Introduction

0.1 Context

Management of environmental resources often requires the analysis of multivariate extreme values. In the classical theory, one is often interested in the behavior of the mean or average of a random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. This average will then be described through the expected value $\mathbb{E}[X]$ of the distribution. The central limit theorem yields, under mild conditions, the asymptotic behavior of the sample mean $\bar{X}$. This result can be used to provide a confidence interval for $\mathbb{E}[X]$ for a level $\alpha \in [0,1]$. But in case of extreme events, it can be just as important to estimate tails probabilities. Furthermore, what if the second moment $\mathbb{E}[X^2]$ or even the mean is not finite? Then the central limit theorem does not apply and the classical theory, carried by the normal distribution, is no longer relevant [Beirlant et al., 2004].

Some extreme events, such as heavy precipitation or wind speed has spatial characteristics and geostatisticians are striving to better understand the physical processes in hand. In geostatistics, we often consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, S a set of locations and $(E, \mathcal{E})$ a measurable state space. We define on this probability space a stochastic process $X = \{X(s), s \in S\}$ with values on $(E, \mathcal{E})$. It is classical to define the following second-order statistic as the variogram (see [Gaetan and Guyon, 2008] Chapter 1.3 for definition and basic properties) :

$$2\gamma(h) = \mathbb{E}[|X(s+h) - X(s)|^2],$$

where $\{X(s), s \in S\}$ represents a spatial and stationary process with a well-defined covariance function. The function $\gamma(\cdot)$ is called the semi-variogram of X. With respect to extremes, this definition is not well adapted because a second order statistic is difficult to interpret inside the framework of extreme value theory or may not even be defined. To ensure that we always work with finite moments quantities, the following type of first-order variogram is introduced by [Cooley et al., 2006]

$$\nu(h) = \frac{1}{2}\mathbb{E}\left[|F(X(s+h)) - F(X(s))|\right],$$

where $F(u) = \mathbb{P}(X(s) \leq u)$ is named as the FMadogram. His link to the pairwise extremal dependence function (Section 4.3 of [Coles et al., 1999]) or the Pickands dependence function ([Pickands, 1981]) make him an interesting quantity to capture the dependence between the extremas of stochastic processes or random variables. Indeed, this quantity may be seen as a dissimilarity measure among bivariate maxima to be used for clustering time-series as shown by [Bernard et al., 2013] or [Bador et al., 2015].

The main drawback of this quantity is that she only focus on the value of the diagonal section of the pairwise extremal dependence function. In the bivariate case, the FMadogram characterize solely the extremal dependence coefficient for random variables X and Y (see Section 8.2.7 of [Beirlant et al., 2004]). To overpass this drawback, [Naveau et al., 2009] introduce the λ-FMadogram defined as,

$$\nu(h, \lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^{\lambda}(X(s+h)) - F^{1-\lambda}(X(s))\right|\right],$$

for every $\lambda \in [0, 1]$.

This quantity characterize the pairwise extremal dependence function outside the diagonal section but also the whole Pickands dependence function ([Marcon et al., 2017]) and contribute to the vast litterature of the estimation of the Pickands dependence function for bivariate extreme value copulas (see for example [Pickands, 1981], [Deheuvels, 1991], [Hall and Tajvidi, 2000] or [Capéraà et al., 1997]). Statisticians may estimate this quantity but the classical results may applied only if the data in hands are clean as possible. This induce that the process of data collection has not been corrupted such as the data table is complete and that the implicit law of the observations is still the same.

Nevertheless, as the volume of data expands, the problem of missing or contaminated data has been increasingly present in many fields of statistical applications. It frequently happens that all of the individuals of a sample of statistical data from a multivariate population are not observed. If a sample be represented in matrix form by allowing the rows to represent the individuals and the columns the variables, then the matrix of the type of sample with which we are concerned is incomplete in that some elements are not present. In dealing with fragmentary samples, it is important to have at hand techniques which will enable the statistician to extract as much information as possible from the data. A useful reference for general parametric statistical inferences with missing data was provided by [Little R.J.A., 1987].

Considering a random sample of incomplete data,

$$(X_t, Y_t, \delta_t), \quad t \in \{1, \dots, T\}, \tag{1}$$

where all the X_t's are observed and $\delta_t = 0$ if Y_t is missing, otherwise $\delta_t = 1$. The simple missing data pattern describe by (1) is basically created by the double sampling or two phase sampling (see chapter 12 of [Cochran, 2007]). Samples like (1) may arise in survival analysis : The study of the duration time preceding an event of interest is considered with series of random censors, which might prevent the capture of the whole survival time. This is known as the censoring mechanism and it arises from restrictions depending from the nature of the study. Typically, they may occur in medicine, with studies of the survival times before the recovery / decease from a specific disease. Another important example is often realized in comparing treament effects of two educational programs. Individuals with lower scores on a preliminary test are more likely to receive the experimental treatment (*i.e.*, a composatory study program), whereas those with higher preliminary scores are more inclined to take the standard control. This phenomenon is well-known as the selection problem and we refer to Chapter 2 of [Angrist and Pischke, 2008] for more details. Beside of missing observations, the process of data might be disturb in a way that innerly deteriorate the quality of some data and one may ask that the estimation process should be robust.

The topic of Robustness in estimation has known an important research activity developed in the 60's and 70's resulting in a large number of publications. For a summary, the interested reader is refered to [Huber, 2011]. Robustness can be seen as an estimation procedure in which both stochastic and approximation error are low (see Section 1.1 from [Baraud et al., 2016]). In other words, an estimator is said to be robust if our model provides a reasonable approximation of the true one and derive an estimator which remains close to the true distribution. In this report, we mean by *robust* as *robust against outlier*, *e.g* the ϵ-contamination model (see [Huber, 1964]), or *robust again heavy-tailed data* where only low-order moments are assumed to be finite for the data distribution. There is no simple relation between the two definitions and the first framework of robustness that we have depicted. We want to propose a robust estimator of the Madogram. In our perspective, we only know [Escobar-Bach et al., 2018] that include the contamination framework in their estimation of the Pickands dependence function in the extreme value theory. To achieve our goal, we leverage the idea of Median-Of-MeaNs (MoN). Intuitively, we replace the linear operator of expectation with the median of averages taken over non-overlaping blocks of the data, in order to get a robust estimate thanks to the median step (see [Lerasle et al., 2019] for a similar idea applied to Kernel).

0.2 Definitions and Notation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be a bivariate random vector with values in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. This random vector has a joint distribution function H and marginal distribution function F and G. A function $C : [0,1]^2 \rightarrow [0,1]$ is called a bivariate copula if it is the restriction to $[0,1]^2$ of a bivariate distribution function whose marginals are given by the uniform distribution on the interval $[0,1]$. Since the work of [Sklar, 1959], it is well known that every distribution function H can be decomposed as $H(x,y) = C(F(x), G(y))$, for all $(x,y) \in \mathbb{R}^2$. This function C characterizes the dependence between X and Y and is called an extreme value copula if and if it admits a representation of the form [Gudendorf and Segers, 2009]

$$C(u,v) = (uv)^{A(log(v)/log(uv))}, \tag{2}$$

for all $u, v \in [0,1]$ and where $A(\cdot)$ is the Pickands dependence function, *i.e.*, $A : [0,1] \longrightarrow [1/2, 1]$ is convex and satisfies $t \vee (1-t) \leq A(t) \leq 1$, $\forall t \in [0,1]$. The upper and lower bound of A has special meanings, the upper bound $A(t) = 1$ corresponds to independence, whereas the lower bound $A(t) = t \vee 1 - t$ corresponds to the perfect dependence (comonotonicity). Notice that, on sections, the extreme value copula is of the form

$$C(u^t, u^{1-t}) = u^{A(t)}. \tag{3}$$

Let $(X_t, Y_t)_{t=1,\dots,T}$ be an *i.i.d.* sample of a bivariate random vector whose underlying copula is denoted by C and whose margins by F, G. For $x, y \in \mathbb{R}$, let $x \wedge y = min(x, y)$ and $x \vee y = max(x, y)$. Let $(b_{t,j})_{t \geq 1, j \in \{1,2\}}$ and $(a_{t,j})_{t \geq 1, j \in \{1,2\}}$ be respectively a sequence of numbers and a sequence of positive numbers. We say that the sequence $(a_{t,1}^{-1}(\bigvee_{t=1}^T X_t - b_{t,1}), a_{t,2}^{-1}(\bigvee_{t=1}^T Y_t - b_{t,2}))$ belongs to the domain of attraction of H, if for all real values x, y (at which the limit is continuous)

$$\mathbb{P}\left(\frac{\bigvee_{t=1}^T X_t - b_{t,1}}{a_{t,1}} \leq x, \frac{\bigvee_{t=1}^T Y_t - b_{t,2}}{a_{t,2}} \leq y\right) \underset{T \to \infty}{\rightarrow} H(x, y).$$

If this relationship hold, H is said to be a multivariate extreme value distribution. We will call by FMadogram the following quantity

$$\nu = \frac{1}{2}\mathbb{E}\left[|F(X) - G(Y)|\right], \tag{4}$$

and the λ-FMadogram by the expression

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^\lambda(X) - G^{1-\lambda}(Y)\right|\right]. \tag{5}$$

A classical estimator of the λ-FMadogram when the margins F, G are unknown is

$$\hat{\nu}(\lambda) = \frac{1}{2T}\sum_{t=1}^T \left|\hat{F}_T^\lambda(X_t) - \hat{G}_T^{1-\lambda}(Y_t)\right| \tag{6}$$

with $\hat{F}_T$ (resp. $\hat{G}_T$) the empirical cumulative distribution function of X (resp. Y). We suppose that we observe sequentially a quadruple defined by

$$(I_t X_t, J_t Y_t, I_t, J_t), \quad t \in \{1, \dots, T\}, \tag{7}$$

where $I_t = 0$ (resp. $J_t = 0$) if X_t (resp. Y_t) is missing, otherwise $I_t = 1$ (resp. $J_t = 1$), *i.e.* at each $t \in \{1, \dots, T\}$, one of both entries may be missing. The probability of observing a realisation partially or completely is denoted by $p_X = \mathbb{P}(I_t = 1) > 0$, $p_Y = \mathbb{P}(J_t = 1) > 0$ and $p_{XY} = \mathbb{P}(I_t = 1, J_t = 1) > 0$. Let us now define the empirical cumulative distribution of X (resp. Y and (X, Y)) in case of missing data,

$$\hat{F}_T(u) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u\}} I_t}{\sum_{t=1}^T I_t}, \quad \hat{G}_T(v) = \frac{\sum_{t=1}^T 1_{\{Y_t \leq v\}} J_t}{\sum_{t=1}^T J_t}, \quad \hat{H}_T(u, v) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u, Y_t \leq v\}} I_t J_t}{\sum_{t=1}^T I_t J_t}. \tag{8}$$

Here, we weight the estimator by the number of observed data which is a natural estimator (if divided by T) of the probabilities of missing. We have all tools in hand to define the *hybrid* *copula estimator* introduced by [Segers, 2014],

$$\hat{C}_T^{\mathcal{H}}(u,v) = \hat{H}_T(\hat{F}_T(u), \hat{G}_T(v)). \tag{9}$$

Given a rate $r_T > 0$ and $r_T \to \infty$ as $T \to \infty$, the normalized estimation error of the hybrid copula estimator is :

$$\mathbb{C}_T^{\mathcal{H}}(u,v) = r_T \left(\hat{C}_T^{\mathcal{H}}(u,v) - C(u,v) \right). \tag{10}$$

In order to propose a robust estimator we will assume that the sample is partitioned into K disjoint subsets $B_1, \dots, B_K$ of cardinalities $n_j := card(B_j)$ respectively, where the partitioning scheme is independent of the data. Let f be a measurable function from $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ to $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, we define the following estimator of $\mathbb{E}[f(X,Y)]$ by

$$\bar{\mathbb{P}}_{n_j} f = \frac{1}{n_j} \sum_{j \in B_j} f(X_j, Y_j).$$

We define the MoN estimator of f as solutions of the optimization problem

$$\hat{f}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} \left| \bar{\mathbb{P}}_{n_j} f - z \right|, \tag{11}$$

which, if we note $med(\cdot)$ the usual univariate median

$$\hat{f}_{MoN} = med(\bar{\mathbb{P}}_{n_1} f, \dots, \bar{\mathbb{P}}_{n_K} f), \tag{12}$$

is a solution of Equation (11).

We will write the generalized inverse function of F (respectively G) as $F^{\leftarrow}(u) = inf\{v \in \mathbb{R} | F(v) \geq u\}$ (respectively $G^{\leftarrow}(u) = inf\{v \in \mathbb{R} | G(v) \geq u\}$) where $0 < u, v < 1$. Given $\mathcal{X} \subset \mathbb{R}^2$, let $l^{\infty}(\mathcal{X})$ denote the spaces of bounded real-valued function on $\mathcal{X}$. For $f : \mathcal{X} \to \mathbb{R}$, let $||f||_{\infty} = sup_{x \in \mathcal{X}} |f(x)|$. Here, we use the abbreviation $Qf = \int f dQ$ for a given measurable function f and signed measure Q. The arrows $\overset{a.s.}{\to}$, $\overset{d}{\to}$ and $\rightsquigarrow$ denote almost sure convergence, convergence in distribution of random vectors and weak convergence in the sense of J.Hoffman-Jørgensen (see Part 1 in the monograph by [van der Vaart and Wellner, 1996]).

This work is organized as follows : In Chapter 1 we state some results on the weak convergence of the estimator of the λ-FMadogram with missing data. To propose a robust estimator of the λ-FMadogram, we leverage the idea of Median-Of-MeaNs (MoN) and state a concentration inequality that this estimator does verify. We also propose a closed formula for the asymptotic variance of the λ-FMadogram for a fixed $\lambda \in [0,1]$.

Chapter 2 will present our results in a finite-sample framework. The asymptotic variance of the normalized estimation error of several models would be drawn with their empirical counterpart obtained through simulation. We also propose a reproduction of the experiment of the λ-FMadogram with a Smith's process as find in [Naveau et al., 2009] and we will explain the augmentation of the Mean Squared Error while h is close to zero. This phenomenon would be also thoroughly explained through simulation and a counterexample.

In Chapter 3 we will present in details the mathematical proof of our statement.

Chapter 1

On the variance of the Madogram

1 Definition of the estimator

For the rest of this report we will assume that the copula C is of extreme value type as defined in Equation (2). Following [Fermanian et al., 2004], to guarantee the weak convergence of our empirical copula process, we introduce the following assumptions.

Assumption A. *(i) The bivariate distribution function H has continuous margins F, G.*

(ii) The derivative of the Pickands dependence function $A'(t)$ exists and is continuous on $(0,1)$.

The Assumption A *(i)* guarantee the uniqueness of the representation $H(x,y) = C(F(x), G(y))$ on the range of (F, G). Under the Assumption A *(ii)*, the first-order partial derivatives of C with respect to u and v exists and are continuous on the set $\{(u,v) \in [0,1]^2 : 0 < u, v < 1\}$. Indeed, we have

$$\frac{\partial C(u,v)}{\partial u} = \begin{cases} \frac{C(u,v)}{u} \left(A(\frac{log(v)}{log(uv)}) - A'(\frac{log(v)}{log(uv)}) \frac{log(v)}{log(uv)} \right), & \text{if } 0 < u, v < 1, \\ 0, & \text{if } v = 0, \quad 0 < u < 1, \end{cases}$$

$$\frac{\partial C(u,v)}{\partial v} = \begin{cases} \frac{C(u,v)}{v} \left(A(\frac{log(v)}{log(uv)}) + A'(\frac{log(v)}{log(uv)}) \frac{log(u)}{log(uv)} \right), & \text{if } 0 < u, v < 1, \\ 0, & \text{if } u = 0, \quad 0 < v < 1, \end{cases}$$

The properties of A imply $0 \leq A(t) - tA'(t) \leq 1$ and $0 \leq A(t) + (1-t)A'(t) \leq 1$ where $t = log(v)/log(uv)$ (see Lemma **??** in Section **??**). Therefore if $v \searrow 0$, then $\partial C(u,v)/\partial u \to 0$ as required.

In the missing data framework given by Equation (7), based on a identical and independent copies $(I_1 X_1, J_1 Y_1, I_1, J_1), \dots, (I_T X_T, J_T Y_T, I_T, J_T)$, we defined the following estimator of the λ-FMadogram

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) = \frac{1}{2 \sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t. \tag{1.1}$$

One may verify that outside the missing data framework, *i.e.* with $p_X = p_Y = p_{XY} = 1$, that $\hat{\nu}^{\mathcal{H}}(\lambda) = \hat{\nu}(\lambda)$. Before going further, let us briefly talk about our estimator. Our estimator defined in (1.1) does not verify $\hat{\nu}_T^{\mathcal{H}}(0) = \hat{\nu}_T^{\mathcal{H}}(1) = 0.25$. But in addition, the variance of our estimator at $\lambda = 0$ or $\lambda = 1$ does not equal 0. Indeed, suppose that we evaluate our statistic at $\lambda = 0$, we thus obtain the following quantity :

$$\hat{\nu}_T^{\mathcal{H}}(0) = \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left(1 - \hat{G}_T(Y_t)\right) I_t J_t$$

In this situation, the sample $(X_t)_{t=1}^T$ is taken account through the indicator's sequence $(I_t)_{t=1}^T$ and induce a variance when estimating. Hence we can force our estimator as in [Naveau et al., 2009] to satisfay these endpoint corrections. This leads to the following definition :

$$\begin{aligned}\hat{\nu}_T^{\mathcal{H}*}(\lambda) = & \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left|\hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t)\right| I_t J_t - \frac{\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{F}_T^{\lambda}(X_t)\} I_t J_t \\ & - \frac{1-\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{G}_T^{1-\lambda}(Y_t)\} I_t J_t + \frac{1}{2}\frac{1-\lambda+\lambda^2}{(2-\lambda)(1+\lambda)} \end{aligned} \tag{1.2}$$

Nevertheless, in missing data framework, the asymptotic behaviour of $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)$ is not the same as $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)$ and they should be studied apart.

In the following proposition, we establish some properties of the λ-FMadogram.

Proposition 1. *Let (X, Y) a $\mathbb{R}^2$-valued random vector of distribution H. We have, for each $\lambda \in [0, 1]$,*

(i) $0 \leq \nu(\lambda) \leq \frac{1}{1+\lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right)$,

(ii) $\nu(0) = \nu(1) = 0.25$, *and if* $\lambda \in (0, 1)$,

$$\nu(\lambda) = \frac{A(\lambda)}{A(\lambda) + \lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right). \tag{1.3}$$

The proof is given in Section **??**.

Remark 1. *The upper bound in* (i) *is exactly the value of the λ-FMadogram when X and Y are independent,* i.e. *when $A(t) = 1$.*

Now, we give some precisions under the missing mechanism,

Assumption B. *We suppose for all $t \in \{1, \ldots, T\}$, the pairs (I_t, J_t) and (X_t, Y_t) are independent, the data are missing completely at random (**MCAR**). Furthermore, we suppose that there exist at least one $t \in \{1, \ldots, T\}$ such that $I_t J_t \neq 0$.*

In order to propose a robust estimator of the FMadogram as defined in Equation (4). We are restricting our analysis to the FMadogram to avoid technical difficulties but the proof would be similar with a discussion according to the value of λ and using that $||x|^{\lambda} - |y|^{\lambda}| \leq |x - y|^{\lambda}$. Intuitively, we replace the linear operator of expectation with the median of averages taken over non-overlaping blocks of the data, in order to get a robust estimate thanks to the median step (see [Lerasle et al., 2019] for a similar idea applied to Kernel). The MoN is one of the mean estimators that achieve a sub-Gaussian behavior under mild conditions. Introduced during the 1980s [Nemirovsky and Yudin, 1983] for the estimation of the mean of real-valued random variables, that is easy to compute, while exhibiting attractive robustness properties.

Let $B_1, \dots, B_K$ a partition of the set $\{1, \dots, T\}$. Denote by $\bar{F}_{n_j}$ (resp. $\bar{G}_{n_j}$) the empirical cumulative distribution for the cumulative distribution of X (resp. Y) computed within block B_j. We propose the following MoN-based madogram estimator

$$\hat{\nu}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} \left| \bar{\nu}_j - z \right|, \tag{1.4}$$

where $\bar{\nu}_{n_j} = \frac{1}{2n_j} \sum_{t \in B_j} |\bar{F}_{n_j}(X_t) - \bar{G}_{n_j}(Y_t)|$. That is, in Equation (11), we take $f(x,y) = |x-y|$ and $\bar{\mathbb{P}}_{n_j} = \bar{C}_{n_j}$ the empirical copula constructed on the block B_j.

Assumption 1. *The sample $((X_1, Y_1), \dots, (X_T, Y_t))$ contains $T - T_o$ outliers drawn according to distribution H, and T_o outliers, upon which no assumption is made.*

In presence of outliers, the key point is to focus on sane blocks, *i.e* on blocks that does not contains a single outliers, since no inference can be made about blocks hit by an outlier. One way to ensure that sane blocks to be in majority is to consider twice more blocks than outliers. Indeed, in the worst case scenario, each outlier contaminate on block, but the sane blocks remains more numerous. Let K_s denote the total number of sane block containing no outliers. In other words, there exists $\delta \in (0, 1/2]$ such that $K_s \geq K(1/2 + \delta)$. If the data are free from contaminations, then $K_s = K$ and $\delta = 1/2$.

We suppose without loss of generality that $n_j = \lceil T/K \rceil$ for every $j \in \{1, \dots, K\}$. Using these notations, we can prove the following deviation bounds for our MoN-based estimator.

2 Main results

Without missing data, the weak convergence of normalized estimation error of the empirical copula process has already been proved by [Fermanian et al., 2004] under the sole Assumption A. This statement make use of previous results on the Hadamard differentiability of the map $\phi : D([0,1]^2) \rightarrow l^{\infty}([0,1]^2)$ which transforms the cumulative distribution function H into its copula function C (see lemma 3.9.28 from [van der Vaart and Wellner, 1996]). With the hybrid copula estimator, we need a technical assumption in order to guarantee the weak convergence of the process $\mathbb{C}_T^{\mathcal{H}}$ (see [Segers, 2014]),

Assumption C. *There exists $\gamma_t > 0$ and $r_t > 0$ such that $r_t \longrightarrow \infty$ as $t \to \infty$ such that in the space $l^{\infty}(\mathbb{R}^2) \otimes (l^{\infty}(\mathbb{R}), l^{\infty}(\mathbb{R}))$ equipped with the topology of uniform convergence, we have the joint weak convergence*

$$\left(r_t(\hat{H}_T - H); r_t(\hat{F}_t - F), r_t(\hat{G}_t - G) \right) \rightsquigarrow (\alpha \circ (F,G), \beta_1 \circ F, \beta_2 \circ G).$$

The stochastic processes α and β_j take values in $l^\infty([0,1]^2)$ and $l^\infty([0,1])$ respectively, and are such that $\alpha \circ F$ and $\beta_j \circ F_j$ have continuous trajectories on $[-\infty,\infty]^2$ and $[-\infty,\infty]$ almost surely.

Under condition A and C (see Theorem **??** in Appendix), the stochastic process $\mathbb{C}_T^{\mathcal{H}}$ converges weakly to the tight Gaussian process S_C where we denote by $S_C(u,v)$ the process defined by,

$$S_C(u,v) = \alpha(u,v) - \frac{\partial C(u,v)}{\partial u}\beta_1(u) - \frac{\partial C(u,v)}{\partial v}\beta_2(v), \quad \forall (u,v) \in [0,1]^2. \tag{1.5}$$

Considering our statistical framework and missing mechanism, [Segers, 2014] shows (in example 3.5) that the processes α, β_1 and β_2 take the following closed form :

$$\begin{aligned}
\beta_1(u) &= p_X^{-1}\mathbb{G}\left(1_{X \le F^{\leftarrow}(u), I=1} - u 1_{I=1}\right), \\
\beta_2(v) &= p_Y^{-1}\mathbb{G}\left(1_{Y \le G^{\leftarrow}(v), J=1} - v 1_{J=1}\right), \\
\alpha(u,v) &= p_{XY}^{-1}\mathbb{G}\left(1_{X \le F^{\leftarrow}(u)} 1_{Y \le G^{\leftarrow}(v), I=1, J=1} - C(u,v) 1_{I=1,J=1}\right).
\end{aligned}$$

Furthermore, we are able to compute their covariance functions. This is summarised in the following lemma and we add some technical details available in last section.

Lemma 1. *The covariance function of the process $\beta_1(u)$, $\beta_2(v)$ and $\alpha(u,v)$ are : for $(u,u_1,u_2,v,v_1,v_2) \in [0,1]^6$,*

$$\begin{aligned}
cov\left(\beta_1(u_1), \beta_1(u_2)\right) &= p_X^{-1}\{u_1 \wedge u_2 - u_1 u_2\}, \\
cov\left(\beta_2(v_1), \beta_2(v_2)\right) &= p_Y^{-1}\{v_1 \wedge v_2 - v_1 v_2\}, \\
cov\left(\beta_1(u), \beta_2(v)\right) &= \frac{p_{XY}}{p_X p_Y}\{C(u,v) - uv\},
\end{aligned}$$

and

$$\begin{aligned}
cov\left(\alpha(u_1,v_1), \alpha(u_2,v_2)\right) &= p_{XY}^{-1}\{C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1,v_1)C(u_2,v_2)\}, \\
cov\left(\alpha(u_1,v), \beta_1(u_2)\right) &= p_X^{-1}\{C(u_1 \wedge u_2, v) - C(u_1,v)u_2\}, \\
cov\left(\alpha(u,v_1), \beta_2(v_2)\right) &= p_Y^{-1}\{C(u, v_1 \wedge v_2) - C(u,v_1)v_2\}.
\end{aligned}$$

We have all tools in hand to consider the weak convergence of the stochastic processes $(\sqrt{T}(\hat{\nu}^{\mathcal{H}}(\lambda) - \nu(\lambda)))_{\lambda \in [0,1]}$ and $(\sqrt{T}(\hat{\nu}^{\mathcal{H}*}(\lambda) - \nu(\lambda)))_{\lambda \in [0,1]}$. To establish such a result, we use empirical process arguments as formulated in [van der Vaart and Wellner, 1996]. This allows us to show the following theorem.

Theorem 1. *Let $\lambda \in [0,1]$. Under Assumptions A, B, C we have the weak convergence in*

$l^\infty([0,1])$ *for the hybrid estimator defined in (1.1), as* $T \to \infty$*,*

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right) \rightsquigarrow \left(\frac{1}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}}, 1) - \beta_1(x^{\frac{1}{\lambda}})dx + \frac{1}{2}\int_{[0,1]} \alpha(1, x^{\frac{1}{1-\lambda}}) - \beta_2(x^{\frac{1}{1-\lambda}})dx \right.$$
$$\left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]}, \tag{1.6}$$

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right) \rightsquigarrow \left(\frac{1-\lambda}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}}, 1) - \beta_1(x^{\frac{1}{\lambda}})dx + \frac{\lambda}{2}\int_{[0,1]} \alpha(1, x^{\frac{1}{1-\lambda}}) - \beta_2(x^{\frac{1}{1-\lambda}})dx \right. \tag{1.7}$$
$$\left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]}.$$

We present also a concentration inequality that the MoN-based estimator of the FMadogram may verified. We thus add some remark on the concentration bound.

Theorem 2. *(Consistency & outlier-robustness of* $\hat{\nu}$*). Under Assumption 1, for any* $\eta \in]0,1[$ *such that* $K = \delta^{-1}log(1/\eta)$ *it holds that with probability* $1-\eta$*,*

$$|\hat{\nu}_{MoN} - \nu| \leq \frac{3}{\sqrt{2}}\frac{log\left(6e2^{\frac{1}{\delta}}\right)}{\delta}\sqrt{\frac{log\,(1/\eta)}{T}}.$$

Remark 2.
- *Dependence on* T *: These finite-sample guarantees show that estimator is robust to outliers, providing consistent estimates with high probability even under arbitrary contamination (affecting less than half of the samples).*
- *Dependence on* δ *: Recall that higher* δ *corresponds to less outliers,* i.e., *cleaner data in which case the bounds above become tighter.*
- *Dependence on* η *: An higher* η *gives a greater bound for which the estimator hold with an greater probability.*

3 Some corrolaries

As an integral of a tight Gaussian process, we know that the two normalized estimation errors follows a centered Gaussian variable for a given $\lambda \in [0,1]$. Furthermore, a few computations are able to give a closed form of the variance of the limiting Gaussian law as an integral of the Pickands dependence function. This is summarized with the following proposition.

Proposition 2. *For* $\lambda \in (0,1)$*, let* $A_1(\lambda) = A(\lambda)/\lambda$*,* $A_2(\lambda) = A(\lambda)/(1-\lambda)$*. Then, the asymptotic variance of* $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)$ *and* $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)$ *has the following closed form*

$$Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)\right) = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \sigma_3^2 + \frac{1}{2}\sigma_{12} - \sigma_{13} - \sigma_{23},$$

$$Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right)\right)=\frac{(1-\lambda)^2}{4}\sigma_1^2+\frac{\lambda^2}{4}\sigma_2^2+\sigma_3^2+\lambda(1-\lambda)\frac{1}{2}\sigma_{12}-(1-\lambda)\sigma_{13}-\lambda\sigma_{23}.$$

The quantities $(\sigma_i^2)_{i\in\{1,2,3\}}$ and $(\sigma_{ij})_{i,j\in\{1,2,3\},i<j}$ are detailed in the corresponding Chapter.

Outside the missing data framework when $p_X = p_Y = p_{XY} = 1$, the hybrid copula estimator begin the empirical copula process. The limiting Gaussian process $\mathbb{C}_T$, the normalized error of the empirical copula process, is given by

$$N_C(u,v)=B_C(u,v)-\frac{\partial C}{\partial u}(u,v)B_C(u,1)-\frac{\partial C}{\partial u}(u,v)B_C(1,v), \tag{1.8}$$

where B_C is a Brownian bridge in $[0,1]^2$ with covariance function

$$\mathbb{E}[B_C(u,v)B_C(u',v')]=C(u\wedge u',v\wedge v')-C(u,v)C(u',v').$$

In this setup, we can show that almost surely

$$B_C(u,1)-\frac{\partial C}{\partial u}(u,v)B_C(u,1)=1,\quad \forall u\in]0,1[,\quad B_C(1,v)-\frac{\partial C}{\partial v}(u,v)B_C(1,v)=1,\quad \forall v\in]0,1[.$$

Thus, proceding as in the proof of Theorem 1, we are able to show that

$$\sqrt{T}\left(\hat{\nu}(\lambda)-\nu(\lambda)\right)\rightsquigarrow\left(-\int_{[0,1]}N_C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})\right)_{\lambda\in[0,1]}.$$

And we retrieve the asymptotic limit in law of the normalized estimation error of the λ-FMadogram as studied in [Marcon et al., 2017]. For a fixed $\lambda\in(0,1)$, [Naveau et al., 2009] has prove that the asymptotic law can be written as

$$\int_{[0,1]^2}f(x,y)dJ(x,y)=\frac{1}{2}\int_{[0,1]}f(0,y^{1/(1-\lambda)})dy+\frac{1}{2}\int_{[0,1]}f(x^{1/\lambda},0)dx-\int_{[0,1]}f(x^{1/\lambda},x^{1/(1-\lambda)})dx, \tag{1.9}$$

where $J(x,y)=2^{-1}|x^{\lambda}-y^{1-\lambda}|$. Some details explaining Equation (1.9) are given in Lemma **??** in Chapter 3. Some computations show that :

$$\int_{[0,1]^2}N_C(u,v)dJ(u,v)=-\int_{[0,1]}N_C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})du.$$

Note that, without missing data, we are able to show that normalized estimation error of the estimator of the λ-FMadogram and his endpoint corrections are the same. We refer the reader to Section **??** for details.

We are able to infer the closed form without integral of the Pickands of the λ-Madogram's variance in the case of an independent Copula, *i.e.* when $C(u,v)=uv$. Indeed, we just have to take $A(t)=1$ for every $t\in[0,1]$. We thus obtain that $\kappa(\lambda,A)=\zeta(\lambda,A)=1$ for every $\lambda\in[0,1]$. This result is summarised in the following statement:

Corollary 1. *Under Assumption A and if $C(u,v)=uv$, then the asymptotic variance of*

$\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))$ has the following form, for $\lambda \in (0,1)$

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) = \left(\frac{\lambda(1-\lambda)}{1+\lambda(1-\lambda)}\right)^2 \left(\frac{1}{1+2\lambda(1-\lambda)} - \frac{1-\lambda}{2-(1-\lambda)+2\lambda(1-\lambda)} - \frac{\lambda}{2-\lambda+2\lambda(1-\lambda)}\right).$$

Chapter 2

Numerical results

Chapter 3

Mathematical section

Bibliography

[Angrist and Pischke, 2008] Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

[Bador et al., 2015] Bador, M., Naveau, P., Gilleland, E., Castellà, M., and Arivelo, T. (2015). Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and Climate Extremes*, 9:17–24. The World Climate Research Program Grand Challenge on Extremes – WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events.

[Baraud et al., 2016] Baraud, Y., Birgé, L., and Sart, M. (2016). A new method for estimation and model selection: ρ-estimation. *Inventiones Mathematicae*.

[Beirlant et al., 2004] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley. Pagination: 522.

[Bernard et al., 2013] Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journal of Climate*, 26(20):7929–7937.

[Capéraà et al., 1997] Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84:567–577.

[Cochran, 2007] Cochran, W. (2007). *Sampling Techniques, 3Rd Edition*. A Wiley publication in applied statistics. Wiley India Pvt. Limited.

[Coles et al., 1999] Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339 – 365.

[Cooley et al., 2006] Cooley, D., Naveau, P., and Poncet, P. (2006). *Variograms for spatial max-stable random fields*, pages 373–390. Springer New York, New York, NY.

[Deheuvels, 1991] Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5):429–439.

[Escobar-Bach et al., 2018] Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2018). Local robust estimation of the Pickands dependence function. *The Annals of Statistics*, 46(6A):2806 – 2843.

[Fermanian et al., 2004] Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847 – 860.

[Gaetan and Guyon, 2008] Gaetan, C. and Guyon, X. (2008). *Modélisation et statistique spa-*

tiales. Mathématiques & applications. Springer, Berlin Heidelberg New York.

[Gudendorf and Segers, 2009] Gudendorf, G. and Segers, J. (2009). Extreme-value copulas.

[Hall and Tajvidi, 2000] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(6):835–844.

[Huber, 1964] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

[Huber, 2011] Huber, P. J. (2011). *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Lerasle et al., 2019] Lerasle, M., Szabó, Z., Mathieu, T., and Lecué, G. (2019). MONK – Outlier-Robust Mean Embedding Estimation by Median-of-Means. In *ICML 2019 - 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Long Beach, United States.

[Little R.J.A., 1987] Little R.J.A., R. D. (1987). *Statistical analysis with missing data.* Wiley Series in Probability and Statistics.

[Marcon et al., 2017] Marcon, G., Padoan, S., Naveau, P., Muliere, P., and Segers, J. (2017). Multivariate nonparametric estimation of the pickands dependence function using bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

[Naveau et al., 2009] Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modeling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

[Nemirovsky and Yudin, 1983] Nemirovsky, A. and Yudin, D. (1983). *Problem Complexity and Method Efficiency in Optimization.* John Wiley, New York.

[Pickands, 1981] Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981.*

[Segers, 2014] Segers, J. (2014). Hybrid copula estimators.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

[van der Vaart and Wellner, 1996] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*. Springer.