Spatial assessment of extreme significant waves heights in the Gulf of Lions

Romain Chailan, Gwladys Toulemonde, Frederic Bouchette, Anne Laurent, Florence Sevault, Heloise Michaud

SPATIAL ASSESSMENT OF EXTREME SIGNIFICANT WAVES HEIGHTS IN THE GULF OF LIONS

Romain Chailan[1], Gwladys Toulemonde[2], Frederic Bouchette[3], Anne Laurent[4], Florence Sevault[5] and Heloise Michaud[6]

In the analysis of coastal hazards, the features of extreme waves are determining information to question the impact of storms to the coast. The spatial behaviour of extreme waves is even more valuable especially since it is sparsely provided. Regarding recent applications in other contexts, a kind of statistical models called max-stable processes is relevant for modelling spatial extreme events. Max-stable processes are extensions of the well-known Generalised Extreme Value (GEV) distribution. Unlike univariate approaches, max-stable processes consider spatial dependence of a phenomenon. Such a modelling also overtakes a standard multivariate approach by providing information continuously over the area studied, even where no observation is available. Relying on such a stochastic modelling, the aim of this study is to discuss the extreme waves hazards in the Gulf of Lions, focusing on their spatial behaviour.

Keywords: extreme waves hazards; extreme value analysis; max-stable processes; spatial extreme modelling; spatial dependence; wave hindcast; Gulf of Lions

INTRODUCTION

Coastal hazards are physical phenomena of many concerns. Infrastructures, environment, activities, population... Either public or private actors need tools to quantify and to describe a key vector of those risks: extreme wave events. It is thus worth finding out reliable methodologies to deal with such extreme conditions. The main difficulty being to extrapolate information from historical observations to anticipate oncoming risks for longest time series. In the specific context of coastal hazards, one may deliver methodologies in view of the spatial behaviour of extreme waves. Spatial extreme behaviour turns out to be a mandatory information since such environmental risks impact not a point but entire areas.

Since several decades, statistical analyses are performed to assess extreme events. Most studies are still based on the well-known and widely accepted Generalized Extreme Value (GEV) distribution. GEV distribution allows to extrapolate return levels – i.e. level of a variable which will be reached only once in average during the corresponding return period – from observed dataset. The first formalism issued was univariate. Considering so, studies are quickly limited while dealing on questioning that requires spatial dependence of a phenomenon. For instance, let us consider a study of the long-term long-shore sediments transport. And let define the joint probability of having the significant wave heights of four littoral sites exceeding their 20 years return level in a month. This quantity must be computed to assess the long-shore transport. Assuming the independence of the extreme events across sites may lead to a bias in the estimation of this quantity. One alternative is to pay attention to the first extension of the univariate GEV distributions: the Multivariate Extreme Value (MEV) distributions. Thanks to MEV distributions, dependence structures between observations sites are assessed and results become more suitable to answer spatial questioning. However, the attractiveness of MEV distributions might be limited since it provides information only on sites where there were observations. In this study we propose to use the so-called max-stable processes, extended also from the former distributions. By construction, max-stable processes still hold the properties provided with both GEV and MEV families of distributions. But max-stable processes overtake them by bringing information continuously over the area studied, even where no observation is available. The performance of such modelling has been shown in applications in other environmental contexts, like for instance the study of heavy snow events in Blanchet and Davison (2011) or heatwaves in Davison and Gholamrezaee (2011). Some investigations on significant waves has been produced by Raillard et al. (2014). Jonathan et al. (2013) present such applications as a promising way to model extreme waves events.

The aim of this study is not to debate the use of other extreme value analyses but rather bring to the front the benefit of considering the dependence structure of extreme. In this sense we present the use of max-stable models to 1) map extreme return levels of wave features and 2) provide accurate joint probabilities. Others

[1]LIRMM-I3M-Geosciences, University of Montpellier 2 / CNRS, France
[2]I3M, University of Montpellier 2 / CNRS, France
[3]Geosciences, University of Montpellier 2 / CNRS, France
[4]LIRMM, University of Montpellier 2 / CNRS, France
[5]CNRM-GAME, Meteo France / CNRS, France
[6]SHOM, France

potential re-uses of max-stable modelling are discussed at the end of the present document.
We apply this methodology to the Gulf of Lions in the western-north Mediterranean Sea. About 400 km of coastlines pertains to this French area. Both local economy – e.g. tourism, fishing, trade – and biodiversity mainly relies on this coastal resources. The Gulf of Lions is highly subjected to storm-waves, which make the area perfectly suitable to demonstrate the interest of the modelling.

The first section presents the realisation of the 52-year hindcast of wave features over our studied area. Then a second section shortly introduces the max-stable processes formalism. In the third section the methodology is developed on the actual case study. We gather some direct applications of the stochastic modelling for risk analysis in the fourth section. Finally the results and limits are discussed in the last section.

A 52-YEAR WAVE HINDCAST

A well-fitted statistical model and therefore the quality of its extrapolation rely on the availability of a representative data set. This becomes a real constraint as far as we consider extreme events. Such events are rare by definition and direct observations are often scarce, both in time and space. It is particularly the case for our area of interest the Gulf of Lions, see Figure 1. Although well known for their abrupt changes, the storm-waves in that area are monitored by only four surface buoys. Obviously the chance to miss extreme waves is high and information to analyse the full area is scarce. Moreover the longest time series available is less than a decade, which is unsatisfying to extrapolate to several decades.
One can use altimeters data sets as an alternative, but their major drawback is the non-regularity of satellites tracks through time and space around the globe. Raillard et al. (2014) deal with such source of data thanks to max-stable processes, but without looking at the spatial dependence structure in the same time. The last way to observe wave data is the use of numerical simulations. Thus the first step of this study was to create an hindcast of wave features for the historical period 1961-2012.

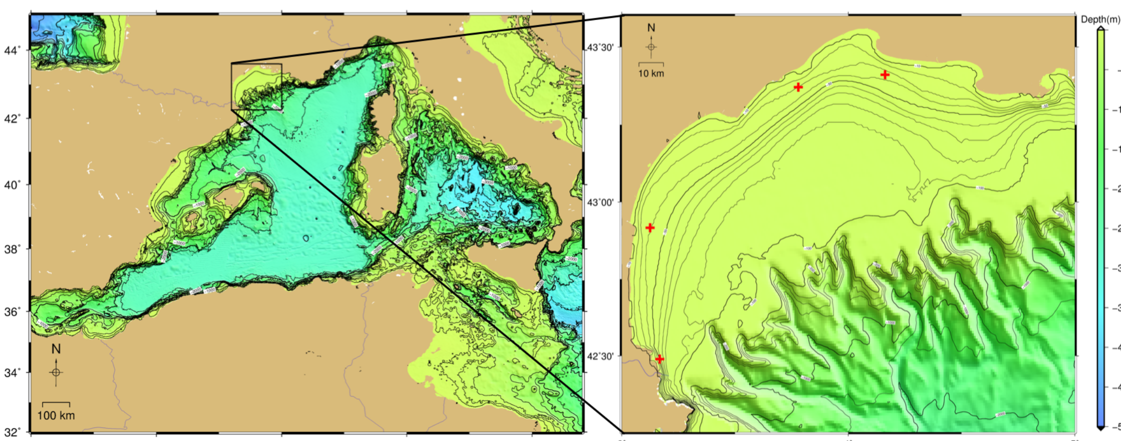


Figure 1: Bathymetry of the north-western Mediterranean sea. Left panel is the full extension of the domain considered for the hindcast. Right panel is the studied area for extremes: the Gulf of Lions. Crosses are locations of surface buoys measuring waves features.

Wave model description

The wave simulation is obtained using the WAVEWATCHIII® (WW3) wave model (Tolman (2014)) in its most recent release (v4.18). WW3 is a third generation wave model resolving the random phase spectral action density balance equation for wavenumber-direction spectra.
The overall domain covers the north-western Mediterranean sea, extending from the Gibraltar's strait to the south of Italy. An unstructured mesh is used to run the simulation. The resolution of the mesh ranges from 1000m to 12km, with specific refinement relative to the bathymetry and a self-designed polygon over the Gulf of Lions. This specification leads to a computational grid of 47086 nodes. An illustration of the grid

is given in Figure 2.
The bathymetry used in this configuration comes from the LEGOS (French Laboratory of Geophysics and

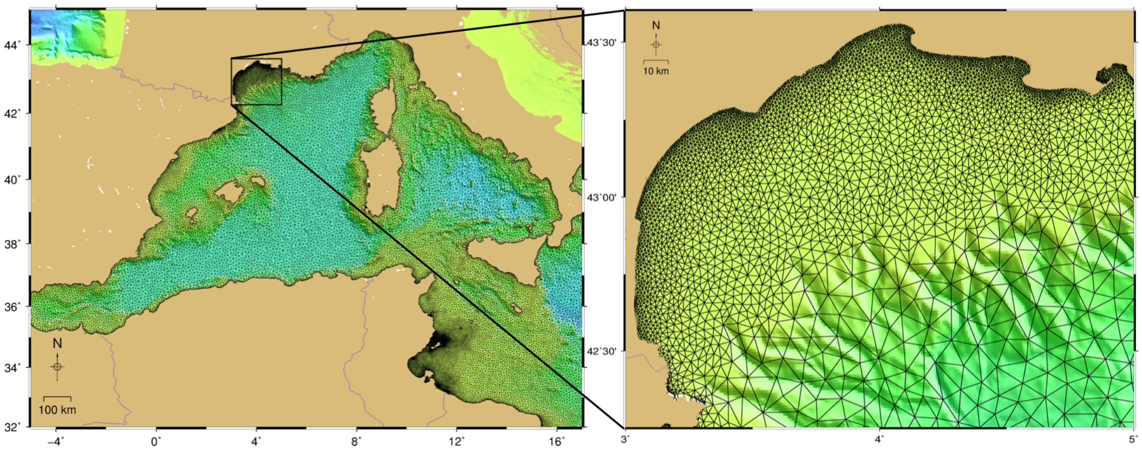


Figure 2: Computational mesh used for the hindcast. It is composed of 47086 nodes with a spatial resolution ranging from 1000m to 12km. Right panel is a zoom of the grid on the area of interest: the Gulf of Lions.

Spatial Oceanography studies). Its overall spatial resolution is of 0.0083 degree.
Over this domain, two sources are used to run the simulation : wind field and surface currents. Both of them come from regional climate re-analyses. The wind field, namely ARPERA re-analyses (Herrmann and Somot (2008)), covers a period of 52 years from 1961 to 2012 at a spatial resolution of 50km and time resolution of 6 hours. The surface currents are issued from a reanalysis called NEMOMED8 (Beuvier et al. (2010)) and has a spatial resolution ranging between 9km to 12km. This reanalysis has a time resolution of 24 hours.
An important remark is that we intentionally do not force waves conditions at open boundaries of the domain. This choice involves to use a great domain extension – the north-western Mediterranean sea – comparing to the area studied – the Gulf of Lions. Indeed, it is reasonable to assume any physical process influencing the waves motion in the Gulf of Lions stems from the former domain extension.

Output dataset

We generate hourly waves features at each node – 47086 in total – of the computational grid for the 52 year period, 1961-2012. Only considering the Gulf of Lions area, it represents 3954 wet sites.
A validation step is performed to measure the quality of the hindcast. The simulated signals (e.g., significant wave height, wave direction, wave period, etc.) are compared to records at the four littoral surface buoys presented above. The results are satisfying, even if some of the extreme events are slightly smoothed out against the observed signals. A short investigation shows that this bias for extreme events often results from an actual bias of the wind input fields. In the sequel we consider the produced data as real observations.
We focus our study on the extreme waves in the Gulf of Lions by stochastically modelling the extreme significant wave heights (Hs). The signal is carefully analysed regarding the spatial dependence behaviour. The inference of such modelling is hard to afford as soon as the number of sites gets important. Therefore we have to decrease the number of sites considered. A Latin Hypercube Sampling (LHS) method is used to randomly select 100 nodes of the computational grid, with respect to a good spatial representation. Only wet sites of the grid are considered. If two sites are close to a distance inferior to 1 km, only one of those sites is conserved in the dataset. This scheme leads us to analyse 97 sites presented in Figure 3.
We restrict our methodology to those 97 sites. However, one may investigate the sensitivity of the approach by increasing the number of sites considered or move their location or both. Such sensitivity analysis would be a valuable add, but remains out of scope for this document.

SPATIAL EXTREME MODELLING

Looking at recent applications in extreme geostatistics (Blanchet and Davison (2011) , Davison and Gholamrezaee (2011)), a new kind of stochastic model suits the need of extreme spatial modelling. They

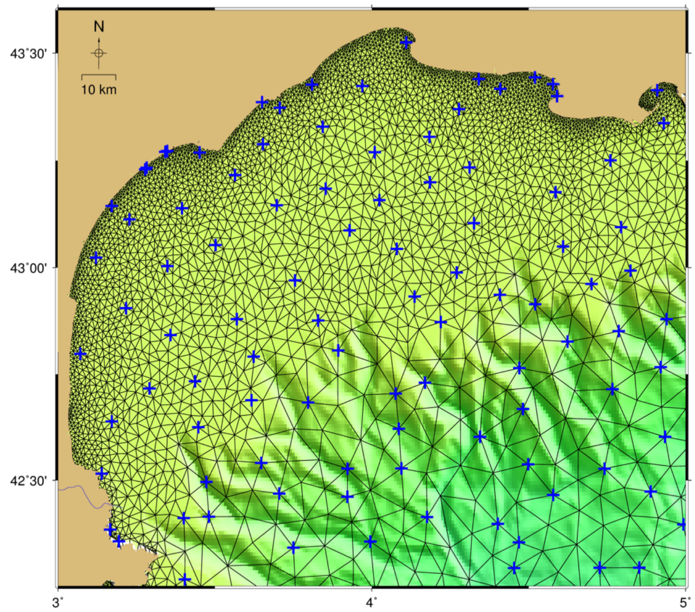


Figure 3: Crosses are sites selected by the LHS algorithm. This set represents stations used for extreme modelling in the following sections.

are based on the so-called max-stable processes. These models have leverage univariate and multivariate extreme value theories to continuously represent the spatial dependence structure of environmental processes. This section introduces the reader to the theory of max-stable processes. In the sequel the notion of random process $\{Z(s), s \in S\}$ will be largely used. We recall that in statistics a random process Z in $S \subset \mathbb{R}^d$ represents a random variables collection indexed by S and taking its values in $E \subset \mathbb{R}^d$. In this document, a process is defined over $S \subset \mathbb{R}^2$ and has real values, which means $E = \mathbb{R}$.

Marginal analysis

Let $Z(s),\ s \in S$, a random variable denoting the maximum of independent and identically distributed (i.i.d) random variables at a site s. From the Extreme Value Theory introduced by Fisher and Tippett (1928) and Gnedenko (1943), we expect that at any site $s \in \chi \subset S$, $Z(s)$ will follow approximately the well-known Generalised Extreme Value (GEV) distribution :

$$P(Z(s) \leq z) \approx GEV_s(z) = \exp\left\{-\left[1 + \xi(s)\left(\frac{z - \mu(s)}{\sigma(s)}\right)\right]_+^{-\frac{1}{\xi(s)}}\right\}, \tag{1}$$

where $\mu(s)$ is a location parameter, $\sigma(s)$ is a scale parameter and $\xi(s)$ a shape parameter. In this representation a_+ corresponds to $\max(a, 0)$. Also, for any site $s \in \chi \subset S$, $\mu(s) \in \mathbb{R}$, $\sigma(s) > 0$ and $\xi(s) \in \mathbb{R}$. The GEV formalism gathers three distributions families : Weibull if $\xi(s) < 0$, Gumbel if $\xi(s) = 0$ and Fréchet if $\xi(s) > 0$.
An essential propriety of any GEV distribution is that the distribution of its maximum is also GEV distributed. This propriety is named max-stability.

While studying environmental phenomenon, marginal fits use to show that $\mu(s)$, $\sigma(s)$ and $\xi(s)$ are spatially varying with covariates. For instance for significant waves heights modelling it could be covariates such as bathymetry, latitude and longitude. Therefore the GEV parameters may evolve as linear functions depending on those covariates. For instance $\mu(s)$ may vary as

$$\mu(s) = \beta_0 + \beta_1 \mathrm{bathy}(s) + \beta_2 \mathrm{lon}(s) + \beta_3 \mathrm{lat}(s). \tag{2}$$

An other alternative is to fit radial-basis splines to model the evolution of the marginal GEV parameters against a covariate. The model for radial-basis splines of order p, p being odd and defined as $p = 2m - 1$ is defined as

$$f(x) = \beta_0 + \beta_1 s + \ldots + \beta_{m-1} s^{m-1} + \sum_{j=1}^{q} \beta_{m+j} \, |s_i - \nu_j|^{2m-1}, \tag{3}$$

with kernels ν_j of the associated radial basis function β_{m+j}.
In the context of numerical modelling and when grids are refined enough, marginal fits at grid points may lead to a good description of the GEV parameters spatial evolution too. Then, a last alternative to cover all the space S is to extrapolate the parameters estimated from those marginal fits to any site $s \in S$.

Marginal transformation

While studying the dependence structure of an extreme spatial process $\{Z(s), s \in S\}$, there is no loss of generality in assuming that its marginal laws can take a particular extreme value distribution. That permits to simplify expressions and definitions of extreme mathematical objects or models. In the sequel, we assume that for all $s \in S$, the random variable $Z(s)$ is Fréchet distributed – i.e. $P(Z(s) \leq z) = \exp(-1/z)$.

Max-stable processes

Univariate extreme value theory requires to work site by site, assuming that all sites are independent from each other. This strong assumption induces a bias in the estimation of several useful quantities, like joint probabilities for instance. Multivariate extreme value theory extends the univariate one, and takes into account the dependence structure between sites. However, these analyses are limited to provide information only at observed sites.
To overtake these drawbacks a continuous spatial modelling was introduced by de Haan et al. (1984) with a new theory extending the univariate and multivariate formalisms: max-stable processes. Still considering $Z_i(\cdot), i = \{1, \ldots, n\}$ to be n-copies of a spatial process of extremes, de Haan et al. (1984) state that the random process $\{Z(x), x \in S\}$ is max-stable if $a_n(x) > 0$ and $b_n(x) \in \mathbb{R}$ defined on S exist such that,

$$\left\{ \max_{i=1,\ldots,n} \frac{Z_i(x) - b_n(x)}{a_n(x)}, x \in S \right\} \stackrel{\mathcal{L}}{=} \{Z(x), x \in S\}. \tag{4}$$

As consequence of this definition any one-dimensional marginal distribution of the process $Z(\cdot)$ satisfies the max-stability property. More specifically they can always be unit Fréchet margins – i.e. GEV – after rescaling and shift. All n-dimensional marginal distributions are MEV.

de Haan et al. (1984) show that the limit process defined as follows

$$\max_{i=1,\ldots,n} \frac{Y_i(x) - b_n(x)}{a_n(x)} \underset{n\to\infty}{\longrightarrow} Z(x), \quad x \in S, \tag{5}$$

belongs to the class of max-stable processes. There are many representations to build such processes. Bacro and Gaetan (2012) recall that two main approaches exist. Smith (1990), Schlather (2002), de Haan and Pereira (2006), use events with a deterministic form but moving randomly in the space. Schlather (2002), Kabluchko et al. (2009), rely on events with a stochastic form but keep the same spatial dependence structure. An illustration is given with the description of one model for each approach.

Smith (1990). The *Gaussian extreme value process* is often used for extreme rainfall modelling. Let us consider $\{(\xi_i, k_i), i \geq 1\}$ a Poisson process on $(0, \infty) \times \mathcal{K}$ with intensity measure of $\xi^{-2} d\xi \times \nu(dk)$, where $\mathcal{K}$ is an arbitrary space and ν a measure defined on $\mathcal{K}$. Moreover let us consider $\{f(k, x), k \in \mathcal{K}, x \in S\}$ a non-negative function satisfying $\int_{\mathcal{K}} f(k, x)\nu(dk) = 1$, for all $x \in S$. Then the so-called storm model is defined by

$$Z(x) = \max_{i=1,\ldots,n} \xi_i f(k_i, x), \quad x \in S. \tag{6}$$

As a basic interpretation, k_i are seen as the centres of the storms and ν their distribution. Each ξ_i represents the intensity of the i^{th}-storm and $\xi_i f(k_i, x)$ the total amount of rainfall for the storm centred on k_i. Finally the *max* operator allows to determine the maximum rain felt after n independent storms.

Schlather (2002) The former model is slightly different from the class introduced in Schlather (2002) which defines stochastic max-stable processes with a same dependence structure of maxima over space. Let $\{(\xi_i), i \geq 1\}$ denote the points of a Poisson process on $(0, \infty)$ with intensity measure of $\xi^{-2} d\xi$ and

$\{W_i(\cdot)\}_{i\geq 1}$ be independent copies of $W(\cdot)$ a stationary process on $\mathbb{R}^d$ (here $d = 2$), with $E(W(x)) = 1$ for all $x \in S$. Then Schlather (2002) states that the random process

$$Z(x) = \max_{i=1,\dots,n} \xi_i W_i(x), \quad x \in S \tag{7}$$

is a stationnary max-stable process with unit Fréchet margins.

Then, Schlather (2002) defines the Extremal Gaussian process as $Z(\cdot)$ from Equation 7 with $W_i(x) = \sqrt{2\pi} \max\{0, \varepsilon_i(x)\}$ and $\varepsilon_i(\cdot)$ a zero-mean stationary Gaussian process with correlation function $\rho(\cdot)$. This model has a simple interpretation: the $\xi_i W_i$ are spatial events having the same dependence structure. They differ in their magnitude ξ_i. The shape of the events may vary if the process W allows it.

Extremal coefficient

The extremal coefficient function $\theta(\cdot)$ introduced by Schlather and Tawn (2003) is a statistical tool to diagnose the dependence in the context of max-stable processes. Roughly speaking, $\theta(h)$ is defined as the measure of the dependence of a pair of sites separated by a distance h. For a max-stable process $Z(\cdot)$ defined as above – i.e. marginal transformation to standard Fréchet scale –, it can be obtained from the bivariate distribution:

$$P(Z(x) \leq z, Z(x+h) \leq z) = \exp\left(-\frac{\theta(h)}{z}\right). \tag{8}$$

From Equation 8 and for any pair $(x, x+h)$ considered, $\theta(h)$ lies between 1 to 2, respectively for perfect dependence to asymptotically independence. The extremal coefficient function depends on the analytical bivariate distribution of the process. Hence, its definition varies whether we consider process from the Smith's representation or Schlather's representation:

- Gaussian extreme value process: $\theta(h) = 2\Phi\left(\frac{1}{2}\sqrt{{}^t h \Sigma^{-1} h}\right)$, where $\Phi(\cdot)$ is the standard normal distribution function.
- Extremal Gaussian process: $\theta(h) = 1 + \sqrt{(1-\rho(h))/2}$, where $\rho(\cdot)$ is the correlation function of the stationary Gaussian process.

Preserving its necessary conditions (see Schlather (2002)), the choice of correlation function ρ allows to be more flexible while fitting the spatial dependence structure. Among them, we later use the *Bessel*, *Powered-exponential*, *Whittle-mattern* and *Cauchy* correlation functions. A representation possible with arbitrary correlation functions parameters is given in Figure 4.

One estimator of the extremal coefficient is known as the madogram Schlather and Tawn (2003). In the sequel we consider one of its extensions, the fmadogram, defined by Cooley et al. (2006) as:

$$\nu_{F(Z)}(h) = \frac{1}{2}\mathbb{E}|F(Z(s+h)) - F(Z(s))| = \frac{1}{2}\frac{\theta(h)-1}{\theta(h)+1}, \ h \in S, \tag{9}$$

which leads to the estimator:

$$\hat{\theta}(h) = \frac{1 + 2\hat{\nu}_{F(Z)}(h)}{1 - 2\hat{\nu}_{F(Z)}(h)}, \tag{10}$$

where $\hat{\nu}$ is an empirical estimator.

A good practice in spatial extreme modelling applications is to plot $\hat{\theta}(h)$ for any pair of locations to feel what kind of dependence structure rules the observed process. To better represent the spatial evolution of the underlying dependence, one can plot estimation of the extremal coefficients between a reference site and all possible pairs, i.e. all $\hat{\theta}(h)_{x_{ref}, x_j}$, with $x_{ref} \neq x_j$, x_{ref} *and* $x_j \in S$. A basic diagnostic is then to interpolate the result on the overall area of interest, as presented in Figure 6.

Model inference and comparison

In statistics, the likelihood function is commonly used to perform the inference of parametric models. By definition it requires the joint density of any associated finite-dimensional mutlivariate distributions. Unfortunately, the full likelihood is generally unreachable in this context. Hence, Padoan et al. (2010) proposed to use a pairwise likelihood function instead. It has the advantage of resting only on any bivariate distributions of the process modelled. By now, composite likelihood has been largely adopted and validated for such inference and the following describes how it works.

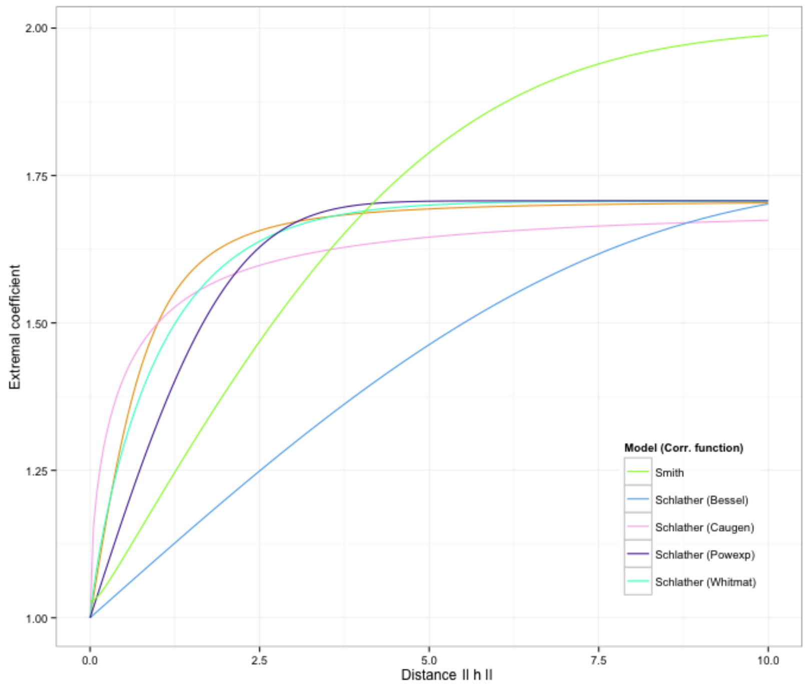


Figure 4: Theoretical extremal coefficient functions for both Smith and Schlather max-stable models. Several correlation functions with arbitrary parameters are used for Schlather's model. The evolution of the dependence along the distance $||h||$ for the Schlather's models often appears sharper at close distance and then bounded. Instead, Smith's model smoothly tends to the asymptotic independence as soon as the distance $||h||$ is large enough.

Let z_{ik} denotes a realisation of the process of maxima $Z(\cdot)$ for the i^{th}-period, at location k. Let us assume $Z(\cdot)$ has standard unit Fréchet margins. The inference step consists in finding the set $\Psi = (\psi_1, \ldots, \psi_p)$ of parameters maximizing the pairwise likelihood, which in log-form is

$$\mathcal{L}(\Psi) = \sum_{i=1}^{n} \sum_{k<l} log f(z_{ik}, z_{il}; \Psi), \tag{11}$$

where $f(\cdot, \cdot)$ is the bivariate density of the unit Fréchet max-stable process. Under suitable conditions, the maximum composite likelihood estimator $\hat{\Psi}$ has a limiting normal distribution as $n \to +\infty$, with mean Ψ and a covariance matrix estimable by $H(\hat{\Psi})^{-1}J(\hat{\Psi})H(\hat{\Psi})^{-1}$. The observed information matrix H and the squared score statistic matrix J are respectively defined by

$$H(\hat{\Psi}) = -\sum_{i=1}^{n} \sum_{k<l} \frac{\partial^2 \log f(z_{ik}, z_{il}; \hat{\Psi})}{\partial \hat{\Psi} \partial \hat{\Psi}^T} \quad \text{and} \tag{12}$$

$$J(\hat{\Psi}) = \sum_{i=1}^{n} \sum_{k<l} \frac{\partial \log f(z_{ik}, z_{il}; \hat{\Psi})}{\partial \hat{\Psi}} \frac{\partial \log f(z_{ik}, z_{il}; \hat{\Psi})}{\partial \hat{\Psi}^T}. \tag{13}$$

Once all models are fitted, the best one must be designated. Akaike Information Criterion (AIC) is a standard measure of the quality of a given model for a dataset. AIC balances the goodness of the fit of a model and its complexity. By definition, it relies on the full likelihood determination. Bearing in mind the inference of max-stable models deals with a pairwise composite likelihood, an other criterion has to be considered. In the sequel, we use the Composite Likelihood Information Criterion introduced by Varin and Vidoni (2005). It is an extension of the former AIC to the composite likelihood context, defined as

$$CLIC = -2\mathcal{L}(\hat{\Psi}) + 2\text{tr}\left\{H(\hat{\Psi})^{-1}J(\hat{\Psi})\right\}, \tag{14}$$

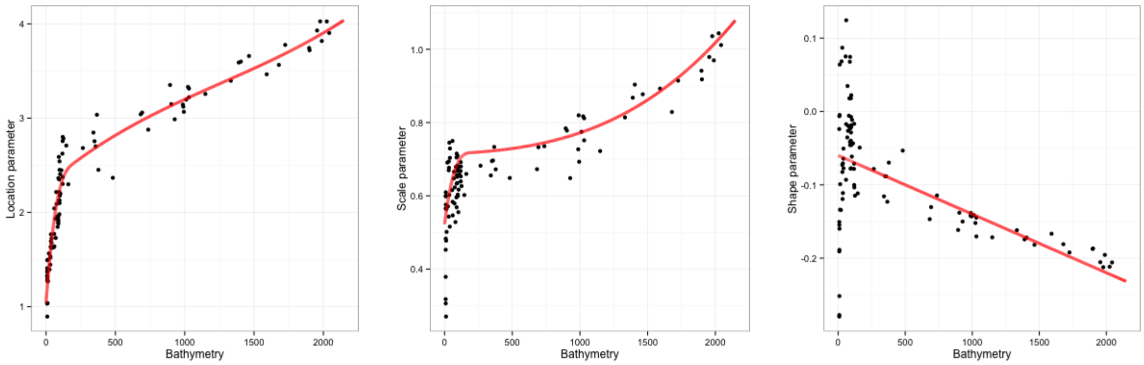
where H and J are still the observed information and the squared score statistic matrices.
By definition, the lower the CLIC the better model quality. We therefore choose the model corresponding to the lower CLIC for applications.

EXTREME SIGNIFICANT WAVES MODELLING IN GULF OF LIONS

Since theoretical aspects of spatial extreme modelling have been presented, this section aims to finally fulfil such analysis on a real data set. The domain S of the case study is the Gulf of Lions. Let $\left\{Z(s), s \in S \subset \mathbb{R}^2\right\}$ denotes a process of monthly maxima of significant wave heights over the Gulf of Lions. In this study, summer months are dropped off from the data set to avoid seasonality effect in the modelling. Hence, the modelling rests on the observations of $n = 52$ (years) $\times$ 8(months) maxima, denoted z_{ik} where $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, 97\}$.
The application consists in firstly transforming observations z_{ik} to unit Fréchet scale. Then, we establish a dependence structure diagnostic thanks to the fmadogram measure. Next, we model the spatial dependence structure with the presented max-stable models. Once selected, the best fitted model is finally used to compute quantities of interest for risks analyses.

Margins transformation. We stated that several techniques are available to define a continuous evolution of the GEV margins parameters through the area of interest. In the one hand and because the computational grid of the hindcast is highly refined, one may transform the data at each location of the mesh and then interpolate the parameters through the space of interest. In the other hand, one can provide smooth functions of those parameters along potential covariates. Here, we do not compare those techniques but choose to work with smooth functions to reason in term of region instead of site-by-site analysis. The later requiring a potential outliers identification routine.
Thus, our analysis is focused on fitting the evolution of the GEV parameters with linear functions or radial basis splines or both. The first approach is to use only linear functions, as presented in Equation 2. Linear functions may be defined for the location parameter, scale parameter, shape parameter or any combinations of those. The second approach is to use radial basis splines to model the evolution of the GEV parameters, as presented in Equation 3. Considering so, 300 combinations of those functions are involved in the estimation of the spatial GEV parameters.
We identify that best results stem from the use of radial basis splines – of order 3 – for both location and scale parameters, while the shape parameter keeps being a linear trend along the bathymetry. These evolutions are illustrated by Figure 5. It appears that the evolution against the bathymetry covariate is much more distinctive than along the longitude and latitude covariates. This can fully be explained by the physical wave process itself: the Gulf of Lions presents an inner continental shelf, which directly controls wave formation and propagation. Also, we can notice that marginal distributions of littoral sites present heavier tails than the distributions of offshore sites. The interpretation is much more ambiguous. The discrepancy may come from the fact that extreme wave spectra are not transferred in the same way as more moderate spectra when waves propagate to the inner continental shelf.

F-Madogram diagnostic. The spatial dependence structure is represented by the evolution of the extremal coefficient, estimated by the fmadogram. We choose a reference-site and proceed to the computation of the fmadogram over all possible pairs available. Then, results are interpolated over all the area. We repeat this procedure for each site selected and obtain fmadograms maps as presented in Figure 6. From those maps several characteristics of the dependence structure can be highlighted.
The Gulf of Lions has a specific orientation, with a south-east open sea boundary. This particular orientation allows swells from along this axis to be more easily generated, which is confirmed by direct observations. As a consequence, the plotted fmadograms reveal an anisotropy along the orthogonal *South-West / North-East* axis. In opposite, the dependence of extreme significant wave heights along the *North-Western / South-East* axis appears to have a clear separation while comparing pairs composed of one littoral site and one offshore site. It might be due to the fetch distance induced by the direction of two of the dominant winds in that region, namely Mistral (northerly) and Tramontane (north-westerly). In those configurations, waves grow while propagating offshore, leaving the littoral sites with a too short fetch for being well formed. Therefore, a weak dependence between littoral sites and offshore sites becomes self-explanatory. An other interpretation may include a more regional explication due to circulation pattern. Indeed the coastline at



(a) Location parameter (b) Scale parameter (c) Shape parameter

Figure 5: Evolution of the GEV parameters along bathymetry covariate. Dots are the marginal estimation of the GEV parameters from the 97 sites. Lines are the continuous functions fitted to those parameters, which smooth the evolution of the parameters for any location of the area.

Model (Corr. function)	Isotropic	CLIC
Smith anisotropic	**NO**	**14975454**
Schlather (any corr. function)	NO	–
Schlather (Bessel)	YES	15106251
Schlather (Cauchy)	YES	15106559
Schlather (Cauchy generalised)	YES	15185107
Schlather (Powered exponential)	**YES**	**15098615**
Smith	YES	15116131
Schlather (Whittle-mattern)	YES	15106561

Table 1: CLIC of models fitted.

the extremities of the Gulf use to block swells coming from South-West and North-East, which in those conditions impact only offshore sites.

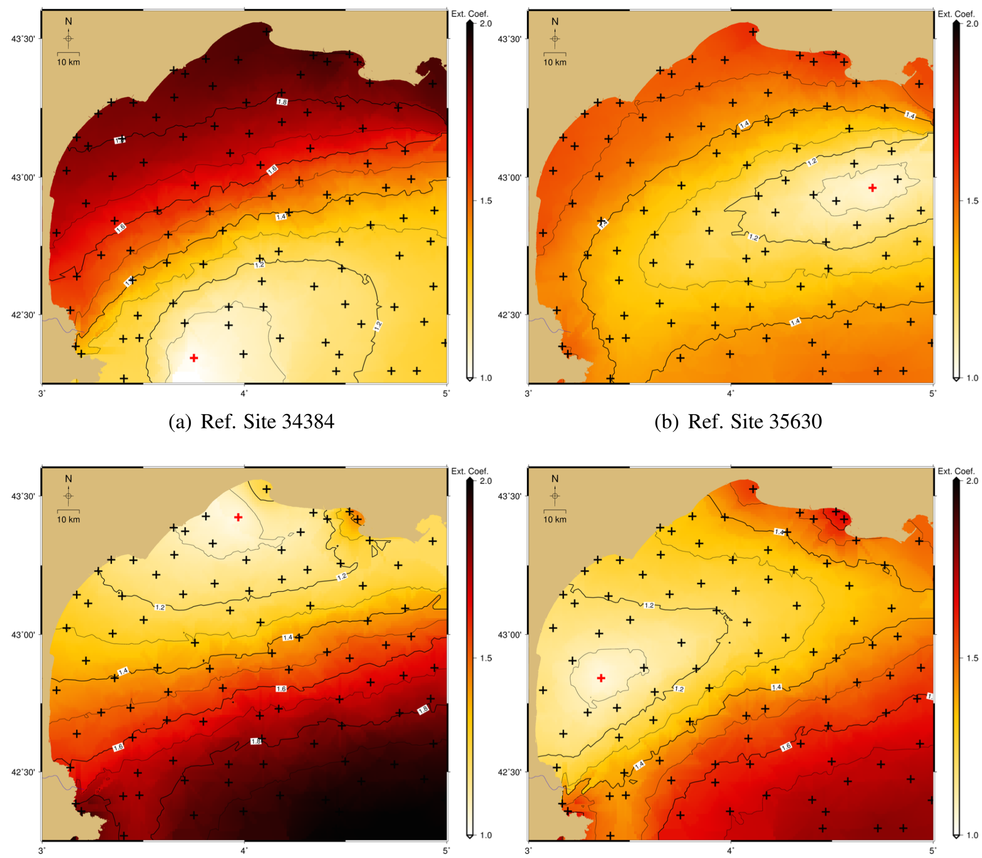
Dependence structure modelling. Having previously defined continuous functions to retrieve the GEV parameters over the area of interest, we transform the marginal data to unit Fréchet scale. To assess the dependence structure, several max-stable models are fitted to these transformed data. The CLIC of the models fitted are reported in Table 1.

This table presents two sections. The first one concerns models fitting anisotropic dependence structures. The second one concerns models able to only fit isotropic dependence structure – see Discussion.

If we only consider isotropic fits, the Schlather model with powered-exponential correlation function outperforms the other models. However the anisotropic Smith model better fits the data than any isotropic model. This result confirms the presence of an underlying anisotropic dependence structure, as discovered thanks to the estimated extremal coefficient maps. Therefore we are going to use the Smith anisotropic fitted model for the following sections.

RESULTS

Extreme return value. In coastal engineering, one of the first quantity of interest in the dimensioning of structures or environmental studies are the return values. From the best fitted model we can seamlessly produce stochastic simulations of monthly maxima processes. These processes implicitly take into account the modelled dependence structure and also the spatial GEV parameters via a back-transformation to original scale. As a real advantage, the information is given at any location of the studied area, where site by site analysis need interpolation to deliver maps of extreme return values. To perform such report, we compute



(a) Ref. Site 34384

(b) Ref. Site 35630

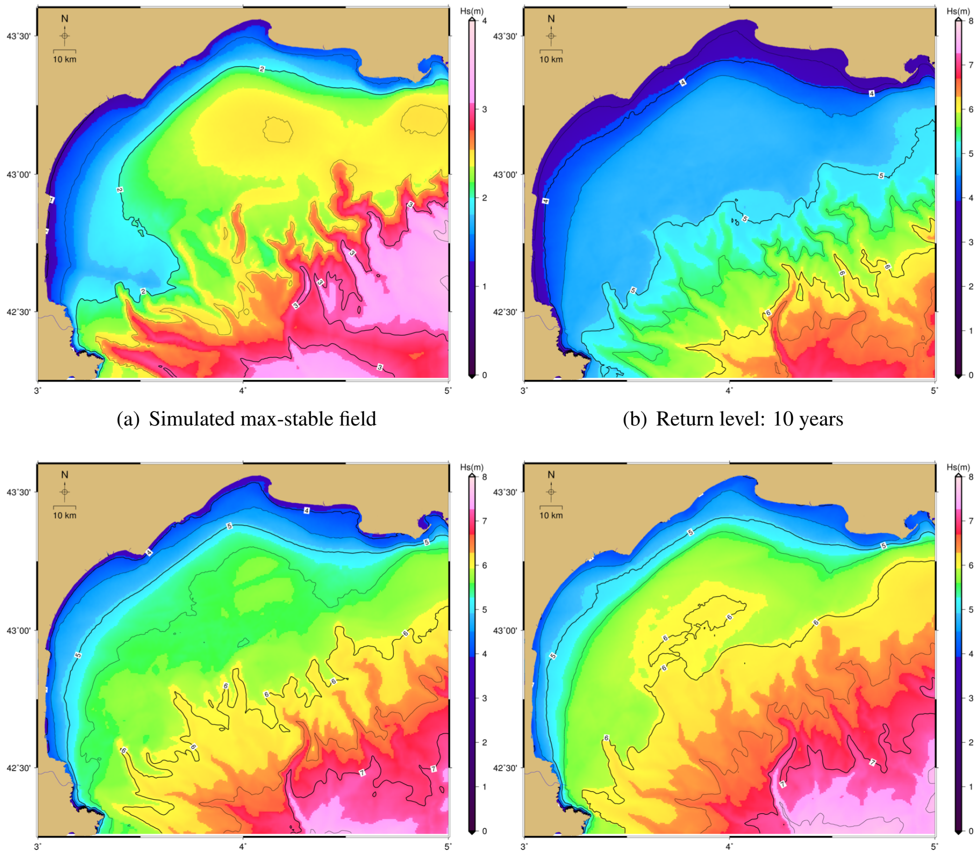
(c) Ref. Site 38497

(d) Ref. Site 35856

Figure 6: Interpolated extremal coefficient estimated by the fmadogram measure between a reference site (red-cross) and all possible pairs. Such map allows to feel the underlying dependence structure within extreme wave events of the Gulf of Lions.

5000 max-stable processes and then empirically derive from them the marginal return values over the Gulf of Lions. The results for 10, 50 and 100 years are presented in figure 7.

Joint probabilities. One quantity of great interest in risk analysis is the survival joint probabilities defined as $P(Z(x) > r_t(x),\ x \in \Omega \subset S)$ with $r_t(x)$ the return level of the t^{th}-period in a site x, i.e. $P(Z(x) > r_t(x)) = 1/t = p$. From a selection of sites $\Omega = (x1, ..., xw) \subset S$, such probability may explain why those sites are – or are not – impacted by waves at the same scale during extreme conditions. Indeed, this probability will determine whether those sites are more likely subject to observe exceedance of their marginal return level in a same period or not. Since sediment transport is calculated from explicit formalisms that require waves features as main input parameters, one can identify some patterns to argue the behaviour of cross-shore or long-shore sediments transport responding to extremes from such probability. For instance, two close sites with a low joint probability of exceedance may traduce a very local behaviour of the extreme waves and therefore an amount of energy significantly different. In the opposite, two sites far away from each other observing a high joint probability of exceedance may represent a more regional behaviour of waves. One can then discuss patterns of sediments transport at a regional scale. If we do not rely on a spatial extreme analysis, one may assume the independence or full dependence of the marginal probability to compute the joint probability of exceedance. If so and assuming margins are transformed in unit Fréchet distribution,



(a) Simulated max-stable field

(b) Return level: 10 years

(c) Return level: 50 years

(d) Return level: 100 years

Figure 7: In top left corner: A simulated process according to the best fitted max-stable model. Elsewhere: Empirical marginal return values computed from simulation of 5000 max-stable processes as presented in (a). These simulations take into account the underlying dependence structure.

the theoretical joint probability of exceedance considering a full dependence of sites is given by

$$\begin{aligned} P(Z(x) > r_t(x),\ x \in \Omega) &= P(Z(x) > r_t,\ x \in \Omega) \\ &= P(Z(x_1) > r_t,\ \ldots,\ Z(x_\omega) > r_t) \\ &= P(Z(x_k) > r_t), \quad x_k \in \Omega \\ &= t^{-1}, \quad \text{by definition,} \end{aligned}$$

while for the total independence it is defined as

$$\begin{aligned} P(Z(x) > r_t(x), x \in \Omega) &= P(Z(x_1) > r_t) \times \ldots \times P(Z(x_\omega) > r_t) \\ &= P(Z(x) > r_t)^{\omega} \\ &= t^{-\omega}, \end{aligned}$$

with ω the number of sites considered.

From several set of sites $\Omega_i \subset \chi$, we compare the joint probability of exceedance $P(Z(x) > r_t(x), x \in \Omega_i)$ computed from 1) the theoretical full dependence case; 2) the theoretical total independent case; 3) the observations – i.e. hindcast dataset – and 4) the best fitted max-stable model through the simulations of 5000 processes. The results are given in Figure 8.

Generally speaking the max-stable model outperforms the other theoretical cases for any set of sites and therefore whatever the distance considered. When considering littoral sites altogether, the model achieve to

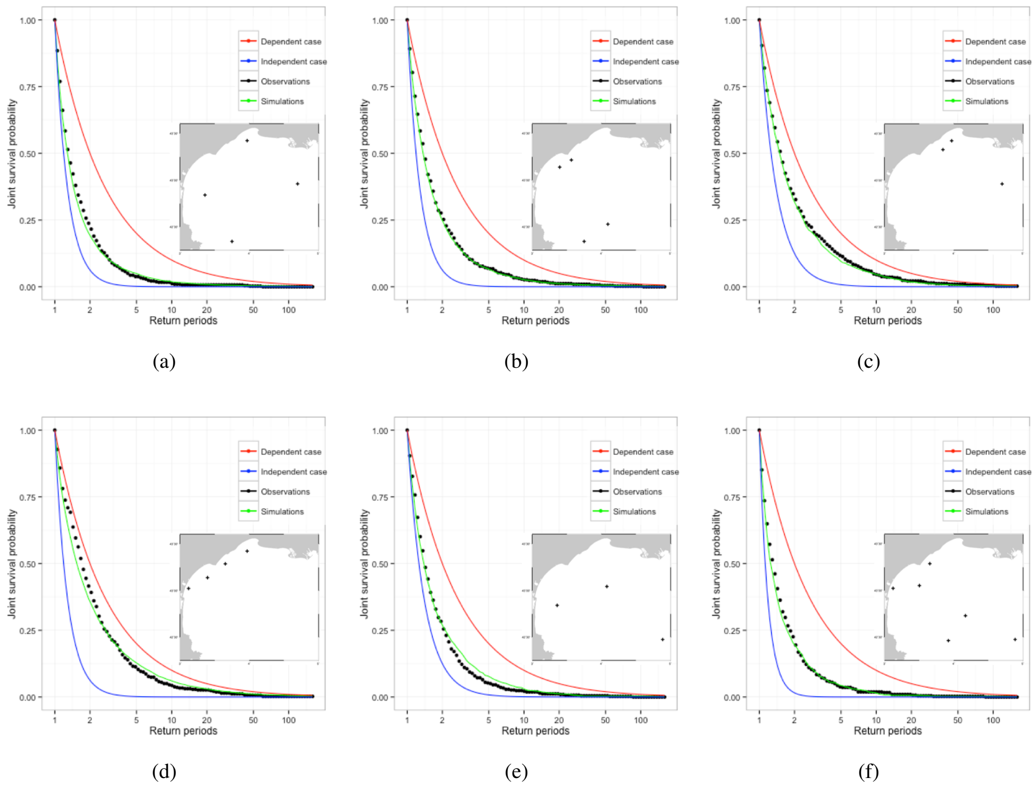


Figure 8: Joint probabilities of exceedance for different sites against return periods. Probabilities from observations are dot points and green curve is the probabilities computed from 5000 max-stable simulations of the best fitted model (Smith anisotropic). The two remaining lines corresponds to theoretical independent and full dependent cases.

represent joint probabilities relatively close to the dependent case, as outlined by the observations. When sites are both picked-up in littoral area and offshore area, the model still seems to represent correctly the dependence of the observations. In that case the joint probabilities of exceedance lies in the very between of the dependent and independent case. This confirms the usefulness of such a model when considering such quantities. Finally, when considering sites at any corner of the Gulf of Lions – i.e. far away from each other –, joint probabilities of exceedance quickly drop to 0. We can interpret this information by the fact that the probability of having really extreme waves over the entire Gulf of Lions in the same period is weak.

Other possibilities. Spatial extreme modelling via max-stable modelling allows many other outcomes. Among those, one can use the best fitted model to determine useful conditional probabilities. This quantity remains directly impacted from the dependence structure and therefore a spatial model is worthwhile. Still using best fitted max-stable model, one can conditionally simulate processes of maxima. Such a process is of great interest to determine spatial behaviour of extremes in a specific configuration, i.e. conditionally to one or many points. An other potential use of this max-stable modelling is to stochastically generate processes and use them "as-is" in physical model addressing long-term questionings.

DISCUSSION

Along this study we alert the reader on the importance of modelling the dependence structure of extreme environmental physical phenomena. The method presented here seems to outperform the univariate and multivariate approaches by dealing with a continuous space. Direct benefits of such modelling are presented. For instance we introduce the possibility of stochastically model extreme events or the capacity to compute accurate joint probabilities. Both quantities are interesting for direct use in risks analyses. How-

ever several limits of the approach can be highlighted from this document.
First of all, such stochastic modelling can not be accurate if the realisations of the random variables studied – here hindcast values – are not representative. Additionally, both GEV margins parameters (with any kind of potential trends) and dependence structure parameters could be estimated in the same procedure. It has the advantages of having a systematic way to compare the model – CLIC selection – , and avoid over-parametrisation of models at the end of the procedure. Facing issues in the inference routine when dealing with all those parameters lead us to do not consider this approach.
Secondly, some drawbacks raise from the assessment of the dependence structure by an anisotropic model since only Smith model is used. Indeed, Smith model natively handles anisotropic cases since it uses a Mahalanobis distance in the definition of the extremal coefficient. Such a distance may be weighted along one axis by providing an unbalanced covariance matrix. In the opposite, the Schlather model relies on euclidean distance. Therefore it does not handle anisotropic cases natively. Instead, we need to apply the model onto a transformed space $S' = VS$, where V is a rotation-transformation $d \times d$ matrix. In that case the estimation of the matrix V parameters need to be included in the overall optimisation procedure. This last step is not yet implemented in the R package of Ribatet et al. (2011). Since our analysis relies on this code, this limitation prevents us to model the observed anisotropy with a Schlather model with any correlation function associated. To face this issue we are under development of our own code. This would be presented in a future work.
Also, we worked on maxima over a period to determine extremes contained in the data. In the same scheme as in the lower order cases, it would be worthwhile to use exceedances over threshold methods to catch more information from the extremes observed.
We identify in a preliminary analysis that one of the most influencing covariates of the significant wave process is their mean-waves direction. Therefore it would be a valuable add to investigate such covariate in the modelling, whether by fixing it or by including it into the model. This last proposition is one of our perspective research. In the same time, we would like to develop a model to handle space-time extreme events, still referring to the formalism of max-stable processes. This would allow us to move from climate scale questioning to storm-event scale.

ACKNOWLEDGEMENTS

This work was performed within the context of both LITTOCMS and MISTRAL-LITTORAL projects. It also takes place in the GLADYS initiative – see www.gladys-littoral.org. The work was partially supported by ANR-McSim and GICC-Miraccle projects. Some computations were also realized with the support of HPC@LR, a Center of Competence in High-Performance Computing from the Languedoc-Roussillon region, funded by the Languedoc-Roussillon region, the Europe and the Université Montpellier 2 Sciences et Techniques. The HPC@LR Center is equiped with an IBM hybrid Supercomputer.

References

J.-N. Bacro and C. Gaetan. A review on spatial extreme modelling. In *Advances and Challenges in Space-time Modelling of Natural Events*, pages 103–124. Springer, 2012.

J. Beuvier, F. Sevault, M. Herrmann, H. Kontoyiannis, W. Ludwig, M. Rixen, E. Stanev, K. Béranger, and S. Somot. Modeling the mediterranean sea interannual variability during 1961–2000: focus on the eastern mediterranean transient. *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C8), 2010.

J. Blanchet and A. C. Davison. Spatial modeling of extreme snow depth. *The Annals of Applied Statistics*, 5(3):1699–1725, 2011.

D. Cooley, P. Naveau, and P. Poncet. Variograms for spatial max-stable random fields. In *Dependence in probability and statistics*, pages 373–390. Springer, 2006.

A. C. Davison and M. M. Gholamrezaee. Geostatistics of extremes Subject collections. *The Royal Society*, (October), 2011.

L. de Haan and T. T. Pereira. Spatial extremes: Models for the stationary case. *The annals of statistics*, pages 146–168, 2006.

L. de Haan et al. A spectral representation for max-stable processes. *The Annals of Probability*, 12(04): 1194–1204, 1984.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, pages 423–453, 1943.

M. J. Herrmann and S. Somot. Relevance of era40 dynamical downscaling for modeling deep convection in the mediterranean sea. *Geophysical Research Letters*, 35(4), 2008.

P. Jonathan, D. Randell, Y. Wu, and K. Ewans. Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Submitted to Ocean Engineering October*, 2013.

Z. Kabluchko, M. Schlather, and L. De Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, pages 2042–2065, 2009.

S. A. Padoan, M. Ribatet, and S. A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, 2010.

N. Raillard, P. Ailliot, and J. Yao. Modeling extreme values of processes observed at irregular time steps: Application to significant wave height. *Annals of Applied Statistics*, pages 1–26, 2014.

M. Ribatet, R. Singleton, and R. Team. Spatialextremes: modelling spatial extremes. *R package version*, pages 1–8, 2011.

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

M. Schlather and J. A. Tawn. A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90(1):139–156, 2003.

R. L. Smith. Max-stable processes and spatial extremes. *Unpublished manuscript, Univer*, 1990.

H. L. Tolman. User manual and system documentation of wavewatch iii® version 4.18. Technical Report 316, 2014.

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