The X-Vine model for serial and cross-sectional extreme propagation

VALENTIN COURGEAU ALMUT E.D. VERAART
Department of Mathematics, Imperial College London
180 Queen's Gate, London, SW7 2AZ, UK

October 31, 2019

Abstract

Understanding extreme events between d random variables plays a crucial role in managing the risks of complex systems as extremes are characterised by their own mechanisms. Existing (semi)parametric approaches rely on providing a framework in which parameter values limit interactions, such as asymptotic (in)dependence, in the extremes territory. Inspired from the conditional approaches and under reasonable technical assumptions, we provide a structure-adapting framework: the Extreme Vine - or X-Vine - model which separates serial and cross-sectional modelling. The former utilises stationary Gamma-distributed latent processes with a threshold parameter whilst the latter is powered by vine copulas. Asymptotic (in)dependence is ensured by the adaptive structure of the vine if one assumes the existence of tail dependence functions. We demonstrate its performance on two datasets on air pollution and energy-weather in the USA using trawl processes for the serial interactions and 3-family regular vines via traditional goodness-of-fit tests and tail coefficient estimation.

Keywords: Trawl process, peaks-over-threshold, generalised Pareto distribution, multivariate extreme value theory, regular vine copulas.

Contents

1 Introduction

Quantifying dependencies between extremes is essential in understanding risky scenarios and chain reactions in applications such as extreme air pollution, weather or hydrology (Dutfoy et al. 2014, Heffernan & Tawn 2004, Renard et al. 2006) or risk management in finance (Hautsch & Herrera 2015). Estimating accurately the probability of extreme propagation through time and variables is believed to be important (Tendijck et al. 2019, Winter & Tawn 2017). In this work, we explore the following problem: given an extreme value at a specific time and in a particular variable, what is the probability that it results in an extreme at a later time in another variable? The current state-of-the-art builds on Heffernan & Tawn (2004) and focuses on encompassing both asymptotic dependence and independence as previous models were limited to the former (Ribatet et al. 2009, Yun 2000) and most models separate serial and cross-sectional modelling (Tendijck et al. 2019, Winter & Tawn 2016, 2017) via k-th order Markov chain to provide time-clustering. We propose a different approach that is even more flexible but which keeps the same essential properties.

This article develops a multivariate extreme value framework using Gamma-distribution latent processes and regular vine copulas - namely, the Extreme-Vine, or *X-Vine*, model. To this purpose, we propose a three-fold presentation of our solution: first, we recall the marginal parametric latent model as originally defined in Noven et al. (2018). Then, inspired from the conditional extreme methodology (Heffernan & Tawn 2004), we harness vine copulas (Czado 2010) and we show how to use their flexibility to compute *Time-Rigged Probability Networks (TRON)*. Those objects constitute a fast and easy way to communicate on the likelihood of serial and cross-sectional propagation of extremes through time and across marginals. Finally, we propose to apply this model on two datasets which we detail below. Although the original reference dates from 2004, it is still very much of interest (Lugrin et al. 2019, Pineau et al. 2018, Tendijck et al. 2019) as it is superior to models based on max-stable processes (see Tendijck et al. (2019) and references therein).

Our contributions are multiple as we present a novel conditional approach to extreme dependency modelling and propose a simple and computationally-efficient probability sequence to communicate on extremes propagation. Note that, unlike traditionally, one need not assume that that all variables are extreme simultaneously nor independence in the extremes in the spirit of recent developments.

We consider extremes as values over a high threshold (i.e. Peaks-Over-Threshold (POT) methodology), the model also makes use of the limiting distribution of exceedances as Generalised Pareto Distribution (Davison & Smith 1990, Pickands 1971) as the threshold gets approaches the endpoint of the distribution. The marginal extreme model is based on a Gamma-Exponential mixture trick as detailed in Bortot & Gaetan (2014) and exploited in Noven et al. (2018) (and later detailed in Courgeau & Veraart (2018)). In general, a Gamma-distributed process is defined and used as the intensity of an Exponential-distributed process. In theory, one could use *any* strictly stationary Gamma-distributed process. For application purposes, we restrict ourselves to *trawl processes* (Barndorff-Nielsen 2011), a family of stationary and infinitely divisible random processes.

The main contribution is a major model extension to the multivariate context. The *Extreme Vine*, or X-Vine, model where d stationary cross-sectional vine copulas are fit, each one conditional on a particular variable being extreme. The X stands both for *extremes* and *cross* since the model considers the serial and cross-sectional dependencies separately. To carry out this analysis, this article unfolds as follows: Section 2 defines the problem tackled in this work and its fundamental set-up within the Peaks-over-Threshold literature. Then, Section 3 introduces the Extreme Vine model, its asymptotic dependence properties and present theoretical elements on vine copulas and Gamma-Exponential processes. The 2-step inference strategy and implementation details will be presented and discussed in Section 4. In addition, we formally define Time-Rigged Occurrence Networks in Section 5 and an application to our model. The article will be concluded by two applications: a 6-dimensional example on London air pollution in Section 6 and a 32-dimensional case study on North-East American Energy-Weather in Section 7. Those experiments corroborate that the X-Vine model captures underlying cross-sectional and serial dependencies as presented and passes standard goodness-of-fit tests. The code used in the case study can be found on GitHub under a public repository `multi-trawl-extremes`.

2 Problem set-up and solution proposal

In this section, we clarify our objective as well as explore the theoretical groundings of our modeling (see Section 2) with regards to asymptotic dependence.

For clarification purposes, we denote by I the set $\{1, \dots, d\}$ where $d > 1$ is the number of random processes considered.

This section is very similar to Section 2, Courgeau & Veraart (2018) since we build on the exact same framework. We propose to start the article with a reminder of the main theory used in the literature to study extreme values followed by an introduction to trawl processes.

2.1 The exceedances process

A common approach to studying extreme value is through *exceedances* of a time series via so-called *Peaks-over-Threshold* (POT) modelling. For a time-series $(Y_t,\ t \geq 0)$, we define the following

Definition 1. *Let $(Y_t,\ t \geq 0)$ be a strictly stationary time-series. We define the exceedances X of Y with respect to the threshold u as*

$$X_t := \max(Y_t - u, 0), \qquad \text{for } t \geq 0.$$

The *Peaks-over-Threshold* methodology is about choosing a threshold u large enough and studying the distributional properties of the exceedance processes thereby created. Pickands (1971) proved under some weak conditions on Y_t that the (scaled) exceedance process with respect to the threshold u converges in distribution to a Generalised Pareto Distribution with parameter (ξ, σ) as $u \to \infty$ if $\xi \geq 0$ (resp. as $u \to \sigma/|\xi|$ if $\xi < 0$). We say that we take u to be large enough for this approximation to hold exactly (see Lugrin et al. (2019) for the error analysis of such approximation).

Notation 2. *We denote the Generalised Pareto Distribution with parameter $(\xi, \sigma) \in \mathbb{R} \times \mathbb{R}_+$ as the distribution with density $f_{GPD}(x|\xi,\sigma) = \sigma^{-1}\left(1 + \xi x/\sigma\right)_+^{-1/\xi - 1}$, where $x_+ := \max(x, 0)$. It is defined for $0 \leq x \leq x_F$ where x_F is the upper endpoint of F_{GPD}. We have that $x_F = \infty$ if $\xi \geq 0$ and $x_F = \sigma/|\xi|$ otherwise, where ∞ and $\sigma/|\xi|$ are called the distribution endpoints in each case. The cumulative distribution function is given by $F_{GPD}(x|\xi,\sigma) = 1 - (1 + \xi x/\sigma)^{-1/\xi}$.*

2.2 Objectives

We are interested in refining our approach to understand dynamics between the extreme values in time (i.e. time clustering of extremes) and across several variables (i.e. cross sectional dependence). For the former, we use any strictly stationary *Gamma*-distributed latent processes to form an extreme value model. Examples of such method can be found in Bortot & Gaetan (2014, 2016) or Noven et al. (2018). Vine copulas (Czado (2010), Stöber & Czado (2014)) are used as a flexible dependence framework to infer extreme propagation likelihood across variables. Bortot & Gaetan (2014) and Noven et al. (2018) showed that for a time-series (Y_t) and a high threshold u, we have

$$Y_t - u \mid \{Y_t > u,\ \Lambda_t\} \sim \mathrm{Exp}(\Lambda_t), \quad \text{and} \quad \mathbb{P}\{Y_t > u | \Lambda_t = \lambda_t\} = \exp\{-\kappa\lambda_t\},$$

where with $\alpha, \beta, \kappa > 0$. Marginally, $\Lambda_t \sim \mathrm{Gamma}(\alpha, \beta)$ with $(\Lambda_t,\ t \geq 0)$ a strictly stationary process. We define the Gamma density as $f(x|\alpha,\beta) = \Gamma(\alpha)^{-1}\beta^\alpha x^{\alpha-1}e^{-\beta x}$ with $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ for $z > 0$. Here, $\kappa > 0$ controls the extreme value probability in the sense that given marginal parameters (α, β), the probability of having an extreme value is $(1 + \kappa/\beta)^{-\alpha}$. Therefore, for each marginal, the univariate hierarchical structure is parametrised by three core parameters $(\alpha, \beta, \kappa) \in \mathbb{R} \times [0, \infty)^2$. Note that using our notation for GPD distribution, we have $\xi = \alpha^{-1}$ and $\sigma = (\beta + \kappa)/|\alpha|$. Formally, we would like to solve the following problem:

Problem 3. *Given d strictly-stationary exceedance processes $((X_t^{(i)},\ t \geq 0),\ i \in I)$, we are interested in estimating the probability of exceedance in $X^{(j)}$ at time $t + h$ above a threshold $w \geq 0$ given an exceedance over a threshold $v \geq 0$ observed in $X^{(i)}$ at time t, for any time $t \geq 0$:*

$$p_{i \to (j,h)}(v, w) := \mathbb{P}\left\{X_{t+h}^{(j)} > w \middle| X_t^{(i)} > v\right\}, \quad \text{for } i, j \in I \text{ and } h > 0.$$

Remark 1. *Note that $t \geq 0$ does not appear in the probability since we assume strict stationarity in both the temporal component and cross-sections.*

When dealing with multivariate extremes, a standard approach was originally presented in Heffernan & Tawn (2004) via conditional extreme values and we adapt it to our setting below.

The original approach attempts to compute the probability of having the event $\{\boldsymbol{X} \in C\}$, where $\boldsymbol{X} = (X^{(1)}, \dots, X^{(d)})$; $C \subset \mathbb{R}^d$ is said to be an *extreme set* such that $\boldsymbol{X} \in C$ if and only if $X^{(i)} > 0$ for some $i \in I$ (Page 499, Heffernan & Tawn (2004)) Define the disjoint subsets for any $i \in I$:

$$C^{(i)} := C \cap \left\{\mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d : F_{X^{(i)}}(x^{(i)}) > F_{X^{(j)}}(x^{(j)}),\ j \in I,\ j \neq i\right\}.$$

where $F_{X^{(j)}}$ is the marginal CDF of $X^{(j)}$ for all $j \in I$. By continuity, they ignore the sets $C \cap \{\mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d : F_{X^{(i)}}(x^{(i)}) = F_{X^{(j)}}(x^{(j)}),\ j \in I,\ j \neq i\}$ for any $i \in I$. It follow that:

$$\mathbb{P}\{\boldsymbol{X} \in C\} = \sum_{i=1}^d \mathbb{P}\left\{X^{(i)} \in C^{(i)}\right\} = \sum_{i=1}^d \mathbb{P}\left\{X^{(i)} \in C^{(i)} | X^{(i)} > v^{(i)}\right\} \mathbb{P}\left\{X^{(i)} > v^{(i)}\right\} \tag{1}$$

where $v^{(i)} = \inf\left\{x^{(i)} : \mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \in C^{(i)}\right\}$ for $k \in I$ (Equation (1.1), Heffernan & Tawn (2004)).

We identify *two main limitations* to this approach: (i) we cannot obtain clear quantification of the cross-sectional interactions of the extremes; (ii) there is no temporal clustering in the extremes as observed in real data (Sections 6 and 7). We attempt to address those issues with the X-Vine model herein introduced.

3 The X-Vine model: time-clustered extreme with regular vines

We define the X-Vine model in its most general form, that is with univariate stationary models comprised of Gamma-distributed latent processes and cross-sectional vine structures conditional on extremes. This said, we detail on the particular forms that are studied in this article in subsequent sections.

3.1 Model definition

Before giving the definition we start with a remark on the notation for model parameters:

Remark 2. *Bortot & Gaetan (2014) and Noven et al. (2018) use the inverse* shape $\alpha = \xi^{-1}$ *and* $\beta = \sigma/|\xi| - \kappa$ *(Notation 2). The absolute Jacobian determinant* $|J|$ *from* $(\alpha, \beta, \kappa, \boldsymbol{\rho})$ *to* $(\xi, \sigma, \kappa, \boldsymbol{\rho})$ *verifies* $|J| = |\xi|^{-3}$*. When considering the latent processes, we mainly use* (α, β)*. However, when discussing the X-Vine model in Sections 3 and 4, we use the established parametrisation* (ξ, σ).

We define the multivariate extreme value model as follows:

Definition 4. *(Extremes Vine model or X-Vine)*
Let $h > 0$ *and* $(X_t^{(i)},\ t \geq 0) \subset [0, \infty)$ *for any* $i \in I$ *be d exceedance processes observed simultaneously. Additionally, suppose the following:*

- *there exist d Gamma*$(1/\xi^{(i)}, \sigma^{(i)}/\xi^{(i)} - \kappa^{(i)})$*-distributed latent processes* $(\Lambda_t^{(i)})$ *for* $i \in I$ *and such that,* $((X_t^{(i)},\ \Lambda_t^{(i)}),\ t \geq 0)$ *follows a univariate latent extreme value model (Section 3.2) with respective parameters vectors* $\boldsymbol{\theta}^{(i)} = (\xi^{(i)}, \sigma^{(i)}, \kappa^{(i)}, \boldsymbol{\rho}^{(i)}) \in \{(x_1, x_2, x_3, \boldsymbol{x}_4) \in \mathbb{R} \times [0, \infty) \times [0, \infty) \times \mathbb{R}^p : x_2 > |x_1| \cdot x_3\}$ *where* $dim(\boldsymbol{\rho}^{(i)}) = p$ *for* $i \in I$.
- *there exist d stationary R-Vine copulas with densities* $(C^h_{|i}(\cdot|\eta_{|i}),\ i \in I)$ *and respective finite-dimensional parameter vector* $\boldsymbol{\eta} = (\boldsymbol{\eta}_{|1}, \ldots, \boldsymbol{\eta}_{|d})$ *such that for any* $j \in I$*:*

$$f\left(X_{t+h}^{(1)}, \ldots, X_{t+h}^{(d)}, X_t^{(j)} | X_t^{(j)} > v^{(j)}\right) = C^h_{|i}\left(X_{t+h}^{(1)}, \ldots, X_{t+h}^{(d)}, X_t^{(i)} | X_t^{(i)} > v^{(i)};\ \boldsymbol{\eta}_{|j}\right) \times \prod_{j=1}^{d} f_{X^{(j)}}\left(X_{t+h}^{(j)}\right) \times f_{X^{(i)}}\left(X_t^{(i)}\right).$$

for any arbitrary $t \geq 0$ *by stationarity.*

Then, we say that the d random processes $(X_t^{(i)},\ t \geq 0)$ *for* $i \in I$ *verify the X-Vine model with marginal parameters* $\boldsymbol{\theta} \subset \mathbb{R}^{(3+p)d}$ *and vine copula* $C^h_{|i}$ *and its parameter vector* $\boldsymbol{\eta}$.

We clarify in the next section what one means by the term *univariate latent extreme value model* in the following subsection.

3.2 Exponential with Gamma-distributed parameters

Bortot & Gaetan (2014) and Noven et al. (2018) used a particular trick where the GPD can be represented as a mixture of exponential processes with Gamma-distributed intensities. More precisely, they define a Gamma-distributed latent process $(\Lambda_t,\ t \geq 0)$ and consider the density of X_t conditional on the value of Λ_t, for $t \geq 0$, as follows

$$f(x_t|\lambda_t) = \begin{cases} 1 - \exp\{-\kappa\lambda_t\}, & \text{if } x_t = 0, \\ \exp\{-\kappa\lambda_t\}\lambda_t \exp\{-x_t\lambda_t\}, & \text{if } x_t > 0, \end{cases} \tag{2}$$

where λ_t is a realisation of $\Lambda_t \sim \text{Gamma}(\alpha, \beta)$ and $\alpha, \beta, \kappa > 0$. This *trick* is well-defined under the assumption that $\alpha > 0$; for $\alpha \leq 0$, see Section 4.1.1.

Notation 5. *We call the exponential with Gamma-distributed parameters mixture as in* (2) *an extreme value model.*

Noven et al. (2018) showed that in the framework, the exceedances $\{X_t|X_t > 0\}$ do have a GPD marginal law. More precisely, they showed that for any $t \geq 0$:

$$\{X_t|X_t > 0\} \sim \text{GPD}(\alpha^{-1}, (\beta + \kappa)/|\alpha|), \quad \text{where } \alpha, \beta, \kappa > 0. \tag{3}$$

which we reformulate into $\{X_t|X_t > 0\} \sim \text{GPD}(\xi, \sigma)$ with $\xi := \alpha^{-1}$ and $\sigma := (\beta + \kappa)/|\alpha|$ following the parametrisation study in Courgeau & Veraart (2018) and Remark 2.

3.3 Asymptotic dependence and independence of extremes

As explained on page 500, Heffernan & Tawn (2004), upper multivariate extremes are either asymptotically dependent or asymptotically independent extremes depending if the tail coefficient

$$\chi^{(i)} := \lim_{x \to x_F^{(i)}} \mathbb{P}\left\{\boldsymbol{X}_{-i} > x\mathbf{1}_{d-1} | X^{(i)} > x\right\}, \tag{4}$$

is either non-zero or zero, respectively. Here, $x_F^{(i)}$ is the i-th upper endpoint and we use $\mathbf{1}_{d-1} \in \mathbb{R}^{d-1}$ a vector of ones, $\boldsymbol{X}_{-i} \in \mathbb{R}^{d-1}$ is the random vector $\boldsymbol{X} := (X^{(1)}, \ldots, X^{(d)})$ without the i-th component.

Classical multivariate extreme value models assume that extremes are either asymptotically dependent or asymptotically independent with the additional assumption that, when $\boldsymbol{X} \in C$, all components should also be extreme at the same time (Heffernan & Tawn 2004, Ledford & Tawn 1996). Recent developments avoid this limitation (Winter & Tawn 2016, 2017). Our model goes in the same direction as vines adapt their structure to accommodate the two different regimes (Section 3.3.2).

3.3.1 An alternative to the Markov Chain approach

Different attempts on time-clustering of extremes using Markov Chains have been introduced in the last twenty years. Smith et al. (1997) uses a first-order Markov Chain under the assumption of that only asymptotic dependence is possible at lag 1 hence at all lags which is a severe constraint. Then it was followed by a k-th order Markov Chain (Ribatet et al. 2009, Yun 2000) under asymptotic dependence only. Those limitations have vanished since then (Tendijck et al. 2019, Winter & Tawn 2017) and allow for both asymptotic dependence and independence.

The model presented in this article presents an alternative using stationary latent processes. We take the specific class of trawl processes allow to take deterministic functions to define the autocorrelation functions. For instance, if we choose the family of exponential trawls, a marginal autocorrelation function is $A(t) = e^{-\rho t}$ for $\rho > 0$. The variety of trawls at hand can be used to model mid to long range dependencies.

Remark 3. *The order of the Markov Chain is an important discussion in the afore-mentioned papers, from partial ACF plots interpretation to statistical estimation. We hereby propose to circumvent this issue using flexible parametric autocorrelation functions (Section 3.4.2).*

We now detail the theoretical justification for the use of vine copulas a scalable dependence structure for multivariate extreme values.

3.3.2 Copulas and tail dependence functions

We detail the necessary assumption of the existence of a conditional survival function and a conditional upper tail dependence function as defined in Joe et al. (2010) which can be summarised as the following:

Summary 6. *The regular vine copula allows for asymptotic dependence and independence in the extremes if there is no (resp. at least one) asymptotically independent bivariate copula in the stack of conditional copulas linking two variables.*

More precisely, copulas have the opposite behaviour as usual multivariate value models (Heffernan & Tawn 2004, Ledford & Tawn 1996, Winter & Tawn 2017) since it is somewhat more difficult to achieve asymptotic dependence than independence with copulas.

Asymptotic independence: as per the following remark from page 265, Joe et al. (2010):

> If some baseline copulas are tail independent then the (D-)vine copula is tail independent. Some margins of the (D-)vine, however, might still be tail dependent. In this situation, whether or not a margin of a (D-)vine copula is tail dependent depends not only on tail dependence of basic linking copulas at level l, $2 \leq l \leq d-1$, but also the rates of approaching zero for conditional tail dependence functions of the baseline copulas with tail independence.

A bivariate copula C is said to verify the (upper) asymptotic linear condition $\partial \bar{C}(u,v)/\partial v \approx uS(v)$, as $u \to 0$ and for a positive continuous bounded function S. The independent copula $C(u,v) = uv$ is an example of such a copula, henceforth upper tail independent. Therefore, using Remark 5, our model allows for asymptotic dependence (resp. independence) if the stack of conditional copulas are *all* upper tail dependent (resp. if there exists a upper tail independent bivariate copula). Even though our model could exploit marginal-specific extreme convergence rates $\boldsymbol{w}_h$ and w_0, we chose the same quantile probability for all marginals in Sections 6 and 7 for simplicity. More precisely, we set $\boldsymbol{w}_h = \mathbf{1}_d$ and $w_0 = 1$.

Asymptotic dependence: recall that $I := \{1, \dots, d\}$ and that $p^{(i)} = F_{X^{(i)}}(0)$ for $i \in I$, the probability of not having an extreme. Suppose the strict stationarity of $(V_{t+h}^{(1)}, \dots, V_{t+h}^{(d)})$ given $E^{(i)} = \{V_t^{(i)} > p^{(i)}\}$ for all $t \geq 0$, in the sense that the conditional distribution of $(V_{t+h}^{(1)}, \dots, V_{t+h}^{(d)}) | V_t^{(i)} = v_t^{(i)}$ only depends on $h > 0$ as temporal argument and this for any $v_t^{(i)} > p^{(i)}$.

Notation 7. *We write $\boldsymbol{V}_h | V_0^{(i)}$ to designate $\boldsymbol{V}_{t+h} | V_t^{(i)}$ for any $t \geq 0$.*

We introduce the necessary notation for this discussion: consider the d-dimensional unconditional (and conditional) copula, denoted C_i^h (respectively $C_{|i}^h$), of $(\boldsymbol{V}_h, V_0^{(i)})$ (respectively $\boldsymbol{V}_h$ given $V_0^{(i)}$). Ignoring the parameters of the copula, we denote its survival function (respectively, conditional survival function) by $\bar{C}_i^h$ (resp. by $\bar{C}_{|i}^h$) and define those as follows:

$$\begin{cases} \bar{C}_i^h(v_h^{(1)}, \dots, v_h^{(d)}, v_0^{(i)}) & := \mathbb{P}\left\{ V_h^{(1)} \geq v_h^{(1)}, \dots, V_h^{(d)} \geq v_h^{(d)}, V_0^{(i)} \geq v^{(i)} \right\}, \\ \bar{C}_{|i}^h(v_h^{(1)}, \dots, v_h^{(d)} | v_0^{(i)}) & := \mathbb{P}\left\{ V_h^{(j)} \geq v_h^{(j)},\ j \in I | V_0^{(i)} \geq v^{(i)} \right\}. \end{cases}$$

Those definitions hold for any $(v_h^{(1)}, \dots, v_h^{(d)}, v_0^{(i)}) \in [0,1]^d$. In addition, denote the conditional upper tail dependence function (Equation (2.6), Joe et al. (2010)), as given by

$$k_{|i}(\boldsymbol{w}_h | w_0^{(i)}) := \lim_{v \downarrow 0} \bar{C}_{|i}^h \left(\boldsymbol{V}_h \geq \mathbf{1}_d - \boldsymbol{w}_h v,\ j \in I \middle| V_0^{(i)} = 1 - w_0^{(i)} v \right)$$

for any $\boldsymbol{w}_h = (w_h^{(1)}, \dots, w_h^{(d)}) \in [0;\infty)^d$, $w_0^{(i)} \in [0, \infty)$ and any $h > 0$. Note, in terms of probability, one has $k_{|i}(\boldsymbol{w}_h | w_0^{(i)}) = \lim_{v \downarrow 0} \mathbb{P}\left\{ V_h^{(j)} \geq 1 - w_h^{(i)} v, j \in I \mid V_0^{(i)} \geq 1 - w_0^{(i)} v \right\}$.

Definition 8. *We say that $\boldsymbol{c}_{|i}$ is (upper) tail dependent if $k_{|i}(\boldsymbol{w}_h, w_0)$ is non-zero for some $(\boldsymbol{w}_h, w_0)$.*

As pinpointed in Section 2, Joe et al. (2010), $\boldsymbol{w} \mapsto k_{|i}(\boldsymbol{w} | w_0^{(i)})$ may be a *sub*-distribution function (i.e. a usual distribution function with potentially some mass at ∞). Therefore, conditional on $E^{(i)} = \{V_t^{(i)} > p^{(i)}\}$, we recreate uniforms on $[0,1]$ marginally using empirical CDF.

Remark 4. *The vector $\boldsymbol{w}_h$ as well as $w_t^{(i)}$ allows variables to become extreme at different rates, which useful in practice. Also, we take conditional copulas and tail dependence functions since we are interested in conditioning on $E^{(i)}$ for $p^{(i)}$ close to 1, for all $i \in I$.*

For this limit to be well-defined, it is usual to suppose that $(Y_t^{(1)}, \dots, Y_t^{(d)})$ has the multivariate regular property for every $t \geq 0$ (Section 2, Joe et al. (2010); Chapter 5, Resnick (2013)).

A key contribution of Joe et al. (2010) concerns the tail dependence functions for vine copulas with Archimedean bivariate copulas that are *regularly varying at 1*. In this context, a real-valued function $g : [0, \infty) \mapsto [0, \infty)$ is said to be regularly varying at 1 with tail index $\gamma > 1$ if $g(1 - 1/(sx))/g(1 - 1/x) \to s^{-\gamma}$ as $x \to \infty$. Suppose we have an Archimedean copula $C(v_h^{(1)}, \dots, v_h^{(d)}, v_0^{(i)}) = \phi\left(\sum_{j=1}^d \phi^{-1}(v_h^{(j)}) + \phi^{-1}(v_0^{(i)}) \right)$, where the inverse Laplace transform ϕ^{-1} is regularly varying at 1 with tail index $\zeta > 1$. Then, Proposition 3.3, Joe et al. (2010) yields that the conditional upper tail dependence function given $w_0^{(i)}$ is such that

$$k_{|i}(\boldsymbol{w}_h | w_0^{(i)}) = \bar{C}_{|i}^{Clayton} \left(1 - k_{|i}(w_h^{(j)} | w_0^{(i)}),\ j \in I; \zeta/(\zeta - 1) \right),$$

where $\bar{C}_{|i}^{Clayton}$ is the survival function of the Clayton copula $\boldsymbol{C}_{|i}^{Clayton}(u^{(1)}, \dots, u^{(d)}; \delta) = \left(\sum_{k=1}^d (u^{(k)})^{-\delta} - (d-1) \right)^{-1/\delta}$ for $\delta > -1$, $\delta \neq 0$. Finally, Equation (3.9), Joe et al. (2010) yields

$$k_{|i}(\boldsymbol{w}_h | w_0^{(i)}) = \left(\sum_{k=1}^d (w_h^{(j)}/w_0^{(i)})^{\zeta} + 1 \right)^{-(\zeta-1)/\zeta} > 0.$$

This implies *asymptotic dependence* in the Archimedean framework (in the sense of Equation (4)). Although it is only presented for a single $(d+1)$-dimensional Archimedean copula but can be extended to vine copulas as proved in Joe et al. (2010).

Remark 5. *In Sections 6 and 7, we use Independent, Gumbel and Clayton copulas, which all verify the regular variation assumption as proved in Chapter 4 (Table 4.1), Nelsen (2006).*

3.4 Choice of the latent process

We highlight that the latent process formulation is very general: the only two requirements are the Gamma distribution and the (strict) stationarity. In principle, one could choose any such process; such as a Markov chain as explored in Bortot & Gaetan (2014, 2016); or with mid to long range dependence via other trawl functions (as opposed to the exponential trawls herein explored).

3.4.1 Transferring the dependence to the latent process

Noven et al. (2018) cast the dependence to the latent process by imposing the independence of distinct observations X_j conditional on values taken by the latent process Λ_j. This shifts the dependence structure of the joint distribution $(X_1, \dots, X_l)$ given $(\Lambda_1, \dots, \Lambda_l)$ for any positive $l \in \mathbb{N}$ onto the latent process itself and it no longer relies on the original *observed* processes

$$f(X_1, \dots, X_l | \Lambda_1, \dots, \Lambda_l) = \prod_{j=1}^{l} f(X_j | \Lambda_j), \qquad \text{for any } l > 0. \tag{5}$$

For $k \in \mathbb{N}$, Equation (5) yields that any k-dimensional density of the stochastic process $(X_t,\ t \geq 0)$ depends on the joint distribution of $(\Lambda_1, \dots, \Lambda_k)$

Remark 6. *Note that we only use a discretised version of this model since the data is always taken to be discrete as per the case studies in Section 6.*

Now, we precise the *particular* choice of trawl processes as a follow-up to Courgeau & Veraart (2018), Noven et al. (2018). One of the main advantages is that one needs a deterministic function is required in addition to the aforementioned two conditions to fully describe a trawl process. Also, we can fully determine the joint distribution $(\Lambda_1, \ldots, \Lambda_k)$.

3.4.2 Trawl processes

We consider the marginal layer of the X-Vine model as part of a class of stationary and infinitely divisible stochastic processes called *trawl processes*, first introduced in Barndorff-Nielsen (2011). We believe that the necessary notions to formally define this model are strictly the same to Noven et al. (2018) and Courgeau & Veraart (2018). We expose the necessary definitions in Appendix A with definitions of infinitely divisibility (ID), independent scattering, Lévy basis and seed involved in this section. The single most important assumption is the *stationarity* of the model.

An essential result is the correspondence between trawl processes and ID laws as per the following proposition (see Rajput & Rosinski (1989) and Barndorff-Nielsen (2011)).

Proposition 1. *For any infinitely divisible law, one can find a trawl process with the same marginal law (but potentially featuring some serial dependence).*

Finally, we define the class of trawl processes as in Barndorff-Nielsen (2011), Courgeau & Veraart (2018), Noven et al. (2018):

Definition 9. *(Trawl processes)*
Let $A \in \mathcal{B}_b(\mathbb{R}^2)$ and consider the collection of sets $(A_t)_{t\in\mathbb{R}}$ defined as $A_t := \{(x, s) : (x, s-t) \in A\}$ for any $t \in \mathbb{R}$, that is A_t is the set A with the second component shifted by t. Let L be an homogeneous Lévy basis. The trawl process $(\Lambda_t)_{t\in\mathbb{R}}$ associated with the set A with respect to the Lévy basis is defined by setting $\Lambda_t := L(A_t)$.

To choose the set A which characterises partly the trawl process (along with the Lévy seed L'), we use both the notions of trawl *functions* and *sets* as follows:

Definition 10. *(Trawl function and trawl set)*
Define a trawl function *d as a non-negative function defined on $]-\infty; 0]$. Additionally, for $t \in \mathbb{R}$ a trawl set indexed at time t is set to be as follows $A_t := \{(x, s) : 0 \le x \le g(s-t),\ s \le t\} \subset \mathbb{R}^2$. In particular, we say that a trawl set is* monotonic *if the underlying trawl function $g(\cdot)$ is itself monotonic.*

A well-studied example for trawl are exponential trawl as presented in Courgeau & Veraart (2018), Dupuis et al. (2019), Noven et al. (2018):

Example 1. *(Exponential trawls)*
An example of monotonic trawls we are going to treat throughout this paper is the so-called exponential trawl. Namely, define the exponential trawl *sets with parameter $\rho > 0$ as the family of sets defined for any $t \in \mathbb{R}$ where $A_t := \{(x, s) : 0 \le x \le \exp(\rho(s-t)),\ s \le t\}$. Here, $g(u) := \exp(\rho u)$ for $u \le 0$. Informally, this family is suitable to model* short-range to mid-range *dependence of a process.*

To benefit from the assumption of Eq. 5, we set $(\Lambda_t, t \ge 0)$ to be a trawl process with Lévy seed L' and trawl set A as in Noven et al. (2018). That is to say that we require

$$L' \sim \text{Gamma}(\alpha/\mu^{leb}(A), \beta),$$

where $\mu^{leb}(A)$ is the Lebesgue measure of the trawl set A. From the trawl process theory, the resulting trawl process $\Lambda_t := L(A_t)$ is Gamma(α, β)-distributed. This equips the model with a flexible time-dependence structure given by the trawl process itself. This entirely specifies by the trawl set A and the Lévy seed L' (see Appendix A or Lemma 2.6, Noven et al. (2018)).

3.5 Choice of copulas

The multidimensional layer of the X-Vine model relies on the adaptive structure that is a *regular vine* copula. More broadly, vine copulas form a field on its own (Czado 2010, Morales Napoles et al. 2010, Nagler et al. 2018, Stöber & Czado 2014) with extensions to Lévy processes in Grothe & Nicklas (2013). The reasoning behind our use of such structure is the ability to grow the number of parameters describing pair interactions as the number of variables increases. Then, information criteria help control the sparsity of the model if necessary whilst ensuring a strong fit. We introduce such objects and then discuss the means at our disposal to transform the data accordingly and then how to perform model selection and inference.

3.5.1 Regular vine copulas using bivariate distributional copulas

We introduce the concept of regular vine copulas (or *R-Vine*) and take the following as definition :

Definition 11. *(Regular vine (R-Vine) as in Stöber & Czado (2014))*
Let $d \ge 2$ be the number of variables. A regular vine tree structure is an ordered sequence of trees $\mathcal{T} := (T_1, \ldots, T_{d-1})$ with $T_m := (N_m,\ E_m)$, $m \in \{1, \ldots, d-1\}$ where N_m is the set of nodes and E_m the set of edges of the m-th tree, such that:

1. *$N_1 := I$ i.e. the first tree has nodes 1,..., d;*
2. *For $m = 2, \ldots, d-1$, $N_m := E_{m-1}$, i.e. the nodes of T_m are the edges of T_{m-1}*
3. *Proximity condition: If, for $i \in \{1, \ldots, d-2\}$, two nodes of T_{m+1} are connected, the corresponding edges in T_m have a common node.*

$(C, \mathcal{T})$ is an regular vine copula if $F = (F_1, ..., F_d)$ is a vector of continuous invertible univariate distribution functions, V is an d-dimensional R-vine tree structure and $C = \{C_e : e \in E_m,\ m = 1, \ldots, d-1\}$ is a set of copulas with C_e being a bivariate copula, a so-called pair-copula.

In theory, each pair copula of the vine is fitted conditional on the uniform bivariate distribution of the previous level. However, it is usual in the literature to make the assumption that we only fit given the uniform univariate marginals themselves to make the computation more tractable.

3.5.2 Existence and uniqueness of vine copulas

A vine copula can be seen as a *proper* hierarchical copula and the celebrated Sklar's theorem still applies in this context as it boils down to the existence and uniqueness of a bivariate copula as original proved in Sklar (1973) and explained in Section 4.2, Czado (2010). The vine copulas made of given bivariate copula families provide a multi-parameter augmented coverage of a subspace of d-dimensional copulas of the given families. As a special case of Archimedean copulas, we quote Section 4.2.2, Czado (2010):

This construction of multivariate distributions and copulas is very general and flexible, since we can use any bivariate copula as building block in the PCC model. In contrast to the extended multivariate Archimedean copulas no restriction to the Archimedean pair-copulas or further parameter restrictions are necessary.

4 Inference

We detail in this section the inference strategies used for the X-Vine model. It will be constructed as a two-layer object: first, we apply independently the latent trawl model from Noven et al. (2018) on each exceedance marginal using (log) pairwise-likelihood maximisation. This accounts for time-clustering of extremes. Then, given the fitted cumulative distribution function, we apply a regular vine model (with either Gumbel, Clayton or independent bivariate copulas). This allows to capture tail dependence across the marginals themselves. We control the fitting performance and the number of parameters included in this second layer using information criterion such as AIC or BIC.

4.1 Marginal model inference

To perform the marginal inference, we use the so-called *pairwise likelihood (PL) approach* as in Bortot & Gaetan (2014), Noven et al. (2018). As the name suggests, it allows to make inference by using pairs which are themselves organised in independent blocks. Following Remark 2, we would like to maximise the log-PL likelihood defined as follows:

$$\ln(f_{PL}(\boldsymbol{\theta};\mathbf{x})) = |J(\boldsymbol{\theta})| \sum_{i=1}^{N-1} \sum_{k=i+1}^{max(i+\Delta,N)} \ln(f(x_i, x_k; \boldsymbol{\theta})), \quad \text{with } \boldsymbol{\theta} = (\xi, \sigma, \kappa, \rho),$$

for which Noven et al. (2018) gave closed form expressions for the pair likelihoods $(x_s, x_t) \to f(x_s, x_t; \boldsymbol{\theta})$ for $s, t \geq 0$. The transformation Jacobian from $(\alpha, \beta, \kappa, \rho)$ to $(\xi, \sigma, \kappa, \rho)$ is $|J(\boldsymbol{\theta})| = |\xi|^{-3}$ as per Remark 2.

Remark 7. *Δ is often referred to as the cluster size or block depth depending if we focus on the inference or interpretation of the model, respectively.*

To speed up the inference, we take the approach of Dupuis et al. (2019) that we detail. First, we take marginal MLEs to estimate $(\xi^{(i)}, \sigma^{(i)})$ for $i \in I$ i.e.

$$\left(\widehat{\xi^{(i)}}_{ML}, \widehat{\sigma^{(i)}}_{ML}\right) = \arg \max_{\xi^{(i)}, \sigma^{(i)}} \sum_{k=1}^{N} \ln \left\{ f_{GPD}\left(x_k^{(i)}; \xi^{(i)}, \sigma^{(i)}\right) \right\}.$$

Under standard regularity conditions (for $\xi > -1/2$, Smith (1985)), those estimates are consistent and jointly normal with a heteroskedasticity and autocorrelation consistent covariance matrix (using the HAC estimator, Newey & West (1986)). Since we have

$$p^{(i)} = \left(1 + \kappa^{(i)}/(\sigma^{(i)}/|\xi^{(i)}| - \kappa^{(i)})\right)^{-1/\xi^{(i)}},$$

we estimate $\kappa^{(i)}$ and denote this estimator $\widehat{\kappa^{(i)}}$. Then, we perform the pairwise-likelihood maximisation for $\boldsymbol{\rho}$

$$\widehat{\boldsymbol{\rho}^{(i)}}_{PL} := \arg \max_{\boldsymbol{\rho}} \left\{ \ln(f_{PL}(\widehat{\xi^{(i)}}_{ML}, \widehat{\sigma^{(i)}}_{ML}, \widehat{\kappa^{(i)}}, \boldsymbol{\rho}; \mathbf{x}) \right\}.$$

Note that this final estimator is dependent on the block depth $\Delta > 1$ that we did not include for clarity. Dupuis et al. (2019) proposes to see this estimation procedure as two-step generalized method of moments (Newey & McFadden 1994) to obtain an estimator for the variance. We bootstrap standard errors using sub-sequences (Patton et al. 2009, Politis & White 2004).

4.1.1 Inferential issues

Courgeau & Veraart (2018) explored the inference of the univariate model and we have exposed a weak identifiability issue when the parametrisation $(\alpha, \beta, \kappa, \boldsymbol{\rho})$ is used. Therefore, we have chosen to use the parametrisation adapted from the GPD parametrisation of Coles et al. (2001). Also, for $\xi = 1/\alpha < 0$, one should use a marginal transformation (see Section 4.3, Courgeau & Veraart (2018)) towards a marginal with positive shape parameter ν. When using such transformation, we impose $\nu = 2$ for second moment finiteness.

4.1.2 Threshold selection

Choosing the right extreme value threshold is a difficult problem on its own. Since we are not trying to solve this problem in this work, we use a threshold selection procedure from Bader et al. (2018) as well as Anderson-Darling (AD) and Cramèr-von Mises (CvM) GPD fit tests with a 5% significant level on the selected thresholds. We would like to thank the author for making their code available. Although the said approach is said to be *automated* we verified each threshold manually with mean-excess plots and marginal hypothesis tests (as per Dutfoy et al. (2014) and not included for conciseness). This was justified by the quote from page 404, Davison & Smith (1990):

> Extreme value analysis should never be performed automatically, without the intervention of the data analyst, because failure to spot unusual features of the data could have serious consequences.

Also, we take the largest exceedance probability available across marginals as yielded by the afore-mentioned procedure for simplicity purposes. Note that the model would work similarly for marginal-specific probabilities of extremes.

4.2 Cross-sectional model inference

We would like to model the dependencies across variables in the $(d+1)$-dimensional vector $(\boldsymbol{X}_{t+h}, X_t^{(i)})$ given that $X_t^{(i)} > 0$ for any $t \geq 0$ separately from the marginals in flexible way. More precisely, we want that the number of dependence parameters growing along with the number of variables involved.

Problem 12. *A limitation of traditional copulas is the incapacity to scale up with the number of variables since those copulas are often parametrised with very few parameters (one or two).*

Regular vine construction of copulas is performed using a *constellation of parametric bivariate dependence functions* (Grothe & Nicklas 2013), see Definition 11.The regular vines pair construction is detailed in Section 4.2.2. Since copulas are designed to capture dependency, a common measure of dependence is Kendall's τ (see Embrechts et al. (2013) for definition). Refer to Stöber & Czado (2014) for formal definitions of a copula.

Remark 8. *A key argument is that we know that there exist structural conditions on vines (Section 3.3.2) under which we can have either asymptotic dependence or independence, similarly to the model from Heffernan & Tawn (2004).*

Example 2. *Three common distributional copulas examples (as found in Czado (2010)) with $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ a vector of uniformly distributed components are as follows*

- *Independent family: $C^{indep}(u^{(1)}, \ldots, u^{(d)}) := \prod_{i=1}^{d} u^{(i)}$. It is the most trivial example.*
- *Clayton family: $C^{Clayton}(u^{(1)}, \ldots, u^{(d)}; \theta) := \max\left\{ \left(\sum_{i=1}^{d} (u^{(i)})^{-\theta} - d + 1\right)^{-1/\theta}, 0 \right\}$. for $\theta \in [-1, \infty)\backslash\{0\}$. We have $\theta = 2\tau/(1-\tau)$ where τ is Kendall's rank correlation coefficient.*

- *Gumbel family:* $C^{Gumbel}(u^{(1)},\dots,u^{(d)};\theta) := \exp\left\{\left\{\sum_{i=1}^{d}\log(u^{(i)})^{-\theta}\right\}^{-1/\theta}\right\}$ *for* $\theta \in [1,\infty)$.

  In this case, we have $\theta = 1/(1-\tau)$.

Remark 9. *To obtain negative dependence between variables, there are so-called* rotated *copulas to capture better positive and negative dependency: to do so, we use the transformation* $u \mapsto 1-u$ *on either one or both variables to rotate the scatter plot and fall back to the case where* $\tau > 0$.

4.2.1 Dealing with the atom at zero

Given our definition of exceedance process, the corresponding CDF will not be continuous at zero as required by the standard theory of (continuous) copulas (Czado 2010). We perform a pre-processing of the data Therefore, the discrete copulas literature proposes different transformations to cope with this problem. The simplest one is the so-called R-distribution transform, from Definition 2.1 Faugeras (2012), which we adapt to our special case and present as a remark:

Remark 10. *Let* $i \in I$. *In our model, we only have a single discontinuity in* $X^{(i)}$ *at zero: we impose to take a uniform sample between* 0 *and* $P\{X_s^{(i)} = 0,\ \forall s \geq 0\}$ *whenever* $X_t^{(i)} = 0$ *for some* $t \geq 0$, *and use the integral transform otherwise. This creates a time series of uniformly sampled values on* $[0,1]$ *which relies on the original ordering of the original observed values from* $(Y_t^{(i)},\ t \geq 0)$ *only for positive exceedance.*

It is crucial to note that we *lose important information* since only a small fraction of the data will be non-zero. To reconstruct a more insightful uniformly distributed random variable, we go back to Heffernan & Tawn (2004) and their semi-parametric integral transform

Definition 13. *Given a time-series* $(Y_t,\ t \geq 0)$, *a threshold* u *and its respective exceedance time-series* $(X_t,\ t \geq 0)$ *verifying the univariate latent trawl model with parameters* $(\alpha, \beta, \kappa, \boldsymbol{\rho})$. *Then, define the following semi-parametric integral transform:*

$$V_t := \begin{cases} \hat{F}_Y(Y_t), & \text{if } X_t = 0, \\ p + (1-p)\cdot F_{GPD(\xi,\sigma)}(X_t), & \text{if } X_t > 0, \end{cases}$$

where $p := \mathbb{P}\{X_t = 0,\ \forall t \geq 0\}$ *and* $y \mapsto \hat{F}_Y(y)$ *the empirical CDF of the marginal of* Y.

This transformation keeps more information in the ranking of responses since we are ultimately interested in obtaining larger uniform values for extreme and smaller uniform values for non-extremal behaviour which this scheme produces. Numerical experiments developed in Sections 6, 7 confirm that it preserves sufficient information to fit vine copulas conditional on an exceedance variable to be non-zero.

4.2.2 Vine structure and model selection

As described in Dissmann et al. (2013), we shall perform three tasks when *fitting* a regular vine: (a) select the vine structure, i.e. pick the unconditioned and conditioned pairs of random variables, (b) choose the bivariate copula family and finally, (c) estimate copula parameters. A greedy approach would select the best fitting vine copula among $d!/2 \cdot 2^{\binom{d-2}{2}}$ possible arrangements (Morales Napoles et al. (2010)) which is intractable at larger dimensions. Therefore, we follow the *sequential regular vine structure selection procedure* from Dissmann et al. (2013) based on Kendall' τ (Kendall (1938), Embrechts et al. (2013)). The vine structure selection is done via Kendall's τ maximisation on spanning trees: one forms a connected graph containing all nodes whilst having

the minimum number of edges with the condition of maximising the sum of Kendall's τ across the edges of the said spanning tree (Algorithm 3.1, Dissmann et al. (2013)). This algorithm is the one implemented in the R package `VineCopula` for R-Vines and will be used in Sections 6 and 7.

4.3 Goodness-of-fit tests

Schepsmeier (2015) present numerous methods to test the goodness of fit for regular vines. We focus on two of them: the first one compares the two matrices $\mathbb{H}(\boldsymbol{\eta}) := \mathbb{E}\left\{\partial^2 l(\boldsymbol{\eta};\mathbf{V})/\partial\boldsymbol{\eta}^2\right\}$ and $\mathbb{C}(\boldsymbol{\eta}) := \mathbb{E}\left\{(\partial l(\boldsymbol{\eta};\mathbf{V})/\partial\boldsymbol{\eta})\cdot(\partial l(\boldsymbol{\eta};\mathbf{V})/\partial\boldsymbol{\eta})^T\right\}$. Those are, respectively, the expected Hessian matrix of the random (vine) copula log-likelihood function $l(\boldsymbol{\eta};V) = \ln C((v^{(1)},\dots,v^{(d)});\boldsymbol{\eta})$ and the expected outer product of the corresponding score function. Here, $\mathbf{V} := (V^{(1)},\dots,V^{(d)})$ is a random vector with copula density $C(\cdot;\boldsymbol{\eta})$. More precisely, White's information matrix test (White 1982) has null hypothesis $H_0 : \mathbb{H}(\boldsymbol{\eta}) + \mathbb{C}(\boldsymbol{\eta}) = 0$ against $H_1 : \mathbb{H}(\boldsymbol{\eta}) + \mathbb{C}(\boldsymbol{\eta}) \neq 0$. See Section 4.1, Schepsmeier (2015) for a full description of the testing statistics. In addition, we use the Empirical Copula Process (ECP) test (Aas & Berg 2013) based on the multivariate Cramér-von Mises statistic (best performing one in Genest et al. (2009), Schepsmeier (2015)) with 2'500 bootstrap samples with $H_0 : C \in \mathcal{C}_0$, where C is the empirical copula and the $\mathcal{C}_0$ is the class of copulas materialised by the bootstrap of the fitted vine copula. An alternative for high-dimensional problems is the *Berg* test Berg & Bakken (2007) with $\alpha = 2$. All the p-values were bootstrapped with 2'500 samples.

4.4 Improving computational efficiency

Since the seminal framework of Heffernan & Tawn (2004) is still used in recent works (Tendijck et al. 2019, Winter & Tawn 2016, 2017), one might ask why we did not pursue with their modelling idea with a temporal extension. Although the approach from Heffernan & Tawn (2004) is very concise and powerful, it requires to compute d pairs of (d-1)-dimensional location-scale functions $(a_{|k}(\cdot), b_{|k}(\cdot))$ as given in Equation (3.8), Heffernan & Tawn (2004) which corresponds to performing an optimisation on $4d(d-1) = O(d^2)$ parameters at once (for $d = 6$, we would have $4d(d-1) = 120$). For the X-Vine model, we have a similar number of parameters in total: $4d$ parameters for the marginals and at most $O(d^2)$ for the R-Vine part (for the first vine layer, we have at most $d-1$ parameters, then $d-2$ for the second layer, etc until the vine apex (with has at most one parameter)). From this perspective, both methods are similar. However, regular vine copulas are constructed *layer-wise*. That is to say that we start with pairs, fit the best bivariate copulas with respect to some goodness-of-fit measure, then fit the bivariate copulas of the resulting *combined* copulas and carry on this process until reaching the apex of the vine copula, see Czado (2010). Usually, the empirical Kendall's τ is used in the vine structure selection process, i.e. coupling of variables (Dissmann et al. (2013)), see Section 4.2.2. This is easily parallelised (as implemented in R packages like `VineCopula`). Compared to that, the approach from Heffernan & Tawn (2004) requires to optimise at least $4(d-1)$ parameters simultaneously for each of the d marginals.

5 Quantifying extreme propagation

To tackle the problem of extreme propagation through time and variables, we define the *Time-Rigged Occurrence Network*, or *TRON*, as the collection of probability maps defined as follows:

Definition 14. *(Time-Rigged Occurrence Network (TRON))*
Given d *stationary exceedance random processes* $(X_t^{(i)},\ t \geq 0)$ *for* $i \in I$ *and a set of thresholds* $\boldsymbol{v} := (v^{(1)},\dots,v^{(d)}) \subset [0,\infty)^d$, *we define the Time-Rigged Occurrence Network (TRON) with*

respect to the vector of thresholds $\boldsymbol{v}$:

$$TRON_{\boldsymbol{X}}(\boldsymbol{v}) := \left\{ p_{i\to(j,h)}(v^{(i)}, v^{(j)}) : i, j \in I,\ h \in [0, \infty) \right\},$$

where $p_{i\to(j,h)}(v^{(i)}, v^{(j)}) := \mathbb{P}\left\{ X_{t+h}^{(j)} > v^{(j)} | X_t^{(i)} > v^{(i)} \right\} = \mathbb{P}\left\{ X_h^{(j)} > v^{(j)} | X_0^{(i)} > v^{(i)} \right\}$. *Here,* $\boldsymbol{X}_t := (X_t^{(1)}, \dots, X_t^{(d)}) \subset \mathbb{R}^d$. *When processes and thresholds are explicitly fixed, the probability* $p_{i\to(j,h)}(v^{(i)}, v^{(j)})$ *is simply referred to as the* $TRON(i, j, h)$ *probability.*

Remark 11. *Usually, we will only consider the case* $\boldsymbol{v} = \boldsymbol{0}_d$.

TRON networks quantify extreme transition probabilities across pairs of variables (see Section 5.1). More precisely, it gives more insights on the dependencies in the extremes than the original method in Heffernan & Tawn (2004) which only estimated the probability of having an extreme in *any* of the variables.

5.1 TRON probabilities with the X-Vine model

We now derive the formulas and numerical recipes to compute TRON probabilities for the X-Vine model:

Theorem 15. *(TRON probabilities computation for X-Vine model)*
Let $(v^{(i)},\ i \in I) \subset [0, \infty)$ *be a set of thresholds. Under the X-Vine model framework (Definition 4), we compute the* $TRON(i, j, h)$ *probability with the formula*

$$p_{i\to(j,h)}(v^{(i)}, v^{(j)}) = \mathbb{E}\left\{ I\left(X_h^{(j)} > v^{(j)}\right) \;\middle|\; \left(X_h^{(i)}, X_h^{(j)}, X_0^{(i)};\ X_0^{(i)} > v^{(i)}\right) \right\},$$

where $\left(X_h^{(i)}, X_h^{(j)}, X_0^{(i)}\right)$ *given* $\left\{X_0^{(i)} > v^{(i)}\right\}$ *is simulated directly from the stationary vine copulas on the* d *observed processes.*
Finally for $i = j$*, the univariate model yields:*

$$p_{i\to(j,h)}(v^{(i)}, v^{(j)}) = \left(1 + (2\kappa + v^{(i)} + v^{(j)}))/\beta\right)^{b_{0,0+h}} \left(1 + (\kappa + v^{(j)})/\beta\right)^{b_{0\backslash h}} \left(1 + (\kappa + v^{(i)})/\beta\right)^{\alpha + b_{h\backslash 0}},$$

where $b_i = -\alpha\mu^{leb}(B_i)/\mu^{leb}(A)$ *for* $i \in \{(0\backslash h), (0, h), (h\backslash 0)\}$ *and* $B_{0\backslash h} = A_0\backslash A_h,\ B_{0,h} = A_0 \cap A_h,\ B_{h\backslash 0} = A_h\backslash A_0$.

Proof. See Appendix B.2. □

Example 3. *As per Example 1,* $b_{h\backslash 0} = b_{0\backslash h} = -\alpha(1 - e^{-\rho h})$*, and* $b_{0,h} = -\alpha e^{-\rho h}$.

We can provide a sort of important sampling procedure to further reduce the variance in the estimation of the probability as follows:

Algorithm 16. *By the properties of the GPD distribution (see Equation* (6) *below), note that* $X_t^{(i)} - v^{(i)} | \{X_t^{(i)} > v^{(i)},\ \Lambda_t\} \sim Exp(\Lambda_t)$*. Hence, the probability* $p_{i\to(j,h)}(v^{(i)}, v^{(j)})$ *can also be computed as follows:*

1. *Simulate* $X_0^{(i)}$ *given that* $\{X_0^{(i)} > v^{(i)}\}$*, say* $x_0^{(i)}$ *is the sample*

2. *Conditional on* $\{X_0^{(i)} = x_0^{(i)}\}$*, sample from the copulas to get samples for* $(X_h^{(1)}, \dots, X_h^{(d)})$*, as described in Bevacqua et al. (2017), Cooke et al. (2015), denoted by* $(x_h^{(1)}, \dots, x_h^{(d)})$.
3. *Repeat this process* $N > 0$ *times and compute* $N^{-1}\sum_{i=1}^{N} I(x_h^{(j)} > v^{(j)})$*, to estimate numerically the probability* $p_{i\to(j,h)}(v^{(i)}, v^{(j)})$ *for each* $j \in I$.

Remark 12. *When conditioning on* $\{X_0^{(i)} > v^{(i)}\}$*, we shall consider the equivalent of the quantile probability* $v^{(j)}$ *in the subset conditional on* $\{X_0^{(i)} > v^{(i)}\}$ *for a realisation to be considered extreme.*

Also, as presented in the next section, the univariate model is well-suited for fast estimation of those probabilities since it yields closed-form expressions for transition probabilities.

5.2 Marginal transition probabilities for the trawl case

Similarly to $\boldsymbol{v}_0$, assume $\boldsymbol{v}_h = (v_h^{(1)}, \dots, v_h^{(d)}) \subset [0, \infty)^d$. Each marginal is assumed to be following a latent trawl extreme value model (see Section 3). This means that we will have, for example, $(\alpha^{(i)}, \beta^{(i)}, \kappa^{(i)}, \rho^{(i)}) \in \mathbb{R} \times [0, \infty)^3$ for each random variable $X^{(i)}$ in the case of exponential trawl (with one trawl parameter). Recall that the Generalised Pareto distribution yields for $v^{(i)} > 0$:

$$\mathbb{P}\left\{ X_t^{(i)} > v^{(i)} \middle| X_t^{(i)} > 0 \right\} = \left(1 + (\kappa^{(i)} + v^{(i)})/\beta^{(i)}\right)^{-\alpha^{(i)}}, \tag{6}$$

for $i \in I$. Additionally, we need the marginal (temporal) transition probabilities given by

$$\mathbb{P}\left\{ X_h^{(i)} > w^{(i)} | X_0^{(i)} > v^{(i)} \right\} \text{ and } \mathbb{P}\left\{ X_h^{(i)} = 0 | X_0^{(i)} > v^{(i)} \right\},$$

which we formalise in the following:

Proposition 2. *(Marginal transition probabilities)*
Let $i \in I$*. Recall that* $(X_t^{(i)},\ t \geq 0)$ *follow the latent trawl extreme value model with parameters* $(\alpha, \beta, \kappa, \boldsymbol{\rho}) \in \mathbb{R}^{3+p}$ *where* $dim(\boldsymbol{\rho}) = p$ *and suppose* $h > 0$*. Then, for* $w^{(i)}, v^{(i)} > 0$*, we have that*

$$\mathbb{P}\left\{ X_h^{(i)} > w^{(i)} \,| X_0^{(i)} > v^{(i)} \right\} = \left(1 + (2\kappa + v^{(i)} + w^{(i)})/\beta\right)^{b_{0,h}} \left(1 + (\kappa + w^{(i)})/\beta\right)^{b_{0\backslash h}} \left(1 + (\kappa + v^{(i)})/\beta\right)^{\alpha + b_{h\backslash 0}}.$$

In a similar way, one has:

$$\mathbb{P}\left\{ X_h^{(i)} = 0 | X_0 > v^{(i)} \right\} = 1 - \left(1 + (2\kappa + v^{(i)})/\beta\right)^{b_{0,h}} \left(1 + \kappa/\beta\right)^{b_{h\backslash 0}} \left(1 + (\kappa + v^{(i)})/\beta\right)^{b_{0\backslash h} + \alpha}.$$

Proof. See Appendix B.1. □

6 Case study: London air pollution

We use CRAN R packages **CDVine** (Brechmann & Schepsmeier (2013)) for core capabilities and **VineCopula** to handle regular vines. The code used in the following case study can be found on GitHub as a public repository `multi-trawl-extremes`.

Remark 13. *All the chemicals reactions mentionned below are described in detail in in the related air pollution literature from the World Health Organization (Krzyzanowski & Cohen 2008).*

6.1 Introduction

In this section, we provide a novel approach to statistical modelling of air pollution with an application of the multivariate latent extreme value model with vine copulas to model the TRON probabilities between six main pollutants presented below. The data was provided by *King's College London Air Quality Network* which provides air pollution data on different timescales and pollutants for many locations in Greater London. Following Krzyzanowski & Cohen (2008), we present the changes in probability of extreme in time and in the six pollutants given an extreme in Ozone (O_3). We also present results given an extreme in Carbon Monoxide (CO) given its strong correlation with other variables.

Our dataset consists of hourly measurements of the six main air pollutants (Ozone (O_3), Nitrate Oxide (NO), Nitrate Dioxide (NO_2), Carbon Oxide (CO), Particulate Matter under 10 microns (PM_{10}) and Sulphur Dioxide (SO_2)) from 1^{st} January 2000 to 31^{st} December 2017 - approximately 157,710 entries. Those ground-level measurements are communicated in $\mu g/m^3$ as often presented in the literature (see public health reviews such as Krzyzanowski & Cohen (2008) or Atkinson et al. (2012)).

6.2 Data preparation

We notice that approximately 10% of the data points are invalid, we remove any row containing invalid entries. As one would expect, we clean data by fitting a linear model using an overall mean, weekly and twice-daily seasonality formulated using *sine* and *cosine* functions, dummy variables for each hour of the day, day of the week and quarter of the year (and removing one of each to ensure identifiability). Using Ordinary Least Squares, we only kept regressors achieving significance p-values below the 5% threshold of a standard t-test using Bonferonni's correction Bonferroni (1936).

Remark 14. *More advanced methods to remove seasonality or trends such as the famous Loess function, or also referred to as STL (Cleveland et al. (1990)), were not necessary in this case and allowed to keep a parametric formulation. In the second case study (Section 7), strong seasonality forced us this tool.*

We take the exceedances on the deseasonalised time-series Y_t and for that we estimate the correct threshold by using the standard mean-excess plot. We find that 95% quantiles are suitable thresholds for each time-series as presented in Table 1:

Agent	Thres	ξ	σ	ρ	κ	Δ	α	β
O_3	62.93	.10 (.038)	2.73 (.105)	.17 (.004)	7.04 (.477)	5	9.63 (.326)	19.2 (1.78)
NO	84.75	.10 (.027)	2.67 (.074)	.15 (.004)	6.89 (.308)	5	9.65 (.208)	18.9 (1.09)
NO_2	89.91	.23 (.004)	2.16 (.106)	.23 (.006)	4.63 (.271)	4	4.24 (.337)	4.52 (1.78)
CO	00.82	.08 (.002)	2.76 (.062)	.27 (.005)	7.29 (.271)	4	7.73 (.206)	13.5 (.964)
SO_2	13.53	.12 (.003)	2.38 (.077)	.17 (.006)	5.94 (.287)	4	11.69 (.189)	25.0 (.926)
PM_{10}	49.22	.32 (.003)	1.84 (.081)	.20 (.006)	3.53 (.331)	4	3.12 (.221)	2.14 (1.20)

Table 1: Threshold values using Bader et al. (2018) for deseasonalised hourly observations and univariate trawl-latent model parameters for London air pollution dataset. In parenthesis are block-bootstraped Monte Carlo standard deviations with 100 samples (see Patton et al. (2009) and the R package np for optimal block size).

Remark 15. *Note that Δ is the block depth, or cluster size, of the pairwise likelihood and was chosen by visual inspection of the QQ-plots and autocorrelation plots.*

Additionally, we observe that the scale of those 6 variables are different. To keep consistency across time-series, we normalise non-zero exceedances using their respective standard deviation

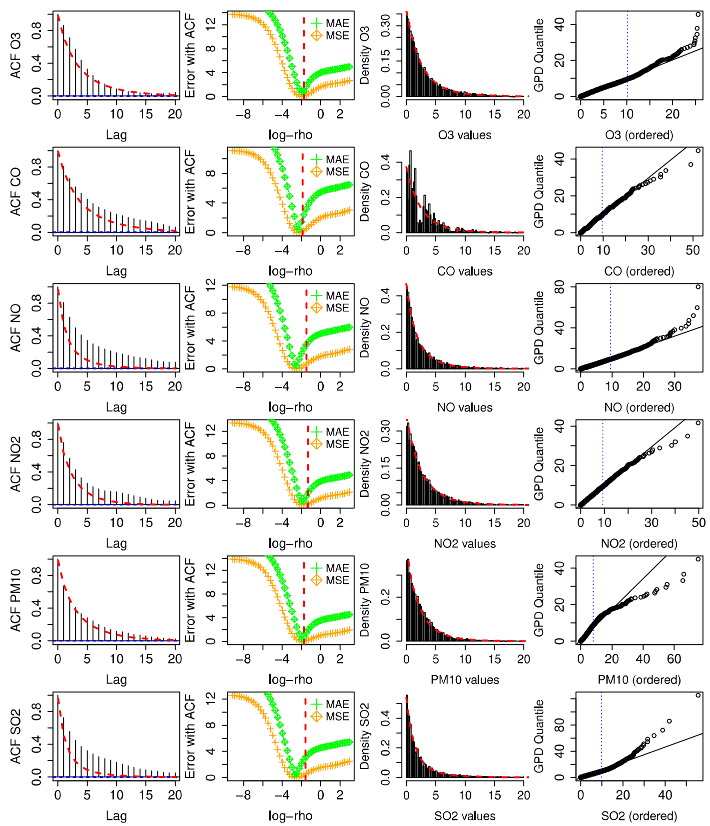

Figure 1: Column-wise: Empirical ACFs (theoretical ACF in red, cf Table 1), MAE and MSE errors between empirical and theoretical ACF (ρ indicated with red vertical line), Histogram of positive exceedances and the marginal fit in red, GPD QQ-plot with blue vertical line indicating the 95% exceedance quantile (i.e. 99.7% quantile for original time series).

to obtain time-series with variance 1. Figure 1 shows that pollutants have very similar auto-correlation dependencies although NO and SO_2 have longer ACF tails. Via 2-step marginal likelihood/pairwise likelihood maximisation from Dupuis et al. (2019), we find a good fit of GPD marginals for exceedances using histograms and QQ-plots (third and fourth columns), Figure 1 below. Indeed, QQ-plots are straight up to the blue lines - those lines indicate the 95% quantiles of positive exceedances, 99.7% quantile of the original dataset. Assuming we can extend the nonlinear expectations approach in Cohen et al. (2018) to GPD, we would need approximately millions of positive exceedances points to go beyond this point. The values of univariate models parameters are as follows:

We used three families of pair copulas: Clayton, Gumbel and Independent (and rotated versions) as they capture (upper) tail dependencies to compute TRON probabilities (as per Section 3.3.2).

6.3 Discussion

Table 1 informs us that the standardisation helped making the variables comparable due to similar model parameters. As noted above, NO and SO_2 have longer ACF tails which can be interpreted from low ρ values. We note that the ρ value for NO_2 is high, probably due to higher (α, β, κ) values. Across variables, ρ is close to minimising both mean squared error and mean absolute errors (Second column) although slightly over-estimating.
Regarding the vine copula conditional on extreme in Ozone (O_3), we note that all three families of copulas are used.

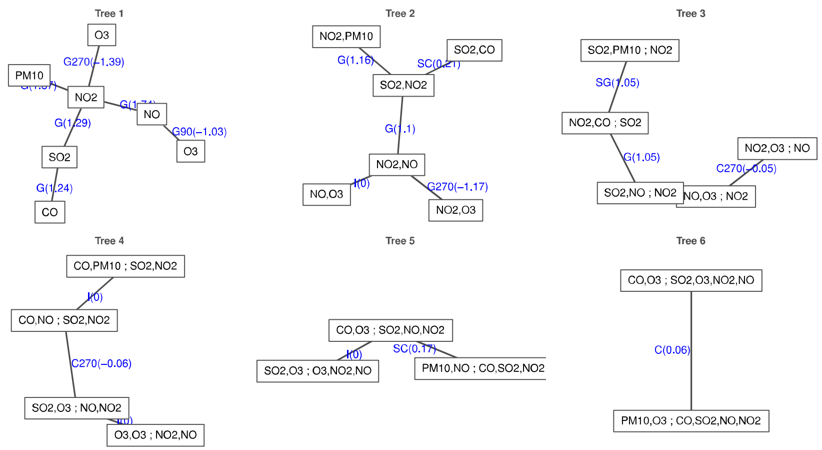

Figure 2: Unfolded vine copula conditional on extreme values in O_3 for horizon $h = 12hrs$ with corresponding extreme time-series denoted 'O3 ex'. In blue are the bivariate copula families followed by the parameter value. C for Clayton, G for Gumbel and I for Independent. SC, SG are 180°-rotated Clayton and Gumbel copulas.

The sequence of intertwined trees (Figure 2) in which Tree 1 has the original variables as nodes and the following trees take as nodes the edges of the previous tree (see Definition 11). Each edge is annotated in blue with the bivariate copula family used as well as the parameter value (here we only used 1-parameter pair copulas). It is interesting to note that although the one presented is conditional on extreme in O_3, this pollutant (either its copy at time t or $t + h$) does not have

a predominant weight in the first tree - however, NO plays this role all the way to Tree 5. We also note the very few independent copulas used suggesting relevant interactions even higher up in the vine. In a similar manner, allowing for rotated copulas gives some necessary flexibility for a good fit given the number of those transformations used throughout the tree structure. Finally, as often with vines, the parameter values get smaller (in the absolute sense) as one goes higher in the tree sequence as the dependencies between (conditional) pairs become very complex.
Recall that the unconditional probability is 5% = .05. A conditional probability close to this value may indicate that the conditioning is unnecessary in this case. Using simulation from the vine copula and Theorem 15, we obtain the numerical estimations of TRON probabilities as presented in Table 2 below.

Conditional on extreme in O3						
h	**O3**	**CO**	**NO**	**NO2**	**PM10**	**SO2**
1	.742 (.039)	.018 (.015)	.000 (.005)	.023 (.012)	.057 (.021)	.034 (.015)
2	.598 (.045)	.018 (.012)	.001 (.003)	.034 (.015)	.057 (.033)	.037 (.012)
3	.469 (.015)	.019 (.012)	.001 (.014)	.046 (.018)	.055 (.021)	.036 (.006)
6	.239 (.042)	.020 (.015)	.006 (.024)	.064 (.021)	.060 (.024)	.034 (.018)
12	.130 (.031)	.017 (.015)	.009 (.009)	.029 (.018)	.050 (.007)	.024 (.015)
24	.133 (.027)	.019 (.015)	.013 (.012)	.028 (.018)	.045 (.024)	.028 (.018)
Conditional on extreme in CO						
h	**O3**	**CO**	**NO**	**NO2**	**PM10**	**SO2**
1	.018 (.012)	.752 (.038)	.326 (.034)	.191 (.012)	.223 (.037)	.273 (.041)
2	.019 (.013)	.629 (.024)	.284 (.012)	.164 (.028)	.203 (.032)	.252 (.016)
3	.023 (.015)	.545 (.036)	.249 (.040)	.142 (.048)	.166 (.011)	.227 (.003)
6	.030 (.054)	.381 (.045)	.183 (.037)	.094 (.027)	.184 (.031)	.187 (.033)
12	.027 (.006)	.235 (.047)	.141 (.034)	.069 (.081)	.137 (.097)	.137 (.012)
24	.016 (.048)	.137 (.011)	.117 (.094)	.059 (.021)	.097 (.090)	.114 (.051)

Table 2: TRON probabilities conditional on extreme in O_3 and in CO for horizons $h \in \{1, 2, 3, 6, 12, 24\}$ (hours) using Theorem 15. Boostraped standard errors in parenthesis using 100'000 independent samples.

Table 2 is useful to see that the X-Vine does capture some dependencies between extremes as hinted by the TRON notation. Indeed, standard deviations being very low, it allows to appreciate the difference between the computed probabilities from the 5% probability it should have been if no dependency was to be observed - at least for most of them. More precisely, 99.7% confidence intervals do not include 5% for any of the pollutants except for PM_{10} in Table 2.

6.4 Summary of results

First, we acknowledge that the standard errors are very different between the two examples from Table 2 and TRON probabilities conditional on O_3 are more reliable; this may indicate that some conditioning make more physical sense than others. As explained in Krzyzanowski & Cohen (2008), Ozone (O_3) takes longer to appear after polluting episodes due to the chemistry involved to form it (via UV rays). This is captured by the model given that TRON probabilities are below 5% in Table 2 - except for O_3 (given by the univariate modelling perspective) - indicating that a spike in O_3 does not involve a future spike in any of the 5 other pollutants since underlying chemical transformations might already have taken place. R-Vines Goodness-of-fit tests (namely, White's test (White 1982) and ECP test (Aas & Berg 2013)) have p-values larger than 0.5 hence failing to reject the null hypothesis of a well-specified model (therefore we chose not include them).

Note that those tests were conducted on a 2'500 rows of data, for each one of the $d = 6$ vines and for all 6 horizons and bootstrapped with 2'500 samples.

However, in the second part of Table 2, TRON probabilities are high for NO, NO_2, PM_{10} and SO_2 across all horizons (although slowly decreasing) which showcases potential mechanics between pollutants since CO appears from non-complete combustion in engines (i.e. directly at the pollution source). As an example of separate mechanics, TRON probabilities $TRON(CO, NO_2, h)$ decrease twice as fast as for $TRON(CO, SO_2, h)$, from .191 down to .059 compared to .273 down .114. Therefore, our model is capable to capture a wide variety of dependencies between variables using a flexible mixed-structure model that seem to corroborate expert knowledge of the field in both serial and cross-sectional dependencies. The use of TRON probabilities - proven to be significantly different from the non-conditional approach of extremes - allows to communicate efficient on the matter. These are then very encouraging results towards deciphering air pollution risk mechanics and ensuring a good fit using standard tests for regular vines.

A note on the univariate fit: non-zero trawl parameter ρ as well as the good marginal fit of GPD on non-zero exceedances. QQ-plots are straight for all variables except PM_{10} for which it bends slightly well after the 95% quantile mark.

7 Case study: East coast weather-energy data

We propose a similar study as in Section 6 with a slight twist. We realise that the TRON probabilities give look-ahead dynamics of extremes but we can also be interested in finding out about the extremes that most likely led to an extreme in a variable of interest - namely, the `AEP` Energy Consumption. This time, we apply the X-Vine model to gain insights on the dynamics involved in the upper spikes of energy consumption on the United States North-East. Again, note that we are *only* interested in upper extremes and a similar study could be carried out for drops in the consumption instead of surges. More precisely, we look into the dynamics involved in the setting of an extreme consumption in `AEP`-managed territory which covers the states of Ohio and West Virginia. Intuitively, extremes in `AEP` and consumption in close areas (like `DUQ`) should increase the probability of extremes and this phenomenon is presented here. We also observe that North-South winds (sine components) of surrounding areas seems to play a preponderant role as well pressure. On the other hand, horizontal components of wind extremes, peaks in temperature and humidity are often related to a drop in extreme probability after the removal of seasonal components.

7.1 Introduction

The dataset is comprised of 5 years (01/10/2012-30/11/2017) worth of hourly data on humidity, pressure, temperature, wind speed and direction in six cities of the American North-East and Vancouver from OpenWeatherMap, under ODbL License (aggregated on `https://www.kaggle.com/selfishgene/h` We couple this with hourly energy consumption from PJM Interconnection LLC (PJM), a regional transmission organization (RTO) in the United States (`https://www.kaggle.com/robikscube/hourly-energy-consumption`). We keep only two variables of the latter dataset, namely the energy consumption as quoted in `AEP` Energy (Ohio and West Virgina) and `DUQ` - Duquesne Light Company (West Pennsylvania). The dataset consists of 32 variables spanned over around 43,000 timesteps, that we standardise to have a standard deviation of one marginally.

7.2 Data preparation

We notice that very few points (less than 5%) are not available and decide to correct of those using interpolation which does not affect the extreme values. Since we have strong seasonality in the

data (temperatures and electricity demand), we needed to perform two successive transformations of the data. The first one using the Loess function (or so-called STL transformation, Cleveland et al. (1990)) with a frequency of 24 (for hourly data), trend window of $24 \times 364/4 = 2,190$ hours to capture quarterly trends and a seasonal window of 24 hours (daily seasonality) and retained only the irregular part. Then, we used a similar linear model as in the previous case study, as detailed in Section 6.2, with dummy variables for hour of the day, day of the week, month, quarter, 24 and 48 hours seasonalities (via cosine / sine functions) and parameters where selected with a 5% significance level via the classical Bonferroni correction for concurrent p-value tests Bonferroni (1936). We take the extremes thresholds to be the 96% quantiles as validated by CvM and AD

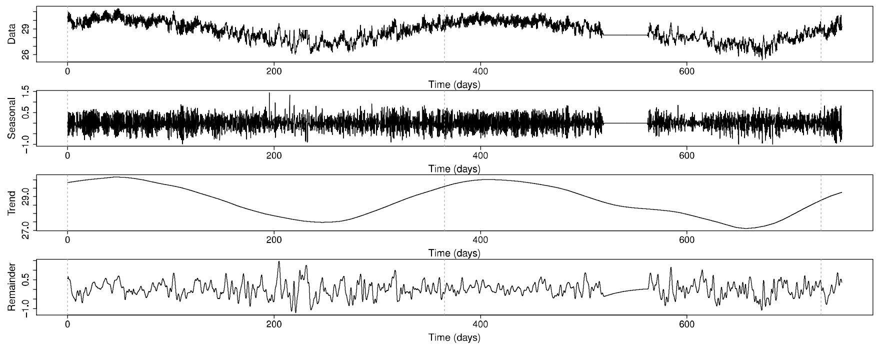

Figure 3: STL Cleaning of the Temperature timeseries in New York City shown on the first two years worth of data: daily seasonal patterns, quarterly trends. Flat parts of data are interpolated points and vertical lines materialise years of data. The x-axis *time* is in days. We keep only the *remainder* part.

tests (Bader et al. 2018).
Regarding the wind data, we have both speed (in m/s) and direction (in degrees from North clockwise) of the wind which unveils an opportunity to combine both to obtain two components, North-South (NS) and West-East (WE), by cosine and sine projections, respectively. Finally, we take the absolute values as we care about in extreme winds in both directions of the NS/WE axis.

7.3 Discussion

We do not claim that our model captures causality but can inform on dependencies outside of the usual correlation toolbox.
Given an extreme in $X^{(i)}$ at time t, we model the probabilities

$$\mathbb{P}\left\{X_0^{(j)} > 0 | X_h^{(i)} > 0\right\}$$

for $j \in \{1, \ldots, 32\}$. For this, we use the (classical) Bayes' formula:

$$\mathbb{P}\left\{X_0^{(j)} > 0 | X_h^{(i)} > 0\right\} = \frac{\mathbb{P}\left\{X_0^{(j)} > 0\right\}}{\mathbb{P}\left\{X_h^{(i)} > 0\right\}} \mathbb{P}\left\{X_h^{(i)} > 0 | X_0^{(j)} > 0\right\} = \mathbb{P}\left\{X_h^{(i)} > 0 | X_0^{(j)} > 0\right\}.$$

since both the numerator and denominator of the fraction are equal to 4%.

Remark 16. *When $i = j$, the univariate latent trawl model is time-symmetric since it is stationary* *in the sense that $X_t^{(i)}|X_{t+h}^{(i)} \overset{d}{=} X_{t+h}^{(i)}|X_t^{(i)}$, again by Bayes' rule for density functions. Therefore, in this case, we can directly interpret the probabilities yielded by the TRON computations.*

7.4 Summary of results

As in Section 6, Berg test found vine copulas conditional on extreme in `AEP` to be a well-specified model with all p-values above 0.1%. This is not surprising given the added flexibility of vine copulas as the number of variables increases.
Table 3 is the complete set of results. Since `AEP` and `DUQ` are so highly correlated (0.90), we expected such a strong positive relationship between an extreme in `DUQ` and one in `AEP`.
A more interesting example is Vancouver which was added as to qualitatively check whether the model was capture artificial links. Here, we observe that spikes in humidity, pressure, temperature, wind in Vancouver did not affect the extreme probability of `AEP` as most of them are between the $3\% - 5\%$ range ($\approx 1\%$ standard deviations).
We also find that pressure spikes in Detroit, Charlotte and Toronto increase extreme probability in (relatively) short-range horizons ($h = 1, 2, 3$ hours) from 4% up to 8% for Toronto up to 14% for Detroit. Then, for longer time horizons, pressures inverts its role, reducing the probability of extreme down to 1% for Toronto. Indeed, large pressure often implies anticyclonic conditions in the short-term which might change within three days.
The most surprising set of variables are temperatures as often taken as a close indicator to energy consumption (Yi-Ling et al. (2014), Zachariadis & Pashourtidou (2007)). Here, none of them increases extreme probability significantly but rather decrease it (for Charlotte and Boston). This said, we think this is due to the existence more qualitative relationships between extremes in `AEP` and other variables than temperature after removal of seasonality; as well as the very short time horizon at which temperature can play a role which is not captured by our hourly dataset as hinted by visual inspections of the vines. Finally, wind speed has an significant impact via its North-South component (for Charlotte, Toronto, New York and Boston) which is sensible since it may be driven by hot airstreams coming from the Gulf of Mexico (such as the Gulf stream itself). On the other hand, West-East wind components have a much smaller impact on extreme probabilities. We interpret this since Ohio and West Virginia have inland positions fairly remote from the ocean shore.

8 Conclusion

Inspired from the univariate latent trawl model for extreme values introduced in Noven et al. (2018), we formulate a novel flexible multidimensional extension using regular vine copulas - the *Extreme Vine* (X-Vine) model. Whilst the marginal serial dependence is ensured by property of the trawl model itself, the multivariate layer provides a powerful cross-sectional dependency structure thanks to (regular) vine copulas - an intertwined collection of bivariate copulas. This also ensures the possibility of both asymptotic dependence and independence as in the model proposed by Heffernan & Tawn (2004). Concurrently, we define *Time-Rigged Occurence Networks (TRON)* as a communication tool to detail dependencies in the extremes across time and variables. We estimate TRON probabilities using Monte Carlo integration or subsequent univariate model on two different datasets (air pollution in Bloomsbury, London, UK; energy/weather on American East coast). In both examples, the vines pass standard goodness-of-fit tests of which we present two main ones - White's and multivariate Kolmogorov-Smirnov tests. We compute and present TRON probabilities conditional on extremes in O_3 and CO to show the variety of dependency obtained for the air pollution dataset. We reverse the analysis with the North-East energy-weather data by investigating variables that led to an increase in extreme probability in an energy consumption. That is, we quantify the most likely phenomenon that led to spikes in temperature in the sense of conditional probabilities.

	horizon	1	2	3	6	12	24	48
Energy Consumption	AEP	.88 (.02)	.79 (.03)	.72 (.03)	.56 (.03)	.45 (.03)	.38 (.03)	.17 (.03)
	DUQ	.34 (.03)	.34 (.03)	.32 (.03)	.26 (.03)	.22 (.03)	.24 (.03)	.12 (.02)
Humidity	VC	.06 (.02)	.05 (.02)	.05 (.02)	.06 (.02)	.05 (.02)	.05 (.02)	.05 (.02)
	DT	.02 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.04 (.01)	.04 (.01)
	CH	.02 (.01)	.02 (.01)	.02 (.01)	.02 (.01)	.01 (.01)	.02 (.01)	.04 (.01)
	TO	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.02 (.01)	.03 (.01)
	NY	.02 (.01)	.02 (.01)	.02 (.01)	.02 (.01)	.01 (.01)	.03 (.01)	.05 (.02)
	BO	.04 (.01)	.04 (.01)	.04 (.01)	.03 (.01)	.03 (.01)	.04 (.01)	.06 (.02)
Pressure	VC	.03 (.01)	.03 (.01)	.02 (.01)	.02 (.01)	.03 (.01)	.05 (.02)	.07 (.02)
	DT	.14 (.02)	.14 (.02)	.14 (.03)	.13 (.02)	.12 (.02)	.10 (.02)	.10 (.02)
	CH	.14 (.03)	.15 (.03)	.15 (.03)	.16 (.03)	.17 (.03)	.15 (.03)	.11 (.02)
	TO	.09 (.02)	.08 (.02)	.09 (.02)	.09 (.02)	.08 (.02)	.08 (.02)	.10 (.02)
	NY	.08 (.02)	.08 (.02)	.09 (.02)	.10 (.02)	.10 (.02)	.13 (.02)	.10 (.02)
	BO	.07 (.02)	.08 (.02)	.08 (.02)	.11 (.02)	.12 (.02)	.13 (.02)	.11 (.02)
Temp.	VC	.06 (.02)	.07 (.02)	.08 (.02)	.08 (.02)	.07 (.02)	.07 (.02)	.06 (.02)
	DT	.05 (.02)	.05 (.02)	.05 (.02)	.05 (.02)	.03 (.01)	.02 (.01)	.02 (.01)
	CH	.00 (.01)	.00 (.01)	.00 (.01)	.00 (.01)	.00 (.01)	.00 (.01)	.00 (.01)
	TO	.07 (.02)	.07 (.02)	.07 (.02)	.08 (.02)	.07 (.02)	.06 (.02)	.04 (.01)
	NY	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.04 (.01)	.03 (.01)
	BO	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.02 (.01)	.03 (.01)	.03 (.01)
Wind Speed North-South	VC	.04 (.01)	.04 (.01)	.04 (.01)	.04 (.01)	.04 (.01)	.04 (.01)	.03 (.01)
	DT	.07 (.02)	.06 (.02)	.06 (.02)	.06 (.02)	.05 (.02)	.05 (.02)	.04 (.01)
	CH	.09 (.02)	.09 (.02)	.08 (.02)	.08 (.02)	.05 (.02)	.03 (.01)	.02 (.01)
	TO	.12 (.02)	.12 (.02)	.12 (.02)	.12 (.02)	.10 (.02)	.09 (.02)	.05 (.02)
	NY	.14 (.02)	.14 (.02)	.14 (.03)	.13 (.02)	.11 (.02)	.07 (.02)	.03 (.01)
	BO	.12 (.02)	.12 (.02)	.12 (.02)	.11 (.02)	.10 (.02)	.08 (.02)	.03 (.01)
Wind Speed East-West	VC	.02 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.03 (.01)	.04 (.01)	.04 (.01)
	DT	.04 (.01)	.04 (.01)	.04 (.02)	.05 (.02)	.05 (.02)	.05 (.02)	.04 (.01)
	CH	.05 (.02)	.05 (.02)	.05 (.02)	.05 (.02)	.03 (.01)	.01 (.01)	.03 (.01)
	TO	.06 (.02)	.06 (.02)	.06 (.02)	.06 (.02)	.06 (.02)	.06 (.02)	.05 (.02)
	NY	.07 (.02)	.07 (.02)	.07 (.02)	.07 (.02)	.04 (.02)	.03 (.01)	.03 (.01)
	BO	.06 (.02)	.07 (.02)	.07 (.02)	.07 (.02)	.06 (.02)	.03 (.01)	.03 (.01)

Table 3: TRON probabilities of having an extreme at t given an extreme in `AEP` at $t+h$. Each row unveils the likelihood of extreme obtained from the vine conditional on an extreme in the corresponding variable of this particular row (e.g. the second row is based on the vine conditional on extremes in `DUQ`). This includes data from Vancouver (VC), Detroit (DT), Charlotte (CH), Toronto (TO), New York City (NY) and Boston (BO). For easier readability, in green (resp. in red) are the variables of interests; that is, generally increasing (resp. decreasing) the likelihood of extremes knowing that the unconditional probability is 4%.

References

Aas, K. & Berg, D. (2013), Models for construction of multivariate dependence–a comparison study, *in* 'Copulae and Multivariate Probability Distributions in Finance', Routledge, pp. 43–64.

Atkinson, R., Cohen, A., Mehta, S. & Anderson, H. (2012), 'Systematic review and meta-analysis of epidemiological time-series studies on outdoor air pollution and health in asia', *Air Quality, Atmosphere & Health* **5**(4), 383–391.

Bader, B., Yan, J., Zhang, X. et al. (2018), 'Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate', *The Annals of Applied Statistics* **12**(1), 310–329.

Barndorff-Nielsen, O. E. (2011), 'Stationary infinitely divisible processes', *Brazilian Journal of Probability and Statistics* pp. 294–322.

Berg, D. & Bakken, H. (2007), 'A copula goodness-of-fit approach based on the conditional probability integral transformation', *h ttp://www. danielberg. no/publications/Btest. pdf (2008-11-01)* .

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M. & Vrac, M. (2017), 'Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in ravenna (italy)', *Hydrology and Earth System Sciences* **21**(6), 2701–2723.

Bonferroni, C. (1936), 'Teoria statistica delle classi e calcolo delle probabilita', *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* **8**, 3–62.

Bortot, P. & Gaetan, C. (2014), 'A latent process model for temporal extremes', *Scandinavian Journal of Statistics* **41**(3), 606–621.

Bortot, P. & Gaetan, C. (2016), 'Latent process modelling of threshold exceedances in hourly rainfall series', *Journal of Agricultural, Biological, and Environmental Statistics* **21**(3), 531–547.
URL: *https://doi.org/10.1007/s13253-016-0254-5*

Brechmann, E. C. & Schepsmeier, U. (2013), 'Modeling dependence with c- and d-vine copulas: The R package CDVine', *Journal of Statistical Software* **52**(3), 1–27.
URL: *http://www.jstatsoft.org/v52/i03/*

Cleveland, R. B., Cleveland, W. S. & Terpenning, I. (1990), 'Stl: A seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics* **6**(1), 3.

Cohen, S. N. et al. (2018), 'Data and uncertainty in extreme risks-a nonlinear expectations approach', *World Scientific Book Chapters* pp. 135–162.

Coles, S., Bawa, J., Trenner, L. & Dorazio, P. (2001), *An introduction to statistical modeling of extreme values*, Vol. 208, Springer.

Cooke, R. M., Kurowicka, D. & Wilson, K. (2015), 'Sampling, conditionalizing, counting, merging, searching regular vines', *Journal of Multivariate Analysis* **138**, 4–18.

Courgeau, V. & Veraart, A. (2018), 'Inference, simulation and application of a latent trawl model for extreme values', *Available at SSRN* .

Czado, C. (2010), Pair-copula constructions of multivariate copulas, *in* 'Copula theory and its applications', Springer, pp. 93–109.

Davison, A. C. & Smith, R. L. (1990), 'Models for exceedances over high thresholds', *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(3), 393–425.

Dissmann, J., Brechmann, E. C., Czado, C. & Kurowicka, D. (2013), 'Selecting and estimating regular vine copulae and application to financial returns', *Computational Statistics & Data Analysis* **59**, 52–69.

Dupuis, D. J., Trapin, L. et al. (2019), 'Ground-level ozone: Evidence of increasing serial dependence in the extremes', *The Annals of Applied Statistics* **13**(1), 34–59.

Dutfoy, A., Parey, S. & Roche, N. (2014), 'Multivariate extreme value theory-a tutorial with applications to hydrology and meteorology', *Dependence Modeling* **2**(1).

Embrechts, P., Klüppelberg, C. & Mikosch, T. (2013), *Modelling extremal events: for insurance and finance*, Vol. 33, Springer Science & Business Media.

Faugeras, O. P. (2012), 'Probabilistic constructions of discrete copulas'.

Genest, C., Rémillard, B. & Beaudoin, D. (2009), 'Goodness-of-fit tests for copulas: A review and a power study', *Insurance: Mathematics and economics* **44**(2), 199–213.

Grothe, O. & Nicklas, S. (2013), 'Vine constructions of lévy copulas', *Journal of Multivariate Analysis* **119**, 1–15.

Hautsch, N. & Herrera, R. (2015), 'Multivariate dynamic intensity peaks-over-threshold models'.

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.

Joe, H., Li, H. & Nikoloulopoulos, A. K. (2010), 'Tail dependence functions and vine copulas', *Journal of Multivariate Analysis* **101**(1), 252–270.

Kendall, M. G. (1938), 'A new measure of rank correlation', *Biometrika* **30**(1/2), 81–93.

Krzyzanowski, M. & Cohen, A. (2008), 'Update of who air quality guidelines', *Air Quality, Atmosphere & Health* **1**(1), 7–13.

Ledford, A. W. & Tawn, J. A. (1996), 'Statistics for near independence in multivariate extreme values', *Biometrika* **83**(1), 169–187.
URL: *https://doi.org/10.1093/biomet/83.1.169*

Lugrin, T., Davison, A. C. & Tawn, J. A. (2019), 'Penultimate analysis of the conditional multivariate extremes tail model', *arXiv preprint arXiv:1902.06972* .

Morales Napoles, O., Cooke, R. M. & Kurowicka, D. (2010), 'About the number of vines and regular vines on n nodes'.

Nagler, T., Bumann, C. & Czado, C. (2018), 'Model selection in sparse high-dimensional vine copula models with application to portfolio risk', *arXiv preprint arXiv:1801.09739* .

Nelsen, R. B. (2006), *An introduction to copulas*, Springer series in statistics, 2nd ed. edn, Springer, New York.

Newey, W. K. & McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Newey, W. K. & West, K. D. (1986), 'A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix'.

Noven, R. C., Veraart, A. E. & Gandy, A. (2018), 'A latent trawl process model for extreme values', *arXiv preprint arXiv:1511.08190. Journal of Energy Markets. Accepted for publication.* .

Patton, A., Politis, D. N. & White, H. (2009), 'Correction to automatic block-length selection for the dependent bootstrap by d. politis and h. white', *Econometric Reviews* **28**(4), 372–375.

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

Pineau, H., Girard, F., Raynaud, S., Prevosto, M. & Raillard, N. (2018), Multivariate extreme analysis methodology in function of structural response, *in* 'ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering', American Society of Mechanical Engineers, pp. V07BT06A014–V07BT06A014.

Politis, D. N. & White, H. (2004), 'Automatic block-length selection for the dependent bootstrap', *Econometric Reviews* **23**(1), 53–70.
URL: *https://doi.org/10.1081/ETC-120028836*

Rajput, B. S. & Rosinski, J. (1989), 'Spectral representations of infinitely divisible processes', *Probability Theory and Related Fields* **82**(3), 451–487.

Renard, B., Lang, M. & Bois, P. (2006), 'Statistical analysis of extreme events in a non-stationary context via a bayesian framework: case study with peak-over-threshold data', *Stochastic Environmental Research and Risk Assessment* **21**(2), 97–112.
URL: *https://doi.org/10.1007/s00477-006-0047-4*

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

Ribatet, M., Ouarda, T. B., Sauquet, E. & Gresillon, J.-M. (2009), 'Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics', *Water Resources Research* **45**(3).

Schepsmeier, U. (2015), 'Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review', *Journal of Multivariate Analysis* **138**, 34–52.

Sklar, A. (1973), 'Random variables, joint distribution functions, and copulas', *Kybernetika* **9**(6), 449–460.

Smith, R. L. (1985), 'Maximum likelihood estimation in a class of nonregular cases', *Biometrika* **72**(1), 67–90.

Smith, R. L., Tawn, J. A. & Coles, S. G. (1997), 'Markov chain models for threshold exceedances', *Biometrika* **84**(2), 249–268.

Stöber, J. & Czado, C. (2014), 'Regime switches in the dependence structure of multidimensional financial data', *Computational Statistics & Data Analysis* **76**, 672–686.

Tendijck, S., Ross, E., Randell, D. & Jonathan, P. (2019), 'A model for the directional evolution of severe ocean storms', *Environmetrics* **30**(1), e2541.

White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica: Journal of the Econometric Society* pp. 1–25.

Winter, H. C. & Tawn, J. A. (2016), 'Modelling heatwaves in central france: a case-study in extremal dependence', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(3), 345–365.

Winter, H. C. & Tawn, J. A. (2017), 'kth-order markov extremal models for assessing heatwave risks', *Extremes* **20**(2), 393–415.
URL: *https://doi.org/10.1007/s10687-016-0275-z*

Yi-Ling, H., Hai-Zhen, M., Guang-Tao, D. & Jun, S. (2014), 'Influences of urban temperature on the electricity consumption of shanghai', *Advances in climate change research* **5**(2), 74–80.

Yun, S. (2000), 'The distributions of cluster functionals of extreme events in a dth-order markov chain', *Journal of Applied Probability* **37**(1), 29–44.

Zachariadis, T. & Pashourtidou, N. (2007), 'An empirical analysis of electricity consumption in Cyprus', *Energy Economics* **29**(2), 183–198.

A Fundamentals on Lévy measures

This part is a direct quote from Section 2.2, Courgeau & Veraart (2018). We hereby define the concepts of random measures, Lévy basis and the infinite divisibility and independent scattering properties. We conclude with Lévy seed and their intervention in the cumulant function.

One of the important objects are random measures defined below where $S \subset \mathbb{R}^p$ is a Borel set and we denote $\mathcal{B}(S)$ the associated Borel σ-algebra. Suppose $p \in \mathbb{N}$. Our trawl processes will be considered using a random measure on the Borel set $S \subset \mathbb{R}^p$ and in particular, consider the set of bounded sets of $\mathcal{B}(S)$ with respect to the Lebesgue measure μ^{leb} as $\mathcal{B}_b(S) := \{A \in \mathcal{S} : \mu^{leb}(A) < \infty\}$. In the rest of this article, we denote by $\dot\cup$ the union of *disjoint* sets.

Definition 17. *We introduce a number of core definitions useful throughout the article:*

1. *A **random measure** M on $(S, \mathcal{B}(S))$ is the collection of $\mathbb{R}$-valued random variables in $\{M(A) : A \in \mathcal{B}_b(S)\}$ such that for any sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{B}_b(S)$ of disjoint elements satisfying $\dot\cup_{n\in\mathbb{N}} A_n \in \mathbb{B}_b(S)$, we have countable additivity, that is, $M\left(\dot\cup_{n\in\mathbb{N}} A_n\right) = \dot\cup_{n\in\mathbb{N}} M(A_n) \quad a.s.$*

2. *A random measure on $(S, \mathcal{B}(S))$ is said to be **stationary** if it is stable by translation, that is, if for any $\boldsymbol{s} \in S$, $n \in \mathbb{N}$ and any finite collection $A_1, \dots, A_n \in \mathcal{B}_b(S) \subset \mathcal{B}(S)$ such that $A_i + \mathbf{s} \subset S$ for $i = 1, \dots, n$, we have $(M(A_1 + \boldsymbol{s}), \dots, M(A_n + \boldsymbol{s})) \stackrel{d}{=} (M(A_1), \dots, M(A_n))$, where $\stackrel{d}{=}$ designates the equality in law.*

3. *The law μ of a random variable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is **infinitely divisible**, or ID, if and only if for any $n \in \mathbb{N}$, there exits a corresponding law μ_n on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu = \mu_n^{*n}$ where μ_n^{*n} denotes the n-fold convolution of μ_n on itself.*

4. *A random measure M on $(S, \mathcal{B}(S))$ is said to be **independently scattered** if for any sequence $(A_n)_{n\in\mathbb{N}} \subset \mathcal{B}_b(S)$ of disjoint elements, $(M(A_n))_{n\in\mathbb{N}}$ is a collection of mutually independent random variables.*

5. *A **Lévy basis** L on $(S, \mathcal{B}(S))$ is an independently scattered, infinitely divisible random measure. Additionally, it is **homogeneous** if it is stationary. A **Lévy measure** is any Borel measure ν on $\mathbb{R}$ such that $\nu(0) = 0$ and $\int_{\mathbb{R}}(1 \wedge x^2)\nu(dx) < \infty$ where $x \wedge y := min(x, y)$.*

A critical result from Rajput & Rosinski (1989) concerns the Lévy-Khintchine formula for homogeneous Lévy bases. Define, for $\xi \in \mathbb{R}$, $\Phi(\xi; L(A)) := \mathbb{E}\left[\exp\{i\xi L(A)\}\right]$ the characteristic function. Let L be an homogeneous Lévy basis on $(S, \mathcal{B}(S))$. Then, the characteristic function verifies that

$$\Phi(\xi; L(A)) = \exp\left\{\mu^{leb}(A)\mathcal{K}(\xi)\right\}, \text{ for any } A \in \mathcal{B}_b(S), \tag{7}$$

where $\xi \mapsto \mathcal{K}(\xi)$ is the cumulant generating function that satisfies (Proposition 2.4, Rajput & Rosinski (1989)):

$$\mathcal{K}(\xi) = i\xi a^*(A) - \frac{1}{2}\xi^2 b^*(A) + \int_{\mathbb{R}} \big(\exp\{i\xi x\} - 1 - i\xi I(|x| \le 1)\big)\nu(dx, A),$$

for some signed measure a^* on $\mathcal{B}_b(S)$ and measure b^* on $\mathcal{B}_b(S)$ and $\nu(\cdot,\cdot)$ being the so-called generalised Lévy measure i.e. such that $\nu(\cdot, A)$ is a Lévy measure on $\mathbb{R}$ for any fixed $A \in \mathcal{B}_b(S)$ and $\nu(dx, \cdot)$ a measure on $\mathcal{B}_b(S)$ for any fixed dx. Note that $z \mapsto I(|z| \le 1)$ being the indicator function on the unit disc. Barndorff-Nielsen (2011) proposes a striking observation: by definition L is infinitely divisible, therefore, there always exists a particular Lévy basis L' - referred to as the *Lévy seed* such that:

$$\Phi(\xi; L(A)) = \exp\left\{\mu^{leb}(A)\mathcal{K}(\xi; L'(A))\right\}, \quad \forall A \in \mathcal{B}_b(S),$$

which showcases that the law of $L(A)$ is characterised by the Lesbegue measure of the set A, $\mu^{leb}(A)$, and the law of the associated Lévy seed L', for any fixed $A \in \mathcal{B}_b(S)$. It is known some correspondence between trawl processes and ID laws as per the following proposition:

Proposition 3. *For any infinitely divisible law, one can find a trawl process with the same marginal law (but potentially featuring some serial dependence).*

B Proofs

Remark 17. *In those short proofs, we use the notation from Noven et al. (2018) (itself carried over from Bortot & Gaetan (2014)), that we set $\alpha := \xi^{-1}$ for simplicity.*

B.1 Proof of Proposition 2 (transition probabilities)

Proof. Recall the independent scattering of Lévy measures that yields the decomposition $X_t \stackrel{d}{=} L(A_{t+h,t}) + L(A_{t\setminus t+h})$ and $X_{t+h} \stackrel{d}{=} L(A_{t+h,t}) + L(A_{t+h\setminus t})$. Using Fubini's theorem, we can compute the following probability:

$$\begin{aligned}
&\mathbb{P}\{X_{t+h} > v, X_t > u\} = \\
&\int_v^{+\infty}\int_u^{+\infty} \mathbb{E}\left\{\Lambda_t e^{-(\kappa+w_t)\Lambda_t}\Lambda_{t+h}e^{-(\kappa+w_{t+h})\Lambda_{t+h}}\right\} dw_t dw_{t+h} \\
&= \mathbb{E}\left\{\int_v^{+\infty}\int_u^{+\infty} \Lambda_t e^{-(\kappa+w_t)\Lambda_t}\Lambda_{t+h}e^{-(\kappa+w_{t+h})\Lambda_{t+h}} dw_t dw_{t+h}\right\} \\
&= \mathbb{E}\left\{e^{-(\kappa+u)\Lambda_t}e^{-(\kappa+v)\Lambda_{t+h}}\right\} \quad \text{since } |\Lambda_t| < \infty \; a.s. \\
&= \mathbb{E}\Big\{e^{-(\kappa+u)L(A_{t\setminus t+h})}e^{-(2\kappa+u+v)L(A_{t,t+h})} \\
&\qquad \times e^{-(\kappa+v)L(A_{t+h\setminus t})}\Big\} \\
&= \mathbb{E}\left\{e^{-(2\kappa+u+v)L(A_{t,t+h})}\right\}\mathbb{E}\left\{e^{-(\kappa+u)L(A_{t\setminus t+h})}\right\}\mathbb{E}\left\{e^{-(\kappa+v)L(A_{t+h\setminus t})}\right\} \\
&= \left(1 + \frac{2\kappa+u+v}{\beta}\right)^{b_{t,t+h}}\left(1 + \frac{\kappa+u}{\beta}\right)^{b_{t+h\setminus t}}\left(1 + \frac{\kappa+v}{\beta}\right)^{b_{t\setminus t+h}}
\end{aligned}$$

We conclude that:

$$\mathbb{P}\{X_{t+h} > v | X_t > u\} = \left(1 + \frac{2\kappa+u+v}{\beta}\right)^{b_{t,t+h}}\left(1 + \frac{\kappa+u}{\beta}\right)^{b_{t+h\setminus t}+\alpha}\left(1 + \frac{\kappa+v}{\beta}\right)^{b_{t\setminus t+h}}.$$

Finally:

$$\begin{aligned}
&\mathbb{P}\{X_{t+h} = 0, X_t > u\} = \\
&\int_u^{+\infty}\mathbb{E}\left\{\Lambda_t e^{-(\kappa+w_t)\Lambda_t}\left(1 - e^{-\kappa\Lambda_{t+h}}\right)\right\}dw_t \\
&= \mathbb{E}\left\{e^{-(\kappa+a)\Lambda_t}\right\} - \mathbb{E}\left\{e^{-(\kappa+u)\Lambda_t - \kappa\Lambda_{t+h}}\right\} \\
&= \left(1 + \frac{\kappa+u}{\beta}\right)^{-\alpha}\left\{1 - \left(1 + \frac{2\kappa+u}{\beta}\right)^{b_{t,t+h}}\left(1 + \frac{\kappa}{\beta}\right)^{b_{t+h\setminus t}}\left(1 + \frac{\kappa+u}{\beta}\right)^{b_{t\setminus t+h}+\alpha}\right\}
\end{aligned}$$

Therefore, this yields:

$$\mathbb{P}\{X_{t+h} = 0 | X_t > u\} = 1 - \left(1 + \frac{2\kappa+u}{\beta}\right)^{b_{t,t+h}}\left(1 + \frac{\kappa}{\beta}\right)^{b_{t+h\setminus t}}\left(1 + \frac{\kappa+u}{\beta}\right)^{b_{t\setminus t+h}+\alpha}.$$

□

B.2 Proof of Theorem 15 (X-Vine TRON probabilities)

Proof. The case $i = j$ is treated in Proposition 2. We calculate directly:

$$\begin{aligned}
p_{i\to(j,h)}(u^{(i)}, u^{(j)}) &= \mathbb{P}\left[X^{(j)}_{t+h} > u^{(j)} | X^{(i)}_t > u^{(i)}\right] \\
&= \int_{u^{(j)}}^{\infty} f_{\left\{X^{(j)}_{t+h}|X^{(i)}_t > u^{(i)}\right\}}\left(x^{(j)}_{t+h}\right)dx^{(j)}_{t+h} \\
&= \int_{u^{(j)}}^{\infty}\int_0^{\infty} f_{\left\{X^{(j)}_{t+h}, X^{(i)}_{t+h}|X^{(i)}_t > u^{(i)}\right\}}\left(x^{(j)}_{t+h}, x^{(i)}_{t+h}\right)dx^{(j)}_{t+h}dx^{(i)}_{t+h} \\
&= \mathbb{E}\left[I\left(X^{(j)}_{t+h} > u^{(j)}\right)\Big|\left(X^{(j)}_{t+h}; X^{(i)}_{t+h}, X^{(i)}_t > u^{(i)}\right)\right].
\end{aligned}$$

A reformulation using a two-step sampling approach is given by definition:

$$\begin{aligned}
&f_{\left\{X^{(j)}_{t+h}, X^{(i)}_{t+h}|X^{(i)}_t > u^{(i)}\right\}}\left(x^{(j)}_{t+h}, x^{(i)}_{t+h}\right) = \\
&f_{\left\{X^{(j)}_{t+h}|X^{(i)}_{t+h} = x^{(i)}_{t+h}, X^{(i)}_t > u^{(i)}\right\}}\left(x^{(j)}_{t+h}\Big|x^{(i)}_{t+h}\right)\cdot f_{\left\{X^{(i)}_{t+h}|X^{(i)}_t > u^{(i)}\right\}}\left(x^{(i)}_{t+h}\right)
\end{aligned}$$

□