Received: 23 May 2018 | Revised: 11 April 2019 | Accepted: 12 April 2019

DOI: 10.1002/env.2575

RESEARCH ARTICLE

WILEY

Model-based inference of conditional extreme value distributions with hydrological applications

R. P. Towe[1] | J. A. Tawn[2] | R. Lamb[3,4] | C. G. Sherlock[2]

[1]School of Computing and Communications, Lancaster University, Lancaster, UK

[2]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

[3]JBA Trust, Skipton, UK

[4]Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence
R. P. Towe, School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK.
Email: r.towe@lancaster.ac.uk

Funding information
Innovate UK, Grant/Award Number: KTP009454; Engineering and Physical Sciences Research Council

Abstract

Multivariate extreme value models are used to estimate joint risk in a number of applications, with a particular focus on environmental fields ranging from climatology and hydrology to oceanography and seismic hazards. The semi-parametric conditional extreme value model of Heffernan and Tawn involving a multivariate regression provides the most suitable of current statistical models in terms of its flexibility to handle a range of extremal dependence classes. However, the standard inference for the joint distribution of the residuals of this model suffers from the curse of dimensionality because, in a d-dimensional application, it involves a $d-1$-dimensional nonparametric density estimator, which requires, for accuracy, a number points and commensurate effort that is exponential in d. Furthermore, it does not allow for any partially missing observations to be included, and a previous proposal to address this is extremely computationally intensive, making its use prohibitive if the proportion of missing data is nontrivial. We propose to replace the $d-1$-dimensional nonparametric density estimator with a model-based copula with univariate marginal densities estimated using kernel methods. This approach provides statistically and computationally efficient estimates whatever the dimension, d, or the degree of missing data. Evidence is presented to show that the benefits of this approach substantially outweigh potential misspecification errors. The methods are illustrated through the analysis of UK river flow data at a network of 46 sites and assessing the rarity of the 2015 floods in North West England.

KEYWORDS

copula, dependence modelling, missing values, multivariate extreme value theory, spatial flood risk assessment

1 | INTRODUCTION

Widespread flooding, such as the events of winter 2015/2016 in the UK, demonstrates the importance of understanding the likelihood of multiple locations experiencing extreme river flows. During these events, 43,000 homes were left without power and the estimated damages totalled £1.3–1.9 billion (Environment Agency, 2018). For flood risk management and

insurance purposes, we are interested in understanding the joint probability of events such as those observed in winter 2015/2016 and the likely nature of events that are even more extreme.

Let R_i represent the river flow at gauge i at a given time with corresponding location $\mathbf{s}_i$. Consider n independent and identically distributed realisations of the variable $\mathbf{R} = (R_1, \ldots, R_d)$, with this variable representing the joint behaviour of river flows at d gauges recorded over a given time period. From observations of these variables, we are interested in estimating marginal and joint probabilities. For example, for assessing the rarity of the December 5th, 2015, event in North West England, let v_i be the measured flood value in this event for the ith gauge in the region. Then, we need to know about marginal risk assessment at gauge i, through estimating the probabilities $\mathbb{P}(R_i > v_i)$, $i = 1, \ldots, d$, and for joint risk assessment, the probability $\mathbb{P}(\mathbf{R} \in A)$, where $A = \{\mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{R}^d : r_i > v_i, i = 1, \ldots, d\}$. More generally, we are interested in estimating the probability $\mathbb{P}(\mathbf{R} \in A)$, where the set $A \subset \mathbb{R}^d$ is extreme for at least one component, say, R_i of $\mathbf{R}$, so that for all $\mathbf{r} \in A$, $r_i > q_i$ with q_i being a high quantile for variable i.

For modelling spatial multivariate extremes data, the most widely used approach uses max-stable processes (Asadi, Davison, & Engelke, 2015; Davison, Padoan, & Ribatet, 2012). However, max-stable processes imply a strong form of extremal dependence, termed asymptotic dependence, in which the largest values at each site, over different events, can occur in the same particular flood event. The assumption of asymptotic dependence is probably reasonable for local-scale studies, such as in a mesoscale river basin. However, for larger scale studies, such as widespread studies across regions of the UK, this dependence assumption is highly restrictive as the largest values at different sites are unlikely to be occur in a single event.

Recent developments in statistical modelling of hydrological extremes allow us to now place such widespread events into a probabilistic framework (Keef, Svensson, & Tawn, 2009; Keef, Tawn, & Lamb, 2013; Lamb et al., 2010). Underpinning such methods is the theory of multivariate conditional extremes of Heffernan and Tawn (2004). This approach is able to handle the required mixture of both asymptotic dependence and asymptotic independence (a weaker form of extremal dependence than asymptotic dependence; see Section 2.2 for both extremal dependence structures that are identified in river flow data. Their conditional dependence model is formed through a semiparametric regression with parametric components describing variation in the means and the variances of the joint conditional distribution, and the joint distribution of the multivariate residuals being estimated empirically. The parametric components determine the core extremal dependence features, such as whether subsets of the variables are asymptotically dependent or asymptotically independent, and model across the range of possible dependence structures.

For hydrological applications, the method needs to be able to handle high dimensions (typically for 10–1, 000 sites); give realistic simulations of multivariate extreme events; enable the estimation of the risk of events, which are simultaneously rare at all and/or many sites; and allow covariates to be incorporated. Direct application of the Heffernan and Tawn (2004) method fails when dealing with any one of these issues, let alone being able to address all of these aspects in one analysis. The key problem with Heffernan and Tawn (2004) is that empirical multivariate residual modelling suffers from the curse of dimensionality, which along with its restriction to its reliance on the previously observed residuals, means that extrapolations to rarer events correspond to relocated and rescaled versions of past events. These events have poor coverage over the extremal regions of the sample space in high-dimensional studies and so lead to inefficient inference.

An additional complication that hydrological applications bring is that of missing data. Here, we assume the data to be missing at random. Data are likely to be missing when gauges are installed at different times or gauges become faulty. The Heffernan and Tawn (2004) approach, with its empirical residual distribution model, can only be applied for a d-dimensional problem when all components of the d-dimensional variable are observed. One approach would be to only analyse complete vector observations. This approach is highly restrictive, for example, when considering the whole of the UK river network, with $\sim 1,000$ gauges, which were considered as part of the National Flood Resilience Review (Tawn, Shooter, Towe, & Lamb, 2018), no concurrent observations are observed at all locations, and hence, lead to highly inefficient inference about extreme events. An alternative approach, proposed by Keef, Tawn, and Svensson (2009), is to replace these missing data, via infilling all the missing residuals with jointly generated multiple samples for the distribution of missing residuals given the observed residuals. This approach, which assumes a Gaussian copula for the joint distribution of missing and observed residuals only and treats fully observed variables empirically, is hugely computationally intensive when the amount of missing data is nontrivial. Critically, it fails to address all the other problems with the Heffernan and Tawn (2004) method that are described above.

Instead, in this paper, the full residual distribution is modelled semiparametrically: One-dimensional kernel-smoothed distribution functions capture the marginal behaviours of the observed residuals and a Gaussian copula is used for their dependence structure (Joe, 2014). Although this change in approach may at first seem rather small, it has major

implications for the applicability of the method, in that it addresses all the problematic issues of Heffernan and Tawn (2004), as well as handling large volumes of missing data efficiently. The primary reasons for its success are that it removes the problems of the curse of dimensionality and the choice of copula is flexible and parsimonious. Of course, there is a cost to be incurred by this modelling approach, as there is no theoretical motivation to support this assumption. However, here, we show plenty of evidence to suggest that the Gaussian copula is suitable for modelling the residual copula structure, mainly as it plays a secondary role in capturing the extremal dependence relative to the Heffernan and Tawn (2004) regression parameters. It is important to have strong diagnostic tools to assess departures from this model and a clear understanding of the effects of misspecification. This paper is the first that looks carefully at these aspects and finds that there are substantial improvements from the added flexibility and the more efficient use of the data on the estimation of probabilities of rare events.

The Heffernan and Tawn (2004) model is explained briefly in Section 2, with the extensions that we propose and their connections with previously adopted Gaussianity assumptions given in Section 3. The methodology for testing the validity of our proposed approach, including dealing with missing data, is detailed in Section 3.2. The comparisons with existing approaches to handle missing values are presented in Section 3.3. A generic simulation algorithm for the proposed conditional extreme value model and techniques for estimating probabilities of extreme joint events are given in Section 4. Then, examples of the proposed methodology are given in Sections 5 and 6 for simulated and observed data, respectively. The methodology is applied to study widespread flooding in North West England; the success of the different methods is compared through estimated probabilities of joint flood risk. The paper finishes with a discussion that considers ways in which the model can be made more parsimonious. Throughout this paper, all vector algebra is to be interpreted as being componentwise.

2 | THE HEFFERNAN AND TAWN MODEL

2.1 | Marginal model

The model for the marginal distributions of $\mathbf{R}$ has two components, separated using the predetermined threshold level u_i for variable $R_i (i = 1, \ldots, d)$. For a univariate random variable R_i, asymptotic theory considers the distribution of excesses over a threshold of u_i, scaled by some function $c(u_i) > 0$, that is, $\mathbb{P}\left(c(u_i)(R_i - u_i) \geq r | R_i > u_i\right)$, with $r > 0$; if this converges to a nondegenerate limit as u_i tends to the upper endpoint of the distribution of R_i; then, the limit distribution can only be the generalised Pareto distribution (GPD; Pickands, 1971). If it is assumed that this limit model holds exactly for some large-enough threshold u_i, it follows that

$$\mathbb{P}\left(R_i \geq r | R_i > u_i\right) = \left[1 + \xi_i (r - u_i)/\sigma_i\right]_+^{-1/\xi_i}, \quad \text{for } r > u_i, \tag{1}$$

with the scale parameter $\sigma_i > 0$, the shape parameter $\xi_i \in \mathbb{R}$, and the notation $[r]_+ = \max(r, 0)$ (Davison & Smith, 1990). Above the threshold, the GPD is adopted. For those points below the threshold u_i, there is no theoretical justification for any particular model choice, so instead, a kernel-smoothed empirical cumulative distribution function $\tilde{F}_i(r)$ of R_i is used. Thus,

$$F_i(r) = \begin{cases} \tilde{F}_i(r), & \text{for } r \leq u_i, \\ 1 - \phi_{u_i}\left[1 + \xi_i (r - u_i)/\sigma_i\right]_+^{-1/\xi_i}, & \text{for } r > u_i, \end{cases} \tag{2}$$

where $\phi_{u_i} = 1 - \tilde{F}_i(u_i)$ is the probability of an exceedance above the threshold u_i.

Estimating (σ_i, ξ_i) for each gauge separately can lead to inefficient inference as the spatial coherence and dependence of R_i over gauges suggest that (σ_i, ξ_i) and (σ_j, ξ_j) should be more similar when gauges i and j are closer together. Methods such as the covariate hierarchical/latent variable models that spatially smooth the GPD parameters have been developed by Cooley, Nychka, and Naveau (2007) and Cooley and Sain (2010). These models are ideal in the generation of marginal quantile maps as they share information from neighbouring sites to reduce any uncertainty in the estimation of quantiles. As the focus of this paper is on dependence modelling, we restrict ourselves to separate marginal fits but recognise that this typically can be improved upon.

To help estimate the dependence structure of the random variable $\mathbf{R}$, the data are transformed componentwise to a variable $\mathbf{Y} = (Y_1, \ldots, Y_d)$, with common Laplace margins, via the transform

$$Y_i = \begin{cases} \log\left\{2F_i(R_i)\right\}, & \text{for } F_i(R_i) < 0.5, \\ \log\left\{2\left[1 - F_i(R_i)\right]\right\}, & \text{for } F_i(R_i) \geq 0.5, \end{cases} \tag{3}$$

for $i = 1, \ldots, d$, where F_i is given in Equation (2). The transformation to Laplace margins means that $\mathbb{P}(Y_i > y+v | Y_i > v) = \mathbb{P}(Y_i < -(y+v) | Y_i < -v) = \exp(-y)$ for $y > 0$ and $v > 0$. Therefore, the marginal random variables of $\mathbf{Y}$ now have exponential upper and lower tails. This is a minor deviation from the Heffernan and Tawn (2004) approach, as they transform to Gumbel margins, but the use of Laplace margins unifies the handling of positive and negative dependence (Keef, Papastathopoulos, & Tawn, 2013).

2.2 | Introduction to extremal dependence properties

Extremal dependence properties need to be studied for all combinations of the variables as, unlike for multivariate Gaussian distribution, not all dependence is determined by the set of pairwise dependences. Therefore, consider $C \in 2^D$ with $|C| \geq 2$ and $D = (1, \ldots, d)$; then, define a measure of extremal dependence for variables $\{R_i; i \in C\}$ by

$$\chi_C = \lim_{p \to 1} \mathbb{P}(F_i(R_i) > p,\ i \in C) / (1-p) = \lim_{v \to \infty} \mathbb{P}(Y_i > v,\ i \in C)\, 2\exp(v),$$

where F_i is the marginal distribution function of R_i. If $\chi_C > 0 (\chi_C = 0)$, the variables in C are jointly asymptotically dependent (asymptotically independent). Here, $\chi_C > 0$ means that extreme events can occur simultaneously over all sites in C, whereas if $\chi_C = 0$, such events are impossible for the set of sites C. Clearly, for $B \subset C$, it is possible that $\chi_C = 0$ and $\chi_B > 0$, but if $\chi_B = 0$, then $\chi_C = 0$. Thus, it is possible to have asymptotic dependence locally but asymptotic independence over all sites.

If a copula model is used, the extremal dependence structure is predetermined by the choice of the copula before the model is fitted. For example, the class of bivariate extreme value distribution copulas has $\chi_{1,2} > 0$ (unless the variables are independent) and the class of multivariate Gaussian copula, with parameters $\{\rho_{i,j}; i \neq j \in D\}$, has $\chi_C = 0$ (unless $\rho_{ij} = 1$ for all $i, j \in C$ for all $C \in 2^D$ with $|C| \geq 2$). Other standard copula models typically can only handle one of the two classes of extremal dependence (Heffernan, 2000). As both of the extremal dependence classes are typically observed in extreme river flow data sets (see Keef et al., 2009; Tawn et al., 2018), a standard copula approach is almost never sufficiently flexible. Instead, like with univariate extremes, we appeal to asymptotic formulations to motivate a class of models specific to the tail region. These models allow any possible combination of feasible χ_C values for $C \in 2^D$.

2.3 | Extremal model for conditional dependence

After making the transformation given in Equation (3), the extremal behaviour of the joint tail of the random variable $\mathbf{Y}$ can now be determined. The approach models $\mathbf{Y}$ given that at least one of its elements is extreme, that is, given that $\max(\mathbf{Y}) > v$ for large v, where v is a dependence threshold.

First assume that $Y_1 > v$; then, the joint distribution of the $(d-1)$ remaining variables $\mathbf{Y}_{-1} = (Y_2, \ldots, Y_d)$ is modelled conditional on Y_1 being above v. The approach is motivated by the following asymptotic formulation studied by Heffernan and Tawn (2004) and Heffernan and Resnick (2007). The underlying idea is to see how $\mathbf{Y}_{-1}$ behaves as Y_1 gets large. In order to avoid nondegeneracy of the limiting conditional distribution of $\mathbf{Y}_{-1}$ as Y_1 tends to its upper end point, it is sensible to look for a componentwise location-scale transformation of $\mathbf{Y}_{-1}$ using functions of Y_1. As dependence between Y_1 and each component of $\mathbf{Y}_{-1}$ may be different, these location-scale transformations need to have the flexibility to be different for each component. This leads to the assumption that there exist normalising functions, $\mathbf{a}(.) : \mathbb{R} \to \mathbb{R}^{d-1}$ and $\mathbf{b}(.) > \mathbf{0} : \mathbb{R} \to \mathbb{R}_+^{d-1}$, such that the following limit probability holds for $y > 0$:

$$\lim_{v \to \infty} \mathbb{P}\left(\frac{\mathbf{Y}_{-1} - \mathbf{a}(Y_1)}{\mathbf{b}(Y_1)} \leq \mathbf{z},\quad Y_1 - v > y \mid Y_1 > v \right) = \exp(-y)\, G(\mathbf{z}), \tag{4}$$

where the joint distribution function $G(\mathbf{z})$ is nondegenerate in each margin and has no mass for any margin at infinity. The first term in the limit given in Equation (4) arises from the fact that Y_1 follows a standard Laplace distribution. The second term in the limit characterises the behaviour of $\mathbf{Y}_{-1} | Y_1 > v$ in terms of the limiting distribution function $G(\mathbf{z})$ along with the location $\mathbf{a}(.)$ and scale $\mathbf{b}(.)$ functions. It is assumed that the normalisations of the variables $\mathbf{Y}_{-1}$ and Y_1 are independent in the limit. This last assumption parallels that in classical point process models for multivariate extremes and regularly varying distributions (Coles & Tawn, 1991; Resnick, 2013), with radial and angular representations being assumed to be independent in the limit as the radial variable tends to infinity. Heffernan and Tawn (2004) show that formulation (4) holds for all standard copula models.

As a result of Equation (4), $G(\mathbf{z})$ is the limiting conditional distribution of

$$\mathbf{Z} = \frac{\mathbf{Y}_{-1} - \mathbf{a}(Y_1)}{\mathbf{b}(Y_1)}, \quad \text{given } Y_1 > v \quad \text{as } v \to \infty, \tag{5}$$

where $\mathbf{Z} \sim G$, and we call $\mathbf{Z}$ the residual of the conditional extreme value model. The result of the limits given in Equations (4) and (5) is that $\mathbf{Z}$ and Y_1 are independent given that $Y_1 > v$ in the limit as $v \to \infty$. Similar limits, with potentially different $\mathbf{a}(.)$, $\mathbf{b}(.)$, and G, hold for $\mathbf{Y}_{-j}|Y_j > v$ for any $j = 2, \ldots, d$. Joining together these d different conditionals, we have a model for the joint tail behaviour of $\mathbf{Y}$, when at least one component is large.

Under weak assumptions on the joint distribution of $\mathbf{Y}$, Heffernan and Resnick (2007) show that componentwise $\mathbf{a}(\cdot)$ and $\mathbf{b}(\cdot)$ must be regularly varying functions satisfying certain constraints, which for Laplace margins corresponds to each of the components of $\mathbf{a}$ (respectively $\mathbf{b}$) being regularly varying functions of index 1 (respectively less than 1). Heffernan and Tawn (2004), Keef et al. (2013), and Papastathopoulos and Tawn (2016) found that, although different classes of extremal dependence have different forms for $\mathbf{a}(.)$ and $\mathbf{b}(.)$, they all can be well approximated in a simple parametric form, which is the dominant power term of the regularly varying functions, that is, excluding the slowly varying function. For Laplace margins, this form simplifies to

$$\mathbf{a}(y) = \boldsymbol{\alpha} y \ \text{ and } \ \mathbf{b}(y) = y^{\boldsymbol{\beta}}, -\mathbf{1} \leq \boldsymbol{\alpha} \leq \mathbf{1} \ \text{ and } -\boldsymbol{\infty} < \boldsymbol{\beta} < \mathbf{1} \tag{6}$$

with $\boldsymbol{\alpha} = (\alpha_2, \ldots, \alpha_d)$ and $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_d)$. When $(\alpha_i, \beta_i) = (1, 0)$ for all $i \in C_{-1} \subset D \setminus \{1\}$, then if $C = C_{-1} \cup \{1\}$, it follows that $\chi_C > 0$ and the variables indexed by C are asymptotically dependent. Similarly, if $\alpha_i < 1$ for any $i \in C_{-1}$, then $\chi_C = 0$ and the variables indexed by C are asymptotically independent. Thus, $\boldsymbol{\alpha}$ controls the collections of variables, which are asymptotically dependent with variable Y_1. It is clear therefore that this model captures all the possible sets of asymptotically independent and dependent variables as set out in Section 2.2. This unification of the parametric forms for all dependence classes enables flexible efficient statistical modelling unlike with standard parametric copula modelling.

Heffernan and Tawn (2004) assume that limit (4) holds exactly above a sufficiently large dependence threshold v and that the normalising functions are given by the parametric forms (6). This leads to the following model:

$$\mathbf{Y}_{-1} = \boldsymbol{\alpha} Y_1 + Y_1^{\boldsymbol{\beta}} \mathbf{Z}, \quad \text{for } \ Y_1 > v, \tag{7}$$

where $-\mathbf{1} \leq \boldsymbol{\alpha} \leq \mathbf{1}$ and $-\boldsymbol{\infty} < \boldsymbol{\beta} < \mathbf{1}$ and $\mathbf{Z} \sim G$, where G is a marginally nondegenerate distribution function and the $\mathbf{Z}$ is independent of Y_1. There is no general theoretically justified family of distributions G for the multivariate residuals $\mathbf{Z}$, so Heffernan and Tawn (2004) assumed that $\mathbf{Z}$ has marginal finite means and variances $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$, respectively, where $\boldsymbol{\mu} = (\mu_2, \ldots, \mu_d)$ and $\boldsymbol{\sigma} = (\sigma_2, \ldots, \sigma_d)$. As a result, the following expressions for the conditional expectation and variance of $Y_i|Y_1 = y$ can be determined for $y > v$ and $i = 2, \ldots, d$:

$$\begin{aligned} \mathbb{E}[Y_i|Y_1 = y] &= \alpha_i y + y^{\beta_i} \mu_i, \\ \mathbb{V}\text{ar}[Y_i|Y_1 = y] &= (y^{\beta_i} \sigma_i)^2. \end{aligned} \tag{8}$$

Heffernan and Tawn (2004) model the joint distribution of $\mathbf{Z}$ nonparametrically using an empirical joint distribution, with the specific form of this model presented in Section 2.4.

So far, we have presented the behaviour of $\mathbf{Y}|Y_1 > v$ for large v, or equivalently $\mathbf{Y}|Y_i > v$ for an arbitrary $i \in D$, but we really want the behaviour of $\mathbf{Y}|\max(\mathbf{Y}) > v$. This conditional behaviour can be derived from the set of distributions of $\mathbf{Y}|Y_i > v$ for $i \in D$. As the conditioning variable changes to Y_i, the norming functions $\mathbf{a}(\cdot)$ and $\mathbf{b}(\cdot)$ and the limiting distributions G all change with i. We can piece together results from a series of models of the form above. A limitation of this set of models is that self-consistency is not ensured unless specific constraints on these different normalisation and distribution functions are made. A lack of self-consistency may lead to inconsistencies when joint exceedance probabilities are estimated, with the results depending on the choice of a conditioning variable. Heffernan and Tawn (2004) review ways of avoiding this problem with partitioning the sample space, and Liu and Tawn (2014) discuss a number of approaches to reduce this problem. In this paper we will, however, largely look at the individual conditional distributions, that is, $\mathbf{Y}|Y_i > v$ for $i \in D$ and not overall joint tail inference.

2.4 | Inference

The dependence parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the Heffernan and Tawn (2004) model are estimated through pairwise maximum pseudolikelihood for the n_v pairs with $Y_1 > v$. The pseudolikelihood $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ for inference for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is constructed under the temporary working assumption that

$$G(\mathbf{z}) = \prod_{i=1}^{d} \Phi\left(\frac{z_i - \mu_i}{\sigma_i}\right),$$

that is, independent Gaussian distributions. Hence,

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \propto \prod_{i=2}^{d} \prod_{j=1}^{n_v} \frac{1}{y_{ij}^{\beta_i} \sigma_i} \exp\left\{ -\frac{\left(y_{ij} - \alpha_i y_{1j} - \mu_i y_{1j}^{\beta_i} \right)^2}{2\left(y_{ij}^{\beta_i} \sigma_i \right)^2} \right\}; \tag{9}$$

here, $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $-1 \leq \alpha_i \leq 1$, and $-\infty < \beta_i < 1$, where y_{ij} denotes component $i = 1, \ldots, d$ for the jth exceedance of v by Y_1. The maximum pseudolikelihood estimates $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_2, \ldots, \hat{\alpha}_d)$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_2, \ldots, \hat{\beta}_d)$ are found, by jointly maximising Equation (9), with $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

Now, we present the Heffernan and Tawn (2004) modelling and inference for the joint distribution of the residuals. This is where our inference approach outlined in Section 3 differs. Firstly, the temporary working assumption of independent Gaussianity of the components of $\mathbf{Z}$ used in the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is discarded. With the fitted values of these parameters, there are n_v observed exceedances of v by Y_1, denoted by $y_{1j}, j = 1, \ldots, n_v$. The associated vectors of residuals are $\{\mathbf{z}^{(j)}, j = 1, \ldots, n_v\}$, where $\mathbf{z}^{(j)} = \left(z_{2j}, \ldots, z_{dj}\right)$ with its component associated with Y_i given by

$$z_{ij} = \frac{y_{ij} - \hat{\alpha}_i y_{1j}}{y_{1j}^{\hat{\beta}_i}}, \quad \text{for } y_{1j} > v, \quad \text{where } j = 1, \ldots, n_v, \quad i = 2, \ldots, d. \tag{10}$$

Heffernan and Tawn (2004) estimate the joint distribution function G through the empirical joint distribution function of these residuals $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(n_v)}$. Extrapolation from the model comes from (7), with larger events arising when Y_1 is larger than the observed events. Due to the independence of Y_1 and $\mathbf{Z}$, for $Y_1 > u$, all simulated events are of the form $\mathbf{Y}_{-1} = (\boldsymbol{\alpha} y + y^{\boldsymbol{\beta}} \mathbf{z}^{(j)})$, for $y > v$ and $j = 1, \ldots, n_v$. This leads to simulated events on Laplace margins being shifted and rescaled versions of past events. Thus, the extrapolation is restricted to n_v sets of 1-dimensional extrapolations, which clearly do not span the required extrapolation space, particularly when n_v is small relative to d.

3 | NEW MODELLING FEATURES

3.1 | Semiparametric inference for G

We model the joint residual distribution G by a semiparametric joint distribution model with 1-dimensional kernel-smoothed marginal distribution functions and a Gaussian copula (Joe, 2014). Let $\hat{G}_i(z)$ be the kernel-smoothed distribution function for observations of Z_i; then,

$$\hat{G}_i(z) = \frac{1}{n_v} \sum_{j=1}^{n_v} \Phi\left(\frac{z - z_{ij}}{h_i} \right), \quad \text{where } i = 2, \ldots, d \tag{11}$$

with $h_i > 0$, being the bandwidth (Silverman, 1986) and z_{ij}, given by expression (10), corresponding to the ith component of the jth residual vector when $Y_1 > v$. The kernel-smoothed distribution provides flexibility as it allows smooth interpolation between observed data points as well as some limited extrapolation, and critically, it leads to a nondeterministic extrapolation of past events. Our model for the joint distribution function G is then

$$G(\mathbf{z}) = \Phi_{d-1}\left(\Phi^{-1} \hat{G}_i(z_i), i = 2, \ldots, d; \Sigma \right), \tag{12}$$

where $\mathbf{z} = (z_2, \ldots, z_d)$, and Φ and $\Phi_{d-1}(., \Sigma)$ are the cumulative distribution functions of a standard univariate Gaussian and a standard $(d-1)$-dimensional Gaussian with correlation matrix Σ with (i,j)th element ρ_{ij} with $i \neq j = 2, \ldots, d$. The use of the componentwise probability integral transformation gives

$$\mathbf{Z}^N = \left(Z_2^N, \ldots, Z_d^N \right) = \left\{ \Phi^{-1}\left(\hat{G}_i(Z_i) \right), \quad i = 2, \ldots, d \right\}.$$

Our copula assumption (12) then corresponds to $\mathbf{Z}^N$ being a $(d-1)$-dimensional standard Gaussian distribution with the correlation matrix Σ giving a relationship between the residuals, which is fully determined by its bivariate marginals. Furthermore, the Gaussian copula is chosen because it is computationally feasible in high dimensions and is closed to marginalisation and conditioning. The Gaussian copula has an asymptotically independent extremal dependence structure (Ledford & Tawn, 1996); however, this property is not restrictive as the joint tails of $\mathbf{Z}$ are not vital for determining the joint tails of $\mathbf{Y}_{-1}|Y_1$ as that distribution is a mixture over Y_1, for $Y_1 > v$, so even independent $\mathbf{Z}$ can lead to $\mathbf{Y}_{-1}|Y_1 > v$ being asymptotically dependent. See Section 3.3 for details of how to estimate Σ.

Unlike the standard Heffernan and Tawn (2004) approach, the residuals are no longer restricted to the sample as the kernel smoothing allows both interpolation and limited extrapolation of the residuals, and the Gaussian copula enables new combinations of $\mathbf{Z}$ to occur.

3.2 | Tests of the Gaussian copula assumption

A formal test to check whether the copula is fairly close to being Gaussian is required to avoid the residual joint model being applied inappropriately. For assessing pairwise dependence, visual inspections of the residual distribution is sometimes sufficient; however, this comparison fails to assess the importance of higher order dependence. In order to assess the full dependence structure, we adopt the methods of Bortot, Coles, and Tawn (2000) for assessing the Gaussian copula in joint tail regions.

Consider the set of independent and identically distributed observations of $\mathbf{Z}^N$, which follows a $(d-1)$-dimensional multivariate Gaussian distribution with correlation matrix Σ. The square of the Mahalanobis distance is defined by

$$T = \mathbf{Z}^N \Sigma^{-1} \left(\mathbf{Z}^N\right)'. \tag{13}$$

Then, T follows a χ^2_{d-1} distribution with $\mathrm{E}[T] = d-1$ and $\mathrm{Var}[T] = 2(d-1)$. In reality, there are missing (at random) values in the observations of the residual variable $\mathbf{Z}^N$ and the percentage of missing values is not consistent across locations. Therefore, the test statistic T has to be adapted to account for the different record lengths of data. First, let $\mathbf{1}_i = (1_{2,i}, \ldots, 1_{d,i})$ be a $(d-1)$-dimensional vector with $1_{j,i} = 0 (1_{j,i} = 1)$ if $Z^N_{j,i}$ is missing (observed), respectively. Consider a particular vector $\mathbf{Z}^N_i$ with missing vector $\mathbf{1}_i$, where d_i elements of $\mathbf{Z}^N_i$ are observed, that is, $d_i = \mathrm{sum}(\mathbf{1}_i)$ with $0 \leq d_i \leq d-1$, then $\mathbf{Z}^N_i \sim \mathrm{MVN}(0, \Sigma_i)$, where $\Sigma_i = \mathbf{1}_i \Sigma \mathbf{1}_i'$ with $\dim(\Sigma_i) = d_i \times d_i$. By defining

$$T_i = \mathbf{Z}^N_i \Sigma^{-1}_i \left(\mathbf{Z}^N_i\right)',$$

it follows that T_i has a $\chi^2_{d_i}$ distribution with $\mathrm{E}[\chi^2_{d_i}] = d_i$ and $\mathrm{Var}[\chi^2_{d_i}] = 2d_i$. We can define the adapted test statistic of Gaussianity to be

$$T^* = \frac{1}{\sqrt{n_v}} \sum_{i=1}^{n_v} \frac{T_i - d_i}{\sqrt{2d_i}}, \tag{14}$$

where n_v is the number of observations of $\mathbf{Z}^N$. If a particularly large value of T^* is observed, then there is a deviation away from the assumption of multivariate normality. The sampling distribution of T^* under the null hypothesis for a given pattern of missing data is easily derived by Monte Carlo methods but has been constructed to have $\mathbb{E}(T^*) = 0$ and $\mathbb{V}\mathrm{ar}(T^*) = 1$ under the null hypothesis of the Gaussian copula whatever the missingness pattern, provided $\min(d_1, \ldots, d_{n_v}) \geq 1$ and Σ is known.

3.3 | Handling missing values

Heffernan and Tawn (2004) only consider vectors of complete observations so with any missing data, the method will be highly inefficient. The data-usage efficiency can be defined as $100 \sum_{i=1}^{n} \mathbb{1}(d_i = d-1)/n$ with $\mathbb{1}$ being the indicator function and d_i as defined in Section 3.2. Keef et al. (2009) developed a strategy to replace each missing variable by a sample of m replicates generated from a $d-1-d_i$-dimensional Gaussian approximation for the conditional distribution of the missing $\mathbf{Z}^N_i$ given the observed $\mathbf{Z}^N_i$ elements for all i with $d_i < d-1$. This approach has major computational problems when more than a small number of missing values are present as it requires $w \sum_{i=1}^{n_v} (d-1-d_i)$ simulations, where w needs to be reasonably large to remove Monte Carlo noise, for example, $w \in (100, 1000)$. This approach is subsequently referred to as the infill approach.

We propose using our Gaussian copula model to give a statistically and computationally efficient approach. Equation (12) is used to transform the $\mathbf{Z}$ variables, on their original margins, to $\mathbf{Z}^N$ on Gaussian margins. Concurrent pairs

of observations of $\mathbf{Z}^N$ are used to estimate the correlation parameters provided that a datum exists for a given Z_i^N and Z_j^N pair. This gives the following estimated correlation matrix $\hat{\Sigma}$, with (i,j)th entry of $\hat{\rho}_{i,j}$ being

$$\hat{\rho}_{ij} = \frac{\sum_{k=1}^{n_v} 1_{i,k}1_{j,k}(z_{i,k} - \bar{z}_i)(z_{j,k} - \bar{z}_j)}{\sqrt{\sum_{k=1}^{n_v} 1_{i,k}1_{j,k}(z_{i,k} - \bar{z}_i)^2 \sum_{k=1}^{n_v} 1_{i,k}1_{j,k}(z_{j,k} - \bar{z}_j)^2}},$$

with $\bar{z}_i = \sum_{k=1}^{n_v} 1_{i,k}z_{i,k} / \sum_{k=1}^{n_v} 1_{i,k}$ and similarly for $\bar{z}_j$. When there are no concurrent data for the pair (i,j), that is, $\sum_{k=1}^{n_v} 1_{i,k}1_{j,k} = 0$, then a covariate model or prior information can be used to give an estimate. As the correlation matrix is estimated for nonoverlapping data sets, there is a possibility that the resulting estimated correlation matrix Σ is not positive semidefinite. However, there are eigen-decomposition methods that can solve this problem by giving the nearest positive-definite matrix $\tilde{\Sigma}$ to $\hat{\Sigma}$ that maintains unit diagonals (Franklin, 2012).

3.4 | Connections with other models

There have been some Gaussian assumptions made in other work using the Heffernan and Tawn (2004) model, but that differs from what is proposed here. In the original Heffernan and Tawn (2004) paper for the inference of the regression parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, a pseudolikelihood is constructed with independent Gaussian residuals, but for subsequent inference on $\mathbf{Z}$, this assumption was then dropped. Thus, there is in fact no overlap with the approach in Heffernan and Tawn (2004). Motivated by early findings in this paper, in a spatial setting, Tawn et al. (2018) assume that $\mathbf{Z}$ is a realisation from a Gaussian process at a set of sites, so there they make an assumption of marginal Gaussianity for $\mathbf{Z}$ in addition to the Gaussian copula we assume. In that paper, there is no discussion on how to assess the Gaussian copula model or why it may be appropriate. This is what this paper does.

There is a question of whether our model is reasonable at all. In fact, $\mathbf{Z}$ is multivariate Gaussian for two very widely used copulas. Specifically, it arises for the asymptotic dependent multivariate extreme value Hüsler–Reiss distribution copula (Hüsler & Reiss, 1989) with $(\alpha_i, \beta_i) = (1, 0)$ for all $i = 2, \ldots, d$ (see Engelke, Malinowski, Kabluchko, & Schlather, 2015), and for the asymptotically independent Gaussian copula with $(\alpha_i, \beta_i) = (\rho_{1i}^2, 1/2)$ for all $i = 2, \ldots, d$ (see Heffernan & Tawn, 2004).

4 | SIMULATION ALGORITHM AND JOINT EVENT ESTIMATION

4.1 | Simulation of extreme events

The procedure to simulate from our model for $\mathbf{R}$, assuming that its first component is large, is an adaptation of the algorithm in Heffernan and Tawn (2004) and Jonathan, Ewans, and Randell (2013). Firstly, we define $q_{i,p}$ as the pth quantile of R_i; thus, $F_i(q_{i,p}) = p$. The aim is then to simulate $\mathbf{R}|R_1 > q_{1,p}$. On Laplace margins, this corresponds to simulating $\mathbf{Y}|Y_1 > v_p$, where $v_p = \log[2(1-p)]$. Here, we assume p is sufficiently large so that $v_p > v$, where v is the dependence threshold described in Section 2.3.

The steps of the simulation procedure are outlined as follows.

1. Simulate $\mathbf{Z}^N$ from a standard $(d-1)$-dimensional Gaussian distribution with correlation matrix $\hat{\Sigma}$, as defined in (12).
2. Transform $\mathbf{Z}^N$ marginally through a 1-dimensional kernel-smoothed distribution function to produce a sample of residuals $\mathbf{Z} = (Z_2, \ldots, Z_d)$, that is, $Z_i \sim \hat{G}_i^{-1}(\Phi(Z_i^N))$ for $i = 2, \ldots, d$.
3. Independent of $\mathbf{Z}^N$, draw a value of the conditioning variable Y_1 from a standard exponential distribution above v_p, for example, $Y_1 = v_p + Y_1^*$, where $Y_1^* \sim \text{Exp}(1)$.
4. Derive the simulated value of the conditioned variates $\mathbf{Y}_{-1}$, which is a function of Y_1, $\mathbf{Z}$ and the estimated dependence parameters $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, via
$$\mathbf{Y}_{-1} = \hat{\boldsymbol{\alpha}} Y_1 + Y_1^{\hat{\beta}} \mathbf{Z}, \quad \text{for } Y_1 > v_p.$$
This gives a sample of $\mathbf{Y} = (Y_1, \mathbf{Y}_{-1})$ with $Y_1 > v_p$.
5. The inverse of the probability integral transform, as given in Equation (3), can be used to transform $\mathbf{Y}$ back to its original margins of $\mathbf{R} = (R_1, \ldots, R_d)$, with $R_1 > q_{1,p}$.

In the simulation of spatially consistent extreme events, we want to ensure that events are simulated conditional on $\mathbf{R}$ being extreme for at least one location. We adopt the model of Keef et al. (2013) that generates an extreme event conditional

on the event $\{\max(F_1(R_1), \ldots, F_d(R_d)) > p\}$ with p near 1, or equivalently $\{\exists i = 1, \ldots, d : R_i > q_{i,p}\}$. After transformation to Laplace margins, this corresponds to simulating $\max(Y_1, \ldots, Y_d) > v_p$. To be able to simulate from this conditional distribution using the previous algorithm for simulating from $\mathbf{Y}|Y_1 > v_p$, we need to determine the conditioning gauge for each event. The approach is to first simulate $I^p = \arg\max\{\mathbf{Y} \mid \max\{Y_1, \ldots, Y_d\} > v_p\}$, with

$$\begin{aligned}\mathbb{P}(I^p = j) &= \frac{\mathbb{P}(Y_j = \max(Y_1, \ldots, Y_d), Y_j > v_p)}{\sum_{k=1}^{d} \mathbb{P}(Y_k = \max(Y_1, \ldots, Y_d), Y_k > v_p)} \\ &= \frac{\mathbb{P}(Y_j = \max(Y_1, \ldots, Y_d) \mid Y_j > v_p)}{\sum_{k=1}^{d} \mathbb{P}(Y_k = \max(Y_1, \ldots, Y_d) \mid Y_k > v_p)},\end{aligned}$$

where here each of these conditional probabilities can be estimated from our models for $\mathbf{Y}|Y_k > v$, for $k = 1, \ldots, d$. Finally, if $I^p = j$, then apply the above algorithm for $\mathbf{Y}|Y_1$ with the index 1 replaced by j and this point is rejected if $\max(\mathbf{Y}_{-j}) > Y_j$, that is, Steps 1–5 need repeating until, for the selected gauge, j, we have $\max(\mathbf{Y}_{-j}) < Y_j$.

4.2 | Estimation of joint extreme events

In many applications, such as the design of flood defence schemes or assessing potential flood losses over an insurance portfolio, interest lies in accurately estimating the probability of rare events across a number of spatial locations or environmental hazards. The Monte Carlo methods described in Section 4.1 are the most effective way to estimate many extreme events. However, as was noted in Section 1, there are major limitations with these methods for events that are rare relative to the marginal probability for the conditioning variable. Estimation of these probabilities requires a more careful analysis, which we can achieve for the first time here due to our semiparametric residual distribution model choice. We will illustrate the estimation for both these types of events.

Firstly, consider an event A that is extreme in the sense that at least R_1 is extreme. Then, there exists a value of p, near 1 such that $A \subset [q_{1,p}, \infty) \times (\infty, \infty)^{d-1}$. It follows that

$$\begin{aligned}\mathbb{P}(\mathbf{R} \in A) &= \mathbb{P}(R_1 > q_{1,p})\mathbb{P}(\mathbf{R} \in A \mid R_1 > q_{1,p}) \\ &= (1-p)\mathbb{P}(\mathbf{R} \in A \mid R_1 > q_{1,p}).\end{aligned}$$

An estimate of this joint probability is given by

$$\hat{\mathbb{P}}(\mathbf{R} \in A) = (1-p)\sum_{t=1}^{\ell} \mathbb{1}(\tilde{\mathbf{R}}_t \in A)/\ell,$$

where $\tilde{\mathbf{R}}_1, \ldots, \tilde{\mathbf{R}}_\ell$ are independent and identically distributed values simulated from $\mathbf{R}|R_1 > q_{1,p}$, and ℓ is the number of the simulations. However, if $\{R_i; i \in C\}$, with $1 \in C$, is asymptotically independent, then as $\chi_C = 0$, the conditional probability that is being estimated by the Monte Carlo methods above is near zero if $A \subset \prod_{i \in C}(q_{i,p}, \infty)$. For sets such as A, it is better to exploit the Gaussian copula structure and express the result through an integral for which standard numerical integration methods can be used. Specifically, for $A = \prod_{i \in D}(q_{i,p_i}, \infty)$, with p_1 near 1, the model gives

$$\begin{aligned}\mathbb{P}\left(R_1 > q_{1,p_1}, \ldots, R_d > q_{d,p_d}\right) &= \mathbb{P}\left(Y_1 > y_1, \ldots, Y_d > y_d\right) \\ &= \int_{y_1}^{\infty} \mathbb{P}\left(\mathbf{Y}_{-1} > \mathbf{y}_{-1} | Y_1 = s\right) f_{Y_1}(s) ds \\ &= \int_{y_1}^{\infty} \mathbb{P}\left(\hat{\boldsymbol{\alpha}} Y_1 + Y_1^{\hat{\boldsymbol{\beta}}} \mathbf{Z} > \mathbf{y}_{-1} | Y_1 = s\right) \frac{1}{2}\exp(-s) ds \\ &= \int_{y_1}^{\infty} \mathbb{P}\left(\mathbf{Z} > \frac{\mathbf{y}_{-1} - \hat{\boldsymbol{\alpha}} s}{s^{\hat{\boldsymbol{\beta}}}} | Y_1 = s\right) \frac{1}{2}\exp(-s) ds \\ &= \int_{y_1}^{\infty} \mathbb{P}\left(\mathbf{Z}^N > \Phi^{-1}\left(\tilde{\mathbf{G}}\left(\frac{\mathbf{y}_{-1} - \hat{\boldsymbol{\alpha}} s}{s^{\hat{\boldsymbol{\beta}}}}\right)\right) | Y_1 = s\right) \frac{1}{2}\exp(-s) ds \\ &= \int_{y_1}^{\infty} \bar{\Phi}_{d-1}\left(\Phi^{-1}\left(\tilde{\mathbf{G}}\left(\frac{\mathbf{y}_{-1} - \hat{\boldsymbol{\alpha}} s}{s^{\hat{\boldsymbol{\beta}}}}\right)\right), \tilde{\Sigma}\right) \frac{1}{2}\exp(-s) ds, \end{aligned} \tag{15}$$

where $\tilde{\mathbf{G}}(\mathbf{z}) = (\tilde{G}_2(z_2), \ldots, \tilde{G}_d(z_d))$, $\mathbf{y}_{-1} = (y_2, \ldots, y_d)$ with y_i being the p_ith quantile of a Laplace distribution, and $\bar{\Phi}_{d-1}(., \Sigma)$ is the joint survivor function of the standard multivariate Gaussian variable with correlation matrix Σ. This

result allows us to reduce the complexity of the $(d-1)$-dimensional integral calculation of rare event probabilities through the direct evaluation of the multivariate Gaussian joint survivor function and a 1-dimensional integral.

5 | SIMULATION STUDY

To assess the performance of our proposed Gaussian copula approach, for modelling the joint distribution of the residuals in the conditional multivariate extremes model, we undertake a simulation study to compare it with the empirical approach of Heffernan and Tawn (2004) and with an approach using a multivariate kernel density estimate

$$\hat{G}(\mathbf{z}) = \frac{1}{n_v}\sum_{i=1}^{n_v}\Phi_{d-1}\left(\mathbf{z}|\mathbf{z}_i, \mathbf{H}\right), \tag{16}$$

where the ith kernel is Gaussian with mean $\mathbf{z}_i$, $\mathbf{H}$ is a positive definite bandwidth matrix (Wand & Jones, 1994), and $\{\mathbf{z}_1, \ldots, \mathbf{z}_{n_v}\}$ are the observed residuals. The methods are compared via their estimation of the probability

$$\gamma_d = \mathbb{P}\left(R_1 > q_{1,p}, \ldots, R_d > q_{d,p}\right) \tag{17}$$

with $p = 0.99, 0.998$, and 0.999.

Data are simulated from a symmetric multivariate extreme value logistic distribution (Tawn, 1990), with dependence parameter $\delta \in (0, 1]$ with the lower and upper limits for δ corresponding in perfect dependence and independence, respectively. For the symmetric logistic distribution and a given dimension d, the true probability of Equation (17) is $\gamma_d = \sum_{m=0}^{d}\binom{d}{m}(-1)^m p^{m^\delta}$. For all $\delta < 1$, the variables are asymptotically dependent, that is, $\chi_D > 0$, and hence, parameters of the Heffernan and Tawn (2004) model are $\boldsymbol{\alpha} = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$. Furthermore, for this distribution, the true copula for $\mathbf{Z}$ is not Gaussian, so our model gives a misspecification. We consider $d = 5, 10$, and 20 with $\delta = 0.75$ (results with $\delta = 0.5$ are not reported but are similar) and a sample size of 5,000 with 25 replicated data sets and a 0.98 dependence threshold corresponding to 100 observations being in the joint tail region. Correctly in each case, we find that there is strong evidence to reject the Gaussian copula assumption, at a 5% level, when using the test statistic (14) for each of our simulations. Despite this, we proceed to using the Gaussian copula model to see if this misspecification is important for inference.

Table 1 shows results for $d = 5$, where the regression parameters are both set to their true values and when they are estimated. For this relatively low-dimensional case, all three methods perform broadly similarly both in terms of their point estimates and bootstrap-based 95% confidence intervals, with all intervals containing the truth. Despite its clear misspecification, the Gaussian copula method gives estimates that are closest to the truth in all six cases. In addition, we see that the multivariate kernel approach performs worst (underestimating) in all cases.

Furthermore, note that getting good knowledge of the regression parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is more important than the choice of distributional model for $\mathbf{Z}$. This feature is interesting given that much of multivariate extreme value inference has focussed on assuming asymptotic dependence (fixing the regression parameters) and effectively only estimating $\mathbf{Z}$ in different ways. These results suggest that focus of attention has been misplaced.

Higher dimensional studies, $d = 10$ and 20, are compared in Table 2 with the true regression dependence parameters treated as known to enable easier comparison of the different methods for handling the residuals. The multivariate

TABLE 1 The estimates (with 95% confidence intervals in parenthesis) for the joint event probability $1{,}000\gamma_d$, given in Equation (17), for $d = 5$ with $\delta = 0.75$ for a sample of size 5,000 from the symmetric logistic distribution

Marginal probability	**0.99**	**0.998**	**0.999**
True joint probability	1.80	0.36	0.18
True regression dependence parameters			
Heffernan and Tawn	1.97 (1.50, 2.48)	0.39 (0.30, 0.50)	0.20 (0.15, 0.25)
Multivariate kernel	1.63 (1.22, 2.03)	0.32 (0.24, 0.41)	0.16 (0.12, 0.20)
Gaussian copula	1.90 (1.44, 2.30)	0.38 (0.28, 0.46)	0.19 (0.14, 0.23)
Estimated regression dependence parameters			
Heffernan and Tawn	1.38 (1.08, 1.87)	0.18 (0.05, 0.26)	0.07 (0.01, 0.12)
Multivariate kernel	1.10 (0.85, 1.45)	0.13 (0.04, 0.22)	0.06 (0.01, 0.10)
Gaussian copula	1.46 (1.03, 2.00)	0.20 (0.07, 0.31)	0.09 (0.02, 0.14)

TABLE 2 The estimates (with 95% confidence intervals in parenthesis) for the joint event probability $1{,}000\gamma_d$, given in Equation (17), for $d = 10$ and 20 with $\delta = 0.75$ for a sample of size 5,000 from the symmetric logistic distribution

Marginal probability	**0.99**	**0.998**	**0.999**
	d=10		
True joint probability	0.39	0.28	0.14
True regression dependence parameters			
Heffernan and Tawn	1.49 (0.98, 1.87)	0.30 (0.20, 0.37)	0.15 (0.10, 0.19)
Multivariate kernel	0.79 (0.52, 1.05)	0.16 (0.10, 0.21)	0.08 (0.05, 0.11)
Gaussian copula	1.34 (1.00, 1.65)	0.27 (0.20, 0.33)	0.13 (0.10, 0.17)
	d=20		
True joint probability	1.15	0.23	0.11
True regression dependence parameters			
Heffernan and Tawn	1.09 (0.84, 1.50)	0.22 (0.17, 0.30)	0.11 (0.10, 0.19)
Gaussian copula	1.12 (0.81, 1.33)	0.22 (0.16, 0.27)	0.11 (0.08, 0.13)

kernel approach is now clearly failing when $d = 10$ and becomes increasingly computationally expensive as d increases and so is omitted from the $d = 20$ study. The empirical approach of Heffernan and Tawn (2004) and the Gaussian copula approach perform broadly similarly as well, though again where they differ the Gaussian method works best. Therefore, even in this case with clear misspecification, the proposed Gaussian copula is at least very highly competitive relative to the existing method. It should be noted that, when there is either no misspecification or there are missing data, the Gaussian copula approach substantially outperforms the empirical approach of Heffernan and Tawn (2004); see Section 6.2.2 for an example of this.

6 | RIVER FLOW APPLICATIONS

6.1 | Data

We apply the proposed semiparametric conditional extreme value model to daily mean measurements of river flow data from the National River Flow Archive (NRFA) to answer questions typically proposed by flood risk managers. Gauges from the north west region of England were selected and the locations of these are given in Figure 1; on average, each gauge has record length of approximately 30 years. This region has one of the better spatial coverages of data in the UK. The proportion of missing values in the data is relatively low. The region exhibits varying spatial characteristics, for example, due to changing soil types and elevation, the behaviour is likely to be very different in Cumbria compared with, say, Manchester (in the north and south of the region, respectively). The data set was selected as it has been used for previous spatial flood risk assessments (Lamb et al., 2010; Tawn et al., 2018; Towe, Tawn, Lamb, Sherlock, & Liu, 2016), and it is a region badly affected by the 2015 floods, discussed in Sections 1 and 6.4. For the data, we discuss how our proposed methodology can aid in producing better inferences for rare events at much reduced computational cost and with minimal risk of misspecification error.

In Section 6.2, we will illustrate all of the steps of the methodology with a basic case study of 10 sites and then undertake to a full application to 46 gauges in Section 6.3. We see the 10-site study as important as it lets us look carefully at some of the features of the modelling/inference without getting lost in the volume of the data. In particular, we can look at what happens when large portions of the data are missing. To help investigate how our methods work in the basic case study, we estimate probabilities of extreme events for two data sets. The original data set, denoted by F, has 1% missing (0.5% are missing conditional on the first site being large), with a missingness pattern that allows use of Heffernan and Tawn (2004) and the infill approach of Keef et al. (2009). The second data set, denoted by M, has 28% removed to missing status in such a way that no complete observations are available (30% are missing conditional on the first site being large). In both of the analyses, the conditioning site is the same and can be identified by the triangle in Figure 1. For the full application considering 46 gauges, in Section 6.3, 2% of the data are missing.

6.2 | Basic case study

6.2.1 | Assessing the Gaussian copula

First, we use the original data set to assess our modelling assumptions for these data. Conditioning on R_1 being large, we focus on studying the behaviour of $\mathbf{Z}^N$, the residuals after the marginal transformation to standard Gaussian margins.

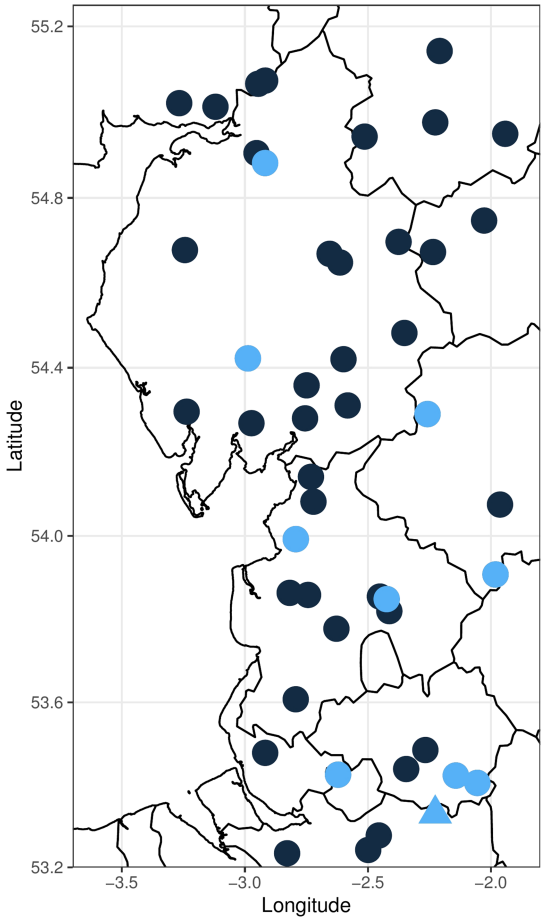

FIGURE 1 Locations of the 46 daily mean river flow gauges situated in the north west of England. The subset of 10 gauges is shown, given in light blue. The conditioning station used in estimation of probability $\tau_{m,p}$, defined in Equation (18), is represented by a triangle

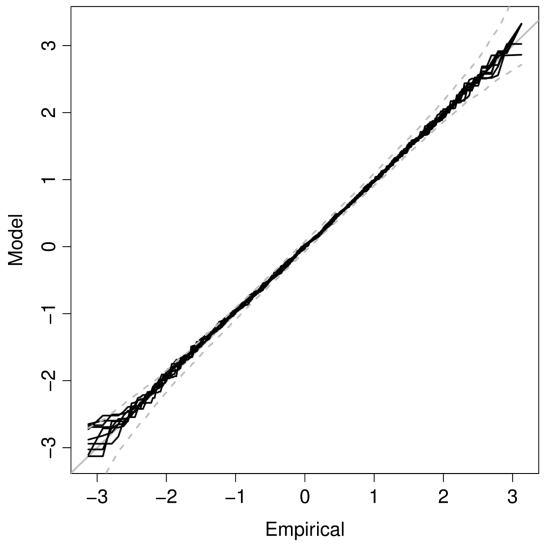

FIGURE 2 Pooled marginal QQ plots of $\mathbf{Z}^N = (Z_2^N, \ldots, Z_{10}^N)$

A check of the assumption of standard Gaussian margins is given in Figure 2; the empirical quantiles of a standard Normal are plotted against those of the residuals $\mathbf{Z}^N$ with this being a pooled QQ plot over all margins and replicates of $\mathbf{Z}^N$. The different lines in Figure 2 for each respective margin of $\mathbf{Z}^N$ show that there is no significant deviation away from the line of equality; therefore, the marginals satisfy the assumptions for the proposed Gaussian copula model.

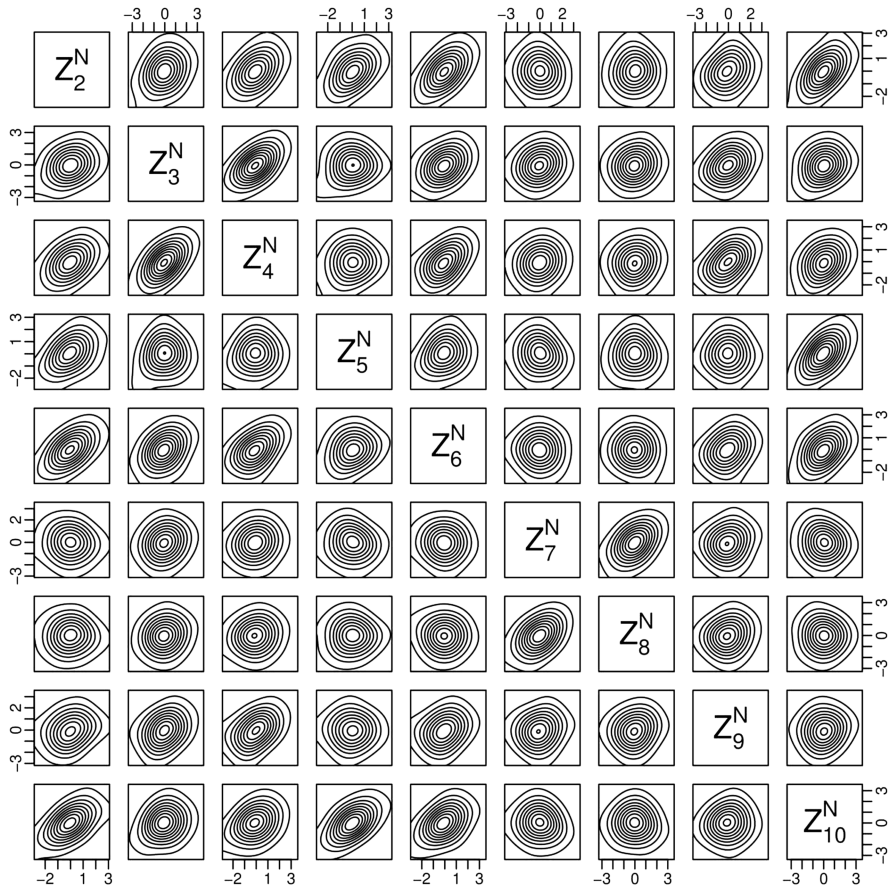

FIGURE 3 Pairwise kernel density estimates for $\mathbf{Z}^N = (Z_2^N, \ldots, Z_{10}^N)$

Pairwise bivariate kernel density estimates for $\mathbf{Z}^N$ can be seen in Figure 3. From a visual inspection, the pairwise dependence seems close to Gaussianity; although in a couple of pairs such as (Z_3^N, Z_5^N), there does seem to be departure away from the expected elliptical contours. Figure 3 does not help us assess any higher order dependence, and as a result, the test for Gaussianity (as given in Section 3.2) is performed to test the assumption of a Gaussian copula more rigorously. The test statistic is calculated using the methodology given in Section 3.2. The p value is calculated to be equal to 0.29, which is greater than the significance level of 0.05. Therefore, the assumption of a Gaussian copula seems reasonable.

Some benefits of the Gaussian copula approach are that the new method is able to interpolate and extrapolate the observed residuals giving simulated events, which are not simply deterministic functions of observed events. A comparison of these features of the Heffernan and Tawn (2004) and Gaussian copula approaches is illustrated in Figure 4. Under these two approaches, Figure 4 shows data and simulations of (top) $Y_2|Y_1 > v_p$ and (bottom) $(Y_2, Y_3)|Y_1 > v_p$, both for $p = 0.99$. From the top row, our proposed approach is seen to give a continuous distribution for $Y_2|Y_1$ with slightly more variation in $Y_2|Y_1 > v_p$. This additional variation, which seems realistic given the extremal behaviour of the observed data set, is due to the use of kernel-smoothed marginal distribution functions for $\mathbf{Z}^N$. Similarly, from the bottom row, it can be seen that the simulated joint residuals can differ from observed values, due to the Gaussian copula assumption. Collectively, these new features lead to the simulation of a more realistic joint sample with our proposed approach than that from the Heffernan and Tawn (2004) model.

6.2.2 | Conditional probabilities for flood risk management

In many flood risk management cases, interest lies in determining the spatial extent of any given flood event. One common risk measure that flood managers are interested in is the probability that given a site, say, Site 1, exceeds its pth quantile that there are then at least m other sites that also exceed their respective pth quantile, that is,

$$\tau_{m,p} = \mathbb{P}(\#(j = 2, \ldots, d : R_j > q_{j,p}) \geq m \mid R_1 > q_{1,p}) = \mathbb{P}(Y_{(m)} > v_p \mid Y_1 > v_p), \tag{18}$$

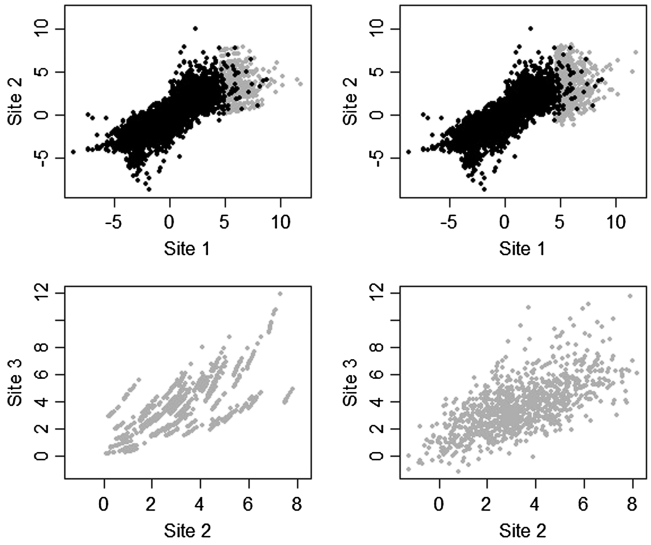

FIGURE 4 Top row: observed (black) and simulated (grey) joint behaviour of Site 1 and Site 2, given that an extreme event is observed at Site 1. Bottom row: observed (black) and simulated (grey) joint behaviour of Site 2 and Site 3, given that an extreme event is observed at Site 1. Left: the existing method. Right: our proposed method. In all of the figures, the data are shown after transformation to standard Laplace margins

TABLE 3 The estimates (with 95% confidence intervals in parenthesis) for the conditional probability $100\tau_{m,T}$, given in Equation (18), with m=5 using the original (F) and 28% missing data (M)

Probability	Heffernan and Tawn (F)	Infill (F)	Gaussian copula (F)	Infill (M)	Gaussian copula (M)
$\tau_{5,100}$	4.3 (0.0, 12.2)	4.3 (0.1, 11.9)	4.1 (0.1, 12.7)	4.0 (0.6, 15.5)	4.5 (0.3, 14.5)
$\tau_{5,500}$	2.9 (0.0, 9.6)	2.9 (0.0, 9.7)	3.0 (0.0, 10.4)	2.4 (0.2, 13.7)	3.1 (0.1, 13.1)
$\tau_{5,1000}$	2.5 (0.0, 8.3)	2.5 (0.0, 8.9)	2.4 (0.0, 9.5)	2.0 (0.1, 12.5)	2.6 (0.1, 12.3)
$\tau_{5,10000}$	1.6 (0.0, 6.8)	1.6 (0.0, 6.8)	1.6 (0.0, 7.8)	1.0 (0.0, 10.9)	1.7 (0.0, 10.1)

Note. The T is the probability that corresponds to a specific annual return period. The Heffernan and Tawn (2004) column corresponds to the conditional extreme value model fitted to all of the data. The modelled infill column refers to the missing values being modelled and infilled into the observed data.

$m = 1, \ldots, d-1$, where $Y_{(m)}$ is the mth largest value of $(Y_2, \ldots, Y_d)$. Probabilities $\tau_{m,p}(m = 1, \ldots, d-1)$ are useful as they give a clear insight into the spatial extent of a flooding event. If the pth quantile is the level of flood defence at all sites, the probability of exactly m other sites being flooded, given Site 1 floods, is $\tau_{m,p} - \tau_{m+1,p}$.

For $\tau_{m,p}$, given in Equation (18) with $m = 5$, in Table 3, we provide a point estimate and associated 95% confidence intervals, obtained by using the parametric bootstrap for a range of return periods. These estimates are compared using the Heffernan and Tawn (2004) method with two missing value methods (the infill method of Keef et al., 2009, and our proposed Gaussian copula method). The two data sets denoted F and M are considered; see Section 6.1.

For data set F, all three methods produce very similar estimates. This is not surprising for the Heffernan and Tawn (2004) and infill methods as for 99% of the data these methods are identical. However, for the Gaussian copula, we are using the modelled residual copula for all the data that are extreme at the conditioning site, and therefore, to find that the estimate varies so little from that of Heffernan and Tawn (2004) is particularly pleasing. For the F data, confidence intervals for both the missing data methods are largest due to a combination of the additional Monte Carlo uncertainty and residual marginal distribution smoothing in the respective methods. Here, only 1% of the data were missing, so we would not expect to see any clear improvement in using these missing data methods, which use all partially observed components unlike in the Heffernan and Tawn (2004) method.

For data set M, it is impossible to obtain estimates from the Heffernan and Tawn (2004) approach due to there being no observations being made concurrently. What is pleasing to see here is that the two missing data methods give broadly similar estimates to those from data set F. In particular, the Gaussian copula model gives estimates that are very close to those using the F data sets for all events in Table 3, whereas for the infill method, the estimates are less self-consistent for the rarer of these events. The confidence intervals of the two methods are approximately the same, which is to be expected as both model the missing values by using a Gaussian copula but handle the computation in different ways. Naturally, the confidence intervals for the M data are larger than the equivalent ones for the F data.

TABLE 4 The estimates (and integration numerical error) for the conditional probability $\tau_{m,p}$, given in expression (15), with $m = 9$

Probability	Estimate	Numerical error
$\tau_{9,100}$	7.34×10^{-5}	6.00×10^{-8}
$\tau_{9,500}$	1.19×10^{-5}	1.26×10^{-8}
$\tau_{9,1000}$	8.58×10^{-7}	1.60×10^{-8}
$\tau_{9,10000}$	2.51×10^{-12}	1.90×10^{-14}

Note. This table uses the same return periods as in Table 3.

A critical feature is that the Gaussian copula approach is computationally much quicker even in this basic case. Specifically, the time to get the point estimates using the Gaussian copula is 30% less than the infill method (assuming $\omega = 100$), and this efficiency gain improves dramatically as both the number of sites and the proportion of missing data increase.

The probabilities in Table 3 were estimated through simulation. However, if we were interested in all sites being above a given return level, this corresponds to $m = 9$ in Equation (18). This probability is incredibly computationally expensive to estimate through Monte Carlo simulation; however, the methods developed in Section 4.2 can provide us with an estimate that avoids Monte Carlo noise, as it is obtained using the formulation (15) divided by p, with $d = 10$. Table 4 provides estimates of the $\tau_{9,p}$ for the same return periods as in Table 3 along with the corresponding numerical integration error.

6.3 | Large-scale study

Here, the entirety of the north west region of England is considered; this equates to 46 sites in our study. The first modelling step is to fit the conditional extreme value model of Heffernan and Tawn (2004) conditioning on each of the 46 gauges in turn. For each of these 46 models, the estimates of the dependence parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are obtained along with the residuals $\mathbf{Z}$ of the model.

The residuals $\mathbf{Z}^N$ of the model are tested to determine whether they can be characterised by using a Gaussian copula. For each conditioning gauge, in turn, the sampling distribution of the test statistic T^*, as given in Section 3.2, is obtained through Monte Carlo simulation and a p value for a Gaussian copula is derived. Figure 5 shows a histogram of the p values with all of the 46 p values above the 5% significance level. Therefore, we can conclude that there is no evidence against modelling the residual distribution with a Gaussian copula. Given this conclusion, it seems reasonable to use the model-based Gaussian copula for the multivariate residual component of the conditional extreme value model of Heffernan and Tawn (2004).

We can use these models to make extrapolations using the Monte Carlo methods given in Section 4.1. These simulations maintain the extremal dependence structure of the observed data set but will also generate events that are larger and more varied than those we have already observed. Two such examples are shown in Figure 6 with these illustrating how the spatial structure of an event varies depending on where in the region the event is extreme. The two events have

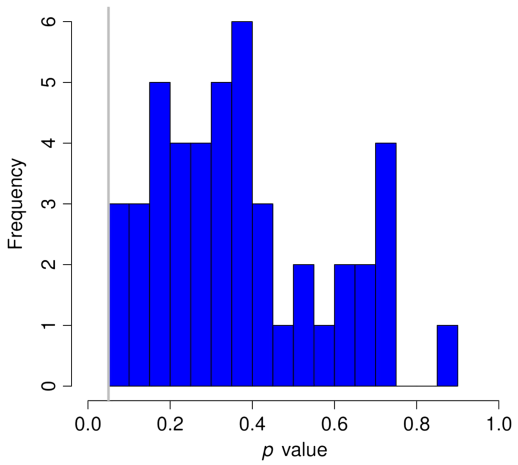

FIGURE 5 Histogram of test statistic p values for the hypothesis of a Gaussian copula for $\mathbf{Z}^N$ under 46 different conditioning sites

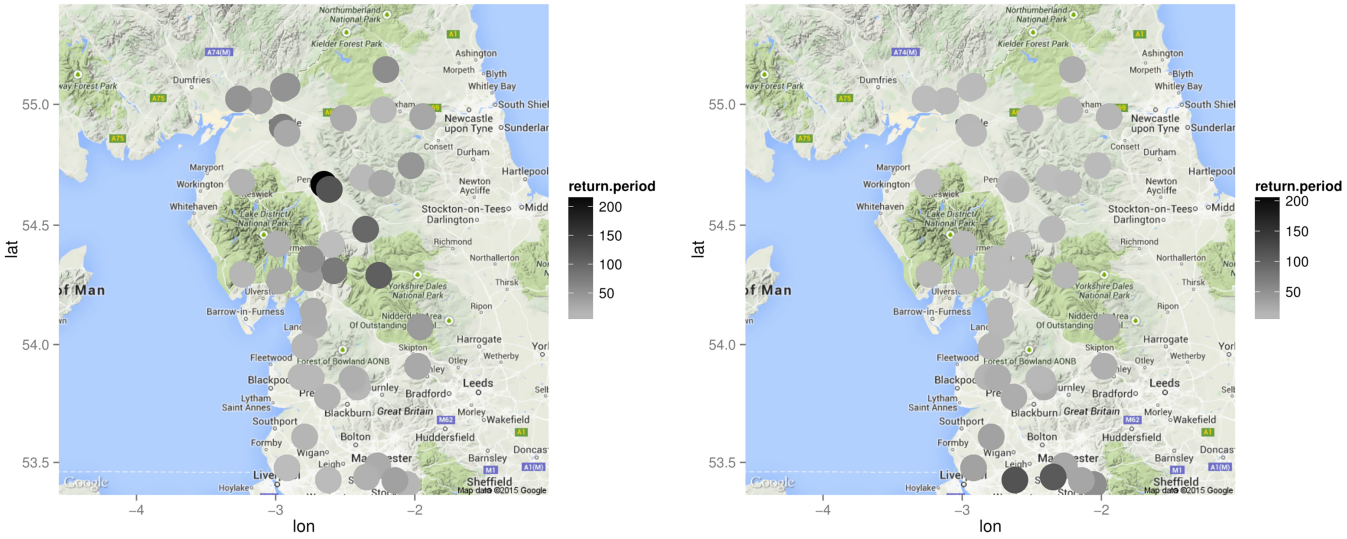

FIGURE 6 Two realisations (on the return period scale) from the proposed model, where the conditioning gauge observes at least a 1-in-100 year event. The conditioning sites are (left) in Cumbria and (right) close to Manchester

been selected to be extreme at two different sites in the region, in Cumbria and Manchester, in the north and south of the region, respectively. In Figure 6a, when the conditioning location is in Cumbria, there is a much wider spatial impact, than in Figure 6b, for an event near Manchester. This reflects that, when we condition on Cumbria being extreme, relative to Manchester being extreme, the associated $\boldsymbol{\alpha}$ parameters are larger over many more sites, so the spatial extremal dependence is stronger and extreme events in the north of the region are more widespread than those in the south of the region.

To further study the varying spatial characteristics of extreme flood events, a conditioning site is selected to have an extreme event and the distribution of the number of other gauges that are also extreme is estimated. This estimated distribution is derived for the same two conditioning sites as in Figure 6. Here, the probability of exactly m other gauges is $\tau_{m,p} - \tau_{m+1,p}$, and this is estimated for three return periods. Estimates of $\tau_{m,p} - \tau_{m+1,p}$ are compared in Figure 7 for the two conditioning gauges. There is a clear difference in these estimated probabilities. The estimates show that there is greater clustering of flood events when conditioning on the Cumbria site being large. However, some of this clustering could be explained by the fact there is a higher density of gauges in this region. Furthermore, the estimates decay to

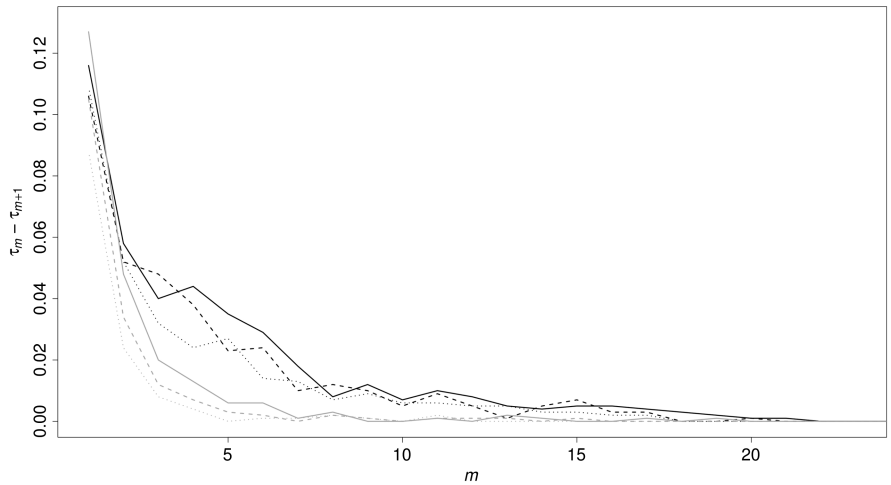

FIGURE 7 Distribution of the number m of other sites that are extreme, given that the condition site is extreme: the grey lines, conditioning on gauge 69017 near Manchester, and black lines conditioning on gauge 74001 in Cumbria. The solid, dashed, and dotted lines correspond to observing a 100, 1,000, and 10,000 year event at the respective conditioning site

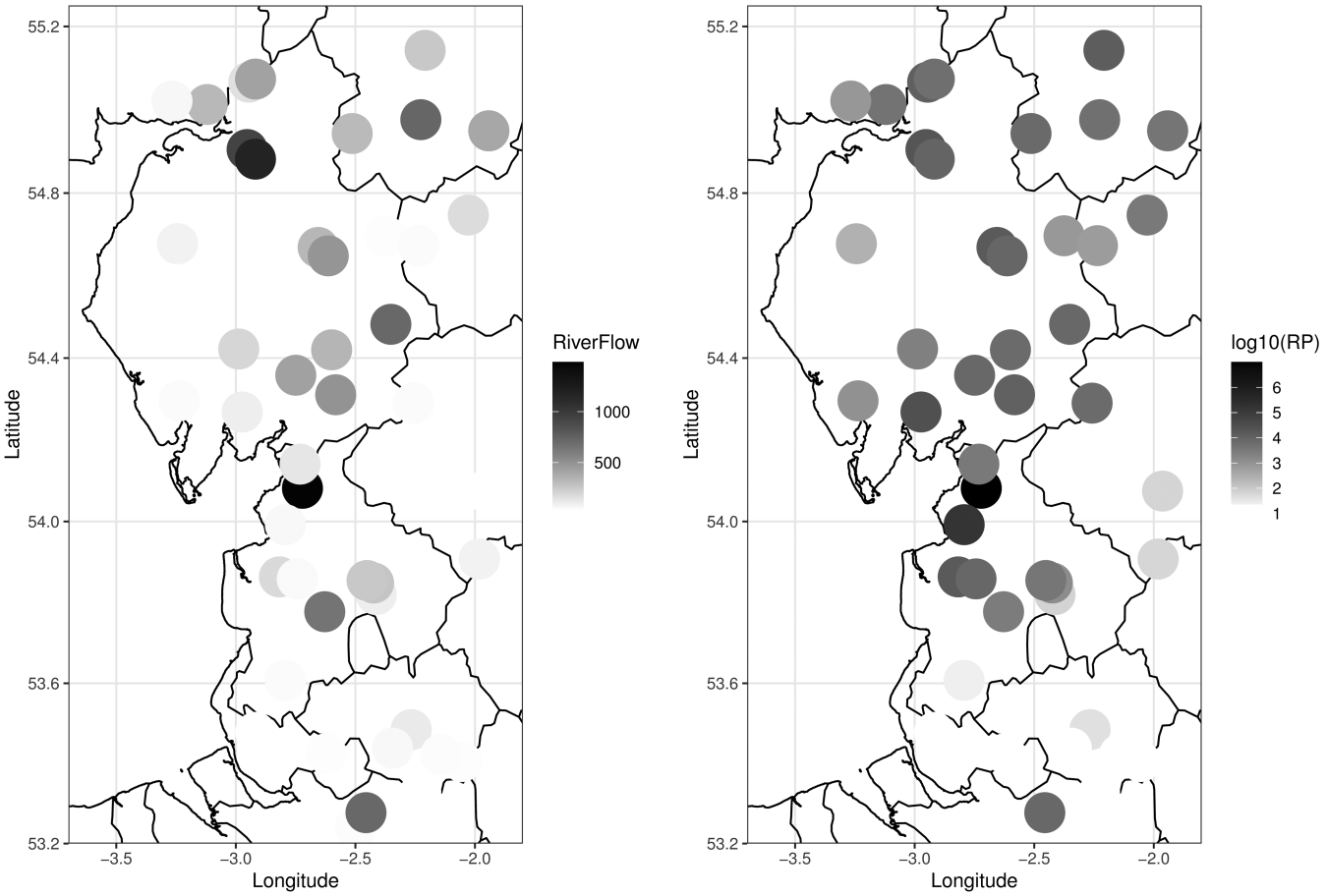

FIGURE 8 Left: observed daily mean river flows measured in m^3s^{-1} from the December 5th, 2015. Right: the corresponding marginal return periods for those observed daily mean river flows plotted on the log scale

zero, for $m > 1$, at different rates; thus, events become more localised as they become more extreme, due to asymptotic independence.

6.4 | Determining the rarity of the storm Desmond event

The methodology is used to determine the rarity of river flows that were observed on the December 5, 2015, storm Desmond event. This estimate is derived from the daily mean river flow data discussed in Section 6.1 with the results presented on a daily scale. The observed daily mean river flows are shown in Figure 8a with the largest values observed near Lancaster and Carlisle. However, when we determine the associated estimated marginal return periods, with inference using the GPD tail model (2), the river flow observed near Lancaster is found to be the most extreme, as shown in Figure 8b. The marginal observational probability for the Lancaster gauge is estimated to be 3.6×10^{-5}. Figure 8b shows that the event was particularly rare over all Cumbria and northern Lancashire, but it was extreme at only one of the gauges near Manchester in the south of the study region.

In order to determine the probability $\mathbb{P}(R_1 > q_{1,p_1}, \ldots, R_d > q_{d,p_d})$ of jointly observing river flows over the region, which are worse than the 2015 event, we use both the empirical Heffernan and Tawn (2004) residual approach and our Gaussian copula approach with the joint probability given by the integral (15). We illustrate the calculations by separately taking the conditioning gauge to the Cumbrian gauge, shown in Figure 6a, and the Lancaster gauge, identified by Figure 8b. Using the Cumbrian gauge, we estimate the joint probability to be $< 1.60 \times 10^{-12}$ and 3.70×10^{-9} using the respective methods, whereas these respective estimates become 9.50×10^{-10} and 8.00×10^{-9} using Lancaster. When conditioning on the Cumbrian gauge, we can only bound the joint probability using the empirical Heffernan and Tawn (2004) residual approach as we get no events as extreme as that observed at Lancaster in 10^8 events simulated, where all of which exceed the observed 2015 event at the Cumbrian gauge. In contrast, the Gaussian copula approach gives estimated probabilities

that are stable with respect to the conditioning gauge and are computationally efficient in contrast to the existing approach for such an extreme and widespread event.

7 | DISCUSSION

Through using semiparametric model-based inference, this paper has shown how the methodology of Heffernan and Tawn (2004) can be extended to produce more efficient inferences, particularly as the dimension of the multivariate problem increases. Our approach proposed improvements in the inference of the residual distribution of the Heffernan and Tawn (2004) model, via kernel-smoothed marginal distributions and using a Gaussian copula. These methods also help in terms of computational and statistical efficiency in dealing with the problem of missing data that is commonly encountered in environmental data sets.

Our proposed Gaussian copula approach has a downside in that a different correlation matrix Σ is required for each conditioning site. Thus, for d sites, there are $d\binom{d-1}{2}$ correlation parameters to estimate, that is, $O(d^3)$ parameters. As a result, it seems sensible to determine whether there are any known relationships that can help to make the model parsimonious. An approach suggested by a referee was to adopt a semiparametric specification method similar to that of de Carvalho and Davison (2014), whereby the different residual densities are interlinked via a tilting term, that is,

$$\log\left(\frac{g_i(\mathbf{z})}{g_1(\mathbf{z})}\right) = \gamma_i + \mathbf{z}^T\delta_i, \quad \text{for } i = 2, \ldots, d \tag{19}$$

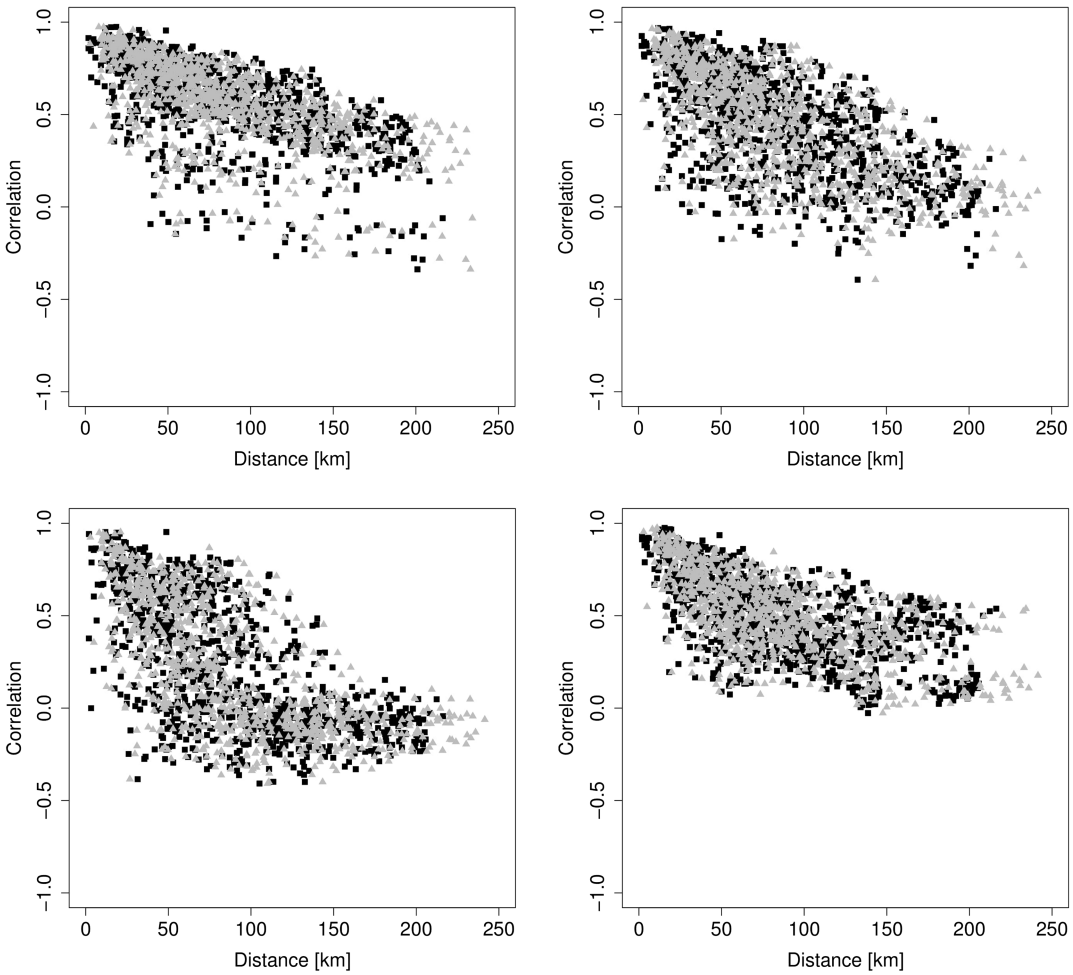

FIGURE 9 The correlation of each pair of residuals against Euclidean (black) and hydrological (grey) distance. Panels of four conditioning gauges from the National River Flow Archive: (a) 68003 (south of Manchester), (b) 69017 (western side of the Peak District), (c) 71001 (river Ribble), and (d) 74001 (a small catchment in the Lake District). A, Gauge 68003. B, Gauge 69017. C, Gauge 71001. D, Gauge 74001

with $g_i(\mathbf{z}) = dG_i(\mathbf{z})/d\mathbf{z}$, where G_i is the limiting distribution in expression (4) when conditioning on variable Y_i being large, and with (γ_i, δ_i) being constants. If condition (19) holds, the number of parameters reduces to $O(d^2)$. Unfortunately, this formulation does not appear to be appropriate for our residual data either before or after standardisation to Gaussian marginals. An alternative $O(d^2)$ approach would be to use a stationary Gaussian process to explain $\mathbf{Z}^N$ (Tawn et al., 2018), but that requires the process to be modelled in an appropriate space. In standard environmental studies, the Euclidean distance metric between sites is used to explain spatial dependence. However, as shown by Keef et al. (2009) and Asadi et al. (2015), Euclidean distance is not always sufficient for capturing the dependence between river flow gauges. The more appropriate distance metric is to consider for the hydrological distance, which is defined as the distance between centroids of the associated catchments for each site. This takes into account that two gauges that spatially might be far apart in fact are similar in nature as they lie within the same catchment.

In order to determine whether this factor could be used to simplify the correlation matrix, four conditioning sites were selected with differing spatial locations and catchment areas. Conditional on location k, the estimates of the correlation between Z_i and Z_j (for sites $\mathbf{s}_i$ and $\mathbf{s}_j$) given Y_k is large, denoted by $\rho_{ij|k}$ for $i,j \neq k$, were plotted as a function of both the Euclidean $\|(\mathbf{s}_i, \mathbf{s}_j)\|_E$ and hydrological $\|(\mathbf{s}_i, \mathbf{s}_j)\|_H$ distance for each pair. This comparison of the correlation and distance metrics can be seen in Figure 9. As expected as the distance between pairs of sites increases, the correlation tends to decrease. Interestingly, there is no substantial difference between the explanatory capabilities of Euclidean and hydrological distance. Anomalous behaviour can be seen in Figure 9a; as for one of the sites, the residual correlation with all other sites is approximately equal to zero. This site is close to conditioning gauge 68003; therefore, the Heffernan and Tawn (2004) model has explained all of extremal behaviour at this gauge, with the other sites. This illustrates that $\rho_{ij|k}$ will depend on $\mathbf{s}_k$ as well. Other known hydrological characteristics could also be used to explain the residual dependence structure; these include variables such as the catchment responsiveness as well as the soil type. For example, a chalk catchment is slower to respond to heavy rainfall events than a catchment in North West England (Boorman, Hollis, & Lilly, 1995). Generalising these features is difficult as we are trying to simplify the correlation of unexplained behaviour of the extremes rather than of the observed process itself.

This paper has shown that the proposed Gaussian copula model for the joint residual distribution of the Heffernan and Tawn (2004) model is ideal for classes of asymptotically dependent and asymptotically independent distributions. A simulation study shows in low- and high -dimensional examples the benefits of the proposed approach over other alternatives for both missing and nonmissing data problems, as well as under misspecification of the Gaussian copula. A case study of river flow data shows the benefits of the method for assessing the risk of an event similar to the storm Desmond event. An analogous analysis using existing methods would have been both incredibly computationally expensive and numerically sensitive to the choice of conditioning variable to estimate using existing methods.

ACKNOWLEDGEMENTS

Towe's research was supported by Jeremy Benn Associates Ltd, and Innovate UK KTP009454 and EP/P002285/1 (The Role of Digital Technology in Understanding, Mitigating and Adapting to Environmental Change). We thank Ye Liu (HR Wallingford) for the helpful discussions. We also thank the referees for their comments and suggestions. The daily mean river flow data were obtained through the National River Flow Archive.

ORCID

R. P. Towe https://orcid.org/0000-0002-2111-6972

REFERENCES

Asadi, P., Davison, A. C., & Engelke, S. (2015). Extremes on river networks. *The Annals of Applied Statistics*, *9*(4), 2023–2050.

Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). *Hydrology of soil types: A hydrologically-based classification of the soils of United Kingdom*. Oxfordshire, UK: Institute of Hydrology.

Bortot, P., Coles, S. G., & Tawn, J. A. (2000). The multivariate Gaussian tail model: An application to oceanographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(1), 31–49.

Coles, S. G., & Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *53*(2), 377–392.

Cooley, D., Nychka, D., & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, *102*(479), 824–840.

Cooley, D., & Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural Biological, and Environmental Statistics*, *15*(3), 381–402.

Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, *27*(2), 161–186.

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *52*(3), 393–442.

de Carvalho, M., & Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, *109*(506), 764–776.

Engelke, S., Malinowski, A., Kabluchko, Z., & Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*(1), 239–265.

Environment Agency. (2018). *Estimating the economic costs of the 2015 to 2016 winter floods* (Technical Report). Bristol, UK: Environment Agency.

Franklin, J. N. (2012). *Matrix theory*. North Chelmsford, MA: Courier Corporation.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes*, *3*, 279–290.

Heffernan, J. E., & Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, *17*(2), 537–571.

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546.

Hüsler, J., & Reiss, R. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, *7*(4), 283–286.

Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press.

Jonathan, P., Ewans, K., & Randell, D. (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, *79*, 22–31.

Keef, C., Papastathopoulos, I., & Tawn, J. A. (2013). 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. *Journal of Multivariate Analysis*, *115*, 396–404.

Keef, C., Svensson, C., & Tawn, J. A. (2009). Spatial dependence in extreme river flows and precipitation for Great Britain. *Journal of Hydrology*, *378*(3–4), 240–252.

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, *24*(1), 13–21.

Keef, C., Tawn, J. A., & Svensson, C. (2009). Spatial risk assessment for extreme river flows. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*(5), 601–618.

Lamb, R., Keef, C., Tawn, J. A., Laeger, S., Meadowcroft, I., Surendran, S., … Batstone, C. (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, *3*(4), 323–336.

Ledford, A. W., & Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, *83*(1), 169–187.

Liu, Y., & Tawn, J. A. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, *127*, 19–35.

Papastathopoulos, I., & Tawn, J. A. (2016). Conditioned limit laws for inverted max-stable processes. *Journal of Multivariate Analysis*, *150*, 214–228.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, *8*(4), 745–756.

Resnick, S. I. (2013). *Extreme values, regular variation and point processes*. New York, NY: Springer.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Boca Raton, FL: CRC Press.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, *77*(2), 245–253.

Tawn, J. A., Shooter, R., Towe, R. P., & Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, *28*, 39–58.

Towe, R. P., Tawn, J. A., Lamb, R., Sherlock, C., & Liu, Y. (2016). Improving statistical models for flood risk assessment. *E3S Web of Conferences*, *7*, 01011.

Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. Boca Raton, FL: Chapman & Hall/CRC.

How to cite this article: Towe RP, Tawn JA, Lamb R, Sherlock C. Model-based inference of conditional extreme value distributions with hydrological applications. *Environmetrics*. 2019;30:e2575. https://doi.org/10.1002/env.2575