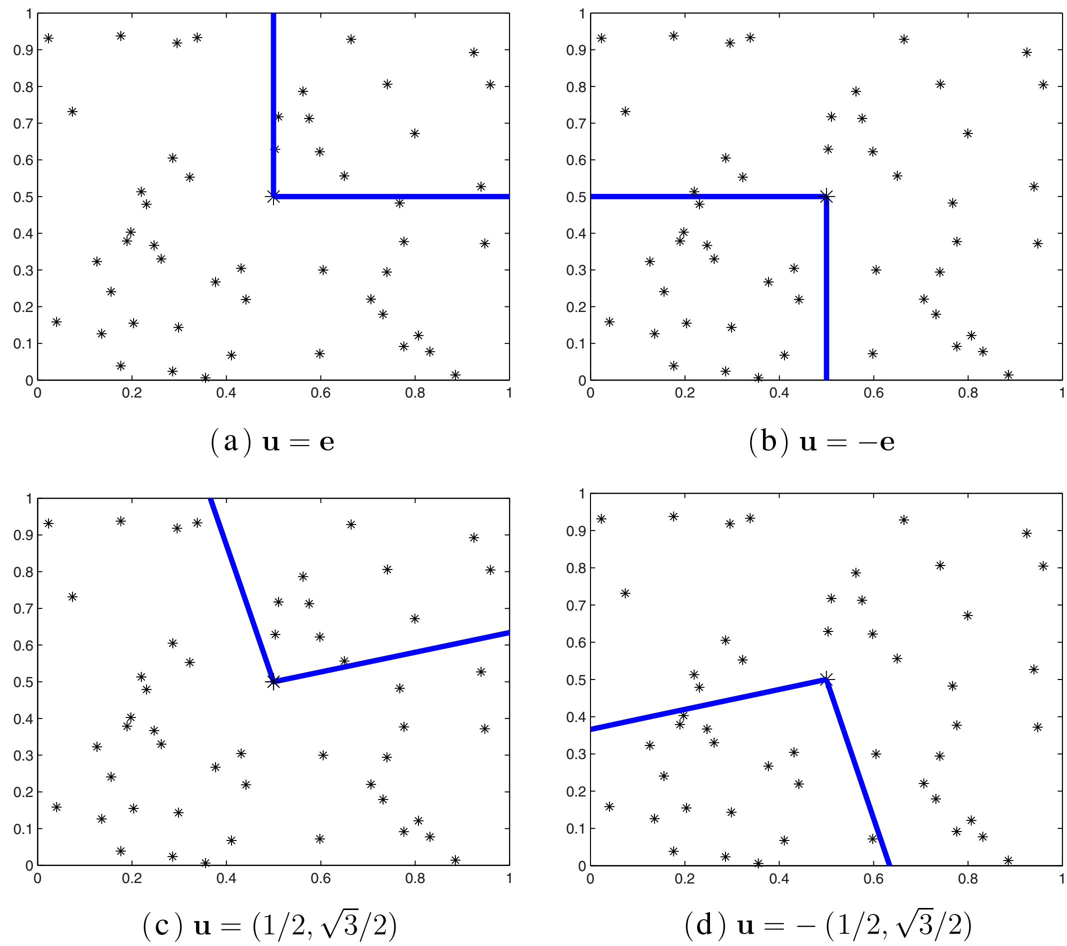
such that $T_{\mathbf{u}}$ and $T_{\mathbf{e}}$ are triangular matrices with positive diagonal elements and $Q_{\mathbf{u}}$ and $Q_{\mathbf{e}}$ are orthogonal matrices. Note that these decompositions are unique due to both the full rank of $M_{\mathbf{u}}$ and $M_{\mathbf{e}}$ and the hypothesis of positive diagonal elements in $T_{\mathbf{u}}$ and $T_{\mathbf{e}}$ (see, e.g., Horn and Johnson, 2013, Theorem 2.1.14, p. 89).

Also, the first columns on $Q_{\mathbf{u}}$ and $Q_{\mathbf{e}}$ are the same as in $M_{\mathbf{u}}$ and $M_{\mathbf{e}}$; that is, $\mathbf{u}$ and $\mathbf{e}$, respectively. Therefore, $Q_{\mathbf{e}}\mathbf{e}_1 = \mathbf{e}$ and $Q_{\mathbf{u}}\mathbf{e}_1 = \mathbf{u}$, and thus, $\left(Q_{\mathbf{e}}Q'_{\mathbf{u}}\right)\mathbf{u} = \mathbf{e}$. This decomposition motivates the definition of the QR-oriented orthant where the rotation matrix is unique.

Definition 1. The QR-oriented orthant with vertex $\mathbf{x}$ in direction $\mathbf{u}$ is an oriented orthant satisfying $R_{\mathbf{u}} = Q_{\mathbf{e}}Q'_{\mathbf{u}}$, hereafter denoted by $\mathfrak{C}_{\mathbf{x}}^{\mathbf{u}}$.

Figure 1 shows examples of the divisions in the bivariate plane that can be performed using the concept of QR-oriented orthant for different directions. Note that the direction $\mathbf{u} = \mathbf{e}$ generates the rotation matrix $R_{\mathbf{u}}$ equal to the identity matrix.

If $n = 1$ (univariate setting), there are only two possible directions $\{-1, 1\}$ and the corresponding orthants at vertex x are the intervals $\{(-\infty, x], [x, \infty)\})$, respectively. Then, in terms of probability, they represent the valuation of the distribution and survival functions in x. But, when $n > 1$, note that the values of the distribution and survival functions at some point $\mathbf{x}$ correspond to the probability of the QR-oriented orthants with vertexes in directions $-\mathbf{e}, \mathbf{e}$, respectively. In the multivariate extremes literature, there are many studies that use those functions as a natural way to extend different procedures from the univariate extreme analysis (e.g., De Michele et al., 2005; Di Bernardino et al., 2015; Embrechts & Puccetti, 2006; Salvadori & De Michele, 2004).



(a) $\mathbf{u} = \mathbf{e}$

(b) $\mathbf{u} = -\mathbf{e}$

(c) $\mathbf{u} = (1/2, \sqrt{3}/2)$

(d) $\mathbf{u} = -(1/2, \sqrt{3}/2)$

FIGURE 1 Examples of QR-oriented orthants with the same vertex but different directions

However, infinite directions are possible when $n > 1$, which motivates the directional approach, because more important than using the distribution and survival functions for a random vector $\mathbf{X}$, could be using directly the probability measure of the random vector to describe the extremes properly. To clarify ideas, one can think in the bivariate setting and a random vector $\mathbf{X}$ with negative dependence. Then, it seems more convenient to use the complementary part of the division of the plane than the pair of directions $\{-\mathbf{e},\mathbf{e}\}$, that is, to use the directions given by $\left\{\left(-1/\sqrt{2},1/\sqrt{2}\right),\left(1/\sqrt{2},-1/\sqrt{2}\right)\right\}$ (e.g., Belzunce et al., 2007; Chebana and Ouarda, 2011b), hence the importance of the directional approach. Hereafter, we call *classical directions* to the collection of 2^n orthants that divide naturally the hyperplane, that is, the collection of unitary n-dimensional vectors with components in $\{-1,1\}$. Now, we can introduce the necessary tools to attain the main purposes of our work, after motivation of the directions.

A directional multivariate quantile of a random vector $\mathbf{X}$ at level α in direction $\mathbf{u}$ is defined as

$$\mathcal{Q}_{\mathbf{X}}(\alpha,\mathbf{u}) = \partial\left\{\mathbf{z}\in\mathbb{R}^n : \mathbb{P}\left[\mathfrak{C}_{\mathbf{z}}^{\mathbf{u}}\}\right] \leqslant \alpha\right\} \tag{3}$$

where $0 < \alpha < 1$, and ∂ means the boundary of a set. Once a value of α is fixed (near to 0 for extreme value analysis), $\mathcal{Q}_{\mathbf{X}}(\alpha,\mathbf{u})$ divides the space into two sets:

- *The upper*α*-level set in direction* $\mathbf{u}$:

$$\mathcal{U}_{\mathbf{X}}(\alpha,\mathbf{u}) = \left\{\mathbf{z}\in\mathbb{R}^n : \mathbb{P}\left[\mathfrak{C}_{\mathbf{z}}^{\mathbf{u}}\}\right] < \alpha\right\} \tag{4}$$

- *The lower* α*-level set in direction* $\mathbf{u}$:

$$\mathcal{L}_{\mathbf{X}}(\alpha,\mathbf{u}) = \left\{\mathbf{z}\in\mathbb{R}^n : \mathbb{P}\left[\mathfrak{C}_{\mathbf{z}}^{\mathbf{u}}\}\right] > \alpha\right\}. \tag{5}$$

These sets motivate the definition of extreme related to the pair $(\alpha,\mathbf{u})$ as those points exceeding the threshold given by the hyper-curve $\mathcal{Q}_{\mathbf{X}}(\alpha,\mathbf{u})$, that is, we characterize the *extreme events* as those points belonging to the associated *upper level set*. The *risky points* are the ones belonging to *the directional multivariate quantile* $\mathcal{Q}_{\mathbf{X}}(\alpha,\mathbf{u})$, and the *non-risky points* are those in the *lower level set*. That is,

$\mathbf{z}$ is a directional extreme related to $(\alpha,\mathbf{u}) \Longleftrightarrow \mathbf{z}\in\mathcal{U}_{\mathbf{X}}(\alpha,\mathbf{u})$.

Note that expressions 3–5 with $\mathbf{u} = \mathbf{e}$ and values of α close to zero are the multivariate extension of the univariate quantile definition based on the survival function. Now, if we rewrite those expressions in terms of the pair $(1-\alpha,-\mathbf{u})$ and reversing the inequalities, we obtain the corresponding quantile extension related to the distribution function. However, these two alternatives are not equivalent for dimension $n \geqslant 2$ unlike the univariate case. Such duality can be also seen in the approaches AND and OR defined in De Michele et al. (2007), or the UPPER and LOWER differentiation given in Embrechts and Puccetti (2006); Cousin and Di Bernardino (2013). But, without loss of generality, we have decided to implement the extreme detection analysis in terms of the survival analogy, because a key relationship can be established between the extremes given by 4 and those associated to the arguments $(1-\alpha,-\mathbf{u})$ reversing the inequalities (see Corollary 4.3 in Torres et al., 2015); that is,

$$\underline{\mathcal{U}}_{\mathbf{X}}(\alpha,\mathbf{u}) := \left\{\mathbf{z}\in\mathbb{R}^n : \mathbb{P}\left[\mathfrak{C}_{\mathbf{z}}^{-\mathbf{u}}\}\right] > 1-\alpha\right\} \subset \mathcal{U}_{\mathbf{X}}(\alpha,\mathbf{u}), \tag{6}$$

Then, in terms of risks, relation 6 allows us to consider risk aversion; that is, we would expect more extreme events that corresponds to a conservative position. Now, we describe a nonparametric procedure to estimate the extreme thresholds, that is, the directional multivariate quantiles, as well as, the lower and upper level sets for a dataset.

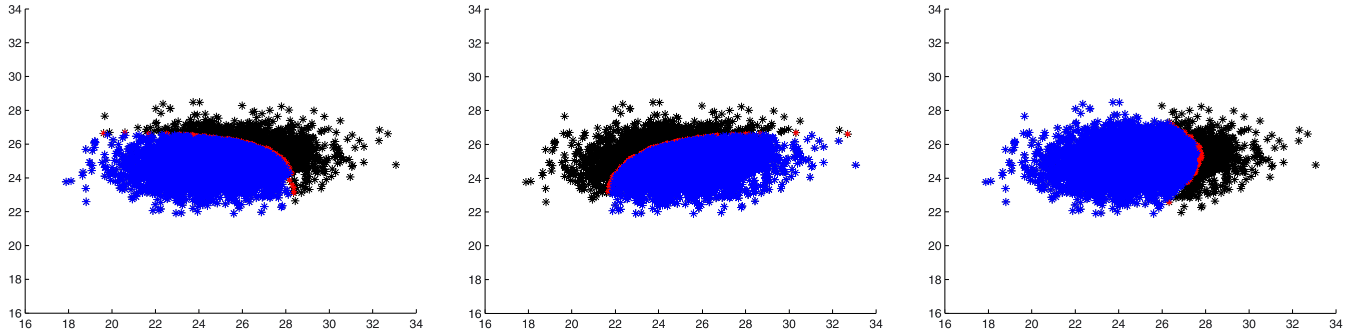
2.2 | Nonparametric procedure and suitable direction of analysis

As we mentioned in the Introduction, one of the contributions of this paper is to provide a nonparametric algorithm to estimate the quantiles. It is remarkable that most of the references that deal with the multivariate extreme identification problem are based on copula procedures that have inherent weaknesses due to the complex process of parameter estimation and the absence of computational feasibility in high dimensions. Therefore, we try to improve these issues by introducing a pseudo-algorithm based on the empirical distribution in order to get the level sets we are interested in. Firstly, we fix a preliminary notation:

- $\mathbf{X}_m := \{\mathbf{x}_1,\cdots,\mathbf{x}_m\}$, sample data from the random vector $\mathbf{X}$,
- $\mathbb{P}_{\mathbf{X}_m}[\cdot]$ is the empirical probability law of $\mathbf{X}_m$,
- $\hat{\mathcal{Q}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u}) := \left\{\mathbf{x}_j : \left|\mathbb{P}_{\mathbf{X}_m}\left[\mathfrak{C}_{\mathbf{x}_j}^{\mathbf{u}}\right]-\alpha\right| \leqslant h\right\}$ the sample quantile curve with a slack h, avoiding an empty set of estimated quantiles.
- $\hat{\mathcal{U}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u}) := \left\{\mathbf{x}_j : \mathbb{P}_{\mathbf{X}_m}\left[\mathfrak{C}_{\mathbf{x}_j}^{\mathbf{u}}\right] < \alpha-h\right\}$ the sample upper α-level set with a slack h,
- $\hat{\mathcal{L}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u}) := \left\{\mathbf{x}_j : \mathbb{P}_{\mathbf{X}_m}\left[\mathfrak{C}_{\mathbf{x}_j}^{\mathbf{u}}\right] > \alpha+h\right\}$ the sample lower α-level set with a slack h.

Once defined the direction of analysis and the parameter α, it is possible to estimate the directional multivariate quantile and the level sets through the following pseudo-algorithm:

Input: $\mathbf{u}$, α, h and the multivariate sample $\mathbf{X}_m$.
  for $i = 1$ to m
    $P_i = \mathbb{P}_{\mathbf{X}_m}\left[\mathfrak{C}_{\mathbf{x}_i}^{\mathbf{u}}\right]$,
    If $|P_i-\alpha| \leq h$
      $\mathbf{x}_i \in \hat{\mathcal{Q}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u})$,
    end
    If $P_i < \alpha-h$
      $\mathbf{x}_i \in \hat{\mathcal{U}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u})$,
    end
    If $P_i > \alpha+h$
      $\mathbf{x}_i \in \hat{\mathcal{L}}^h_{\mathbf{X}_m}(\alpha,\mathbf{u})$,
    end



(a) $\mathbf{u} = \mathbf{e}$ (b) $\mathbf{u} = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ (c) $\mathbf{u} =$ First PCA

FIGURE 2 Directional extremes at $\alpha = 5\%$

As output, we get an estimation of the directional α-quantile, or in other words, the hyper surface of thresholds in the selected direction of analysis, $\hat{\mathcal{Q}}^h_{\mathbf{X}_m}(\alpha, \mathbf{u})$. We also obtain those points belonging to the non-risky level set, $\hat{\mathcal{L}}^h_{\mathbf{X}_m}(\alpha, \mathbf{u})$, and the extreme level set, $\hat{\mathcal{U}}^h_{\mathbf{X}_m}(\alpha, \mathbf{u})$. An example is presented in Figure 2, where simulated data from a bivariate normal distribution with $\boldsymbol{\mu} = [25{,}25]$, $\sigma_1^2 = 4$, $\sigma_2^2 = 1$, $\rho_{1,2} = 0.15$ is considered. We show the three sets of observations, the *directional* 5% — *quantiles* in red, the *upper* 5% — *level set* or *extreme level-set* in black and the *lower* 5%—*level set* or *non-risky level set* in blue. We have used three different directions: the classical direction $\mathbf{e}$ (survival distribution), the complementary bivariate direction $\left(-1/\sqrt{2}, 1/\sqrt{2}\right)$ and the direction given by taken the first vector of its PCA. One can observe how the identification of extremes varies according to changes in the direction in which the data is analyzed, for the same level $\alpha = 5\%$.

Notice that different contexts or phenomena could lead to consider different particular directions of interest. For instance in portfolio theory, the direction given by the portfolio weights of investments is of particular interest because it takes into account the losses due to the composition of the investment in the portfolio (see Laniado et al., 2012; Torres et al., 2015). On the other hand, researchers in environmental science could consider important other directions more related to the phenomenon of analysis.

In any case, we want to motivate here an interesting way to obtain a relevant direction of analysis by considering the PCA based on the original available data, which is an important statistical multivariate tool that describes the information about variability of the data jointly considered. It is well known that the first component provides the direction that accumulates the maximum amount of uncertainty of the data by the strongest linear combination representing the behavior of the system. This is a statistical property of the first PCA, but other possible choices can be more appropriated depending on the problem under consideration. We have tested this direction as a good candidate for the analysis in the applications considered in the paper.

3 | EXTREMES BASED ON COPULAS AND THE DIRECTIONAL APPROACH

The importance of copulas have been recognized due to their capacity to capture the dependence structure of a set of random variables. Copulas are also a tool to construct families of multivariate distributions. In addition, copulas move in a compact support that guarantees theoretical advantages. Recall for example the capability to simulate data through copulas as we show in the case study of flood risk at a dam. Therefore, this section is devoted to introduce the directional approach to detect extremes when the dataset is modeled using copulas, which is the third contribution of the paper.

A n-copula C is a multivariate distribution function in $[0,1]^n$ with marginals equally distributed as $[0,1]$—uniforms. Also, if we consider the associated survival distribution, we get the survival copula $\bar{C}$. For example, when $n = 2$, C and $\bar{C}$ are linked as follows,

$$\bar{C}(v_1, v_2) = v_1 + v_2 - 1 + C(1 - v_1, 1 - v_2). \tag{7}$$

The importance of modeling through copulas is due to Sklar's theorem (see Theorem 1 in the Appendix.), because any *joint survival function* $\bar{F}$ (*joint distribution function* F) can be obtained through its *marginal survivals* $\bar{F}_i$, $i = 1, \dots, n$ (*marginal distributions* F_i) and the *survival copula* $\bar{C}$ (*copula* C). This representation of the models makes more feasible to obtain closed or approximated expressions for $\mathcal{Q}_{\mathbf{X}}(\alpha, \mathbf{e})$ in 3, $\mathcal{U}_{\mathbf{X}}(\alpha, \mathbf{e})$ in 4 and $\mathcal{L}_{\mathbf{X}}(\alpha, \mathbf{e})$ in 5. Thereby, in terms of survival copulas, Equations 3–5 for $\mathbf{u} = \mathbf{e}$ become the following,

$$\begin{aligned}\mathcal{Q}_{\mathbf{X}}(\alpha, \mathbf{e}) \equiv \{\mathbf{x} \in \mathbb{R}^n \quad &\text{such that} \quad x_i = \bar{F}^{-1}_{X_i}(v_i); \\ i = 1, \dots, n; \quad &\mathbf{v} \in [0,1] : \quad \bar{C}(\mathbf{v}) = \alpha\},\end{aligned} \tag{8}$$

$$\begin{aligned}\mathcal{U}_{\mathbf{X}}(\alpha, \mathbf{e}) \equiv \{\mathbf{x} \in \mathbb{R}^n \quad &\text{such that} \quad x_i = \bar{F}^{-1}_{X_i}(v_i); \\ i = 1, \dots, n; \quad &\mathbf{v} \in [0,1] : \quad \bar{C}(\mathbf{v}) < \alpha\}.\end{aligned} \tag{9}$$

$$\begin{aligned}\mathcal{L}_{\mathbf{X}}(\alpha, \mathbf{e}) \equiv \{\mathbf{x} \in \mathbb{R}^n \quad &\text{such that} \quad x_i = \bar{F}^{-1}_{X_i}(v_i); \\ i = 1, \dots, n; \quad &\mathbf{v} \in [0,1] : \quad \bar{C}(\mathbf{v}) > \alpha\}.\end{aligned} \tag{10}$$

Most of the studies dealing with extremes detection in terms of copulas are based on definitions similar to 8–10 (e.g., Grimaldi & Serinaldi 2013; Salvadori & De Michele, 2004), which are focused on the direction given by the survival function (or those associated to the parameters $(1-\alpha, -\mathbf{e})$, which are considering distribution functions). However, Chebana and Ouarda (2011b) used the directions $\left\{\left(1/\sqrt{2}, -1/\sqrt{2}\right), \left(-1/\sqrt{2}, 1/\sqrt{2}\right)\right\}$ when negative dependent bivariate models are considered. Indeed, they consider copulas associated to a random vector, but rotated 90° and 270°, which can be done taking advantage of the relationships between the corresponding copula C and the following expressions,

$$C_{90}(v_1, v_2) = u_1 + C(v_1, 1-v_2) \quad \text{and}$$
$$C_{270}(v_1, v_2) = v_2 + C(1-v_1, v_2).$$

These considerations highlight the need to include directions in the copula approach. Thus, the goal of this section is to include the general directional setting to the copula approach and to describe a directional extreme detection method based on copulas, although we will also show the drawbacks of the procedure with some illustrative simulations. The following result shows how the directional approach can be implemented using copulas.

Proposition 1. Let $\mathbf{u}$ be fixed, then the directional quantiles and the associated upper and lower level sets of a random vector $\mathbf{X}$ (defined in 3–5) are the same as those obtained by applying the copula method (summarized in 8–10)) to the random vector $R_{\mathbf{u}}\mathbf{X}$, where $R_{\mathbf{u}}$ is the rotation matrix in 1.

Proof. First note that any analysis using the information of the survival or the distribution functions for a random vector $\mathbf{X}$ through copulas is equivalent to the analysis in the set of classic directions $\{\mathbf{e}, -\mathbf{e}\}$, that is, the copula quantile analysis is always done in those directions. Moreover, once $\mathbf{u}$ is fixed, Sklar's theorem (see Appendix) provides the following relationships between the random vector $R_{\mathbf{u}}\mathbf{X}$ and the copulas $C_{\mathbf{u}}, \bar{C}_{\mathbf{u}}$ for any pre-rotation $R_{\mathbf{u}}$,

$$\mathbb{P}\left[R_{\mathbf{u}}\mathbf{X} \in \mathfrak{C}_{\mathbf{x}}^{\mathbf{e}}\right] = \bar{F}_{R_{\mathbf{u}}\mathbf{X}}(\mathbf{x}) = \bar{C}_{\mathbf{u}}\left(\bar{F}_{[R_{\mathbf{u}}\mathbf{X}]_1}(x_1), \cdots, \bar{F}_{[R_{\mathbf{u}}\mathbf{X}]_n}(x_n)\right), \tag{11}$$

$$\mathbb{P}\left[R_{\mathbf{u}}\mathbf{X} \in \mathfrak{C}_{\mathbf{x}}^{-\mathbf{e}}\right] = F_{R_{-\mathbf{u}}\mathbf{X}}(\mathbf{x}) = C_{\mathbf{u}}\left(F_{[R_{-\mathbf{u}}\mathbf{X}]_1}(x_1), \cdots, F_{[R_{-\mathbf{u}}\mathbf{X}]_n}(x_n)\right), \tag{12}$$

where $\bar{F}_{[R_{\mathbf{u}}\mathbf{X}]_i}(x_i), F_{[R_{-\mathbf{u}}\mathbf{X}]_i}(x_i), i = 1, \dots, n$ are respectively the marginal survival and distribution functions of the rotated random vector $R_{\mathbf{u}}\mathbf{X}$.

Hence, we get the directional level sets 3–5 by applying the inverse of the rotation $R_{\mathbf{u}}$ to the elements belonging to the sets defined in Equations 8–10 where the copula modeling of $R_{\mathbf{u}}\mathbf{X}$ has been used. All this thanks to Property 3.8 in Torres et al. (2015) and relationship (11), (the result also holds through (12) when the alternative definition based on joint distributions is used). □

As a conclusion, the directional analysis can be done theoretically using copula models but over the pre-rotated random vector. An example to illustrate Proposition 1 on the bivariate field is provided below. Indeed, $n = 2$ can be considered as the foundation of the nesting copula procedures used in the literature to confront the problem of large dimensions: *nested copula method* (see De Michele et al., 2007; Grimaldi & Serinaldi 2006) and *pair-copula construction*, also called the *vine copula method* (see Grimaldi & Serinaldi, 2013), see Appendix for more details.

Let $\mathbf{X} = (X_1, X_2)$ be a bivariate vector with Gaussian survival marginals $\bar{F}_1, \bar{F}_2$ with parameters μ_1, σ_1^2 and μ_2, σ_2^2, respectively. We also assume that $\mathbf{X}$ satisfies a Gaussian survival copula $\bar{C}$ with Pearson's correlation coefficient ρ. Note that,

$$\begin{aligned} \mu_i &= \int_{-\infty}^{\infty} x_i dF_i = \int_0^1 F_i^{-1}(u_i) du_i \\ \sigma_i^2 &= \int_{-\infty}^{\infty} (x_i - \mu_i)^2 dF_i = \int_0^1 (F_i^{-1}(u_i) - \mu_i)^2 du_i, \end{aligned} \tag{13}$$

where $F = 1 - \bar{F}$, for all $i = 1{,}2$ and denote the covariance matrix of $\mathbf{X}$ by,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

It is well known that the Gaussian copula is closed under orthogonal transformations. Then, for any direction $\mathbf{u}, R_{\mathbf{u}}\mathbf{X}$ also holds a Gaussian survival copula $\bar{C}^{\mathbf{u}}$ with Pearson's correlation coefficient given by

$$\rho^{\mathbf{u}} = \left[R_{\mathbf{u}}\Sigma R_{\mathbf{u}}'\right]_{12}, \tag{14}$$

and Gaussian survival marginals $\bar{F}_1^{\mathbf{u}}, \bar{F}_2^{\mathbf{u}}$ with parameters

$$\begin{aligned} \mu_1^{\mathbf{u}} &= \left[R_{\mathbf{u}}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right]_{11}, \sigma_1^{2\ \mathbf{u}} = \left[R_{\mathbf{u}}\Sigma R_{\mathbf{u}}'\right]_{11} \quad \text{and} \\ \mu_2^{\mathbf{u}} &= \left[R_{\mathbf{u}}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right]_{21}, \sigma_2^{2\ \mathbf{u}} = \left[R_{\mathbf{u}}\Sigma R_{\mathbf{u}}'\right]_{22}, \end{aligned} \tag{15}$$

where $[\cdot]_{ij}$ is the i,j position in a matrix.

Now, we fix the parameters $\mu_1 = 5$, $\sigma_1^2 = 25$, $\mu_2 = 10$, $\sigma_2^2 = 1$ and $\rho = 0.2$ to illustrate the extreme detection through copulas. Figure 3 summarizes the results. The three top plots describe the procedure in the classical direction$\mathbf{e}$ for $\alpha = 1\%$, and the three bottom plots describe the results for the same α considering the first PCA direction given by the model, which in this case refers to the direction $\mathbf{u} = (1, 0.0083)$ representing the principal axis of the elliptical distribution. Figure 3a shows the simulated data from the Gaussian model previously described, Figure 3b plots the copula space of the data (Gaussian copula) and the theoretical α-quantile (red), the lower (blue), and upper (black) level sets following the Equations 8–10. Finally, Figure 3c shows the corresponding results once the original space of the data is recovered through the inverse of the marginal survivals (all the colors have the same meaning as in Figure 3b. In a similar way, but for the

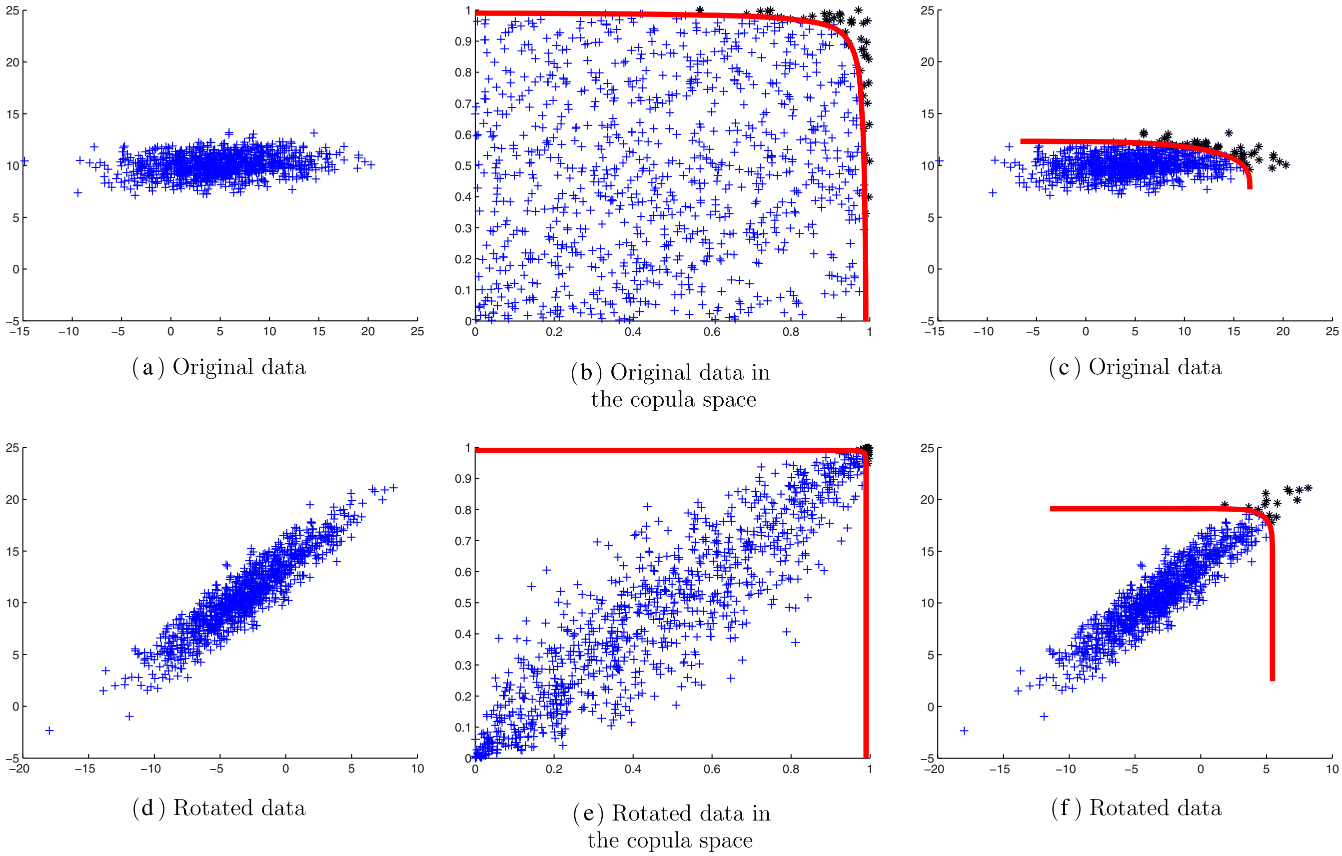


FIGURE 3 Gaussian model. Top: theoretical results in direction **e** and bottom: theoretical results in direction **e** for the rotation of the data given by the first PCA direction

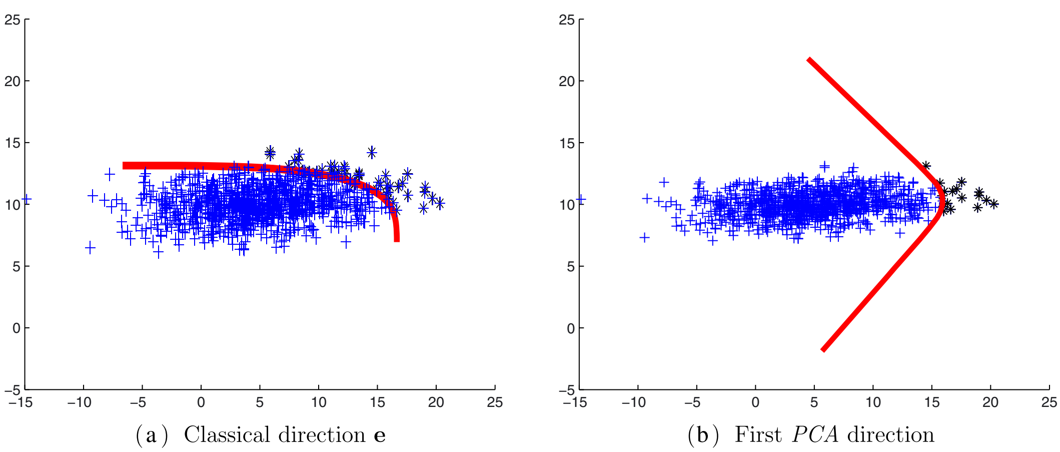


FIGURE 4 Gaussian model. Comparison of the identification of extremes in directions **e** and first PCA (black points)

first PCA direction, Figure 3d shows the pre-rotated data due to the given direction **u**, Figure 3e plots the copula space of the rotated data and the extreme detection in the copula space, and Figure 3f displays the extremes in the rotated space after applying the inverse of the rotated survival marginals. In order to compare the results in both directions, Figure 4a shows the extremes considering the direction **e** and Figure 4b shows extremes in the first PCA direction undoing the rotation $R_{\mathbf{u}}$. Graphically, the differences in the two directions are obvious and the extremes detected using the first PCA direction look more realistic because they are more congruent with the shape of the data.

4 | REAL CASE OF STUDY: FLOOD RISK AT A DAM

Salvadori et al. (2011) presented a three-dimensional (3D) model to describe floods occurring at Ceppo Morelli dam, located in Piedmont region, northwestern Italy. In that work, the following three variables are considered: maximum annual peak (Q in m^3/s), maximum annual volume (V in $10^6 m^3$), and initial water level (L in *m.a.s.l.*) in the reservoir before the flood event. The model that links all the variables was estimated using a copula approach to capture the correlation structure and generalized extreme value (GEV) distributions to describe the marginal behavior of Q and V, while a nonparametric normal kernel for L. However, for simplicity in the calculations, the simulation has been made using GEV for all the marginals. Then, the model was finally completed through Sklar's theorem and nested copula procedures, (see Appendix). Figure 5 shows the scatterplot and the 3D plot of the dataset used in Salvadori et al. (2011).

The water level in Ceppo Morelli dam, L, before the flood event is the result of the rules used by the manager at the dam. Generally, they try to keep the water level as high as they can, in order to maximize the energy production. In any case, it is independent from the flood occurrence, and consequently flood peak and flood volume. This is the motivation from the authors to use the specifications of the model that we provide in Table 1, there we can find the GEV distributions fitted for each variable with the corresponding parameters of locationϵ, scale β and shape γ and the copula model to recover the joint distribution of $\mathbf{X} = (Q,V,L)$. The pair (Q, V) has associated a Gumbel copula with positive dependence, the pairs, (Q, L) and (V, L) are modeled using the copula product. Finally, the flood copula model is given by C_{QVL} after a nesting procedure.

The authors have used the quantile surfaces associated to this model to extend the notion of return period to the multivariate setting. Assuming the previous model for the random vector $\mathbf{X}$ as appropriate, we now perform a Monte Carlo simulation with a large sample size to compare the multivariate extreme detection between the classical direction $\mathbf{e}$ (direction of the survival function) and the direction given by the first PCA.

Figure 6 presents the cross-sections of 1,000 observations simulated from the copula model and Figure 7 shows the corresponding scatterplot and 3D plot of the simulated data using the GEV distributions for the marginals and Sklar's theorem to reconstruct dam behavior. Then, once the sample is generated, the extreme identification is made following the nonparametric approach at level $\alpha = 1\%$ in the two directions previously mentioned. Figure 8a illustrates the analysis considering the classical direction $\mathbf{e}$, and Figure 8b presents the extremes obtained considering the first PCA direction $\mathbf{u} = (0.9988, 0.0486, -0.0014)$ given by the original dataset collected since 1937 to 1994 at Ceppo Morelli dam. Both plots draw the lower or non-risky 1%—level sets in blue, the directional 1%—quantiles in red and the upper or extreme 1%—level sets in black.

Note that the number of extremes identified in direction $\mathbf{e}$ is significantly greater than using the first PCA direction. Such a number of extremes seems excessive when a small value of α is considered. The improvements obtained in the first PCA direction are remarkable graphically.

To obtain more evidence of the advantages of the directional approach, we generate (Q, V, L) triplets as inputs to operate the reservoir routing, analyzing the stress and reliability of the dam after long-time horizons of 1,000 years

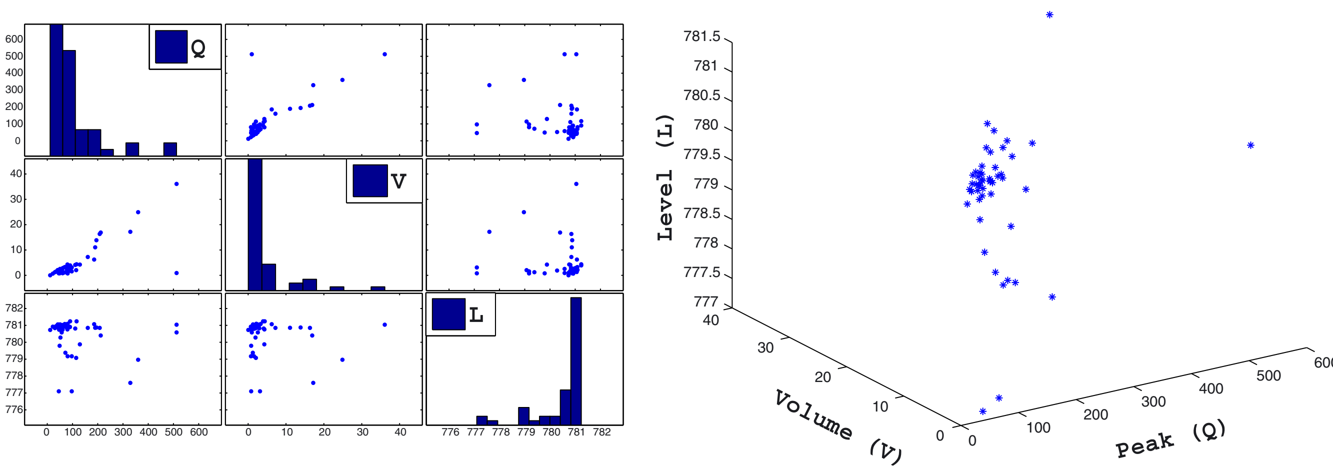


FIGURE 5 Flood risk at a dam: cross-sections and 3D plot of the dataset from Ceppo Morelli dam

TABLE 1 Model description given by Salvadori et al. (2011), changing to a GEV distribution the modelization of L

Q	$\boldsymbol{v_1 = F_Q(q)}$	**GEV with $\epsilon = 59.358 m^3 s, \beta = 36.203 m^3 s, \gamma = 0.368$**
V	$v_2 = F_V(v)$	GEV with $\epsilon = 1.7231 m^3, \beta = 1.5246 m^3, \gamma = 0.6149$
L	$v_3 = F_L(l)$	GEV with $\epsilon = 780.6261 m, \beta = 0.7623 m, \gamma = -1.5476$
QV	$C_{QV}(v_1, v_2)$	Gumbel copula with $\theta = 3.1378$
QVL	$C_{QVL} = v_3 C_{QV}(v_1, v_2)$	Nesting using copula product

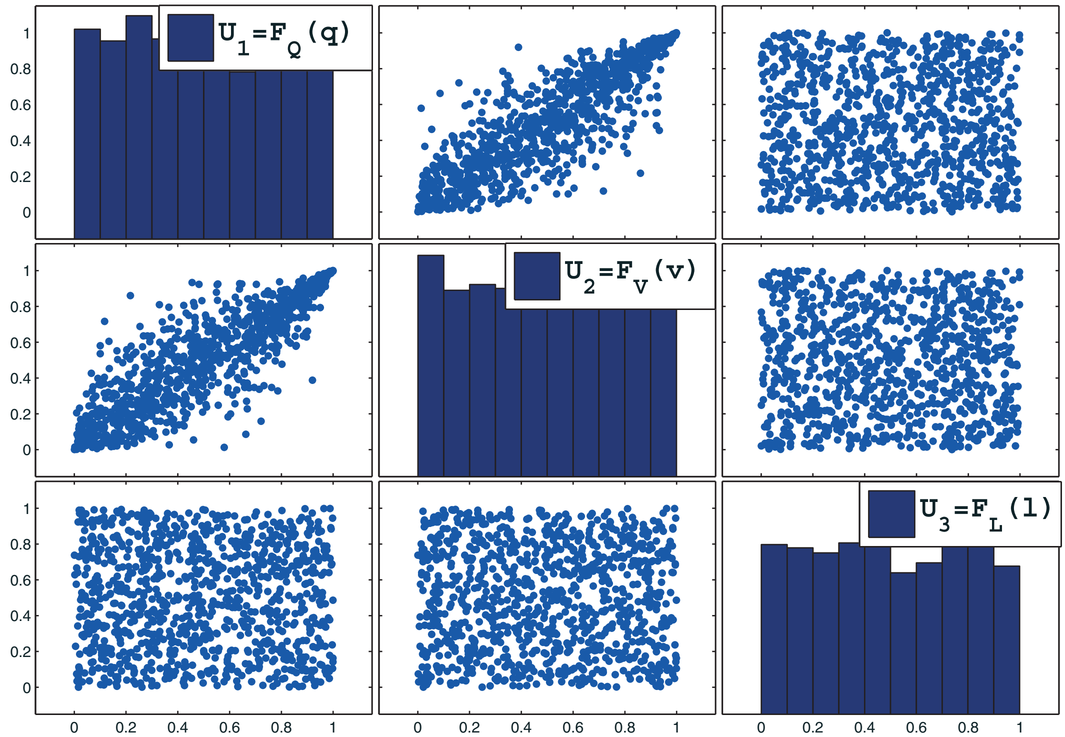


FIGURE 6 Flood risk at a dam: cross-sections of the simulated sample from the copula model

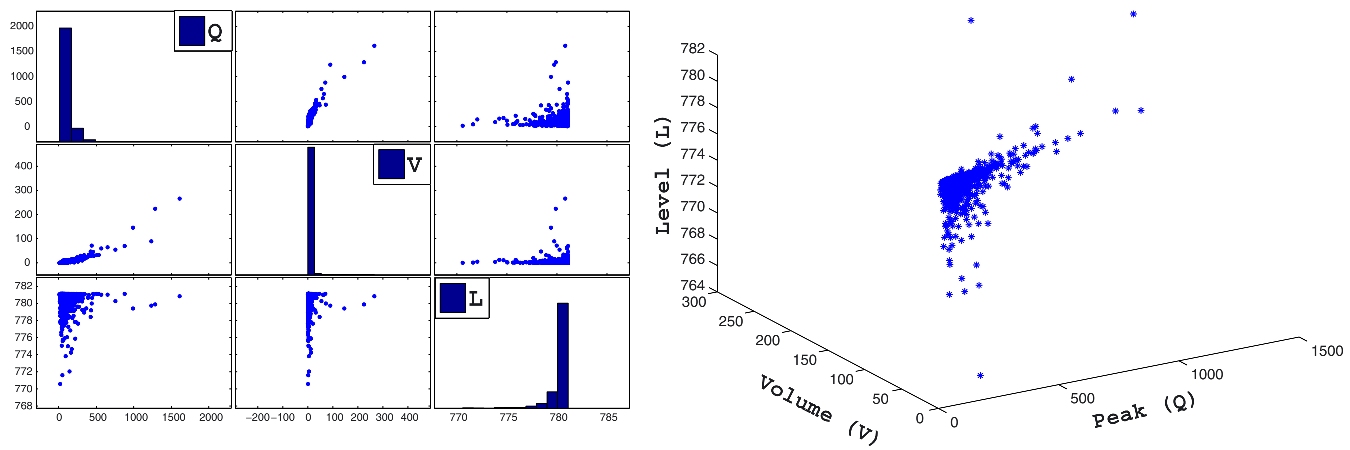
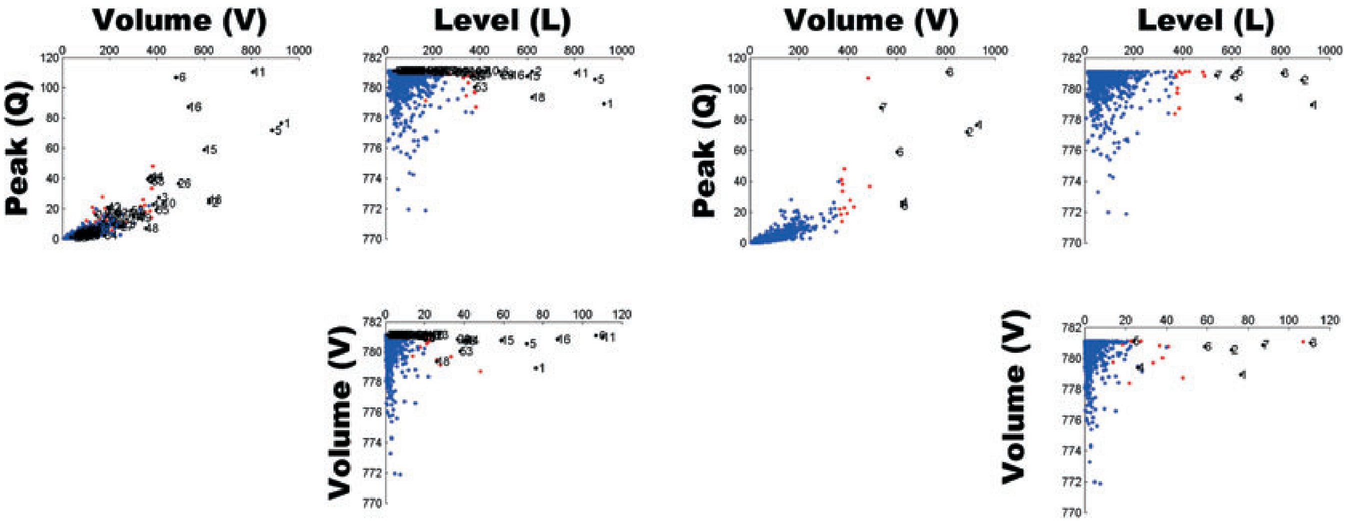


FIGURE 7 Flood risk at a dam: cross-sections and 3D plot of the simulated data for the model of floods

long. This was done similarly to De Michele et al. (2005). In particular, each couple (Q, V) is transformed into a triangular flood hydrogragh of volume V and maximum peak Q, with base time $T_b = 2V/Q$, time of rise $T_p = T_b/2.67$, and time of recession $T_r = 1.67T_p$, (see, e.g., Chow et al., 1988, p. 229). L is the water level in the dam at the beginning of the flood event. We operate the reservoir routing of flood hydrographs (see for details Bras, 1990, p. 475–478) considering as outlets only the uncontrolled spillways and checking if the spillways are capable of disposing the flood events without overtopping the dam crest.

Figure 9 presents two examples of the results after the simulation of dam behavior. In the images, it is possible to see the level of the dam spillway (781.50 m a.s.l.) that is the virtual line drawn between the maximum levels occurred (red points) and the initial levels (blue points). Also shown are the lines defining the maximum level (782.50 m a.s.l.) and the dam crest (784.00 m a.s.l.). Therefore, all the points between the maximum regulation level and the dam crest are considered as risky events, and those points above the dam crest are considered catastrophic events. We have done 100 simulations where each simulation spans 1,000 years and the conclusion in all of them is that the PCA directional analysis captures better the critical events, that is, the union of the sets of points given by the risky events and the extreme or catastrophic events. Meanwhile, the classical direction **e** identifies a huge number of such events.



(a) $\mathbf{u} = \mathbf{e}$ (b) $\mathbf{u}$ =First PCA

FIGURE 8 Flood risk at a dam: directional extremes at $\alpha = 1\%$, *(y-axis correspond to column variable names, x-axis correspond to row variable names)*

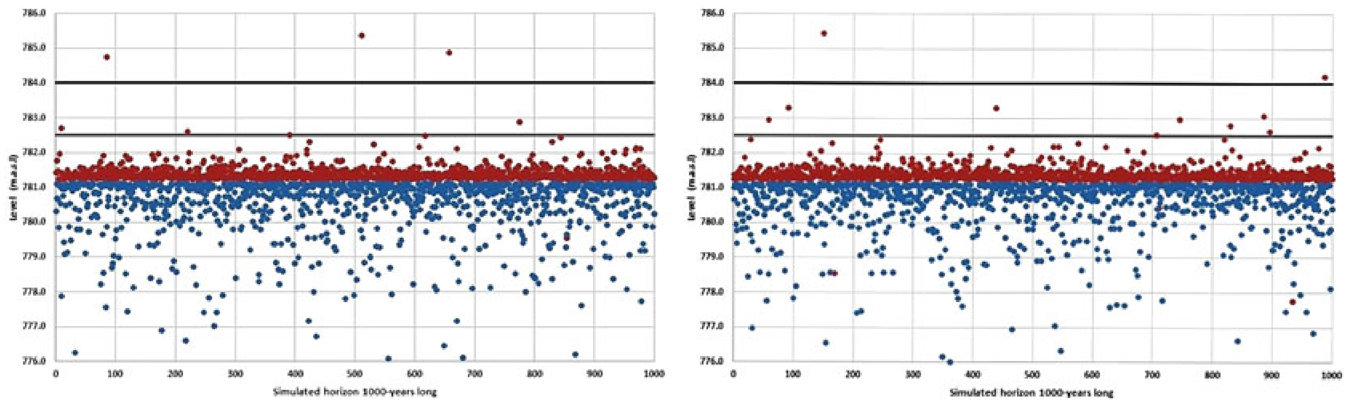


FIGURE 9 Flood risk at a dam: two (out of 100) examples of 1000-year final levels (red) given simulated initial levels (blue), and annual peaks and volumes

Table 2 summarizes average indexes over the 100 simulated samples analyzed in the two directions with $\alpha = 1\%$. Specifically, the table describes (a) the false positive ratio, which is the number of observations bad identified as critical over the total number of critical identifications; (b) the true positive ratio, which is the number of critical values correctly identified over the total number of real critical values from the dam routing simulation. (c) The extremes detection ratio, which is the number of observations identified as critical over the total number of observations. (d) The true extremes ratio, which is the number of real critical values over the total number of observations.

The table shows that both directions identify correctly all the critical values with a 100% of *true positives ratio*, but the analysis using the first PCA direction reduces significantly the *false positives ratio* of detection to 52.49%, compared to 91.74% in the classic direction. Also observe the small 1.77% exceeding in the *extreme detection ratio* given by the first PCA direction with respect to the *true extremes ratio* 0.83%, in comparison with the critical detection in the classic direction, which has a huge number of exceedances with a 10.17% *extremes detection ratio*.

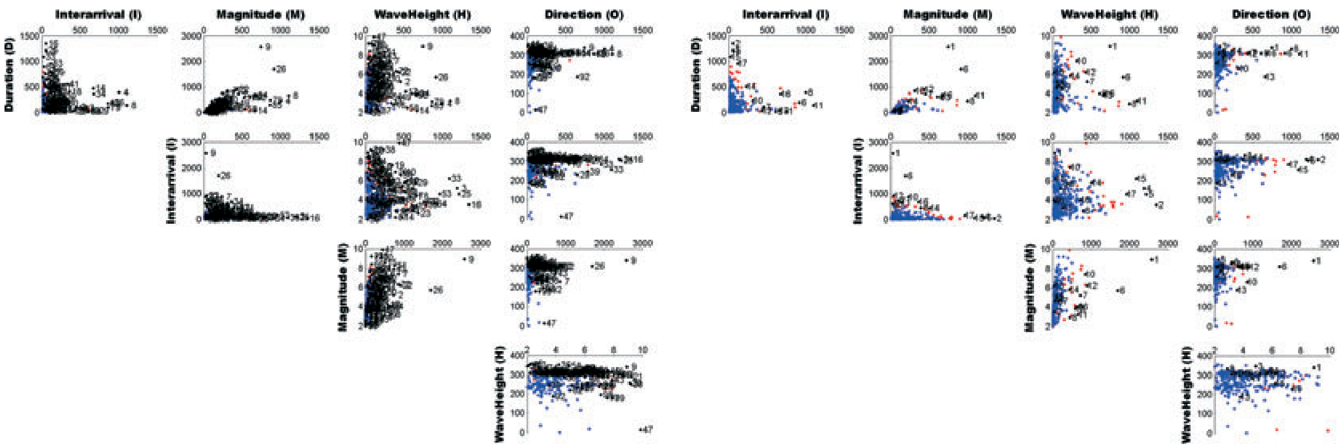
TABLE 2 Flood risk at a dam: results of the directional extreme analysis

	Classical direction	PCA direction
False positives ratio	91.74%	52.49%
True positives ratio	100%	100%
Extremes detection ratio	10.17%	1.77%
True extremes ratio	0.83%	0.83%

5 | REAL CASE OF STUDY: SEA STORMS

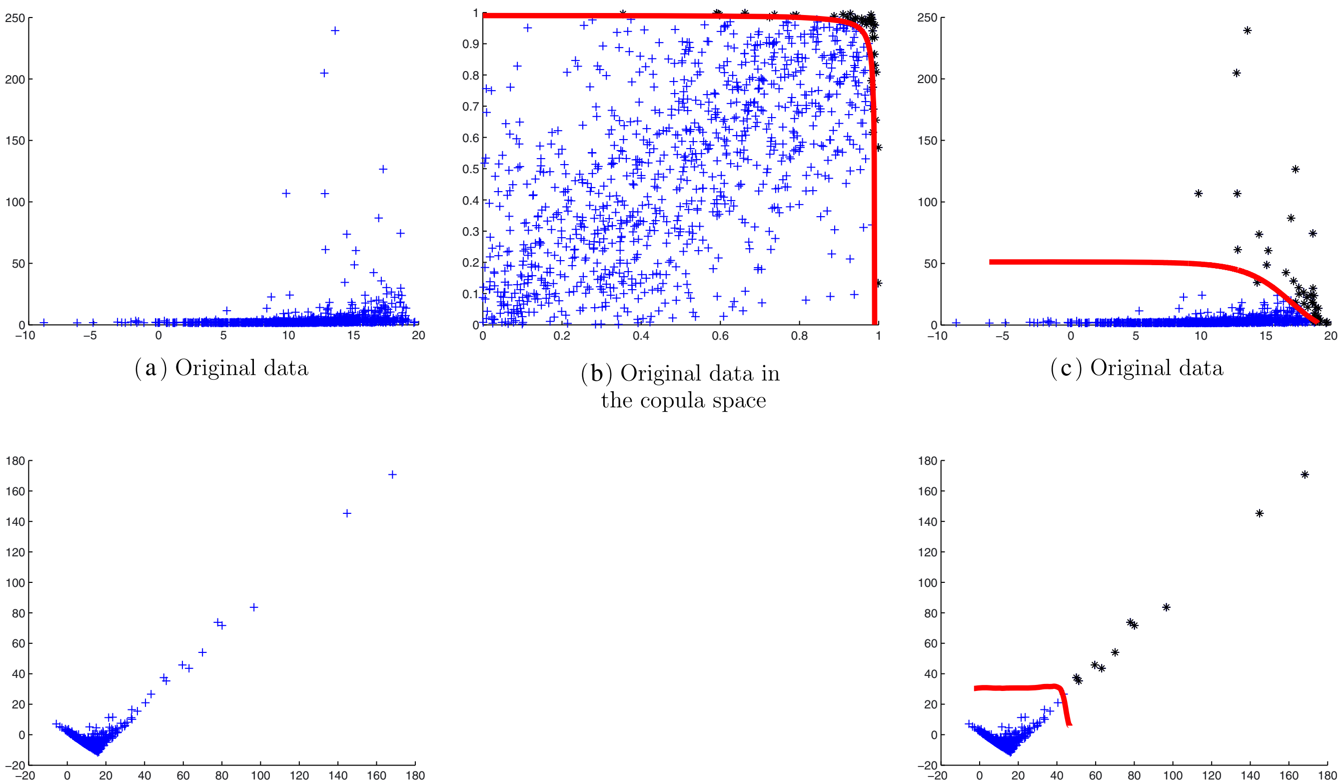
This case study is based on a dataset of sea storms that are described by five variables. This dataset has been studied in De Michele et al. (2007) and was collected by a wave buoy at Alghero (Sardinia, Italy) for a period of 12 years: from July 1, 1989 to October 31, 2001. The variables considered in the study are wave height (H in meters), storm duration (D in hours), storm magnitude (M in m * hr), storm direction (O in degrees), and storm inter-arrival time (I in hours), which records the period of calm between two successive storms. It is assumed that sea storms can be considered independent and homogeneous events. A sea storm occurs when the wave height crosses upwards of 2 m and ends when the wave height stays below 2 m for at least six consecutive hours. Specifically, the dataset counts a total of 415 sea storms during the considered period.

Our objective in this case study is to identify those risky events with our directional proposal in this five-dimensional setting, comparing the analysis in the two directions proposed in the previous case study, the classical direction **e** and the first PCA direction, which in this case is equal to (0.5181, − 0.0255, 0.8549, 0.0025, 0.0058). It indicates that



(a) $\mathbf{u} = \mathbf{e}$ (b) $\mathbf{u}$ =First PCA

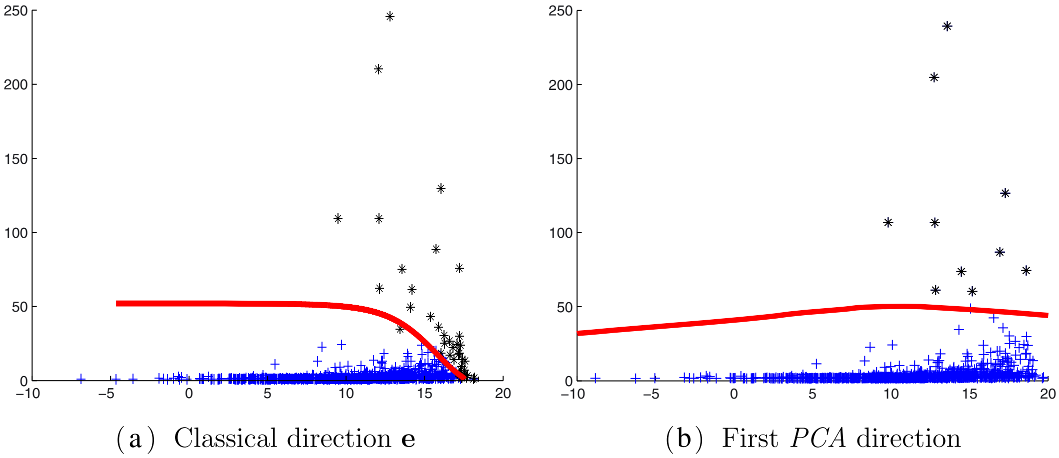
FIGURE 10 Sea storms: directional extremes for the case study sea storms at $\alpha = 1\%$, *(y-axis correspond to column variable names, x-axis correspond to row variable names)*



(a) Original data (b) Original data in the copula space (c) Original data

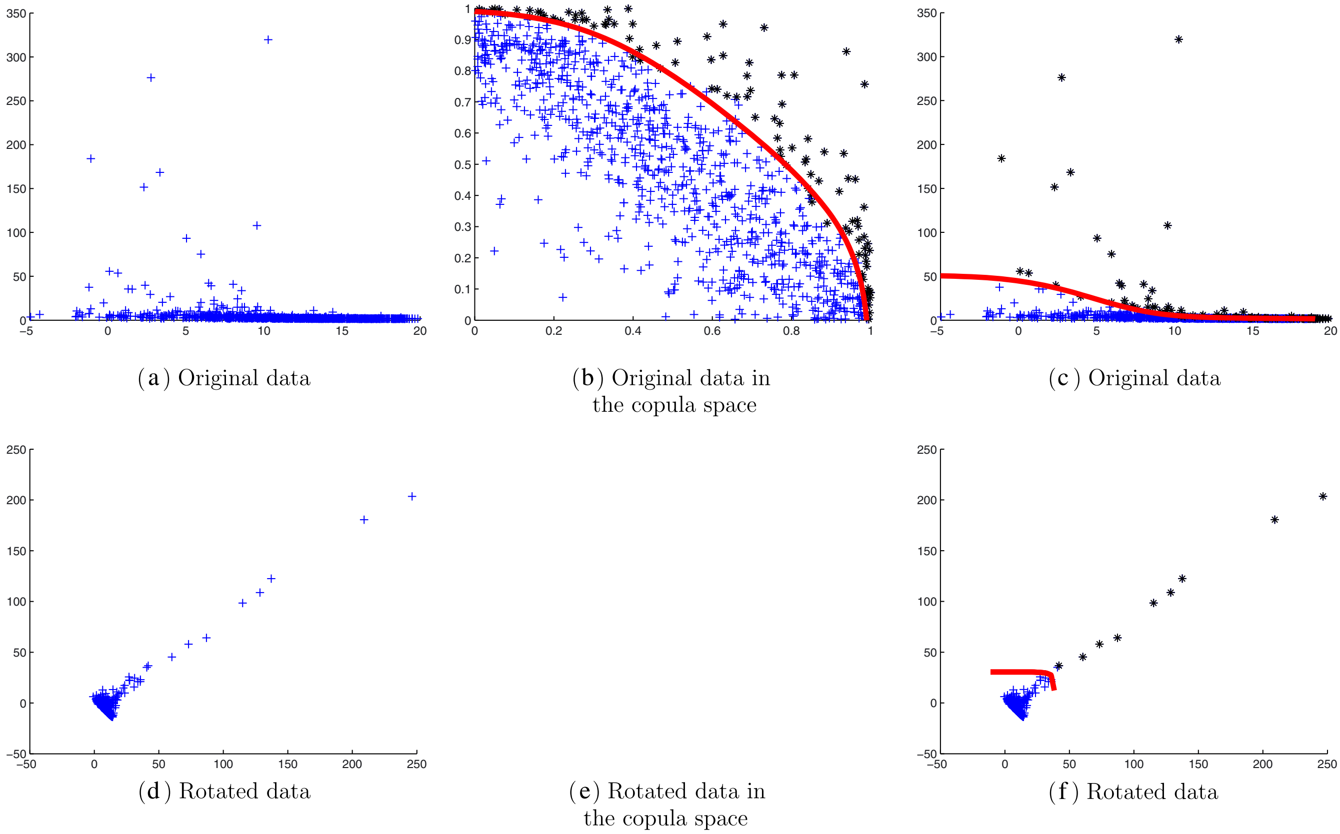
(d) Rotated data (e) Rotated data in the copula space (f) Rotated data

FIGURE 11 Frank copula model with positive dependence. Top: theoretical results in direction **e** and b ottom: nonparametric approach in direction **e** for the rotation of the data given by the first PCA direction



(a) Classical direction **e** (b) First *PCA* direction

FIGURE 12 Frank copula model with positive dependence. Comparison of the identification of extremes in directions **e** and first PCA (black points)



(a) Original data (b) Original data in the copula space (c) Original data (d) Rotated data (e) Rotated data in the copula space (f) Rotated data

FIGURE 13 Frank copula model with negative dependence. Top: theoretical results in direction **e** and bottom: nonparametric approach in direction **e** for the rotation of the data given by the first PCA direction

Magnitude and *Duration* are the more relevant variables while *inter-arrival time* has negative value due to the fact that lower values of the variable increases the risk of storms. Figure 10 shows the cross-sections of the sea storms dataset, where the left plot presents the identification of extremes in direction**e** for $\alpha = 1\%$(black points) and right plot shows the extremes associated with the first PCA direction for the same α (black points). In the same way as in the previous section, the visualization of extremes is more acceptable when the first PCA direction is used.

6 | WEAK POINTS OF THE DIRECTIONAL APPROACH USING COPULAS

The Gaussian copula is a toy example where the directional approach can be theoretically extended to the classical copula procedure. However, the usual fact is that the knowledge of the copula and the marginals associated to a random vector **X** does not imply knowing the copula and the marginals over a rotation of the random vector. Therefore, a disadvantage of the directional copula approach is that it increases the

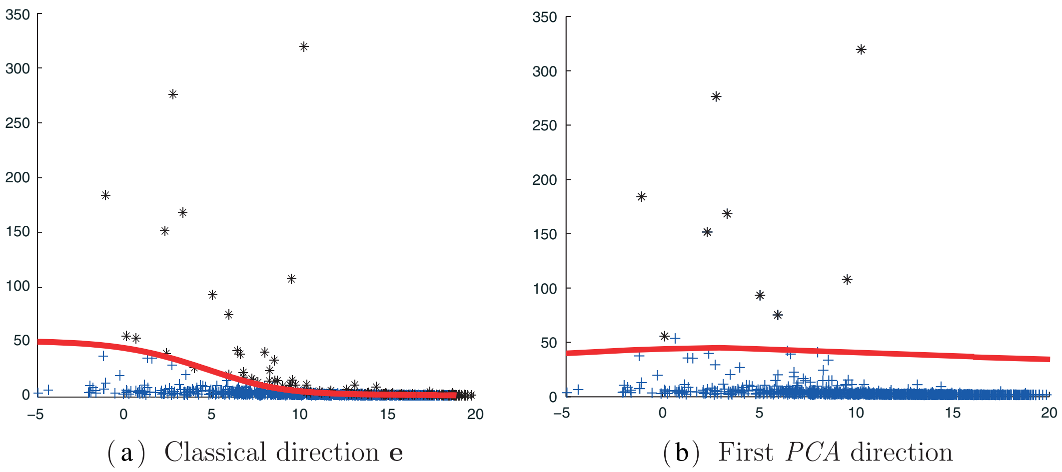


FIGURE 14 Frank copula model with negative dependence. Comparison of the identification of extremes in directions **e** and first PCA (black points)

computational cost when one decides to consider another direction of analysis different from **e**.

For example, let us consider a *Frank* copula and marginal distributions belonging to the GEV family (see Appendix for more details about the formulations of a Frank copula and GEV distributions). Firstly, we have assumed positive dependence in the model by setting a Frank survival copula with dependence parameter$\theta = 5$ and GEV marginals with parameters $\beta_1 = 5$, $\epsilon_1 = 10$, $\gamma_1 =- 1/2$, $\beta_2 = 1/2$, $\epsilon_2 = 2$ and $\gamma_2 = 1$. Figure 11a–c show the classical theoretical procedure used with copulas for $\alpha = 1\%$ and direction **e** with the same meaning as in Figure 3a–c. However, Figure 11d–f plot the analysis for the same α, but using the nonparametric approach in direction **e** over the pre-rotated data under the rotation $R_{\mathbf{u}}$ given by the first PCA direction $\mathbf{u} = (0.0639,0.998)$. Figure 11d shows the data in the rotated space, Figure 11e is empty due to the absence of theoretical evidence of the copula after the rotation of the data. Note that a possibility to fill the empty figure is to apply the nonparametric directional procedure presented in Section 2.2, but to the nonparametric copula of the rotated data (see Capéraà et al. (1997)), because Proposition 1 guarantees the theoretical equivalence. However, the directional approach has the advantage that the extremes can be obtained without considering the copula space of the rotated data as is shown in the identification presented in Figure 11f. To compare the detected extremes, Figure 12 displays in black those points considered as extremes in both directions, once the rotation of the data is undone in the case of the first PCA direction. The large number of points identified as extremes in the case of the classical direction **e** with $\alpha = 1\%$ can be observed, when many of these identified observations could be considered as regular observations. Meanwhile using the first PCA direction, the number of extremes is considerably reduced and they appear more reasonable. To conclude this section, we consider a model with negative dependence. In this case, α is again 1%, but the parameter of dependence in the Frank survival copula is $\theta = -8$ and we use the same GEV marginals as in the previous example. Figure 13 shows the outputs in the same framework as Figure 11, and Figure 14 shows the contrast between the classical and the first PCA directions for the detection of extremes. Once again, we can observe a better pattern of extreme recognition by considering the alternative direction of analysis $\mathbf{u} = (-0.0633,0.998)$, the first PCA direction.

7 | CONCLUSIONS

In this paper, we propose a directional multivariate extreme identification procedure based on the notion of directional multivariate quantile. A nonparametric implementation feasible in high dimensions is also presented. We have proposed a directional inclusion to the classical extreme detection procedure based on copulas. We have highlighted the advantages and disadvantages of the directional nonparametric approach and the directional copula procedure, and we have analyzed simulated and real scenarios where the advantages of using different directions to detect extremes is evident.

Specifically, PCA has been tested as a method to choose a suitable direction of analysis that offers a reasonable number of points identified as extremes, but more importantly, the locations of those identifications are more in the "*atypical zone*," if one looks at the cloud of observations and its shape. However, it is well known that the PCA is very sensitive to skewed data, data with heavy univariate tails or outliers. It would be very interesting to carry out a sensitivity analysis of the directional extreme detection method respect to these scenarios or to consider a more appropriate directions such as robust PCA (see, e.g., Candès et al., 2011). Anyway, if the dimension of the data is high, Donoho (2000) states that the usual PCA is better than any robust alternative due to the *blessing of dimensionality*.

ACKNOWLEDGEMENTS

The authors would like to thank the editor and referees for their valuable comments. This research was partially supported by a Spanish Ministry of Economy and Competitiveness grant ECO2015-66593-P. The third author wishes to thank to El Departamento Administrativo de Ciencia y Tecnología, COLCIENCIAS, for the financial support under the Convocatoria 656, 2014, Es Tiempo de Volver as well as to express his thanks to the hospitality of El Departamento de Ingeniería de la Organización, Universidad Nacional de Colombia sede Medellín and to the research group, Modelamiento y Análisis Energía Ambiente Economía (MAEAE) led by Professor Sergio Botero-Botero.

REFERENCES

Bras, R. (1990). *Hydrology: an introduction to hydrologic science*. Reading, Mass., USA: Addison Wesley Publishing Company.

Belzunce, F., Castaño, A., Olvera-Cervantes, A., & Suárez-Llorens, A. (2007). Quantile curves and dependence structure for bivariate distributions. *Computational Statistical & Data Analysis*, *51*, pp. 5112–5129.

Capéraà, P., Fougères, A., & Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, *84*(3), pp. 567–577.

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the ACM (JACM)*, *58*(3), p. 11.

Cascos, I., & Molchanov, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and stochastics*, *11*(3), pp. 373–397.

Chebana, F., & Ouarda, T. (2011a). Multivariate extreme value identification using depth functions. *Envirometrics*, *22*, pp. 441–455.

Chebana, F., & Ouarda, T. (2011b). Multivariate quantiles in hydrological frequency analysis. *Envirometrics*, *22*, pp. 63–78.

Chow, V., Maidment, D., & Mays, L. (1988). *Applied hydrology*. Singapore: McGraw-Hill.

Cousin, A., & Di Bernardino, E. (2013). On multivariate extensions of value-at-risk. *Journal of Multivariate Analysis*, *119*, pp. 32–46.

De Michele, C., & Salvadori, G. (2003). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research*, *108*(D2), p. 4067.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., & Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, *10*, pp. 50–57.

De Michele, C., Salvadori, G., Passoni, G., & Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering*, *54*, pp. 734–751.

De Michele, C., Salvadori, G., Vezzoli, R., & Pecora, S. (2013). Multivariate assessment of droughts: Frequency analysis and dynamic return period. *Water Resources Research*, *49*, pp. 6985–6994.

Di Bernardino, E., Fernández-Ponce, J., Palacios-Rodríguez, F., & Rodíguez-Griñolo, M. (2015). On multivariate extensions of the conditional value-at-risk measure. *Insurance: Mathematics and Statistics*, *61*, pp. 1–16.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pp. 1–32.

Embrechts, P., & Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, *97*, pp. 526–547.

Fernández-Ponce, J., & Suárez-Llorens, A. (2002). Central regions for bivariate distributions. *Austrian Journal of Statistics*, *31*, pp. 141–156.

Fraiman, R., & Pateiro-López, B. (2012). Quantiles for finite and infinite dimensional data. Rutcor Research Report 25.

Grimaldi, S., & Serinaldi, F. (2006). Asymmetric copulas in multivariate flood frequency analysis. *Advances in Water Resources*, *29*, pp. 1155–1167.

Grimaldi, S., & Serinaldi, F. (2013). Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, *17*, pp. 1281–1296.

Gupta, S., & Manohar, C. (2005). Multivariate extreme value distributions for random vibration applications. *Journal of engineering mechanics*, *131*(7), pp. 712–720.

Hallin, M., Paindaveine, D., & Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L1 optimization to halfspace depth. *Annals of Statistics*, *38*, pp. 635–669.

Heffernan, J., & Tawn, J. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *66*(3), pp. 497–530.

Hegerl, G., Zwiers, F., Kharin, S., & Stott, P. (2004). Detectability of anthropogenic changes in temperature and precipitation extremes. *Journal of Climate*, *17*, pp. 3683–3700.

Horn, R., & Johnson, C. (2013). *Matrix analysis* (2nd Ed.). Cambridge, UK: Cambridge University Press.

Kim, T., Valdés, J., & Yoo, C. (2003). Nonparametric approach for estimating return periods of droughts in arid regions. *Journal of Hydrologic Engineering*, *8*, pp. 237–246.

Kong, L., & Mizera, I. (2012). Quantile tomography: Using quantiles with multivariate data. *Statistica Sinica*, *22*(4), pp. 1589–1610.

Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions: Theory and applications*. London: Imperial College Press.

Laniado, H., Lillo, R., Pellerey, F., & Romo, J. (2012). Portfolio selection through an extremality stochastic order. *Insurance: Mathematics and Economics*, *51*, pp. 1–9.

Laniado, H., Lillo, R., & Romo, J. (2010). Multivariate extremality measure. Working Paper: Statistics and Econometrics Series 08, Universidad Carlos III de Madrid.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd edn). New York, USA: Springer-Verlag.

Salvadori, G. (2004). Bivariate return periods via 2-copulas. *Methodology & Statistics*, *1*, pp. 129–144.

Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research*, *40*(12), W12511 pp. 1–7.

Salvadori, G., De Michele, C., Kottegoda, N., & Rosso, R. (2007). *Extremes in nature: An approach using copulas*. Dordrecht, Germany: Springer-Verlag.

Salvadori, G., De Michele, C., & Durante, F. (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, *15*, pp. 3293–3305.

Salvadori, G., Durante, F., & De Michele, C. (2013). Multivariate return period calculation via survival functions. *Water Resources Research*, *49*, pp. 2308–2311.

Salvadori, G., & De Michele, C. (2015). Multivariate real-time assessment of droughts via copula-based multi-site hazard trajectories and fans. *Journal of Hydrology*, *526*, pp. 101–115.

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

Serfling, R. (2002). Generalized quantile processes based on multivariate depth function, with applications in nonparametric multivariate analysis. *Journal of Multivariate Analysis*, *83*, pp. 232–247.

Shiau, J. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment*, *17*, pp. 42–57.

Torres, R., Lillo, R., & Laniado, H. (2015). A directional multivariate value at risk. *Insurance: Mathematics and Economics*, *65*, pp. 111–123.

Zhang, L., & Singh, V. (2006). Bivariate flood frequency analysis using the copula method. *Journal of Hydrologic Engineering*, *11*, pp. 150–164.

How to cite this article: Torres R, De Michele C, Laniado H, Lillo RE. Directional multivariate extremes in environmental phenomena. *Environmetrics*. 2017;28:e2428. https://doi.org/10.1002/env.2428

APPENDIX

Theorem 1. (Sklar's theorem)

Sklar's theorem: Let F be a n-dimensional distribution function with marginals $F_1,\cdots,F_n$. Then, there exists a n-copula C such that for all $\mathbf{x} \in \bar{\mathbb{R}}^n$,

$$F(x_1,\cdots,x_n) = C(F_1(x_1),\cdots,F_n(x_n)). \quad \text{(A1)}$$

If $F_1,\ldots,F_n$ are continuous distribution functions, then C is unique. On the other hand, C is only defined on $Ran(F_1)\times\cdots\times Ran(F_n)$. Conversely, if C is a n-copula and $F_1,\ldots,F_n$ are distributions functions, then the function F defined by A1 is a n-dimensional distribution function with marginals $F_1,\ldots,F_n$.

Also, it is possible to express this result linking the joint survival function of the random vector through a survival copula, with the survival marginals by the equation,

$$\bar{F}(x_1,\cdots,x_n) = \bar{C}(\bar{F}_1(x_1),\cdots,\bar{F}_n(x_n)). \quad \text{(A2)}$$

In the literature, there are many classes or families of copulas and they have become a powerful tool for modeling practical situations in the multivariate framework where there is relevant joint information. For a deeper discussion of copula theory, we refer the reader to Nelsen (2006) and Salvadori et al. (2007). The formulation of the two families of bivariate copulas used in the paper is the following:

Gaussian Copula: The Gaussian copula is given by the expression

$$C(v_1,v_2) = \Phi_\rho\left(\Phi_1^{-1}(v_1),\Phi_2^{-1}(v_2)\right), \quad \text{(A3)}$$

where Φ_ρ is a bivariate standard Gaussian distribution with Pearson's correlation coefficient ρ, Φ_1^{-1} and Φ_2^{-1} the pseudo-inverse of Gaussian univariate distributions with parameters μ_1, σ_1^2 and μ_2, σ_2^2, respectively.

The Frank copula: An Archimedean copula with the following bivariate expression:

$$C_\theta(v_1,v_2) = -\frac{1}{\theta}ln\left(1+\frac{(e^{-\theta v_1}-1)(e^{-\theta v_2}-1)}{e^{-\theta}-1}\right), \quad \text{(A4)}$$

where $\theta \in \mathbb{R}/\{0\}$. We also summarize some concepts of the nested copula methods. The more elemental nesting procedure is through copula product, which describes the copula of independent random variables. Its expression in the bivariate setting is $C(v_1, v_2) = v_1 \times v_2$. Other more sophisticated methods are the Archimedean nesting procedures and the vine copula, also called pair-copula method.

Archimedean nesting procedures: Archimedean copulas are characterized by the representation,

$$C(v_1,\ldots,v_n) = \varphi^{-1}(\varphi(v_1)+\cdots+\varphi(v_n)),$$

where φ is the generator function, monotone in each component on $(0,\infty)$, and φ^{-1} is its the pseudo-inverse. This representation allows to define a nesting procedure (not necessarily in bivariate terms) through a hierarchical structure,

$$C_0(v_i,\cdot) = \varphi_0\left(\varphi_0^{-1}(v_i)+\varphi_0^{-1}(\cdot)\right), \quad \text{(A6)}$$

where the argument $(\cdot)$ in A6 is replaced by another Archimedean copula (again, not necessarily bivariate one), such as

$$C_{jk}(v_j,v_k) = \varphi_{jk}\left(\varphi_{jk}^{-1}(v_j)+\varphi_{jk}^{-1}(v_k)\right).$$

Then, the nested Archimedean copula is obtained as follows

$$\begin{aligned}C(v_i,v_j,v_k) &= C_0\left(v_i, C_{jk}(v_j,v_k)\right)\\ &= \varphi_0\left(\varphi_0^{-1}(v_i)+\varphi_0^{-1}\left(\varphi_{jk}\left(\varphi_{jk}^{-1}(v_j)+\varphi_{jk}^{-1}(v_k)\right)\right)\right).\end{aligned} \quad \text{(A8)}$$

Vine copulas: Under the hypothesis of Sklar's theorem, for every bivariate copula C, a bivariate copula density c exists given by

$$c(v_1,v_2) = \frac{\partial C(v_1,v_2)}{\partial v_1\partial v_2}.$$

This implies

- Joint density,

$$f(x_1,x_2) = c(F_1(x_1,F_2(v_2)) \times f_1(x_1) \times f_2(x_2), \quad \text{(A10)}$$

- Conditional density,

$$f(x_2|x_1) = c(F_1(x_1,F_2(v_2)) \times f_2(x_2). \quad \text{(A11)}$$

Then, a density $f(x_1,\ldots,x_n)$ can be represented as a product of a pair-copula densities and marginal densities. For instance, in dimension $n = 3$,$f(x_1, x_2, x_3) = f_{3|1,2}(x_3|x_1,x_2)\times f_{2|1}(x_2|x_1)\times f_1(x_1)$ and replacing properly A10 and A11, we get the representation. Vine method is more flexible than Archimedean nesting procedures, because we can select bivariate copulas from a wide range of parametric families.

Finally, we summarize a set of univariate distributions quite useful in extreme value analysis (see Kotz and Nadarajah (2000)), which is called GEV. Then a distribution belongs to the GEV family if it follows the structure:

GEV distributions:

$$F_X(x) = \begin{cases}\exp\left\{-\left[1+\gamma\left(\frac{x-\epsilon}{\beta}\right)\right]^{-\frac{1}{\gamma}}\right\} & \text{if}\gamma \neq 0,\\ \exp\left\{-\exp\left\{-\left(\frac{x-\epsilon}{\beta}\right)\right\}\right\} & \text{if}\gamma = 0,\end{cases} \quad \text{(A12)}$$

where $\beta > 0$, ϵ, $\gamma \in \mathbb{R}$ are the scale, location, and shape parameters, respectively. For a good introduction to univariate extreme analysis as well as the multivariate approach using copulas and the applications, we refer to Salvadori et al. (2007).