Modelling extreme values of processes observed at irregular time step. Application to significant wave height.

Nicolas Raillard, Pierre Ailliot, Jian-Feng Yao

Modelling extreme values of processes observed at irregular time step. Application to significant wave height.

Nicolas Raillard[1,2,3], Pierre Ailliot[1], Jian-feng Yao[4]

Septembre 16, 2011

[1] Laboratoire de Mathématiques, Université de Bretagne Occidentale, Brest, France.
[2] Laboratoire d'Océanographie Spatiale, IFREMER, France.
[3] Laboratoire de Mathématiques de Rennes, Université de Rennes 1, Rennes, France.
[4] Department of Statistics & Actuarial Sciences, The University of Hong Kong, Pokfulam, Hong-Kong.

Abstract

The distribution of extremes such as flood peaks, maximum wave height or minimum daily returns over annual or other time intervals is of common interest to many disciplines including the natural and social sciences. This work is motivated by the analysis of extreme values from times series of significant wave heights observed in North Atlantic. One of these time series exhibits missing data (buoy data) and another one irregular time sampling (satellite data). This situation is frequent when considering environmental data sets and new statistical methods are needed to analyze the extremal behavior of such time series. The method proposed in this work consists in assuming that the behavior of the process above a high threshold is well approximated by a max-stable process which parameters are estimated by maximizing a composite likelihood function. The consistency of these estimates is established. Then, using an extensive set of simulated time series, we assess the finite-sample behavior of our estimates for small to medium sample sizes and compare them to other available estimation methods proposed in the literature for analyzing the extremal behavior of stochastic processes on the basis of standard validation statistics. Finally, a detailed study of significant wave height data is performed. It is shown that the proposed methodology may be used to estimate characteristics of extreme significant wave height at any location in the ocean from altimeter satellite data.

Keywords: Extreme values, time series, max-stable process, composite likelihood, consistency, irregular time sampling, significant wave height, altimeter

1 Introduction

Extreme events have become a major concern in risk management or engineering and appropriate statistical methods are needed to derive estimations of the extremal properties of various phenomena from complex data sets. For example, the design of marine structure depends on the extreme waves that they may face and mainly three sources of data can be used to estimate the extreme quantiles of the significant wave height (Hs) distribution :

- *Reanalysis data* which provide long time series (typically a few decades) at regular time step and without missing values but which tend to smooth out extreme values.

- *Buoy data* which generally give more accurate observations but on shorter time period (typically a few years with missing values) and have a poor spatial distribution.
- *Satellite data* which also provide accurate observations of the wave height over the last 20 years. However, the time series obtained by selecting all the satellite data available at a given location exhibits a complex irregular time sampling depending on the number and the tracks of the operating satellites.

The motivation of this work is to develop statistical methods for analyzing the extremal properties of Hs based on such data sets. In particular, the method that we propose can be used for estimating various characteristics of the extremal behavior of processes (high quantiles, return period, storm durations,...) observed at regular or irregular time steps whereas the other existing methods are not appropriate in the last case.

Two methods are commonly used in the literature for the statistical analysis of extreme events (see e.g. [9], [4], [3], [8], and references therein). The first one, generally referred as the block maxima method, relies on probabilistic results which suggest using the generalized extreme-value (GEV) distribution for modeling the maximum of a large number of identically distributed random variables. The main drawback of this approach is the waste of data induced by taking the maximum over a large block, typically one year for meteorological applications, before fitting the GEV distribution. Hence another approach, generally referred as Peaks Over Threshold (POT), consists in keeping all the observations above a threshold which is chosen high enough in order to ensure that the distribution of the excesses above this threshold is well approximated by a generalized Pareto distribution (GPD). A problem when using POT approach for time series of dependent data is that clusters of consecutive dependent exceedances are generally observed, especially when the time-lag between successive observations is smaller than the characteristic duration of extreme events and thus some changes are needed in practice when using POT method in this context. Usually, the first step consists in identifying clusters ("declustering" step) before fitting a GPD distribution to the sample of clusters maxima. This methodology also leads to waste data since only maxima within each cluster are used to fit the GPD and relies on arbitrary rules for declustering the data which are even more difficult to chose in presence of missing values or irregular time sampling. Hence another approach, initially proposed in [20], consists in keeping all exceedances and model the dependence structure between neighboring excesses as a first order Markov chain which transition kernel is derived from bivariate extreme value theory. This method has been applied successfully to various meteorological time series (see e.g. [16]).

In Section 2 of this paper we propose an alternative approach for modeling the dependence structure above a high threshold in which the time series of exceedances is assumed to be a realization of a censored max-stable process. Although the methodology is not restricted to a specific model of max-stable processes, we focus on the continuous-time model proposed in [19] which has a simple meteorological interpretation and can easily handle observations available at irregular time steps.

Parameter estimation is discussed in Section 3. Since the likelihood function is not tractable, we propose to estimate the unknown parameters by maximizing a composite likelihood function and we prove theoretical results which indicate that these estimates are consistent. We also discuss the properties of the estimates for small to medium size samples, comparable to those typically available in usual applications, using simulations.

In Section 4, we validate the methodology on time series simulated using different classical time series models. The fitted censored max-stable process can be easily simulated and this allows to estimate various quantities of interest for the applications, such as quantiles, return periods or characteristics of the sojourns above high threshold using Monte-Carlo simulations.

The results are compared to the ones obtained using the most classical approaches mentioned above.

One advantage of our approach is that it can easily handle time series with missing values or irregular sampling. This is illustrated in Section 5 on Hs data in North Atlantic where the methodology is used to compare the extremal properties of reanalysis, buoy and satellite data.

Conclusions and key findings are given in Section 7.

2 Censored max-stable process

In this section we introduce an original model for describing the extremal behavior of a time series of dependent observations. In order to motivate this model, we focus in Section 2.1 on the simpler case of independent and identically distributed (iid) random variables and discuss the interpretation of the POT method in terms of censoring. The generalization to time sequences of dependent observations is discussed in Section 2.2.

2.1 Threshold models and censoring in the independent case

Probably the most classical approach for modelling the extremal properties of an iid sample $(X_1, ..., X_n)$ consists in using the "block maxima" approach. It relies on probabilistic results originating in [10] which suggest approximating the distribution of $M_n = \max_{i=1,...,n} X_i$ by a GEV distribution. The GEV distribution has the following cumulative distribution function (cdf)

$$F(x;\mu,\sigma,\xi) = \begin{cases} \exp\{-[1+\xi\frac{x-\mu}{\sigma}]^{-\frac{1}{\xi}}\} & (\xi \neq 0) \\ \exp\{-\exp[-\frac{x-\mu}{\sigma}]\} & (\xi = 0) \end{cases} \quad (1)$$

defined for x such that $1+\xi\frac{x-\mu}{\sigma} > 0$ with parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$. For practical applications, the data are blocked into blocks of equal length and a GEV distribution is fitted to the sample obtained by keeping the maxima within in each block. The choice of the size of the blocks is critical in practice. As concerns environmental time series, the GEV distribution is generally fitted to the time series of annual maxima in order to remove seasonal effects. This leads to an important waste of data and the three parameters of the GEV distribution have to be estimated based on small samples which size corresponds to the number of years of observation with no or few missing values (a few decades in the best cases). Although many methods have been proposed in the literature to provide estimates which have a good behavior on small samples (see [2] and references therein), it remains an important issue for practical applications.

The POT approach is the classical alternative to the block maxima approach. It is less wasteful of data since it keeps all the data above a high threshold u which is chosen such that the conditional distribution $\mathbb{P}[X_i \leq x | X_i > u]$ is well approximated by a GPD with cdf

$$G(x;\mu,\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi} & (\xi \neq 0) \\ 1-\exp[-\frac{x-\mu}{\sigma}] & (\xi = 0) \end{cases}$$

defined for $x \geq \mu$ such that $\left(1+\xi\frac{x-\mu}{\sigma}\right) \geq 0$ with $\mu = u$ and parameters $\sigma > 0$ and $\xi \in \mathbb{R}$. Again, the use of the GPD is motivated by probabilistic results and various methods have been proposed in the literature for estimating the two parameters of the GPD based on the sample of exceedances as well as to choose u, although the latter is a more difficult problem to handle (see [6]). Once u is chosen, the most usual method for estimating the unknown parameters consists

in maximizing the likelihood function given by

$$\begin{aligned} L(\lambda,\sigma,\xi;X_1,...,X_n) &= \lambda^{N_u}(1-\lambda)^{n-N_u}\prod_{i=1|X_i>u} g(x_i;u,\sigma,\xi) \\ &= \prod_{i=1|X_i\leq u}\lambda\prod_{i=1|X_i>u}(1-\lambda)g(x_i;u,\sigma,\xi) \end{aligned} \tag{2}$$

where $\lambda=\mathbb{P}(X_i\leq u)$, N_u is the number of observations below the threshold u and $g(x;\mu,\sigma,\xi)$ is the probability density function (pdf) of the GPD. It is well known that the conditional distribution of the exceedances of a GPD above an arbitrary threshold is also a GPD and this allows to interpret (2) as the likelihood of an iid sample of a GPD censored at the threshold u. More precisely, let $(\tilde{X}_1,\tilde{X}_2,...,\tilde{X}_n)$ be an iid sample of a GPD with parameter $(\mu,\tilde{\sigma},\xi)$ and consider the censored random variable

$$Y_i = u\mathbb{1}_{[\tilde{X}_i\leq u]}+\tilde{X}_i\mathbb{1}_{[\tilde{X}_i>u]} = \begin{cases} u & \text{if } \tilde{X}_i\leq u \\ \tilde{X}_i & \text{if } \tilde{X}_i>u \end{cases}$$

where the threeshold u belongs to the support of the GPD distribution. We have $P(Y_i=u)=\lambda$ with $\lambda=G(u;\mu,\tilde{\sigma},\xi)$ and, for $x>u$,

$$\begin{aligned} \mathbb{P}(Y_i\geq x) &= \mathbb{P}(\tilde{X}_i\geq u)\mathbb{P}(\tilde{X}_i\geq x|\tilde{X}_i\geq u) \\ &= (1-G(u;\mu,\tilde{\sigma},\xi))\frac{1-G(x;\mu,\tilde{\sigma},\xi)}{1-G(u;\mu,\tilde{\sigma},\xi)} \\ &= (1-\lambda)(1-G(x;u,\sigma,\xi)) \end{aligned}$$

with $\sigma=\tilde{\sigma}\left(1+\xi\frac{u-\mu}{\tilde{\sigma}}\right)$ and thus (2) is the likelihood of $(Y_1,...,Y_n)$. Finally, the assumptions made when using the POT approach are equivalent to assuming that the original sample $(X_1,...,X_n)$ satisfies

$$u\mathbb{1}_{[X_i\leq u]}+X_i\mathbb{1}_{[X_i>u]} = u\mathbb{1}_{[\tilde{X}_i\leq u]}+\tilde{X}_i\mathbb{1}_{[\tilde{X}_i>u]} \tag{3}$$

for all $i\in\{1,...,n\}$ where $\left(\tilde{X}_1,...,\tilde{X}_n\right)$ is an iid sample of a GPD.

We will see below that this interpretation of the POT approach in terms of censoring may have advantages for modeling purpose. From a numerical point of view, it can be viewed as a reparametrization of the likelihood function. Maximizing (2) over (λ,σ,ξ) leads to the following estimate for λ, $\widehat{\lambda}=\frac{N_u}{n}$ and this estimate has the desirable property to be easy to interpret and be independent of the estimates of the two other parameters. On the other hand, maximizing (2) over $(\mu,\tilde{\sigma},\xi)$ leads to a more complicated 3-dimensional optimization problem and estimates which are correlated and this can be problematic for some applications (see [17]).

Although the GPD distribution is the most usual approximation for the tail of the distribution when modeling the exceedances over a high threshold , other approximations have been proposed in the literature. In particular, we have

$$G(x;\mu,\sigma,\xi)\approx F(x;\mu,\sigma,\xi)$$

for "high" values of x and this suggests that similar results will be obtained if we model the distribution of $\tilde{X}_i$ by a GEV distribution instead of a GPD. We performed various tests on simulated samples which confirms that both approximations lead to similar results in practice.

It has been shown that the tail approximations discussed above are still valid for dependent sequences under mild conditions (see [12]) and this justify the use of both the block maxima and POT approaches for analyzing the extremal behavior of a time series. One difficulty when

using POT approach in this context is that clusters of consecutive dependent exceedances are generally observed whereas the likelihood function (2) is the true likelihood of the sample only if the exceedances are independent. The most usual method is then to apply a declustering step and keep only the maxima within each cluster in order to obtain a sample of approximatively independent exceedances. It leads to waste data and thus degrade the quality of the estimates but also to lose information on the dynamics of the process inside the clusters. This may be problematic for applications sensitive to the within-cluster behavior. A method which generalizes the principle of censoring for dependent sequences and permits to keeps all exceedances above the selected threshold is proposed in the next section.

2.2 Censored max-stable process

We now consider a sample $(X_{t_1}, ..., X_{t_n})$ of a stochastic process $\{X_t\}$ observed at time $(t_1, ..., t_n)$. It is generally assumed in the literature that the observations are available at regular time step (i.e. $t_{i+1} - t_i = t_{j+1} - t_j$ for all $(i, j) \in \{1, ..., n-1\}$) but we would like to have a method which is flexible enough to deal with irregular time sampling (see Section 5 for practical motivations). We propose to analyse the extremal behaviour of such data set by extending the POT approach discussed above and model the distribution of the process $\{X_t\}$ over a high threshold u by a censored max-stable process. The theory of max-stable processes (see [7], [8]) is a natural generalization of the traditional univariate max-stable theory which was used above to motivate the choice of the GEV distribution in the iid case. Several families of max-stable process have been proposed in the literature (see e.g. [19, 18]), but hereafter we will focus on the specific *Gaussian extreme value process* introduced in [19] although the methodology is flexible enough to deal with other models.

More precisely, we assume that

$$u\mathbb{1}_{[X_t \leq u]} + X_t \mathbb{1}_{[X_t > u]} = u\mathbb{1}_{[\tilde{X}_t \leq u]} + \tilde{X}_t \mathbb{1}_{[\tilde{X}_t > u]} \tag{4}$$

holds for all t where u is a fixed threshold and $\{\tilde{X}_t\}$ is a stationary Gaussian extreme value process with parameter $\theta = (\mu, \sigma, \xi, \nu) \in (-\infty, +\infty) \times (0, +\infty) \times (-\infty, +\infty) \times (0, +\infty)$ defined below.

- The marginal distribution of $\{\tilde{X}_t\}$ is a GEV distribution with parameter (μ, σ, ξ). With this assumption, the process $\{Z_t\}$ obtained by applying the following marginal transformation

$$Z_t = -\frac{1}{\log(F(\tilde{X}_t; \mu, \sigma, \xi))} \tag{5}$$

  is a stationary process with unit Fréchet (i.e. GEV distribution with parameter $(1, 1, 1)$) marginal distribution

- We further assume, following [19], that

$$Z_t = \max\left\{\frac{\zeta_i}{\nu\sqrt{2\pi}} \exp\left(-\frac{(s_i - t)^2}{2\nu^2}\right)\right\}$$

  where $\{(\zeta_i, s_i), i \geq 1\}$ denote the points of a Poisson process on $(0, \infty) \times \mathbb{R}$ with intensity measure $\zeta^{-2} d\zeta \times ds$.

We focus on Gaussian extreme value processes because they have a nice meteorological interpretation (see [19]) and provide a flexible class of models. The parameters (μ, σ, ξ) are related

to the marginal distribution and can be interpreted respectively as location, scale and shape parameters, whereas the parameter ν is related to the temporal structure of the process and may be interpreted as the typical duration of the storms. More precisely, we have (see [19]):

$$\mathbb{P}(Z_{t_1} \leq z_{t_1}, Z_{t_2} \leq z_{t_2}) = F_Z(z_{t_1}, z_{t_2}; \nu) = \exp\left[-V(z_{t_1}, z_{t_2}; \nu)\right], \tag{6}$$

where

$$V(z_{t_1}, z_{t_2}; \nu) = \frac{1}{z_{t_1}}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_{t_2}}{z_{t_1}}\right) + \frac{1}{z_{t_2}}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_{t_1}}{z_{t_2}}\right) \tag{7}$$

with $a = \frac{|t_1 - t_2|}{\nu}$ and Φ the cdf of the standard normal distribution. The limit cases $\nu \to 0$ and $\nu \to +\infty$ corresponds respectively to independence and perfect dependence.

Applying the inverse marginal transformation leads to the following bivariate cdf for the Gaussian extreme value process $\{\tilde{X}_t\}$:

$$\begin{aligned} F_{\tilde{X}}(\tilde{x}_{t_1}, \tilde{x}_{t_2}; \theta) &= \mathbb{P}(\tilde{X}_{t_1} \leq \tilde{x}_{t_1}, \tilde{X}_{t_1} \leq \tilde{x}_{t_2}) \\ &= \exp\left[-\frac{1}{z_{t_1}}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_{t_2}}{z_{t_1}}\right) - \frac{1}{z_{t_2}}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{z_{t_1}}{z_{t_2}}\right)\right] \end{aligned} \tag{8}$$

with $a = \frac{|t_1 - t_2|}{\nu}$ and $z_{t_i} = \frac{-1}{\log F(\tilde{x}_{t_i}; \mu, \sigma, \xi)}$.

The distribution of (Y_{t_1}, Y_{t_2}), where $Y_t = u\mathbb{1}_{[\tilde{X}_t \leq u]} + \tilde{X}_t\mathbb{1}_{[\tilde{X}_t > u]}$ is the censored Gaussian extreme value process, has the following bivariate pdf

$$p_Y(y_{t_1}, y_{t_2}; \theta) = \begin{cases} F_{\tilde{X}}(u, u; \theta) & \text{if } y_{t_1} = u \text{ and } y_{t_2} = u, \\ \frac{\partial F_{\tilde{X}}}{\partial \tilde{x}_{t_1}}(y_{t_1}, u; \theta) & \text{if } y_{t_1} > u \text{ and } y_{t_2} = u, \\ \frac{\partial F_{\tilde{X}}}{\partial \tilde{x}_{t_2}}(u, y_{t_2}; \theta) & \text{if } y_{t_1} = u \text{ and } y_{t_2} > u, \\ \frac{\partial^2 F_{\tilde{X}}}{\partial \tilde{x}_{t_1} \partial \tilde{x}_{t_2}}(y_{t_1}, y_{t_2}; \theta) & \text{if } y_{t_1} > u \text{ and } y_{t_2} > u, \end{cases} \tag{9}$$

with respect to the product measure $m \times m$, where $m(dx) = \delta_u(dx) + dx$ is the measure obtained by mixing the Dirac measure at u with the Lebesgue measure.

Similar approximations, motivated using probabilistic results from the bivariate extreme value theory, have already been proposed in the literature for modeling the bivariate distribution of neighboring exceedances (see [20], [16] and references therein). In these papers, it is further assumed that the censored process is a Markov chain observed at regular time step. With these assumptions the likelihood function can be derived from the bivariate distribution of successive observations and optimized to compute the maximum likelihood estimates. In our case, the likelihood function is not tractable and an alternative estimation strategy is needed. This is discussed in the next section.

3 Parameter estimation

In this section, we address the problem of estimating the parameters of the censored Gaussian extreme value process introduced in the previous section. In Section 3.1, we introduce the composite likelihood function which will be maximized to define the maximum composite likelihood estimates. Then, in Section 3.2, we prove various results related to the asymptotic properties of these estimates before illustrating these results using simulations in Section 3.3.

3.1 Composite likelihood functions

In this section $(y_{t_1}, ..., y_{t_n}) \in (u, +\infty)^n$ denotes a realization of a Gaussian extreme value process $\{Y_t\}$ with unknown parameter $\theta^* = (\mu^*, \sigma^*, \xi^*, \nu^*)$ censored at the threshold $u \geq 0$ and observed at times $(t_1, ..., t_n)$. There is no known tractable expression for the joint distribution of such sample and, as a consequence, the maximum likelihood estimates can not be computed. We have seen however in the previous section that the marginal and bivariate distributions have tractable expressions and this suggests replacing the likelihood by the composite likelihood functions introduced below (see e.g. [13, 23, 22, 5]).

- The *independent likelihood* function defined as

$$IL(\theta; y_{t_1}, ..., y_{t_n}) = \prod_{i=1}^{n} p_Y(y_{t_i}; \theta). \tag{10}$$

  where $p_Y(y_t; \theta)$ is the pdf of the marginal distribution of Y_{t_i}, with respect to the measure m. It is given by

$$p_Y(y_t; \theta) = \begin{cases} F(u; \mu, \sigma, \xi) & \text{if } y_t = u \\ f(y_t; \mu, \sigma, \xi) & \text{if } y_t > u \end{cases}$$

  where F and f denotes respectively the cdf and pdf of the GEV distribution. It corresponds to the likelihood function of an iid sample of a censored GEV distribution (see Section 2.1) and does not depend on the parameter ν which describes the dependence structure of the process. We denote MILE the estimates of (μ, σ, ξ) obtained by maximising this function.

- The *pairwise likelihood* function

$$PL(\theta; y_{t_1}, ..., y_{t_n}) = \prod_{i=1}^{n-1} \prod_{j>i} p_Y(y_{t_i}, y_{t_j}; \theta)^{\omega_{t_i,t_j}}. \tag{11}$$

  with $p_Y(y_{t_i}, y_{t_j}; \theta)$ given by (9) and $\omega_{t_i,t_j} \in \{0, 1\}$ indicates if the pair of observation (y_{t_i}, y_{t_j}) contributes to the pairwise likelihood function. This approach has already been considered for time series with regular time sampling (see [22] and references therein). It is generally assumed that

$$\omega_{t_i,t_j} = \mathbb{1}_{[|i-j| \leq K]} \tag{12}$$

  such that only the pairs of observations which are less than K time unit apart are kept to build the pairwise likelihood function. Hereafter we will denote PL_K the corresponding pairwise likelihood function estimates and MPL_KE the estimates obtained by maximizing this function. Keeping only the neighboring observations (i.e. using $K = 1$) has clear computational benefits, since it permits to significantly reduce the number of terms in the product (11), and may also lead to more efficient estimates in practice (see [22] and Section 3.3.). Another strategy would consist in keeping the pairs of observations separated by a time lag smaller than T and take

$$\omega_{t_i,t_j} = \mathbb{1}_{[|t_i - t_j| \leq T]} \tag{13}$$

  This second strategy is similar to the first one when the process is observed at regular time sampling but not in the irregular case. This will be further discussed using simulations in Section 3.3 .

- The *Markovian likelihood* function is defined as:

$$\begin{aligned} ML(\theta; y_{t_1}, ..., y_{t_n}) &= p_Y(y_{t_1}; \theta) \prod_{i=2}^{n} p_Y(y_{t_i} | y_{t_{i-1}}; \theta) \\ &= \frac{\prod_{i=1}^{n-1} p_Y(y_{t_i}, y_{t_{i-1}}; \theta)}{\prod_{i=2}^{n-1} p_Y(y_{t_i}; \theta)} \\ &= \frac{\prod_{i=1}^{n-1} p_Y(y_{t_i}, y_{t_{i-1}}; \theta)}{\prod_{i=2}^{n-1} p_Y(y_{t_i}; \theta)} \\ &= \frac{PL_1(\theta; y_{t_1}, ..., y_{t_n})}{IL(\theta; y_{t_2}, ..., y_{t_n})} \end{aligned} \tag{14}$$

  We denote MMLE the values of θ maximizing this function. When the process is observed at regular time step, we retrieve the Markovian model considered in [20, 16] for the specific bivariate max-stable distribution associated to the Gaussian extreme value process.

From a numerical point of view, we found useful to use a two-stage procedure, where the parameters (μ, σ, ξ) of the marginal distribution are first estimated by maximizing the independent likelihood function, before estimating the dependence parameter ν by maximizing the pairwise likelihood function over ν with the parameters of the marginal distribution being fixed to the values obtained in the first step. It permits to reduce computational time (the independent likelihood function can be evaluated quickly compared to the pairwise likelihood function) and avoid divergence problems which may occur when optimizing the pairwise likelihood function simultaneously over all the parameters with an inappropriate starting point. Eventually, a global optimization of the pairwise likelihood function may be performed to refine the estimates obtained with the two-stage procedure.

3.2 Consistency of MPL$_1$E with known marginal distribution

In this section, we prove results related to the consistency of the MPL$_1$E introduced in the previous section in an idealized situation. More precisely, we assume that there is no censoring ($u = -\infty$), that the process is observed at regular time step and that the marginal parameters (μ^*, σ^*, ξ^*) are known. This latter assumption is not realistic in practice but mimics the second step of the two-stage procedure introduced in the previous section and permits to avoid supplementary difficulties which appear when the support of the distribution depends on the parameters. With these assumptions, we can apply the marginal transformation (5) and assume, without loss of generality, that the process has a unit Fréchet marginal distribution. Finally, we denote $(Z_1, ..., Z_n)$ a Gaussian extreme value process observed at time $t_i = i$ for $i \in \{1, ...n\}$ with parameter $(1, 1, 1, \nu^*)$ and $\hat{\nu}$ the estimate of $\nu^* \in (0, +\infty)$ obtained my maximizing the PL_1 function over ν.

The proof of the following theorem is postponed to Section 6.

Theorem 3.1. *Assume that parameter space is* $\Theta = [\nu_-,\ \nu_+]$ *with* $0 < \nu_- < \nu_+$ *for* ν. *Then* $\hat{\nu}$ *is a consistent estimate of* ν^*.

3.3 Simulation Study

A simulation study was undertaken to assess the accuracy of the estimates introduced in Section 3.1 for practical applications. Random samples were generated from a Gaussian extreme value process with parameters $\mu = 0$, $\sigma = 1$ and $\xi = 0.3$ which corresponds to realistic values for

environmental applications. We performed various experiments to study the impact of the sample size n (see Figure 1), of the dependence parameter ν (see Figure 2), of the threshold u (see Figure 3) and finally of the strategy considered to select the pair of observations which contribute to the pairwise likelihood function (see Figure 4) on the accuracy of the estimates. For each experiment, 200 independent random samples were generated from a Gaussian extreme value process and the various estimates of (μ, σ, ξ, ν) were determined to derive an empirical root-mean-squared error (RMSE). The MILE only provides estimates of (μ, σ, ξ) but we have also computed the estimate of ν considered in Section 3.2 which is obtained by maximizing the PL_1 function over ν with (μ, σ, ξ) fixed to their true values (see the right panels of Figures 1, 2 and 3).

Let us first focus on the case when the time step between successive observations is constant. According to Figure 1 all the estimates seem to be convergent when the sample size increases and we checked empirically that, when multiplied by $\sqrt{n}$, the errors are almost constant and thus that we retrieve the usual speed of convergence. The MILE are clearly the less efficient estimates whereas MPL_1E, MPL_5E and MMLE give similar results. A closer look reveals that MPL_5E is the less accurate of these three estimates. MMLE slightly outperforms MPL_1E in estimating the scale parameter σ and the shape parameter ξ whereas MPL_1E provides the best estimate of the dependence parameter ν. Figure 2 shows that the RMSE of all the estimates decreases with the dependence parameter ν and that it is more difficult to get reliable estimates when there is a stronger dependence between successive observations. The efficiency of MMLE generally deteriorates quicker when ν increases, and it provides the worst estimates of μ, σ and ν when $\nu \geq 1$ but provides the best estimate of ξ for all the values of ν considered in this experiment. Those results indicate that the Markovian approximation may not be appropriate when the dependence in strong. Figure 3 depicts the behavior of the different estimators when censoring occurs. As expected, all of them worsen when the threshold increases and the number of observations decreases. We can notice that the estimator which most suffers from censoring is the MILE, that MMLE outperforms both MPL_1E and MPL_5E in estimating the parameters of the marginal distribution, but MPL_1E again provides the best estimate of the dependence parameter ν.

Figure 4 shows the influence of the windows considered when defining the neighborhood which are taken into account in the pairwise likelihood functions. The Gaussian extreme value process was simulated using an irregular time sampling (the time lags between successive observations were drawn from a uniform distribution on $[0, 2]$) in order to allow a comparison between the two strategies discussed in Section 3.1: the first one consist in using the K closest observations (see Equation (12)) whereas the second one consists in taking into account all the observations falling within less that K time steps (see Equation (13)). The first strategy is found to always be the best. The evolution of RMSE with K differs according to the strategy, since it is increasing for the first strategy, meaning that the best results will be obtained with $K = 1$, but is generally decreasing for the second strategy and the difference between both strategies decreases when K increases. The comparison with MILE and MMLE indicates again that MMLE slightly outperforms MPL_1E for estimating μ, σ and ξ but MPL_1E provides the best estimate of ν.

The results given in this section suggest using MMLE or MPL_1E for practical applications since their RMSE is generally lower than the ones of the other estimates. Both estimates leads to similar computational cost, the first one may provide slightly better estimates when the dependence between successive observations is small, but is clearly less efficient if the dependence is strong. Finally, it seems reasonable to use the MPL_1E for practical applications and we will mainly focus on this estimate in the next sections.

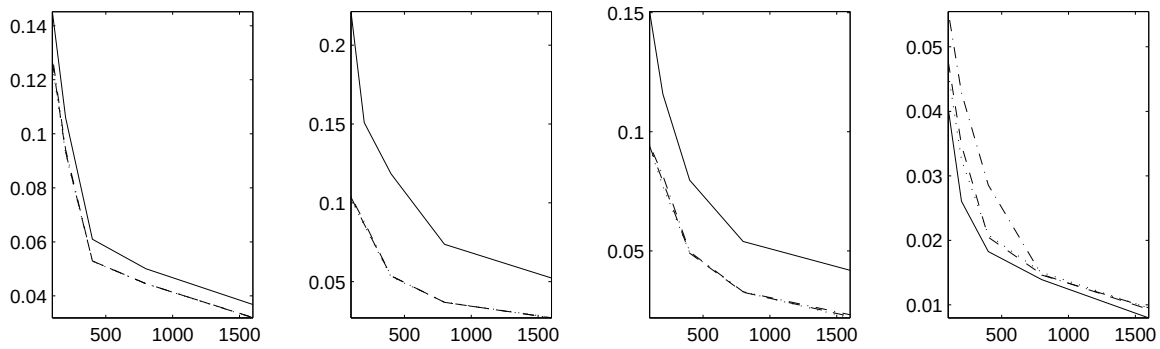

Figure 1: RMSE of the different estimates (y-axis) for various sample size n (x-axis). Results obtained using 200 simulations of a Gaussian extreme value process with parameters values $\mu = 0$, $\sigma = 1$, $\xi = 0.3$ and $\nu = 0.5$, no censoring ($u = -\infty$) and regular time sampling. From left to right panels: estimates of μ; estimates of σ; estimates of ξ; estimates of ν. Solid line: MILE ; dotted line: MMLE ; dashed line: MPL_1E ; dashed-dotted line: MPL_5E.

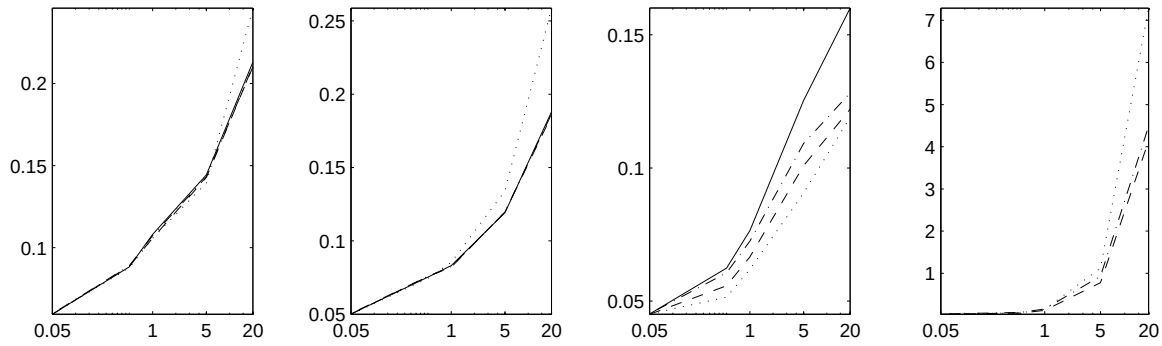

Figure 2: RMSE of the different estimates (y-axis) for various values of ν (x-axis, logscale). Results obtained using 200 simulations of a Gaussian extreme value process with parameters values $\mu = 0$, $\sigma = 1$, $\xi = 0.3$, sample size $n = 300$, no censoring ($u = -\infty$) and regular time sampling. From left to right panels: estimates of μ; estimates of σ; estimates of ξ; estimates of ν. Solid line: MILE ; dotted line: MMLE ; dashed line: MPL_1E ; dashed-dotted line: MPL_5E.

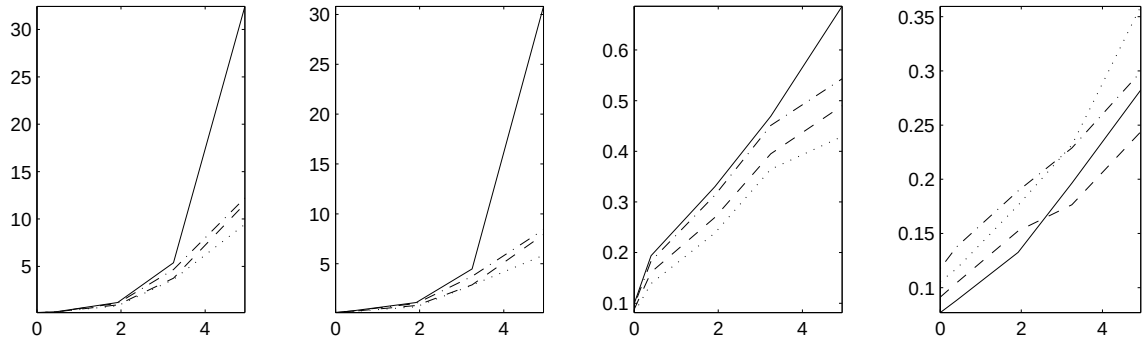

Figure 3: RMSE of the different estimates (y-axis) for various values of the threshold u (x-axis, in terms of quantiles). Results obtained using 200 simulations of a Gaussian extreme value process with parameters values $\mu = 0$, $\sigma = 1$, $\xi = 0.3$, $\nu = 0.5$, sample size $n = 300$ and regular time sampling. From left to right panels: estimates of μ; estimates of σ; estimates of ξ; estimates of ν. Solid line: MILE; dotted line: MMLE ; dashed line: $\mathrm{MPL}_1\mathrm{E}$; dashed-dotted line: $\mathrm{MPL}_5\mathrm{E}$.

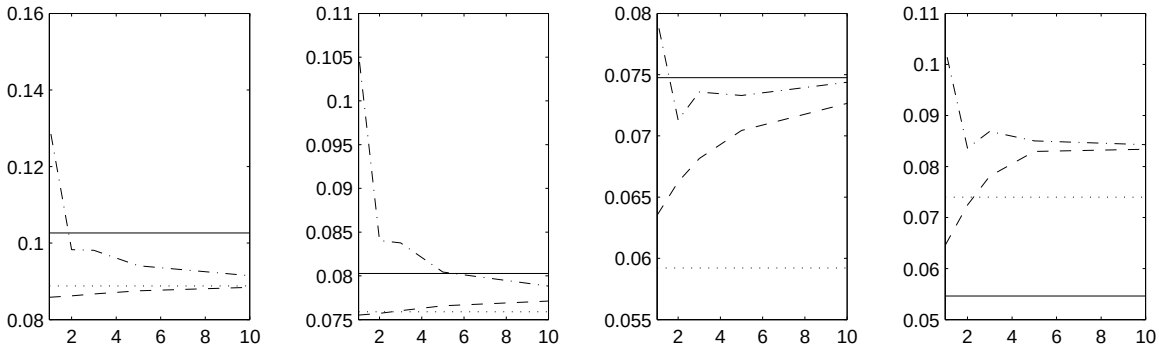

Figure 4: RMSE of the different estimates (y-axis) for various values of K (x-axis). Results obtained using 200 simulations of a Gaussian extreme value process with parameters values $\mu = 0$, $\sigma = 1$, $\xi = 0.3$, $\nu = 0.5$, sample size $n = 300$ and no censoring ($u = -\infty$). The time step between successive observations is drawn from a uniform distribution on the interval $(0, 2)$. From left to right panels: estimates of μ; estimates of σ; estimates of ξ; estimates of ν. Solid line: MILE; dotted line: MMLE ; dashed line: $\mathrm{MPL}_K\mathrm{E}$ with the first weighting strategy (see (12)) ; dashed-dotted line: $\mathrm{MPL}_K\mathrm{E}$ with the second weighting strategy (see (13)).

4 Performance on classical time series models

The lack of data makes it generally difficult to validate models for extreme values when facing real data. In this Section we thus perform a simulation study to check if the proposed methodology is able to catch the extremal properties of some widely used time series models. In Section 4.1, we simulate large samples in order to get parameters estimates with low variance and check if the Gaussian extreme value process provides an appropriate approximation of the extremal behavior of the time series models under consideration. Then, in Section 4.2 we simulate samples with smaller sizes in order to validate the whole methodology in a more realistic context for practical applications.

4.1 Model validation

We have chosen to focus on the following time series models in the sequel:

- **IID**: $\{X_t\}$ is an iid sequence of standard normal distribution.
- **AR(1)**: $\{X_t\}$ is a discrete time stationary process which satisfies

  $$X_t = \alpha X_{t-1} + \sqrt{1-\alpha^2}\epsilon_t$$

  for all t, where $\alpha \in (-1, 1)$ describes the dependence between successive observations and $\{\epsilon_t\}$ is an iid sequence of standard normal distribution. The marginal distribution of $\{X_t\}$ is a standard normal distribution and the extremal index (see [4]) is equal to one (no clustering of extremes). We use the value $\alpha = 0.2$ in the sequel.
- **logARMAX(1)**: $\{X_t\}$ is a discrete time stationary process which satisfies $X_t = \log(U_t)$ where $\{U_t\}$ is a usual ARMAX(1) process defined as

  $$U_t = \max\{(1-\alpha)U_{t-1}, \alpha\epsilon_t\}$$

  for all t, where $\alpha \in (0, 1)$ describes the dependence between successive observations and $\{\epsilon_t\}$ is an iid sequence of unit Fréchet distribution. The log transformation is used to avoid numerical problems which occur when estimating quantities related to heavy tail distributions by simulation: the marginal distribution of $\{U_t\}$ is unit Fréchet whereas the one of $\{X_t\}$ is Gumbel. The extremal index is α. We use the value $\alpha = 0.2$ in the sequel.
- **OU** : $\{X_t\}$ is a continuous time stationary process which satisfies

  $$dX_t = -\alpha X_t dt + \sqrt{2\alpha} dW_t$$

  where $\alpha > 0$ describes the dependence structure and $\{W_t\}$ is a standard Brownian motion. The marginal distribution of $\{X_t\}$ is a standard normal distribution and we retrieve the AR(1) model when the process is sampled at regular time step. We use the value $\alpha = 0.05$ in the sequel and the time lags between successive observations are drawn from a uniform distribution on $[0, 2]$.

For each of these models, we first generate a long realization (equivalent to 100 years with one observation per day) and fit a Gaussian extreme value process to the simulated sequence censored at the 95% quantile by computing the $\mathrm{MPL}_1\mathrm{E}$. Then we compare various characteristics of the original model and the fitted censored Gaussian extreme value process model, namely

- mean number of up-crossings during a given time period (1 year) as a function of the threshold

- mean length of the sojourns above a varying threshold
- mean length of the sojourns below a varying threshold

These quantities have been selected because they summarize important properties of the extremal behavior of the processes and may be important for practitioners. All these quantities were computed using a large number of simulations of both the original time series model (IID, AR(1), logARMAX(1) or OU generated using standard algorithms) and the fitted Gaussian extreme value process (see [18] for more details on simulation algorithms). This is illustrated on Figure 5 which shows realizations of both the AR(1) model and the fitted Gaussian extreme value process whereas Figure 7 permits a more systematic comparison of the behavior of both processes. According to this last figure, the fitted model seems to be able to reproduce both the frequency of up-crossings and time between up-crossings, even for high thresholds, but slightly overestimates the mean length of the clusters above high thresholds. Indeed (see middle panel of Figure 7) for the AR(1) model the mean length of the clusters tends to one when the threshold increases as expected from the theory (no clustering of extremes) whereas the fitted Gaussian extreme value process exhibits some small extremal dependence and thus clustering of extremes. Using a higher threshold for censoring before fitting the Gaussian extreme value process may help improving these results and retrieving extremal independence. According to Figure 6, the results are indeed better for the IID model which is a particular case of the AR(1) model with no dependence between successive between observations and the fitted model seems to be able to catch the extremal properties of an iid sequence of a standard normal distribution. It is not surprising to get similar results for the OU (see Figure 9) and AR(1) models since they are equivalent when the observation step is regular.

Contrary to the other models, the extreme values of logARMAX(1) process tend to cluster. According to Figure 8, the fitted model seems to be able to reproduce both the frequency of up-crossings and the mean length of the cluster, which tends to a limit greater than one as expected from the theory, but seems to overestimate the mean length of the sojourn below high thresholds although the erratic behavior of the curves suggests that the observed differences may be due to the sampling error.

These results indicate that the Gaussian extreme value process is able to catch important properties of the extremal behavior of some usual time series model. Similar results have been obtained on the other models that we have tried.

4.2 Simulation results in a realistic context

In the previous Section, the Gaussian extreme value process was fitted on long realizations to test its ability to describe the extremal behavior of usual time series models. In this Section, we validate the proposed methodology on shorter synthetic time series which length correspond to the amount of data typically available in environmental applications (a few years of data). For each time series model introduced in the previous Section, we have repeated the following numerical experiment 200 times:

- Generate a 5-years sequence (one observation per day) of the reference time series model.
- Fit the censored Gaussian extreme value process on this sequence after censoring at the 95% quantile.
- Compute the 100 years return level for the fitted model. It is defined as the level such that the mean number of **clusters** above this level on a 100 years time period is equal to 1 and is computed by simulating a long realization (1000 years) of the fitted model. This

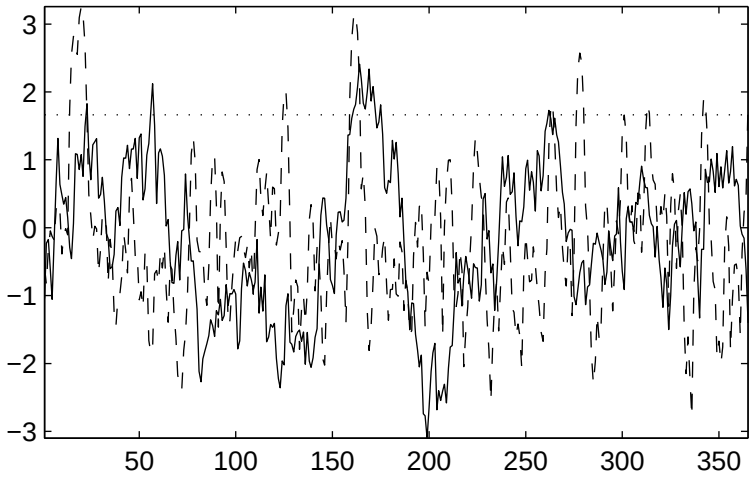

Figure 5: Short samples of the AR(1) model (solid line) and the fitted Gaussian extreme value process (dashed line). The horizontal dotted line is the threshold used for censoring.

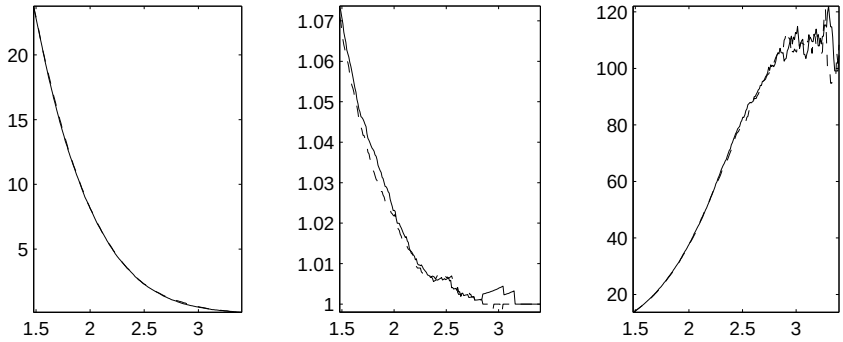

Figure 6: Comparison of the extremal behavior of the IID model (full line) against the fitted Gaussian extreme value process (dashed line). From left to right panels: mean number of up-crossings per year, mean length of the clusters and mean time between consecutive up-crossings as a function of the threshold (x-axis). Results obtained by simulating 1000 years of each model (one observation per day).

quantity was chosen because it is probably the most usual quantity of interest for practical applications. It depends on both the marginal distribution and the dependence structure of the process.

The results are given in Table 1. In every case, the results obtained with the fitted censored Gaussian extreme value process clearly outperform the results obtained with the usual POT method. Both MMLE and $\mathrm{CPL}_1\mathrm{E}$ give similar results for the three models with no extremal dependence (IID, AR(1) and OU) but the results obtained with the $\mathrm{CPL}_1\mathrm{E}$ are clearly superior to the ones obtained with MMLE in terms of accuracy and bias for the logARMAX(1) model. The IID and AR(1) models have almost the same 100 years return period, which is expected from the

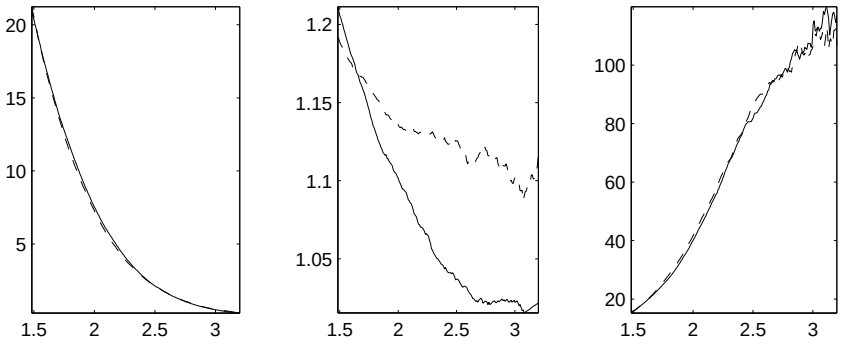

Figure 7: Comparison of the extremal behavior of the AR(1) model (full line) against the fitted Gaussian extreme value process (dashed line). From left to right panels: mean number of up-crossings per year, mean length of the clusters and mean time between consecutive up-crossings as a function of the threshold (x-axis). Results obtained by simulating 1000 years of each model (one observation per day).

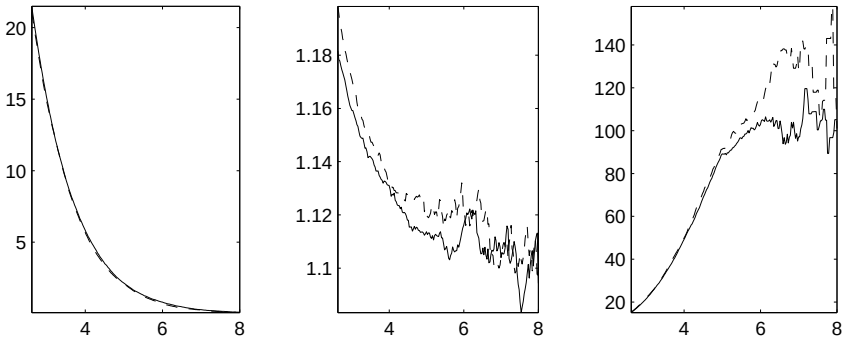

Figure 8: Comparison of the extremal behavior of the logARMAX(1) model (full line) against the fitted Gaussian extreme value process (dashed line). From left to right panels: mean number of up-crossings per year, mean length of the clusters and mean time between consecutive up-crossings as a function of the threshold (x-axis). Results obtained by simulating 1000 years of each model (one observation per day).

theory since they have the same marginal distribution and no extremal dependence. The lower return period of the OU process, which has also the same marginal distribution and no extremal dependence when observed at regular time step, may be due to the irregular time sampling. The extremal dependence of the logARMAX(1) model leads to a higher return value, in comparison with the other models, but also to more important bias and variance in the estimates. This is in adequacy with the simulation results given in Section 3.3 (see Figure 2).

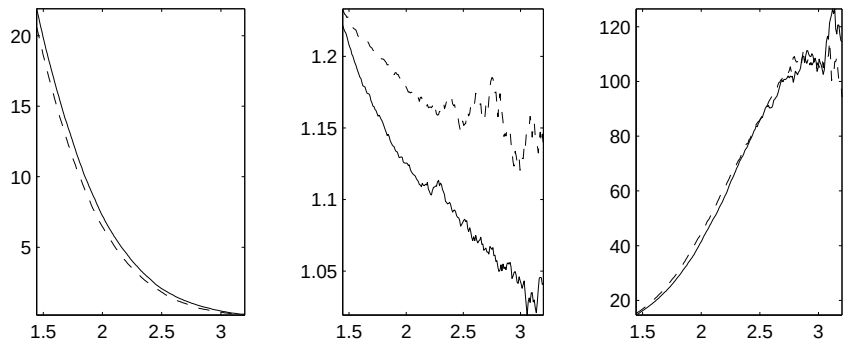

Figure 9: Comparison of the extremal behavior of the OU model (full line) against the fitted Gaussian extreme value process (dashed line). From left to right panels: mean number of up-crossings per year, mean length of the clusters and mean time between consecutive up-crossings as a function of the threshold (x-axis). Results obtained by simulating 1000 years of each model (one observation per day).

Method	IID	AR(1)	logARMAX(1)	OU
True value	4.03	4.03	8.90	3.79
POT	4.18 (3.12—6.02)	4.16 (3.14—5.68)	12.19 (5.31 — 29.96)	3.21 (2.31—5.04)
MMLE	3.89 (3.12—5.15)	3.90 (3.09—4.95)	18.77 (5.88—36.96)	3.62 (2.61—5.05)
CPL_1E	3.84 (3.11—5.17)	3.76 (3.16—4.72)	9.52 (5.54—17.09)	3.62 (2.61—4.98)

Table 1: Mean value of the estimated 100 years return level with 90% fluctuation intervals in parenthesis. Simulation results based on 200 independent 5-years synthetic sequences of each model. A declustering step was applied in the POT method (see [4]).

5 Application to significant wave height

In this section, the proposed methodology is used to analyze the extremal behavior of the three time series of significant wave height (Hs) briefly introduced below.

- **Buoy data**. We focus on data from the buoy Brittany (station 62163), which is part of the UK Met Office monitoring network. It is located at position (47.5 N, 8.5 W) and provides hourly significant wave height data. Buoys are often considered as a reference to provide ground-truth for the other datasets introduced below. In this work, we consider 10 years of data, from 1995 until 2005 (no data available for 2000) and focus on the months of December in order to get rid of the seasonal components. Blocking the data by month leads to waste data and probably to lose important information on extreme events. The development of non-stationary models which include seasonal and interannual components and can be fitted on the whole time series will be the topic of future works. Missing values represent about 7.7% of the data and are generally associated to extreme events (breakdowns generally occur during storms) and this is an important issue when implementing block maxima or POT approaches. Buoys provide accurate information on sea-state conditions but are sparsely distributed over the ocean and there is generally no buoy at the location of interest for a particular application. In such situation, it is important to be able to derive estimates

of the extremal behavior of Hs from the other sources of data introduced below which are available all over the oceans.

- **Reanalysis data**. The ERA-interim dataset provides from a global reanalysis carried out by the European Centre for Medium-Range Weather Forecasts (ECMWF). It can be freely downloaded and used for scientific purposes[1] For this study, we have considered the Hs data available at the location (48 N, 9 W) which is the closest grid point to the buoy Brittany and also extracted the months of December. It leads to a long time series (21 years from 1989 until 2009) which can be analyzed using standard statistical methods since the time sampling is regular (6 hours between successive observations) and there is no missing data. Figure 5 shows both the buoy and ERA-interim time series for the month of December 2005. The agreement is generally good although reanalysis data tends to be smoother than boy data and the quantile-quantile plot (see Figure 11) exhibits some differences between the empirical marginal distributions.

- **Satellite data**. The observations consist of the significant wave height measured at discrete locations along one-dimensional tracks from seven different satellite altimeters which have been deployed progressively since 1991.The dataset and information on it can be freely downloaded[2]. Satellite data exhibits a rather complicated space-time sampling and generally do not provide observations exactly at the location of interest. In order to avoid using interpolation methods, which may smooth the data (see [1]), we have decided to consider the time series obtained by keeping all the closest observations in the tracks which intersect a $1.5^o \times 1.5^o$ box centered on the location of interest (see e.g [24] and [25]). We have retrieved Hs data for 18 months of December, from 1992 until 2009. Figure 5 shows the resulting series for the month of December 2005 at the location of the buoy Britanny. There is fewer data compared to the other datasets and the time sampling is irregular (durations between successive observations ranges from a few minutes to 3 days with a mean value of 19 hours) which prevents from using usual statistical methods (block maxima or POT) for analyzing the extremal behavior of the dataset. Satellite altimeter data have been calibrated using buoys data and Figure 5 shows indeed a good overall agreement between the two time series (see also e.g [15]). Figure 11 suggests however that the empirical distribution of satellite data have an heavier tail than the one of buoy data.

We first focus on reanalysis data in order to avoid estimation problems related to irregular sampling or missing values and detail the practical implementation of the proposed methodology on this particular dataset. We first need to chose a censoring threshold u. It is a crucial choice since u must be chosen high enough to justify the approximation by probabilistic models derived from extreme value theory but not too high in order to keep enough observations to fit the model. A common tool for selecting an appropriate threshold is to fit the model for various thresholds and chose the lowest one such that the estimates are almost stable for any higher threshold value. Indeed, from a theoretical point of view, if the fitted censored Gaussian extreme value process is an appropriate model to describe the behavior of the observed process above a threshold u then it should also be appropriate above higher threshold. However, in practice it is known that is generally difficult to come up to a decision using such diagnostic plots. Figure 12 shows indeed that the values of the parameters do not seem to stabilize when the threshold increases. Comparable results were obtained using similar diagnostic plots when implementing the POT method (see [4] for more details). It may be due to model misspecification or to the estimation error which increases with the threshold since the number of exceedances decreases.

[1] http://data.ecmwf.int/data/

[2] ftp://ftp.ifremer.fr/ifremer/cersat/products/swath/altimeters/waves/

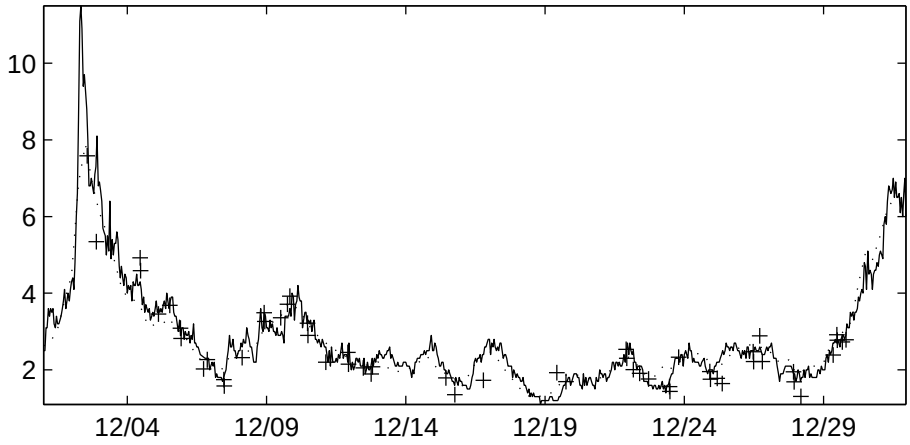

Figure 10: Comparison of the three time series for the month of December 2005. Solid line: buoy data (location (47.5 N, 8.5 W)); dotted line: reanalysis data (location 48 N, 9 W)); plus points: closest satellite observation to the buoy from each satellite track within a 1.5° box.

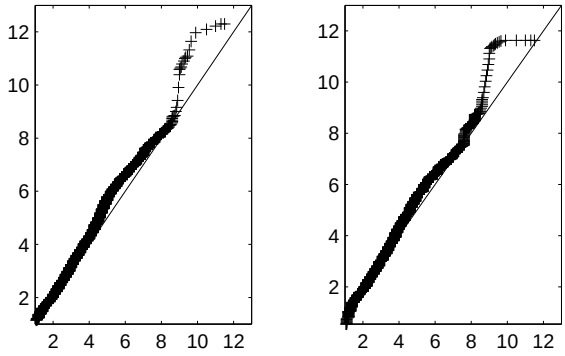

Figure 11: QQ-plot of the sample distribution of the buoy (x-axis) against the empirical distribution of reanalysis data (y-axis on the left panel) and satellite data (y-axis on the right panel)

In this context, adding confidence intervals on Figure 12 would provide helpful information but unfortunately our method does not allow the computation of such intervals. Similar figures were done for buoy and satellite data and led us to select a threshold $u = 6$ meters. The same threshold was used for all the datasets in order to facilitate the comparison.

According to Figure 13, the fitted censored Gaussian extreme value process seems to provide a realistic description of the extremal behavior of the reanalysis data. Indeed, the statistics computed from the data generally lie in the 95% prediction intervals for the fitted model. This indicates that the difference between the statistics computed from the observations and the fitted model may be due to the sampling error and gives confidence in the results obtained when

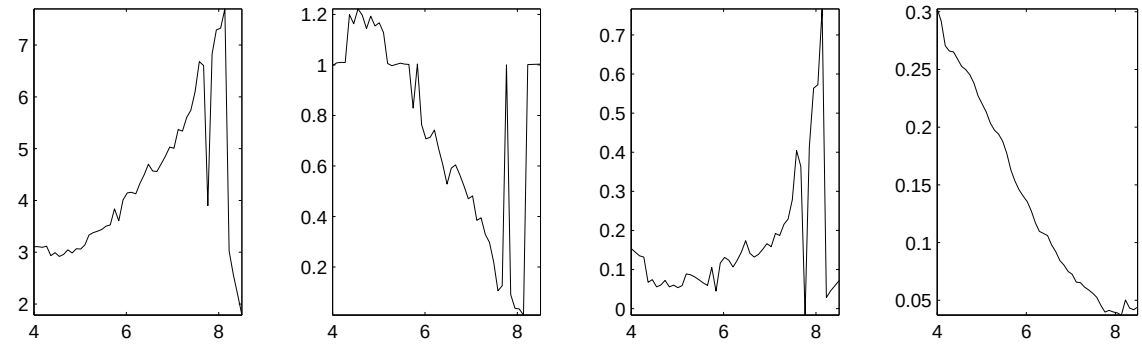

Figure 12: Values of $\mathrm{MPL}_1\mathrm{E}$ as a function of the threshold u (x-axis). From left to right panels: estimate of μ, σ, ξ and ν.

extrapolating the extremal behavior of the data using the model to compute, for example, the 100 years return levels discussed below.

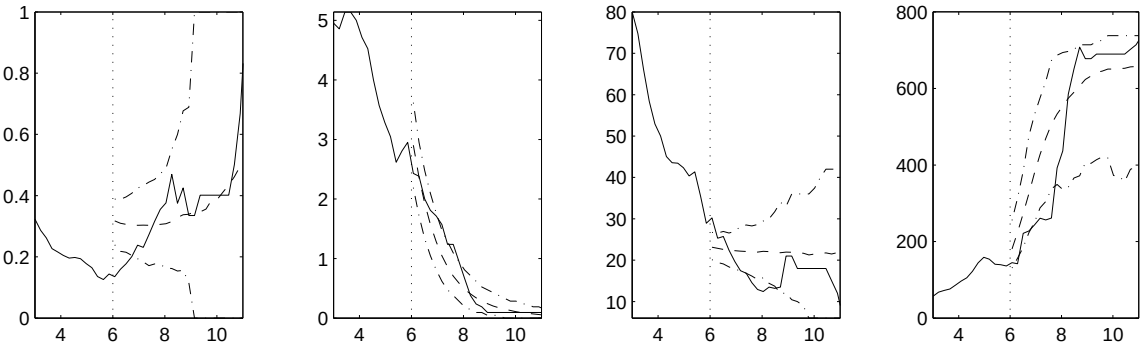

Figure 13: Comparison of extremal behavior of the reanalysis data (full line) and the fitted model (dashed line) as a function of the threshol (x-axis). The dotted lines represents 95% prediction intervals computed by simulating 1000 sequences of the same length than the original data from the fitted model. From left to right panels: estimate of the extremal index (see [12]), number of up-crossings, mean length of the clusters and mean time between consecutive up-crossings.

The censored Gaussian extreme value processes was also fitted to the buoy and satellite time series introduced above by computing the $\mathrm{MPL}_1\mathrm{E}$. For comparison purpose, we have also applied the usual POT method with a declustering scheme to the reanalysis dataset (this method is not readily available for other datasets due to the irregular time sampling). Parameter values are given in Table 2 and Figure 14 provides a graphical comparison of the properties of the censored Gaussian extreme value processes fitted on the various datasets. The parameters values of the models fitted on buoy and satellite time series are relatively close but there are important differences with the ones obtained on reanalysis data. In particular, the estimated value of the extremal index is higher for reanalysis data and this additive extremal dependence is probably related to the smoothness of this dataset which tends to smooth out the short term temporal variability. The values of the shape parameter ξ obtained with the $\mathrm{MPL}_1\mathrm{E}$ indicate

Data source	μ	σ	ξ	ν	Extremal index	$q_{100 \text{ years}}$
ERA-Interim (POT)	3.6328	1.1236	0.0394	NA	0.129	12.47
ERA-Interim (MPL_1E)	4.2002	0.6853	0.1380	0.1341	0.056	14.42
Buoy (MPL_1E)	2.9514	0.9923	0.0533	0.0070	0.213	14.60
Satellite(MPL_1E)	3.6229	1.0913	0.0783	0.0113	0.185	17.96

Table 2: Parameter values for the different datasets. The last column gives the 100 return level.

that the model fitted on reanalysis data also exhibits a marginal distribution with heavier tail, which seems consistent with Figure 11. This seems however contradictory with the results obtained with the POT method on reanalysis data which leads to a smaller value of the shape parameter. However the observed differences in the values of the shape parameter are probably not statistically significant. This is suggested by the following 95% asymptotic confidence interval for ξ, $[-0.22,\ 0.30]$, which has been derived from the information matrix when using the POT method. Figure 14 confirms that the models identified on buoy and satellite exhibits similar characteristics but noticeably differ from the model fitted on the reanalysis dataset. For example, the storm have a typical duration of about 5 hours for the first two datasets whereas the reanalysis dataset identifies storms which mean durations of about 22 hours. Table 2 also gives the 100 years return level, and the results may seem surprising according to the discussion above. Indeed reanalysis and buoy data lead to similar return values (about 14 meters) whereas satellite data leads to a higher value (17 meters). Actually, the return level is a complex function of the four parameters of the Gaussian extreme value process and the differences on the parameter values may compensate to provide similar return levels. Again, the observed differences may not be statistically significant.

Similar results were obtained at other locations where buoys data are available. Buoys and satellite generally lead to identifying model which are close to each other whereas reanalysis data identifies more extremal dependence and longer storms. If we believe that buoys data are a good reference, these results suggest that satellites may provide more accurate information on the extremal behavior of Hs than reanalysis data. In this context, the proposed methodology can be an efficient tool to derive estimates of any quantities of interest, related to the extremal properties of H_s, at any location of the ocean where satellite data are available.

6 Proof of Therorem 3.1

A Gaussian extremal process $\{Z_t\}$ is a moving maxima process as defined in [21]. Using the results given in this paper, we deduce that $\{Z_t\}$ is a stationary unit Fréchet process, continuous in probability, mixing, and hence ergodic. It allows us to use the following theorem which is a straightforward generalisation of Theorem 1.12 in [14]:

Theorem 6.1. *Let $\{Z_i\}_{i=1,\dots,n}$ be a stationary and ergodic process which distribution depends on a parameter $\nu^* \in \Theta$, where Θ is a compact subset of $\mathbb{R}$ and let Q_n be a contrast function defined as*

$$Q_n(\nu) = \frac{1}{n}\sum_{i=1}^{n-1} f(Z_i, Z_{i+1}; \nu),$$

where f is a measurable function with real values and continuous in ν. Suppose that

1. $\mathbb{E}\inf_{\nu\in\Theta} f(Z_1, Z_2; \nu) > -\infty$;
2. $\nu \mapsto \mathbb{E}f(Z_1, Z_2; \nu)$ *has a unique finite minimum at ν^*.*

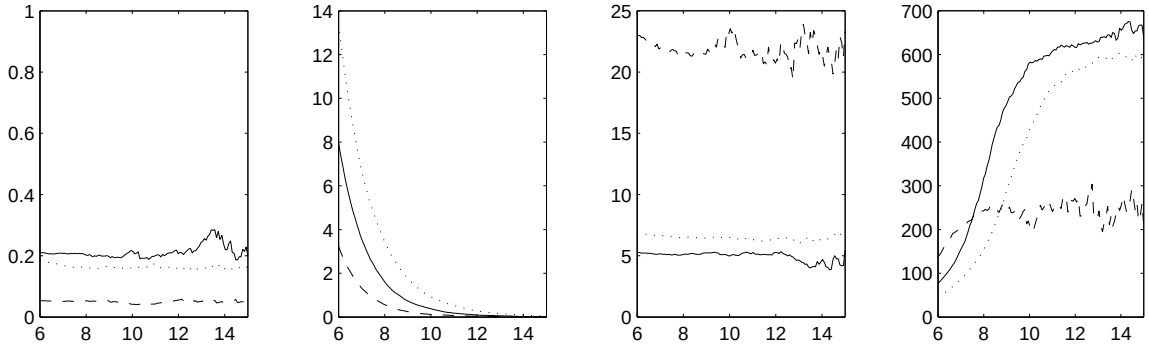

Figure 14: Comparison of the extremal behavior of the models fitted on buoy data (solid line), reanalysis data (dashed line) and satellite data (dotted line). From left to right panels: extremal index, mean number of up-crossings per year, mean length of the clusters and mean time between consecutive up-crossings as a function of the threshold (x-axis). The results were obtained by simulating 1000 years of hourly data with each model.

Then the minimum contrast estimator $\hat{\nu}_n = argmin_{\nu\in\Theta} Q_n(\nu)$ *is strongly consistent:*

$$\lim_{n\to\infty} \hat{\nu}_n = \nu^* \ a.s.$$

We will use this theorem with

$$f(Z_1, Z_2; \nu) = -\log p(Z_1, Z_2; \nu)$$

where $p(z_1, z_2; \nu)$ denotes the joint pdf of (Z_1, Z_2), $\theta = \nu$, $\Theta = (\nu_-, \nu_+)$ with $0 < \nu_- < \nu_+$. An explicit expression for f is given in Section 6.1. It will be used to prove the properties (1) and (2) of Theorem 3.1 in Sections 6.2 and 6.3 respectively.

6.1 Expression of the constrast

We have

$$p(Z_1, Z_2; \nu) = \frac{\partial^2}{\partial z_1 \partial z_2} F_Z(z_1, z_2; \nu)$$

with $F_Z(z_1, z_2; \nu) = \exp((-V(z_1, z_2; \nu))$ and V defined in (7) and thus

$$f(Z_1, Z_2; \nu) = V(z_1, z_2; \nu) - \log\left(\frac{\partial V}{\partial z_1}(z_1, z_2)\frac{\partial V}{\partial z_2}(z_1, z_2) + \frac{\partial^2 V}{\partial z_1 \partial z_2}(z_1, z_2)\right)$$

V and its derivatives satisfy

$$\begin{aligned}
V(z_1, z_2; \nu) &= \frac{\Phi(a/2 + 1/a \log \frac{z_2}{z_1})}{z_1} + \frac{\Phi(a/2 + 1/a \log \frac{z_1}{z_2})}{z_2} = \frac{\Phi(w)}{z_1} + \frac{\Phi(v)}{z_2} \\
\frac{\partial V}{\partial z_1}(z_1, z_2; \nu) &= -\frac{\Phi(w)}{z_1^2} - \frac{\varphi(w)}{a z_1^2} + \frac{\varphi(v)}{a z_1 z_2} \\
\frac{\partial V}{\partial z_2}(z_1, z_2; \nu) &= -\frac{\Phi(v)}{z_2^2} - \frac{\varphi(v)}{a z_2^2} + \frac{\varphi(w)}{a z_1 z_2} \\
\frac{\partial^2 V}{\partial z_1 \partial z_2}(z_1, z_2; \nu) &= -\frac{v\varphi(w)}{a^2 z_1^2 z_2} - \frac{w\varphi(v)}{a^2 z_1 z_2^2}
\end{aligned}$$

with

$$\begin{aligned} w &= \frac{a}{2} + \frac{1}{a}\log\frac{z_2}{z_1} \\ v &= a - w = \frac{a}{2} + \frac{1}{a}\log\frac{z_1}{z_2} \\ \varphi(w) &= \frac{1}{\sqrt{2\pi}} e^{-a^2/8} e^{-\frac{\log^2(z_1/z_2)}{2a^2}} \sqrt{\frac{z_1}{z_2}} \\ \varphi(v) &= \frac{1}{\sqrt{2\pi}} e^{-a^2/8} e^{-\frac{\log^2(z_1/z_2)}{2a^2}} \sqrt{\frac{z_2}{z_1}} \end{aligned}$$

Taking into account that $\frac{\varphi(w)}{az_1^2} - \frac{\varphi(v)}{az_1z_2} = 0$ leads to the following simplified expressions for the derivatives of V

$$\frac{\partial V}{\partial z_1}(z_1, z_2) = -\frac{\Phi(w)}{z_1^2},$$
$$\frac{\partial V}{\partial z_2}(z_1, z_2) = -\frac{\Phi(v)}{z_2^2},$$

$$\frac{\partial^2 V}{\partial z_1 \partial z_2}(z_1, z_2) = -\frac{\varphi(w)}{az_1^2 z_2} = -\frac{\varphi(v)}{az_2^2 z_1} = \frac{e^{-a^2/8} \exp\left(\frac{\log^2 \frac{z_2}{z_1}}{2a^2}\right)}{(z_1 z_2)^{3/2}}.$$

Finally, we deduce that

$$f(z_1, z_2; \nu) = V(z_1, z_2; \nu) - \log\left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2} + \frac{\varphi(w)}{az_1^2 z_2}\right] \tag{15}$$

6.2 Minoration (1)

We have

$$\inf_\nu\{f(z_1, z_2; \nu)\} \geq \inf_\nu\{V(z_1, z_2; \nu)\} + \inf_\nu\left\{-\log\left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2} + \frac{\varphi(w)}{az_1^2 z_2}\right]\right\}$$

and each term of the right hand term of this expression are treated separately below.

- **Term** $V(z_1, z_2; \nu)$.
  V can be bounded using the Fréchet Bound ([11])

  $$P(Z_1 \leq z_1) + P(Z_2 \leq z_2) - 1 \leq \mathbb{P}(Z_1 \leq z_1, Z_2 \leq z_2) \leq \min\{P(Z_1 \leq z_1), P(Z_2 \leq z_2))$$

  which implies that

  $$\min\left(-\frac{1}{z_1}, -\frac{1}{z_2}\right) \leq V(z_1, z_2; \nu) \leq \exp\left(-\frac{1}{z_1}\right) + \exp\left(\frac{1}{z_2}\right)$$

  and thus

  $$\inf_\nu\{V(z_1, z_2; \nu)\} \geq \min\left(-\frac{1}{z_1}, -\frac{1}{z_2}\right)$$

  The right hand term of the last expression has finite expectation since $\frac{1}{Z_1}$ and $\frac{1}{Z_2}$ have unit exponential distributions.

- **Term** $-\log\left[\frac{\Phi(w)\Phi(v)}{a^2 z_1 z_2}+\frac{\varphi(w)}{a z_1^2 z_2}\right]$.

  We have

$$\begin{aligned}-\log\left[\frac{\Phi(w)\Phi(v)}{z_1 z_2}\nu^2+\frac{\varphi(w)}{z_1^2 z_2}\nu\right] &\geq 1-\frac{\Phi(w)\Phi(v)}{z_1 z_2}\nu^2+\frac{\varphi(w)}{z_1^2 z_2}\nu\\ &\geq 1-\frac{\nu^2}{z_1 z_2}\\ &\geq 1-\frac{\nu_+^2}{z_1 z_2}\end{aligned}$$

  and the Cauchy's inequality implies that the right hand term of the last expression has finite expectation since

$$\mathbb{E}\left[\frac{1}{Z_1 Z_2}\right]\leq\sqrt{\mathbb{E}[1/Z_1^2]\mathbb{E}[1/Z_2^2]}=1$$

6.3 Identifiability

We have to prove that $\nu\mapsto\mathbb{E}f(Z_1,Z_2;\nu)$ has a unique finite minimum at ν^* on $\Theta=[\nu_-\nu_+]$. To this end, denote by P_ν the distribution with density function $p(z_1,z_2;\nu)$. Then,

$$\mathbb{E}_{\nu^*}\left[-\log\frac{p(Z_1,Z_2;\nu)}{p(Z_1,Z_2;\nu^*)}\right]=K(P_{\nu^*},P_\nu),$$

is the Kullback-Leibler divergence between P_{ν^*} and P_ν. We have then $K\geq 0$ and $K=0$ iff $P_\nu=P_{\nu^*}$. As the density functions are positive and continuous in (z_1,z_2), this is equivalent to $p(z_1,z_2;\nu^*)=p(z_1,z_2;\nu)$ for all (z_1,z_2). In particular, $K=0$ implies that for all $z_1=z_2=z>0$, we have:

$$\exp\left[-\frac{2}{z}\Phi\left(\frac{1}{2\nu}\right)\right]\left[\frac{\Phi\left(\frac{1}{2\nu}\right)^2}{z^2}\nu^2+\frac{\varphi\left(\frac{1}{2\nu}\right)}{z^3}\nu\right]=\exp\left[-\frac{2}{z}\Phi\left(\frac{1}{2\nu^*}\right)\right]\left[\frac{\Phi\left(\frac{1}{2\nu^*}\right)^2}{z^2}\nu^{*2}+\frac{\varphi\left(\frac{1}{2\nu^*}\right)}{z^3}\nu^*\right].$$

Letting $z\to 0$ while $z>0$ we see that the exponents in the exponential function must be equal, i.e.

$$\Phi\left(\frac{1}{\nu}\right)=\Phi\left(\frac{1}{\nu^*}\right).$$

Hence $\nu=\nu^*$. The proof is complete.

7 Conclusion

In this paper we propose an original method to analyze the extremal behavior of univariate time series. It was motivated by the need to analyze environmental time series with missing values or irregular sampling but the tests performed on classical time series model indicate that the method also performs well on time series with regular sampling compared to the other methods which have been proposed in the literature. The parameters are estimated by using a composite likelihood method and both theoretical and simulation results indicate that it leads to consistent estimates. Results obtained on Hs data indicate that the proposed methodology could be used

to estimate the extremal behavior of Hs from satellite data and produce climatology of extreme Hs all over the ocean which are more accurate than the ones obtained from reanalysis data.

We believe that our methodology is flexible enough to build extensions which may be useful for practical applications. For example, the methodology could deal with other max-stable processes, such as the Brown-Resnick process which could give more flexibility to the model, include non-stationary components or be extended to a space-time model. This will be the subject of future research.

Various aspects of the proposed methodology need also to be improved. In particular, the theoretical results are incomplete. Consistency is only proven in an idealized situation where some of the parameters are known and asymptotic normality has not been established yet. Solving this last problem would have important practical implications since it would give tools to estimate uncertainties.

Acknowledgment

The authors are indebted to the Laboratoire d'Océanographie Spatiale, IFREMER and to the ECMWF for having provided the data used in this study.

References

[1] P. Ailliot, A. Baxevani, A. Cuzol, V. Monbet, and N. Raillard. Space-time models for moving fields with an application to significant wave height fields. *Environmetrics*, 22(3):354–369, 2011.

[2] P. Ailliot, C. Thompson, and P. Thomson. Mixed methods for fitting the gev distribution. *Water Resources Research*, 47, 2011.

[3] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

[4] S. G. Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.

[5] D. R. Cox. A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 92:729–737, 2004.

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

[7] L. de Haan. A spectral representation for max-stable processes. *Annals of Statistics*, 12(4):1194–1204, 1984.

[8] L. de Haan and A. Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[9] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

[10] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(02):180–190, 1928.

[11] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon*, 14:53–77, 1951.

[12] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York, 1983.

[13] B. G. Lindsay. Composite likelihood methods. In *Statistical inference from stochastic processes*. American Mathematical Society, 1988.

[14] J. Pfanzagl. A characterization of sufficiency by power functions. *Metrika*, 21:197–199, 1974.

[15] P. Queffeulou. Long-term validation of wave height measurements from altimeters. *Marine Geodesy*, 27:495–510, 2004.

[16] M. Ribatet, T. B. M. J. Ouarda, E. Sauquet, and J. M. Gresillon. Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics. *Water Resources Research*, 45(3), March 2009.

[17] P. Ribereau, P. Naveau, and A. Guillou. A note of caution when interpreting parameters of the distribution of excesses. *Advances in Water Resources*, 34:1215–1221, 2011.

[18] M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.

[19] R. L. Smith. Max-stable processes and spatial extremes. Unpublished, 1990.

[20] R. L. Smith, J. A. Tawn, and S. G. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268, 1997.

[21] S. A. Stoev. On the ergodicity and mixing of max-stable processes. *Stochastic Processes and their Applications*, 118(9):1679–1705, 2008.

[22] C. Varin. On composite marginal likelihoods. *Advances in Statistical Analysis*, 92(1):1–28, 2008.

[23] C. Varin and P. Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.

[24] J. Vinoth and I. R. Young. Global estimates of extreme wind speed and wave height. *Journal of Climate*, 24(6):1647–1665, 2011.

[25] W. Wimmer, P. Challenor, and C. Retzler. Extreme wave heights in the north atlantic from altimeter data. *Renewable Energy*, 31(2):241–248, 2006.