Application of Scientific Computing and Statistical Analysis to address Coastal Hazards

Romain Chailan

Collège Doctoral du Languedoc-Roussillon

THÈSE

Pour obtenir le grade de

Docteur

Délivré par l'**Université de Montpellier**

Préparée au sein de l'école doctorale **I2S***
Et des unités de recherche **UMR 5506, 5149, 5243**

Spécialité: **Informatique**

Présentée par **Romain Chailan**
romain.chailan@umontpellier.fr

Application of Scientific Computing & Statistical Analysis to address Coastal Hazards

Soutenue le 23/11/2015 devant le jury composé de :

Liliane Bel, Professeur, AgroParisTech	Rapporteur
Liva Ralaivola, Professeur, Université Aix-Marseille	Rapporteur
Pierre Ailliot, Maitre de Conférence, Université de Brest	Examinateur
Edward Anthony, Professeur, CEREGE	Examinateur
Bijan Mohammadi, Professeur, Université de Montpellier	Examinateur
Anne Laurent, Professeur, Université de Montpellier	Directrice
Gwladys Toulemonde, Maitre de Conférence, Université de Montpellier	Encadrant
Frédéric Bouchette, Maitre de Conférence, Université de Montpellier	Encadrant
Olivier Hess, Docteur, IBM France	Invité

* **I2S:** École doctorale Information Structures Systèmes

Remember that all models are wrong;
the practical question is how wrong do they have to be to not be useful.

– Box G. and Draper N. (1987)

Remerciements

À l'instar de nombreux collègues doctorants, je commence ici par la fin : les remerciements. Ces remerciements sont donnés en fin de parcours, mais méritent leur première place tant les personnes citées ici ont compté et compte pour moi, dans mes activités professionnelles et bien au delà.

La thèse n'est pas un sprint, mais un marathon dans lequel il faut parfois... sprinter. Cette performance d'endurance est indissociable de l'accompagnement et de l'inspiration qui m'ont été fournis, que ce soit du point de vu scientifique ou humain. En ce sens je remercie en tout premier lieu mes encadrants de thèse Anne Laurent, Gwladys Toulemonde et Frédéric Bouchette.

Je remercie également mes rapporteurs, Liliane Bel et Liva Ralaivola, qui ont pris le soin d'étudier, de commenter et enfin d'arbitrer mon travail, depuis la remise du présent document jusqu'à sa soutenance.
Mes examinateurs Pierre Ailliot, Edward Anthony et Bijan Mohammadi ont contribué activement aux échanges enrichissants lors de ma soutenance de thèse. C'est pourquoi je les remercie tout particulièrement.
L'ensemble de ce jury m'a conforté dans ma perception de la recherche et je reste enthousiaste dans ma volonté de travailler avec chacun d'entre eux, à l'interface de leur discipline respective.

Ma thèse s'inscrit dans un contexte industriel (thèse CIFRE), et à ce titre j'ai eu le privilège d'évoluer au sein d'IBM France, au centre de Montpellier. Cette opportunité a pu se réaliser grâce à l'énorme implication de Colin Dumontier, Olivier Hess, Xavier Vasques et Yann de Visme. Je les remercie vivement, et notamment car j'ai grandement apprécié leur accompagnement dans mon projet professionnel avant, pendant et pour l'après thèse.
Je remercie plus généralement mon équipe à IBM, et particulièrement Marie-Angèle pour ses relectures et ses conseils ; Geoffrey, Elsa et Nicolas pour (entre autre) les tirades à n'en plus finir au bureau. Par ailleurs je remercie François et ses gâteaux, Olivier et son entrecôte, Eric et ses soirées FIFA ainsi que Dolu et ses barbecues pour l'aide sur les calculateurs HPC. Je remercie aussi Hervé et sa ponctualité « pauses café » robuste à toute épreuve pour son support réseau, Matthieu et ses versions alpha avec qui ont aime penser à faire bouger les choses et enfin Chris, qui apporte bonne humeur, rire et football contre vents et marées

(sans jeu de mot avec le sujet de la thèse).

À l'Université, j'ai eu la chance de rencontrer et travailler avec des personnes de grande qualité. Je remercie Fabien bien sûr, avec qui j'ai évolué de paire pendant cette thèse. Avec Julien et Cyril, nous resterons la team Mirmidon et je vous remercie pour ces nombreuses heures de travail communes, pour votre implication bien au delà de ce qui était requis. Lise, Manon, Robin, Olivier et tout le laboratoire GM au sens large, je vous remercie pour les échanges, les repas et votre accueil.

Au delà des mes organismes d'accueil, je remercie Héloïse Michaud pour sa disponibilité, son expertise et ses conseils sur la modélisation d'états de mer. Je remercie par ailleurs l'équipe du LEGOS à Toulouse pour ses formations et sa disponibilité.

La vie de doctorant est rythmée par la présence de conférences. Dans mon cas, j'ai pu rencontrer dans ces dernières les très sympathiques Aurélien, Julie et Marc. Ils m'ont accompagné dans ces expériences et les ont égayées : merci à eux.

Autour du cadre professionnel, de nombreuses personnes m'ont soutenu. Je pense notamment au soutien inconditionnel de toute l'équipe IG de Polytech Montpellier et naturellement je les remercie.

Aussi, pour tenir un tel marathon, il faut être préparé mentalement et physiquement. Ici et là, mes nombreuses sessions de kite en compagnie de mes fidèles amis riders Cyril, Olivier et Arnaud ont largement contribué à cette bonne préparation, en plus de mettre un peu de pratique vis à vis du contexte de ma thèse. Merci beaucoup !

Un énorme merci à tous mes amis qui ont eu la patience et le courage de me supporter pendant cette période.
Merci à mes amis de Montpellier : Charlotte (CCC), Florent (Flo), Philippe (Fifou), Ronan (Petit phoque), Max (Bicks sauvage), Chris (Parrain), Maximilien (Max), Ziad (Zouzou), Jean-Baptiste (JBeo) et Aurélien (Auré).
Un grand grand merci aux amis de la team galaxie anglaise sur qui je peux toujours compter : Michel (Michel Michel Michel), Nico (Nicotine) et Alex (Super Flying Matiron).
Un très grand merci à mes amis d'enfance pour leur présence pendant ces trois ans : Kévin (Massu), Greg (Goldi) et Laurent (xlimit).

Je remercie enfin toute ma famille pour son soutien et support tout au long de ces trois années. En particulier, un immense merci à Marion, qui a dû subir au quotidien mes sauts d'humeur et mon stress, et qui a surtout su me conseiller dans tous les bons et mauvais moments, du premier au dernier kilomètre !

Merci à tous !

Résumé

L'étude et la gestion des risques littoraux sont plébiscitées par notre société au vu des enjeux économiques et écologiques qui y sont associés. Ces risques sont généralement réponse à des conditions environnementales extrêmes. L'étude de ces phénomènes physiques repose sur la compréhension de ces conditions rarement (voire jamais) observées.
Dans un milieu littoral, la principale source d'énergie physique est véhiculée par les vagues. Cette énergie est responsable des risques littoraux comme l'érosion et la submersion qui évoluent à des échelles de temps différentes (long-terme ou événementielle).
Le travail réalisé, situé à l'interface de l'analyse statistique, de la géophysique et de l'informatique, vise à apporter des méthodologies et outils aux décideurs en charge de la gestion de tels risques.
En pratique, nous nous intéressons à mettre en place des méthodes qui prennent en compte non seulement un site ponctuel mais traitent les problématiques de façon spatiale. Ce besoin provient de la nature même des phénomènes environnementaux qui sont spatiaux, tels les champs de vagues.
L'étude des réalisations extrêmes de ces processus repose sur la disponibilité d'un jeu de données représentatif à la fois dans l'espace et dans le temps, permettant de projeter l'information au-delà de ce qui a déjà été observé. Dans le cas particulier des champs de vagues, nous avons recours à la simulation numérique sur calculateur haute performance (HPC) pour réaliser un tel jeu de données. Le résultat de ce premier travail offre de nombreuses possibilités d'applications.
En particulier, nous proposons à partir de ce jeu de données deux méthodologies statistiques qui ont pour but respectif de répondre aux problématiques de risques littoraux long-termes (érosion) et à celles relatives aux risques événementiels (submersion).
La première méthodologie s'appuie sur l'application de modèles stochastiques dit max-stables, particulièrement adapté à l'étude des événements extrêmes. En plus de l'information marginale, ces modèles permettent de prendre en compte la structure de dépendance spatiale des valeurs extrêmes. Nos résultats montrent l'intérêt de cette méthode au devant de la négligence de la dépendance spatiale de ces phénomènes pour le calcul d'indices de risque.
La seconde approche est une méthode semi-paramétrique dont le but est de simuler des champs spatio-temporels d'états-de-mer extrêmes. Ces champs, interprétés comme des tempêtes, sont des amplifications contrôlées et bi-variées d'épisodes

extrêmes déjà observés. Ils forment donc des tempêtes encore plus extrêmes. Les tempêtes simulées à une intensité contrôlée alimentent des modèles physiques événementiels à la côte, permettant d'aider les décideurs à l'anticipation de ces risques encore non observés.
Enfin et depuis la construction de ces scenarii extrêmes, nous abordons la notion de pré-calcul dans le but d'apporter en quasi-temps réel au décideur et en temps de crise une aide à la décision sur le risque littoral.
L'ensemble de ce travail s'inscrit dans le cadre d'un besoin industriel d'aide à la modélisation physique : chainage de modèles numériques et statistiques. La dimension industrielle de cette thèse est largement consacrée à la conception et au développement d'un prototype de plateforme de modélisation permettant l'utilisation systématique d'un calculateur HPC pour les simulations et le chainage de modèles de façon générique.
Autour de problématiques liées à la gestion du risque littoral, cette thèse démontre l'apport d'un travail de recherche à l'interface de plusieurs disciplines. Elle y répond en conciliant et proposant des méthodes de pointe prenant racine dans chacune de ces disciplines.

Titre en français

Application du Calcul Scientifique et de l'Analyse Statistique à la gestion du risque en milieu littoral

Mots-clés

— Risques Littoraux
— Analyse en valeurs Extrêmes
— Statistiques de Haute Performance
— Modélisation d'Etats de Mer

Abstract

Studies and management of coastal hazards are of high concerns in our society, since they engage highly valuable economical and ecological stakes. Coastal hazards are generally responding to extreme environmental conditions. The study of these physical phenomena relies on the understanding of such environmental conditions, which are rarely (or even never) observed.
In coastal areas, waves are the main source of energy. This energy is responsible of coastal hazards developed at different time-scales, like the submersion or the erosion.
This work, taking place at the interface between Statistical Analysis, Geophysics and Computer Sciences, aiming at bringing forward tools and methods serving decision makers in charge of the management of such risks.
In practice, the proposed solutions answer to the questionings with a consideration of the spatial dimension rather than only punctual aspects. This approach is more natural considering that environmental phenomena are generally spatial, as the sea-waves fields.
The study of extreme realisations of such processes is based on the availability of a representative data set, both in time and space dimensions, allowing to extrapolating information beyond the actual observations. In particular for sea-waves fields, we use numerical simulation on high performance computational clusters (HPC) to product such a data set. The outcome of this work offers many application possibilities.
Most notably, we propose from this data set two statistical methodologies, having respective goals of dealing with littoral hazards long-terms questionings (e.g., erosion) and event-scale questionings (e.g., submersion).
The first one is based on the application of stochastic models so-called max-stable models, particularly adapted to the study of extreme values in a spatial context. Indeed, additionally to the marginal information, max-stable models allow to take into account the spatial dependence structures of the observed extreme processes. Our results show the interest of this method against the ones neglecting the spatial dependence of these phenomena for risk indices computation.
The second approach is a semi-parametric method aiming at simulating extreme waves space-time processes. Those processes, interpreted as storms, are controlled and bi-variate uplifting of already observed extreme episodes. In other words, we create most severe storms than the ones already observed. These processes simulated at a controlled intensity may feed littoral physical models in

order to describe a very extreme event in both space and time dimensions. They allow helping decision-makers in the anticipation of hazards not yet observed.
Finally and from the construction of these extreme scenarios, we introduce a pre-computing principle in the goal of providing the decision-makers with a real-time and accurate information in case of a sudden coastal crisis, without performing any physical simulation.
This work fits into a growing industrial demand of modelling help. Most notably a need related to the chaining of numerical and statistical models. Consequently, the industrial dimension of this PhD. is mostly dedicated to the design and development of a prototype modelling platform. This platform aims at systematically using HPC resources to run simulations and easing the chaining of models.
Embracing solutions towards questionings related to the management of coastal hazard, this thesis demonstrates the benefits of a research work placed at the interface between several domains. This thesis answers such questionings by providing end-users with cutting-edge methods stemming from each of those domains.

Title in English

Application of Scientific Computing and Statistical Analysis to Address Coastal Hazards

Keywords

— Coastal Hazards
— Extreme Value Analysis
— High Performance Analytics
— Wave Modelling

Rattachement

Laboratoire

LIRMM - Laboratoire d'Informatique, Robotique et Micro-électronique de Montpellier

Adresse

Université de Montpellier
UMR 5506 - LIRMM
CC477
161 rue Ada
34095 Montpellier Cedex 5 - France

Laboratoire

IMAG - Institut Montpelliérain Alexander Grothendiek

Adresse

Université de Montpellier
UMR 5149 - IMAG
CC051
Place Eugène Bataillon
34095 Montpellier Cedex 5 - France

Laboratoire

GM - Géosciences Montpellier

Adresse

Université de Montpellier
UMR 5243 - GM
CC060
Place Eugène Bataillon
34095 Montpellier Cedex 5 - France

Contents

Introduction

Humanity in the 21st century is a consequence of the combination of endless possibilities, at various levels. Among them, exchanges of all kind played a key-role. Exchanges of energy at the big-bang between components of what we currently call the matter, then exchanges of processes, knowledges or wealth between first human beings. These exchanges have brought us and our society to a complex level that even Charles Darwin may have not consider.
An exchange is policed by an **interface**. In Geography, an interface is an area between two limited countries in which people and goods transit. In Software Engineering, it is a piece of code allowing exchanging information between several programs. Regarding coastal hazards, to explore, design and create tools and methods aiming at helping the decision towards the anticipation and management of those crises require also a role of interface. Indeed, many subjects have to interact to give birth of cutting edge solutions able to address these hazards.
Most notably, solutions against **erosion** and **submersion** phenomena (see Figure 1) are questionings of high concerns considering economical and ecological stakes engaged by coastal hazards.

The main factor responsible of such devastating physical phenomena is the amount of energy present in littoral systems. What is remarkable is that seawaves can be physically reduced as a transport of energy. Hence the study of their behaviour is paramount to quantify coastal hazards.
Coastal hazards respond to different time-scales. They may refer to climate scale phenomena or to event scale phenomena. Climate-scale hazards refer to questionings of a long-term effect: where time is aggregated over long periods like several years. This is for instance the time-scale evolution of the **long-term shoreline erosion**. Event-scale hazards refer to questionings where the focus is on the short-term time evolution of the physical phenomenon itself. **Submersion** events are typical candidates of such questionings.
Whatever the kind of hazard and in regard of the complexity of such environmental processes, decision-makers demand tools and methods to help them in their activities of anticipation and management of coastal crises. What it is expected is not only the assessment of common behaviour of those physical processes, but much more about rarely observed or even the unexpected ones, known as **extreme** events.
As a matter of fact, we can state that extreme – rare but strong – events are generally related to severely damaging coastal hazards.

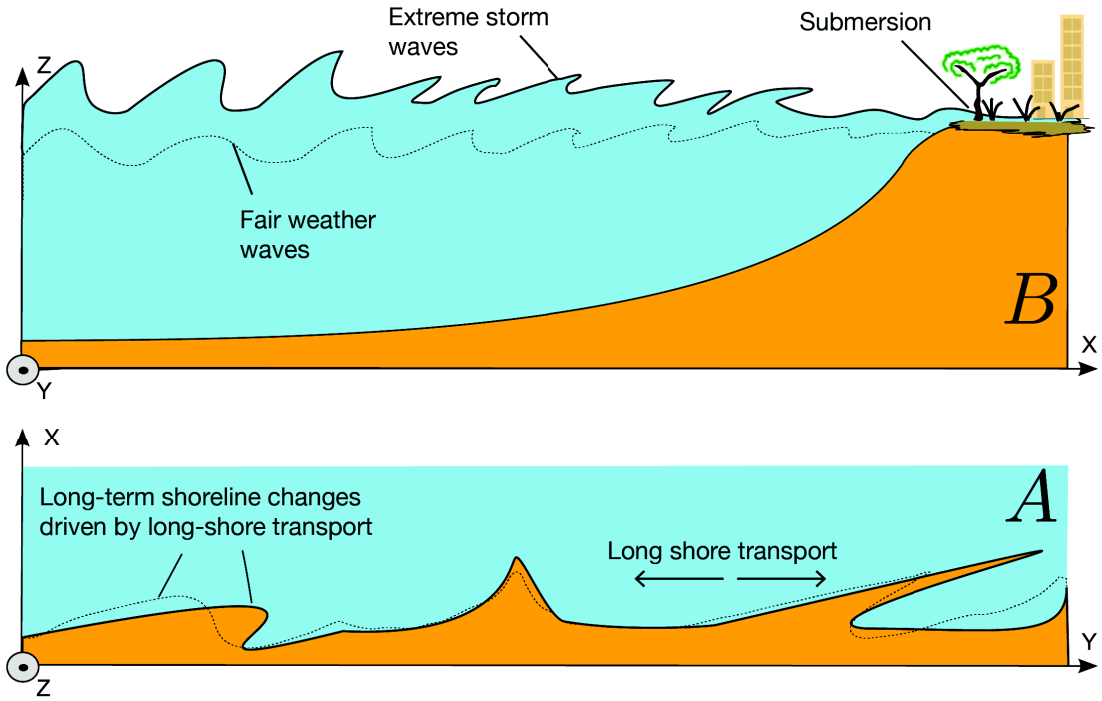


Figure 1: Schematic representation of two coastal hazards examples. A) Long-term shoreline change caused by long-shore transport, which is a climate-scale hazard. B) Submersion caused by extreme storm waves, observed at an event-scale.

Waves supply the major amount of energy that contributes to impact the coast. This impact is therefore relative to waves extremeness. Extremeness of a wave does not simply rely on its height but also on its direction and its period. Therefore it is mandatory to assess the complex behaviour of extreme waves in the final goal of explaining past hazards, managing on-going ones and anticipating up-coming ones. Observations do not suffice to anticipate extreme events. New techniques and tools have to be developed and applied. In particular, to model up-coming extreme events is affordable by the use of stochastic modelling.

In practice, proposed solutions of this document answer to the questionings with a consideration of the spatial dimension rather than dealing with only punctual aspects. This approach is more natural considering that environmental phenomena are generally spatial, as the sea-waves fields.
Since we want to quantify coastal hazards in regards of events that are likely to occur once in several year or in several decade, and from wich information is lacking, stochastic extreme value modelling has to be used. This is a well-accepted theory used to model extreme values. Firstly introduced in the 1930's, this theory becomes nowadays particularly used in geo-statistics when questionings deal with extreme environmental events. The extreme value modelling topic is largely discussed later in this document. At this point it is important to understand the overall goal of such modelling.
Above we highlight the will to anticipate extreme events. They can be events

that are likely to appear but have not yet appeared or events that have not been observed due to a technical lack. The extreme value theory provides a mathematical framework to extrapolate such information from historical time-series. The quality of extrapolation depends on the availability and quality of the observed extreme events. By definition, to observe extreme events is commonly a requirement that is hard to satisfy.
Even with such constraints, this mathematical framework is robust and can provide information up to long return period; A return period being associated to a return level, which is a value reached once in means during its return period interval.

Regarding the presented coastal questionings, to be able to model extreme values contribute to understand the behaviour of events that are likely to cause severe damages to ecological or economical assets of the littoral. Hence two of the scientific challenges addressed in this thesis are to provide statistical extreme value methodologies allowing responding to both a long-term scale questioning and an event-scale questioning. The first one is based on the application of stochastic models so-called max-stable models, particularly adapted to the study of extreme values in a spatial context. Indeed, additionally to the marginal information, max-stable models allow to take into account the spatial dependence structures of the observed extreme processes. Our results show the interest of this method against the ones neglecting the spatial dependence of these phenomena for risk indices computation.
The second approach is a semi-parametric method aiming at simulating extreme waves space-time processes. Those processes, interpreted as storms, are controlled and bi-variate uplifting of already observed extreme episodes. In other words, we create most severe storms than the one already observed. These processes simulated at a controlled intensity may feed littoral physical models in order to describe a very extreme event in both space and time dimensions. They allow helping decision-makers in the anticipation of hazards not yet observed.

To understand the behaviour of a random variable up to its very rare quantities, we need a reliable set of its realisations. There are several ways to observe waves in modern sciences and engineering.

— In first position we can mention surface-buoys. They are instrumented platforms lying on surface of oceans. Surface-buoys are able to measure sea-states in situ before communicating the collected information. The records are accurate but those sensors are relatively expensive to deploy and maintain. Hence their time-series might be short or discontinued or both.
— An alternative to surface-buoys are altimeters. Altimeters are embedded into satellites observing surface oceans all around the earth on moving tracks. Measures are accurate too, but their major drawback is the non-regularity of their tracks through time and space around the globe. This

can make their analyses challenging from a statistical point of view.

— Finally modern sciences provide a more numerical way to observe waves: wave models. These consist in simulating the behaviour of seas and oceans by solving complex physical equations. Those equations rest on atmospheric and ocean boundary conditions, and are solved by numerical schemes. In practice wave-modelling is widely used since the simulated data-sets match both valuable criteria: such data-set contains long-times series and are referenced on a refined-non-moving grid.

All those techniques might be explored further to fully understand their differences and their complementarity regarding the measuring methodology. In particular there are many numerical wave models relying on different concepts that cannot be explored in this introduction. Even though, the reader should notice that data issued by a numerical model come with a degree of incertitude. This incertitude is often higher than the ones coming with measures from surface-buoys or altimeters. This statement remains valid even if every models keep improving regarding their actual goal: to reflect the reality.
A general source of incertitude is the parameterisation of the models by the users. Indeed it is sometimes hard to parameterise a model in order to reproduce the entire spectrum of the observable values. This reflects the limits of the underlying equations or their implementation, or both. Incertitude can also comes from the inputs conditions of the model. Obviously if one input conditions (e.g., Wind) does not reflect the reality, then the data produced from the wave-model will have a high incertitude.
These remarks are important since highest biases generally appear when the modelling focuses on extreme phenomena. Therefore one scientific challenge of this thesis resides on the aim of providing sea-states conditions focusing on three points: accuracy, spatial covering and historical period.

Whatever physical or statistical – or both – models we are using, they have a computational cost. Those models generally require huge computational resources, especially when the severity of events implies to quickly compute information.
One can argue that the explosion of the computational capacity of nowadays computers may solve this issue. However as soon as the computational capacity is growing, models accuracy is raised by taking into account new physics or new co-variables that were computationally unreachable before. Hence they are even more resources demanding. From this remark, we can state that alternatives have to be found to face coastal crises. An idea is to be able to pre-compute extreme scenarios and store their results. Those results are then queried when a new crisis arises to identify the degree of severity of the actual crisis.
This concept brings forward some scientific obstacles. In particular, since the action of computing a scenario is a costly process, we can easily understand that the main difficulty of the proposed solution will stay in the selection of the scenario to (pre-)compute. The second concerns could be the definition of a distance

function between the scenarios, which is a mandatory information to perform any recommendation. The third point may concern a more technical aspect: what is the best architecture to store and query the results of the pre-computations.

This thesis is articulated as follows and summed up in Figure 2. Chapter 1 provides an overview of the extreme value theory. Then Chapter 2 introduces some bases about waves and their modelling. It is followed by Chapter 3, which presents the construction and validation of the historical data-set of sea-states conditions (A). From the construction of this data-set, Chapter 4 presents an application of extreme value modelling in the context of long-term scale scientific questionings (B1). This chapter is followed by the Chapter 5 focused on the construction of a semi-parametric method to simulate space-time extreme processes (B2). This method is demonstrated along an event-scale questioning: to model impacts of the wave energy to the coast during severe extreme events. Leveraging the approach of extreme space-time processes simulation, Chapter 6 introduces the proposed principle of pre-computation aiming at helping the decision-making on event-scale coastal hazards assessment (C). Finally the Chapter 7 demonstrates a platform prototype aiming at easing the chaining of weather physical modelling and environmental statistical – extreme – modelling (D). This platform being a technical base of what we consider the next generation of decision helping tools for coastal hazards.

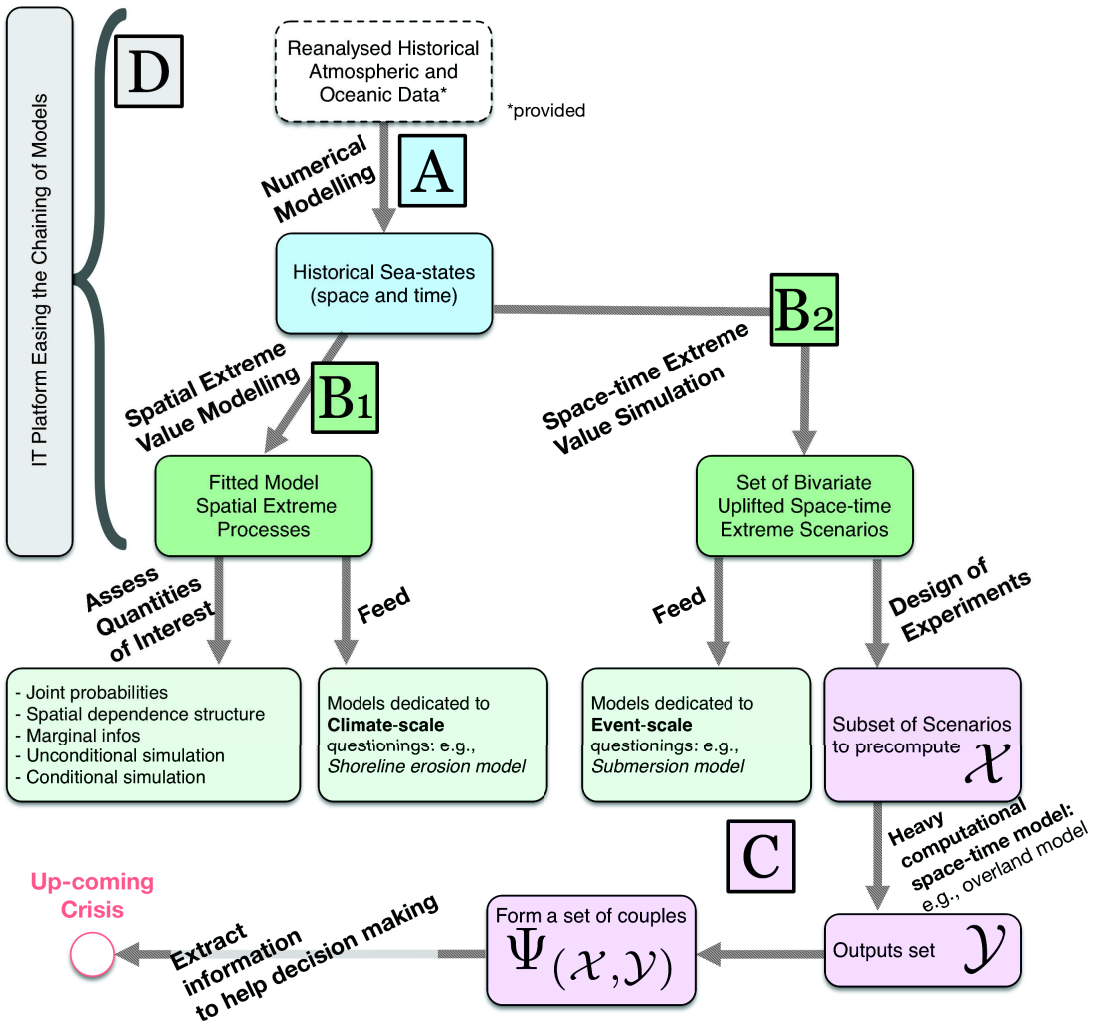


Figure 2: A schematic representation of axes of research of the presented thesis. A) First step is to create a reliable wave data set. B) From this data set, stochastic extreme value modelling are performed. Both in a spatial context (B1), and a space-time context (B2). The latter allowing simulating very extreme scenarios. C) On the basis of the simulated extreme space-time scenarios, a subset is derived. The creation of this subset is based on design of experiments algorithms adapted regarding the questioning. Then, those selected scenarios are precomputed on the targeted (heavy computational) physical model, as for instance an overland flood model. Such computations allow the construction of a set of IO couples, from where information is extracted in case of up-coming crises to ease the decision-making. D) Hydraulic modelling are made user friendly and efficient by the development of an IT platform allowing to access seamlessly huge computational resources and to chain models. This platform might be extended to chain stochastic models as well.

Part I

Overview

Chapter 1

Extreme Value Modelling

Chapter Summary

Stochastic extreme value modelling is identified as a mandatory approach to deal with questionings related to extreme events. This chapter presents an overview of extreme value modelling. The concepts and methods are illustrated with simple examples. Those examples are related to the context of littoral hazard. In particular, the famous sea-state parameter named the significant wave height (Hs) will be largely used [a]. At the end of this chapter, the reader will have all technical and scientific background in extreme value modelling to easily go through the rest of the document and in particular through the statistical applications.

a. At this point the reader should understand that Hs is the most used variable to describe sea-state conditions. Hs is traditionally computed from the mean of the highest third of observed of the waves. This definition is discussed in Chapter 2.

1.1 What is Stochastic Extreme Value Modelling

Endorsed as a real expert of the discipline, Coles (2001) presents the Extreme Value Theory (EVT) as a statistical approach to study the behaviour of a random variable in its extreme realisations. Ones also describe EVT as an approach to obtain information of events appearing in tails of distributions of random variables, as illustrated in Figure 1.1. As a major result, univariate EVT brings a mathematical framework to compute return values associated to long return periods, with confidence intervals. When extended to multivariate contexts, the

main difference (and difficulty) of EVT is to assess the underlying dependence structure of extreme events.
This mathematical approach keeps getting relevance in many disciplines. For instance, insurers adjust their insurance rates regarding the Value at Risk (VaR) – i.e. a return level corresponding to a high quantile – of their financial portfolio (Gilli and Këllezi, 2006). Offshore petroleum stations are designed taking into account wave and wind return-levels of several centenaries to make them durable even when facing extreme storms Li et al. (2013); Ewans and Jonathan (2014). Regarding the littoral hazard context, ones generally use EVT to design ports or seawalls able to face extreme weather conditions.

All these applications need extrapolations of the information from observed levels to unobserved ones. Such extrapolations require statistical models. The following sections present justifications of those models and details on how to model extreme events, from univariate case to more complex cases.

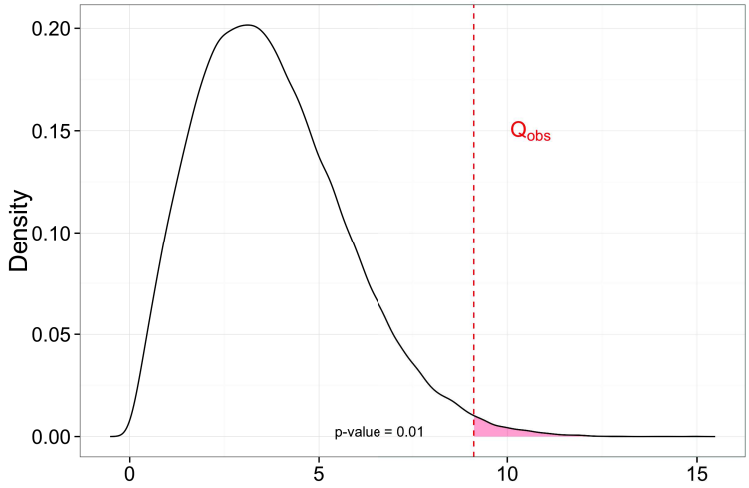


Figure 1.1: A density function of a random variable. EVT is used to extract information from the behaviour of the random variable over a (very) high quantile, as for instance $Q_{obs} = 99\%$ (i.e. the probability p that a random variable exceeds Q_{obs} is $p = 0.01$).

1.2 Univariate Modelling

1.2.1 Maxima Approach

Let us consider a random variable X which may represent the significant wave height (Hs) at a single site s. Let $X_1, X_2, \ldots, X_n$ independent copies of X. For each $1 \leq i \leq n$, X_i corresponds to significant wave heights observed at the day i. Ones are interested in the mean behaviour of X, denoted $\bar{X}_n$ such as $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. The well-known Central Limit Theorem (CLT) de Laplace (1810) states that the distribution of $\bar{X}_n$, correctly normalised, converges to the Normal

law.
In a similar reasoning, EVT introduced by Fisher and Tippett (1928); Gnedenko (1943) is focused on the extreme behaviour of X. It can be seen as the study of $M_n = \max_{1\leq i\leq n}\{X_i\}$. If the full behaviour of X is known then the behaviour of M_n is also known. Unfortunately this is not the case in practice. To face this issue, EVT states that if $(a_n)_{n\geq 0} > 0$ and $(b_n)_{n\geq 0}$ exist such as $(M_n - b_n)/a_n$ converges to a non degenerate distribution $G(\cdot)$, then G is a GEV (Generalised Extreme Value) distribution whose the cumulative distribution function (c.d.f) is given by

$$GEV_{\mu,\sigma,\xi}(x) = \exp\left\{-\left[1+\xi\Big(\frac{x-\mu}{\sigma}\Big)\right]_+^{-\frac{1}{\xi}}\right\}, \tag{1.1}$$

where μ is a location parameter, σ is a scale parameter and ξ is a shape parameter. Also a_+ denotes $\max(a, 0)$, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$. This generalised definition gathers three distribution families: Weibull if $\xi < 0$, Gumbel if $\xi = 0$ and Fréchet if $\xi > 0$ (see Figure 1.2). A noticeable property is that a GEV distribution is invariant while evaluating the distribution of its max – considering eventual affine transformation. This property is named : max-stability.

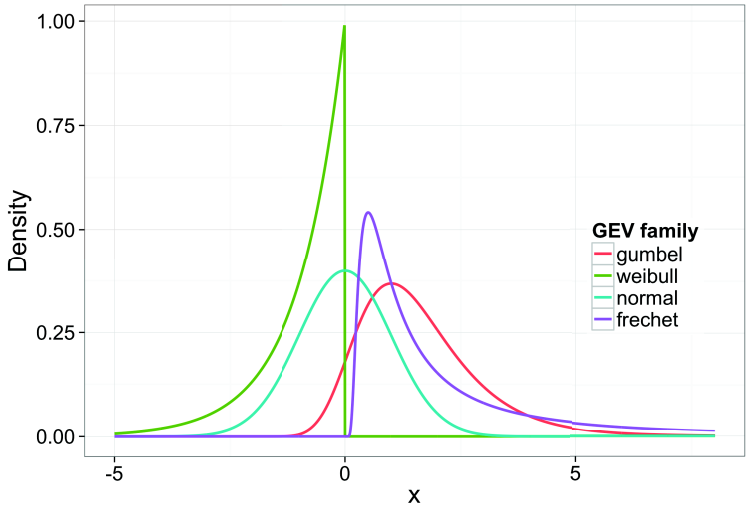


Figure 1.2: Theoretical densities of distributions of the GEV family ($GEV_{\mu,\sigma,\xi}$) and the standard Normal law density $\mathcal{N}(0, 1)$. The Weibull density ($GEV_{-1,1,-1}$) shows a finite endpoint. The Gumbel density ($GEV_{1,1,0}$) accepts to have some realisation in the tail. Finally the Fréchet density ($GEV_{1,1,1}$) has clearly the heaviest tail.

In practice, to fit such a model to data, a first technique consists in splitting the time-series into blocks of the same size. Maxima are extracted from each block. Assuming these maxima follow a GEV distribution, the three parameters of the GEV are estimated to best fit these observations. Several ways of inference can be used to estimate them. Among the inference techniques we can cite the maximum likelihood estimation, which is one of the most used technique. This is the technique used in the sequel.

In the context of univariate extreme value modelling, the most interesting application of a fitted model is to compute the T-block return level and associated confidence intervals, where T denotes the period in year if dealing with yearly maxima (the block size equals to a year). The T-year return value x_T is defined by $GEV_{\mu,\sigma,\xi}(x_T) = 1 - 1/T$. It can be interpreted as the value reached once in average in a period of T years. In other terms, a return value x_T corresponds to the quantile $1 - 1/T$ of the given GEV distribution.

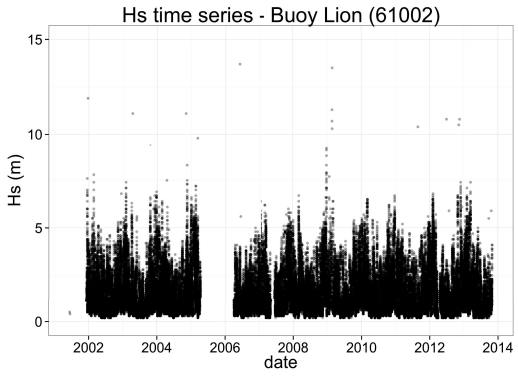
The methodology is presented in the following Example 1.2.1.

Example 1.2.1. *In this example, we process the daily Hs values observed at the Lion's surface-buoy (N* 42.06*, E* 4.64*). The entire time-series (Figure 1.3(a)) stretches December 2001 to November 2013. Yearly maxima are extracted (Figure 1.3(b)) and a GEV model is fitted on. Quality of the fit is evaluated thanks to a Quantile-Quantile plot (Figure 1.3(c)). Such a graph presents the observed values on a plot opposing their position in terms of empirical quantiles against in terms of the theoretical ones, i.e. provided by the fitted model. The more the observations are placed on the diagonal line, the more the fit is good and the extrapolations would be accurate. The return level plot (Figure 1.3(d)) gives what would be the return levels against return periods (i.e. , level reached once in mean for its associated return period). In this last graphic, return values modelled are the red line and can be extrapolated for long return periods. Return periods are given in block size unit, i.e. in years. Dots are the observed events and dashed lines are the associated* 95% *confidence intervals obtained through simulation. With such a fit an estimation of the return level for a return period of 25 years is* $Rl_{25} = 13.60 \pm 1.60$ *(m), where* ± 1.60 *is the confidence interval computed from the asymptotic normality property.*

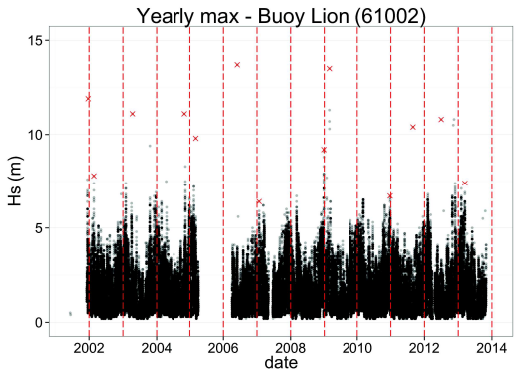
The main drawback of the block maxima approach is that much information is gathered through the time dimension, therefore lost into each block. For instance if we consider annual maxima (as commonly used in environmental applications such as in Example 1.2.1), the time-series may contain two very high and independent values into the same year period. Only the highest is used in the inference process. The second value is lost for inference purpose, even if it contains information about the distribution of extremes.
Alternatives have been proposed to take into account the maximum of information of extreme values. In particular one alternative is to detect the extreme values as exceedances over a high threshold. Details on that method are given in the following Section 1.2.2.

1.2.2 Peaks Over Threshold Approach

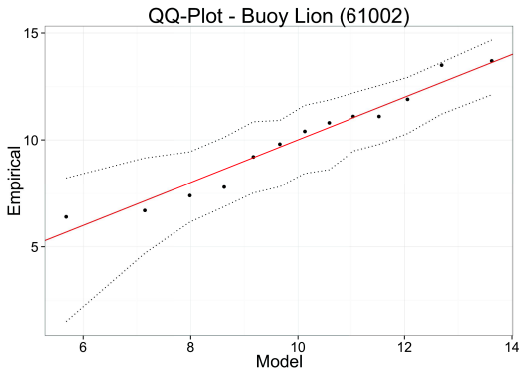
The aim of the peaks over threshold (POT) approach is still the study of the extreme values behaviour. Let us consider the random variable X defined like in the previous section. Pickands (1975) first introduces that for a large enough



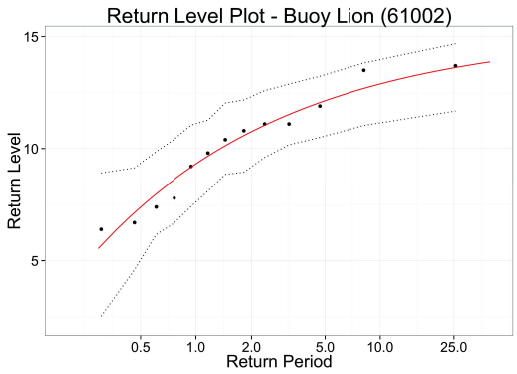
(a) Hs time-series observed at Lion's Buoy (61002).



(b) Extraction of annual maxima.



(c) Quantile quantile plot. Simulated (non-parametric) confidence intervals are given in dashed lines.



(d) Return values plot. Return periods are in block size unit, i.e. years. Simulated (non-parametric) confidence intervals are given in dashed lines.

Figure 1.3: Procedure of a GEV fit with block maxima approach over the significant wave height time-series observed at the Lion's surface buoy.

value u, the c.d.f of $(X-u)$ conditional on $X > u$ can be approximated by:

$$H(x) = 1 - \left(1 + \frac{\gamma x}{\tilde{\sigma}}\right)^{-\frac{1}{\gamma}} \tag{1.2}$$

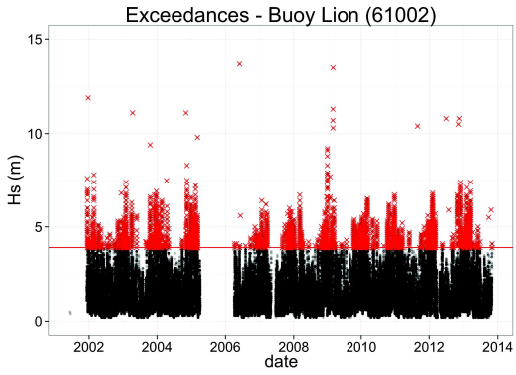
defined on $x : x > 0$ and $(1 + \gamma x/\tilde{\sigma}) > 0$, where $\tilde{\sigma} = \sigma + \gamma(u - \mu)$. The family of distribution corresponding to the Equation 1.2 is called the Generalised Pareto Distribution (GPD).

Maxima M_n correctly normalised converge to a GEV distribution when $n \to \infty$, and excesses over a high thresholds u converge to a GPD one when $u \to x_F$, with $x_F = \sup\{x : F(x) < 1\}$. Both distributions are linked. Indeed the shape parameter γ of the GPD distribution is equal to that of the corresponding GEV distribution ξ.
In practice, one difficulty of the block maxima approach is to select a suitable size of the block in order to do not gather too much information into each block. In the threshold exceedances approach, the main difficulty is to determine what is the threshold to use. Taking a too low threshold means that the asymptotical assumptions are not verified any more and the formalism becomes wrong, implying bias. Taking a too high threshold results in having not enough data to fit the model on, which leads to a high variance.

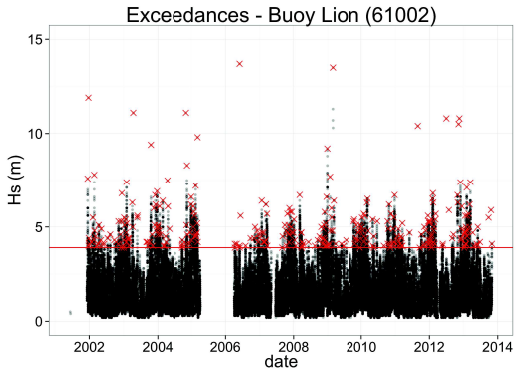
Example 1.2.2. *To illustrate it, let us analyse the data set from Example 1.2.1 with the POT approach. We chose to take the empirical 95% quantile as threshold u. Figure 1.4(a) illustrates the detection of exceedances in red. Since the data set contains hourly Hs observations, observations are not independent and identically distributed (IID) and it is not valid to fit a GPD on all those exceedances (see Subsection 1.2.4). Therefore the data are first declustered to get IID exceedances, as shown in Figure 1.4(b). This leads to use the highest values inside 310 clusters instead of 4631 excesses. Those data are then used to fit a GPD distribution and the resulting fit is evaluated thanks to the quantile-quantile plot in Figure 1.4(c). As for the block maxima approach, the return levels are computed in the goal of extrapolating information to longest time-series than the observed one. With such a fit, an estimation of the return level for a return period of 25 years is $Rl_{25} = 17.50 \pm 6.18$ (m), where ± 6.18 is the confidence interval computed from the asymptotic normality property.*

1.2.3 Others

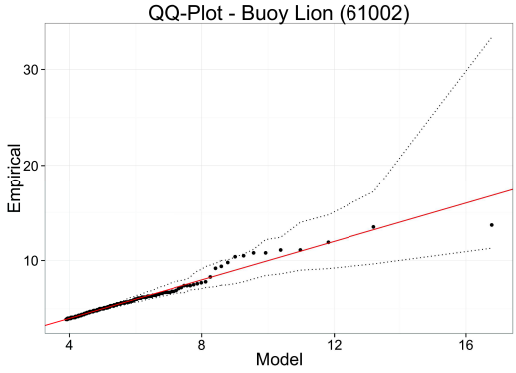
Others approaches exist to model the extreme values. In particular, there is an elegant way of characterizing the extreme value behaviour of a process due to (Pickands, 1971) and known as the Point Process approach. Most notably, it has the advantage of unifying the interpretation of extreme value behaviour from methods introduced before and of enabling a more natural formulation of



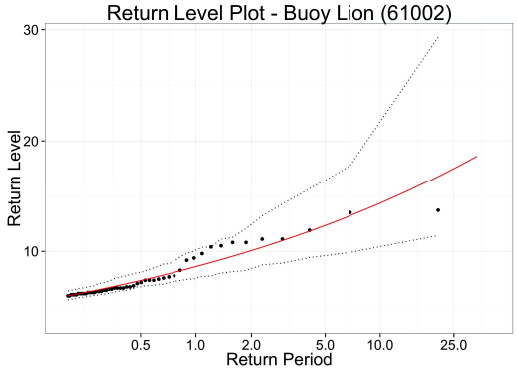
(a) Extracted exceedances over the 95% quantile without declustering at Lion's Buoy (61002).



(b) Extraction of exceedances over the 95% quantile at Lion's Buoy (61002), declustered.



(c) Quantile quantile plot of declustered exceedances. Simulated (non-parametric) confidence intervals are given in dashed lines.



(d) Return values plot of declustered exceedances. Return periods are expressed in years. Simulated (non-parametric) confidence intervals are given in dashed lines.

Figure 1.4: Procedure of a GPD fit with the peaks over threshold approach over the significant wave height time-series observed at the Lion's surface buoy.

non-stationarity in threshold excesses. The theory of Point Process is out of the scope of this document but might be explored in Coles (2001).

1.2.4 Assumptions

The extreme value modelling relies on assumptions of stationarity and independence. However the formalism is flexible enough to deal with non-stationarity or dependent (or both) cases as briefly introduced in Example 1.2.2.
To deal with non-stationary processes, a GEV model can use time-varying expression in the definition of its parameters (e.g., $\mu(t)$,$\sigma(t)$,$\xi(t)$). In case of GPD fits and facing such an issue, an handy solution is to migrate from a fix threshold u to a time-dynamic one $u(t)$. In both cases and in presence of seasonality, one may also choose to restrict the study on a single period of each season.
The other drawback arises from the independence of realisations of the random variable studied. This is a mandatory condition to apply EVT, but in practice it is generally an unrealistic assumption. Hence whether using block maxima approach or peaks over threshold one, those methods have to be able to deal with a short-term dependence between realisations. The basic idea is to generalise the definition of models applied on independent variables to stationary series. When working with block maxima, a theoretical result shows that providing a stationary series $X_1^*, X_2^*, \ldots, X_n^*$ satisfying the $D(u_n)$ condition (Leadbetter, 1983), the maximum of this series (correctly normalised) denoted M_n^* converges to a distribution G_1, if and only if considering the independent series $X_1, X_2, \ldots, X_n^*$ having the same marginal distribution, its maximum denoted M_n (correctly normalised) converges to the distribution G_2 with $G_1^{\theta^*} = G_2$, for a constant $0 < \theta^* \leq 1$. It is remarkable that those distributions are GEV from definition, and θ^* denotes the so-called extremal index. In practice, for a time-series having a weak enough short-term dependency (i.e. $D(u_n)$ condition satisfied), a GEV distribution could be fitted directly on the data. In that case the role of the extremal index is included in the estimation of the GEV parameters.
When working with excesses, similar arguments make the GPD approach valid for the study of stationary series. However by definition of a stationary series excesses tend to cluster, which lead to falsify the initial log-likelihood definition. In practice the user may consider to decluster the time-series. This step permits to avoid considering peaks from the same extreme event, but only take into account the highest value of each cluster, which are independent by definition making the inference process valid.

1.3 Multivariate Modelling

The previous subsections deal with univariate extreme value models. In many environmental questionings extreme values behaviour need to be known at several locations or over an entire space. In the presented context it could be information

required either along a coastline or over a regional coastline area.
Rather than univariate quantities, we now turn attention to multivariate extremes. In particular, the notion of dependence between variables in their extreme realisation is paramount.
For instance, we can be interested in estimating the joint probability of random variables X_{s_i} at two locations s_1 and s_2 to exceed their respective 100-year return value. Considering only univariate approaches applied on each site result in masking the underlying dependence structure between those. If independence is assumed, their joint probability is defined as the product of the margins. If s_1 is nearby s_2, it is trivial to understand that this assumption of independence is generally never reached. Respectively, considering their full dependence would also be a mistake since distant sites could be almost independent even for extreme realisations.

Therefore, to deal with both the dependence structure and the marginal distributions of random variables, multivariate approaches have been developed. Theory of multivariate extremes is well developed, leading to the construction of multivariate extreme models. Such models rely on multivariate approaches, analogues to the presented univariate ones (e.g., block maxima or excess over threshold).

To simplify the understanding of multivariate extremes, let us consider the bivariate case with X_{s_1} and X_{s_2} defining two random variables at two distant locations s_1 and s_2. Each marginal random variable X_{s_i} can be interpreted as the random variable X of the previous univariate section.
We restrict this overview to the extensions of the block maxima and threshold approaches.

1.3.1 Component-wise Maxima

Characterization of MEV model is based on component-wise maxima $M_n = (M_{n,s_1}, M_{n,s_2})$ and can be seen as an extension of the GEV approach. By construction each marginal component of such M_n, correctly normalised, converges to a GEV distribution. For convenience and because the limiting distribution is invariant while transforming marginal distribution, margins are transformed to unit Fréchet distributions, i.e. of c.d.f $F(x) = e^{-1/x}$ for $x > 0$. Placed in the case where marginals are transformed to unit Fréchet and assuming that M_n/n converges to a non-degenerate distribution, then Resnick (1987) states that such a MEV distribution is characterized by :

$$G(x, y) = \exp(-V(x, y)), x > 0, y > 0, \tag{1.3}$$

where

$$V(x,y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega), \tag{1.4}$$

and H is a distribution defined on $[0,1]$ verifying the constraint $\int_0^1 \omega dH(\omega) = 1/2$. As in the univariate case, $G(\cdot)$ is max-stable. Since there is only one constraint, the class of the possible limits is infinite. One way to proceed is to work on parametric sub-families of distributions.

There are many parametric families of models to characterize such distribution G due to the presence of H. Among them we can cite Gumbel (1960); Tawn (1988, 1990); Coles and Tawn (1991). For instance, still in the bivariate case, the logistic-Gumbel model is defined as

$$G(x,y) = \exp\left\{-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)\right\}^{\alpha}, \tag{1.5}$$

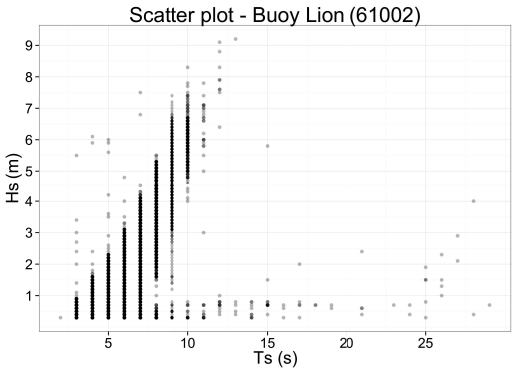
with $x > 0$, $y > 0$ and $0 < \alpha \leq 1$. The parameter α defines here the dependence in the extreme of the two random variables with the total dependence case ($\alpha \to 0$) and the independence one ($\alpha = 1$).

In practice, one way to proceed is to assume that the component-wise maxima vector converges to a distribution issued by one of the multivariate models. Hence a set of multivariate models could be fitted to the data by conventional estimation techniques. Fitted models are then sorted out in terms of goodness of fit. The selection of the best model is performed via the use of statistical criterion, as the Akaike Information Criterion (AIC) [1]. The principal outcomes of multivariate models are for instance the possibility to compute joint probabilities considering the underlying dependence structure or to study the dependence structure itself (e.g., presence of anisotropy). Marginals analyses are still available as well.

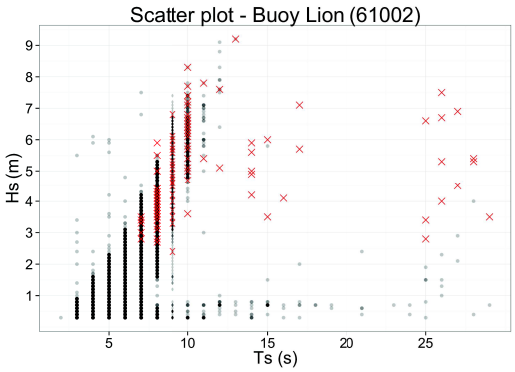
Example 1.3.1. *To illustrate such bivariate modelling, let us consider the bivariate process of both significant wave height (Hs) and mean wave period (Ts) time-series, which are hourly observed at the Lion's surface buoy. From expert advice and to avoid measuring errors, we filter out couple of values where $Hs < 0.2m$. Surface buoys can indeed record high Ts values when simultaneous Hs is too low but in that case we are no more looking at the same physical waves (see next Chapter). As in the univariate block maxima approach, the first step is to detect maxima. Monthly maxima are extracted and are represented in Figure 1.5(b). Something remarkable is that most of the maxima are not directly observed, but are the combination of both Hs and Ts maxima within a month. We choose to fit the eight bivariate models implemented in the* ***evd*** *R Package* [2]*: logistic (Gumbel, 1960), asymmetric-logistic (Tawn, 1990), Husler-Reiss (Hüsler and Reiss, 1989),*

1. 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.

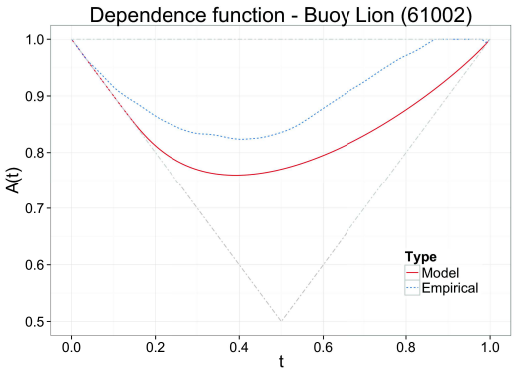
2. https://cran.r-project.org/web/packages/evd/evd.pdf.



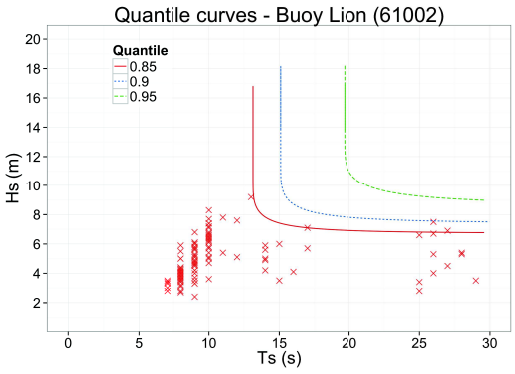
(a) Scatter plot of Hs/Ts times series at Lion's Buoy (61002).



(b) Extraction of bivariate montlhy maxima.



(c) Modelled dependence function (straight line) versus empirical one.



(d) Quantile curves plot representing. Curves represent respectively the quantiles 85%, 90% and 95%.

Figure 1.5: Bivariate extreme value modelling using bilogistic model for Hs and Ts at Lion's surface buoy.

negative logistic (Galambos, 1975), asymmetric negative logistic (Joe, 1990), bilogistic (Smith et al., 1990), negative-bilogistic (Coles and Tawn, 1994), Coles-Tawn (Coles and Tawn, 1991) and the symmetric mixed model (Tawn, 1988). Once fitted, we compare their respective AIC (Akaike Information Criterion) to select the best fitted model. According to this criterion, the best model is the one having the lowest AIC. Indeed, the AIC penalises the likelihood of the model by its number of parameters in order to promote the parsimonious models. The bilogistic fit appears to be the better on the presented data. The Pickands function $A(t)$ *represents the estimated dependence function and lies in the interval* $[0, 1]$. *This function measures the dependence between the two random variables. For the complete dependence* $A(t) = 1$ *and* $A(t) = 0$ *for the full independence (Beirlant et al., 2004). The empirical (estimated) dependence is compared to the modelled dependence in Figure 1.5(c).*
The skewness of the empirical curve reinforces the idea of using a model like the bilogistic one, allowing an asymmetry in the characterisation of the bivariate dependence structure. Another observation is that the model reduces the dependence between the two variables, even if it remains important.
Finally we observe the bivariate quantile curves in Figure 1.5(d). Each quantile curve q_i *issued by the bivariate model delimits the values for which the probability of being simultaneously lower than those maxima-pair is* q_i.

1.3.2 Threshold Excess Model

The univariate peaks-over-threshold approach is extended by two main approaches to the multivariate (here bivariate) context, reviewed in Bacro and Gaetan (2014).
The first approach is to approximate the bivariate distribution $F(x, y)$ when $x > u_x$ and $y > u_y$ for large enough thresholds u_x and u_y, providing that marginal distributions of F are of the form of (1.2).
If those margins are transformed to standard Fréchet distributions, it is possible to show that for large enough u_x and u_y

$$F(x, y) \approx G(x, y) = \exp\{-V(\tilde{x}, \tilde{y})\}, \; x > u_x, \; y > u_y, \tag{1.6}$$

with $\tilde{x}$ and $\tilde{y}$ defined in terms of x and y transformed to standard Fréchet scale. One difficulty to deal with such model is that inference is complicated. Indeed, (1.6) holds only when both x and y are above their marginal threshold. It means that in the other regions it is necessary to censor the likelihood component, which is well detailed for instance in Coles (2001) pages 155-156.
Rootzén and Tajvidi (2006) worked on the other approach. They consider a bivariate distribution of large values as well, but when at least one of the component is large. From their result, we can assume that such distribution is a generalisation of the univariate Generalised Pareto distribution, suggesting the following approximation:

For large thresholds u_x and u_y as

$$P\left(\frac{X-u_x}{\sigma_{u_x}} \leq z_x, \frac{Y-u_y}{\sigma_{u_y}} \leq z_y \,\middle|\, X > u_x \text{ or } Y > u_y\right) \approx H(z_x, z_y), \tag{1.7}$$

with $H(\cdot,\cdot)$ the bivariate generalised Pareto distribution is of the form

$$\begin{aligned} H(z_x, z_y) &= -\frac{1}{\log(G(0,0))} \\ &\times \log \frac{G(z_x, z_y)}{G(\min((z_x, z_y), (0,0))} \end{aligned} \tag{1.8}$$

for all (z_x, z_y) and with G max-stable. The result of Rootzén and Tajvidi (2006) is actually valid to a dimension greater than two.

1.3.3 Limits

In the context of assessing coastal hazards, many questionings are related to entire coastlines or to entire coastal regions. To address those questionings with multivariate approaches could be quite delicate. A complex dependence structure (number of sites >> 3) is hard to handle due to the lack of flexibility of these models. Another drawback of such approaches is that they are restricted to provide information only on the sites of observations. The following section presents one solution to avoid such constraints.

1.4 Spatial modelling

1.4.1 Max-stable Processes

To overtake the multivariate induced drawbacks, a continuous spatial modelling is introduced by de Haan (1984) with a new theory extending both the GEV and MEV formalisms: max-stable processes. Let $Z_i(\cdot)$, $i = \{1, \ldots, n\}$ be n independent copies of a spatial process of extremes. Let $\mathbb{C}(\chi)$ be the space of continuous real functions f defined on $\chi \subset \mathbb{R}^d$. de Haan (1984) states that the random process $\{Z(x),\ x \in \chi\}$ is max-stable if $a_n(x) > 0$ and $b_n(x) \in \mathbb{R}$ defined on $\mathbb{C}(\chi)$ exist such that,

$$\left\{\max_{i=1,\ldots,n} \frac{Z_i(x) - b_n(x)}{a_n(x)}, x \in \chi\right\} \stackrel{\mathcal{L}}{=} \{Z(x), x \in \chi\}. \tag{1.9}$$

As a consequence of this definition any n-dimensional marginal distribution of $Z(\cdot)$ satisfies the max-stability property. More specifically they are MEV. Hence, univariate marginal distributions are GEV and there is no loss of generality in assuming that they can always be unit Fréchet margins after rescaling and shifting.

Let $Y_i(\cdot)$, $i = \{1, \ldots, n\}$ be n independent copies of a spatial process. de Haan (1984) shows that if there exist normalizing functions $a_n(x) > 0$ and $b_n(x) \in \mathbb{R}$ such that

$$\max_{i=1,\ldots,n} \frac{Y_i(x) - b_n(x)}{a_n(x)}$$

converges when $x \in \chi$ and $n \to \infty$ to a non degenerated process $Z(\cdot)$, then the limit process $Z(\cdot)$ belongs to the max-stable processes class.

In the literature, it exists several models 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. To better understand those differences, let us describe one model of each approach in the following paragraphs.

Smith (1990)

The first model described is known as the *Gaussian extreme value process* (Smith, 1990). It is often used for extreme rainfall modelling.
Without loss of generality and because we are dealing with spatial max-stable processes, let us introduce the spaces $S \subset \mathbb{R}^2$ and $\chi \subset \mathbb{R}^2$ used in the sequel.
Let us consider $\{(\xi_i, s_i), i \geq 1\}$ a Poisson process on $(0, \infty) \times S$, with intensity measure of $\xi^{-2} d\xi \times \nu(ds)$, where ν is a measure defined on S. Moreover let us consider $\{f(s, x), s \in S, x \in \chi\}$ a non-negative function, then the so-called storm model, is defined by

$$Z(x) = \max_{i \geq 1} \xi_i f(s_i, x), x \in \chi. \tag{1.10}$$

Under the constraint $\int_S f(s, x)\nu(ds) = 1$, for all $x \in \chi$, $Z(\cdot)$ is a max-stable process with unit Fréchet margins that means a simple max-stable process.
As a basic interpretation, s_i is seen as the centre of the i^{th}-storm situated in the space S and ν their distribution. Each ξ_i represents the intensity of the storm and $\xi_i f(s_i, x)$ the total amount of rainfall for the storm centred on s_i. Finally the max operator allows to determine the maximum rain felt over n independent storms. One example of max-stable process simulated by such a model is illustrated in Figure 1.6.

Schlather (2002)

The former model is slightly different from the class often attributed to Schlather (2002), which is in reality a special case of a more general representation stated by Penrose (1992). In this representation, Schlather (2002) considers max-stable processes with a random shape instead of being defined from a deterministic function.
Let $\{\xi_i, i \geq 1\}$ denote the points of a Poisson process on $(0, \infty)$ with intensity

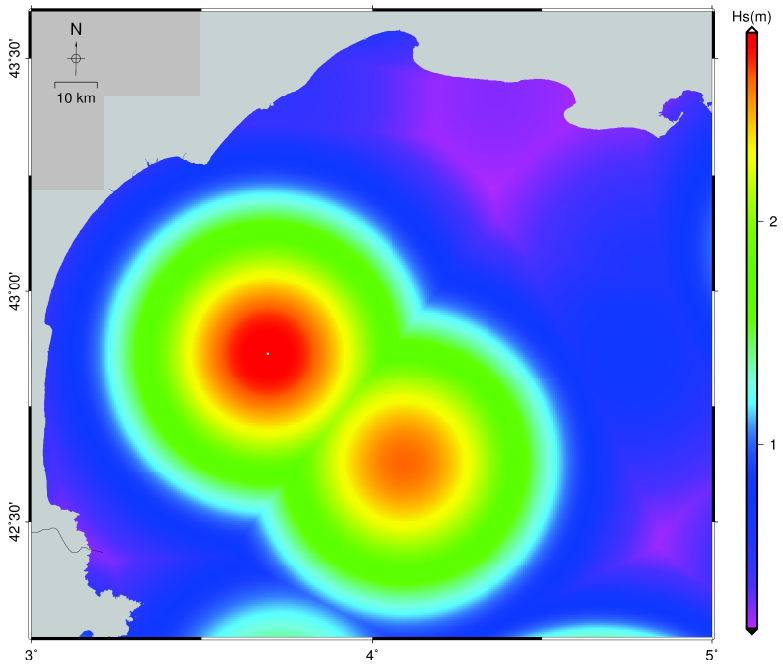


Figure 1.6: A random max-stable process simulated from a Smith's max-stable model. Many possibilities of the function $f(\cdot)$ are available. Functions $f(\cdot)$ are probability density functions defined on $\chi \subset \mathbb{R}^d$, with $d = 2$ in this spatial context. In this example, we chose the multivariate Gaussian density function having the covariance matrix $\Sigma = \begin{pmatrix} 1/12 & 0 \\ 0 & 1/12 \end{pmatrix}$.

measure of $\xi^{-2}d\xi$ and $\{W_i(.)\}_{i\geq 1}$ be independent copies of $W(.)$ a stationary process on $\mathbb{R}^2$, with $E(W(x)) = 1$ for all $x \in \chi$. Then the random process

$$Z(x) = \max_{i\geq 1} \xi_i W_i(x), x \in \chi, \tag{1.11}$$

is a simple max-stable process (Schlather, 2002).

From this definition, Schlather (2002) defines the Extremal Gaussian process as $Z(\cdot)$ with $W_i(x) = \sqrt{2\pi}\max\{0, \varepsilon_i(s)\}$, with $\varepsilon_i(s)$ are IID stationary Gaussian process. This model has a simple interpretation too: the ξW are spatial events having the same dependence structure. They differ in their magnitude ξ. The shape of the events may vary if the process W allows it. One example of max-stable process simulated by such a model is illustrated in Figure 1.7.

Extremal Coefficient

One interest of modelling spatial extremes is on the understanding of the underlying dependence structure of those processes. Generally the dependence between two random variables uses to be computed by means of their correlation. When focusing on extremes, this is inappropriate since such a method measures the dependence about the mean values. Additionally some extreme value distributions do not have moment of order two, making such a computation impossible.

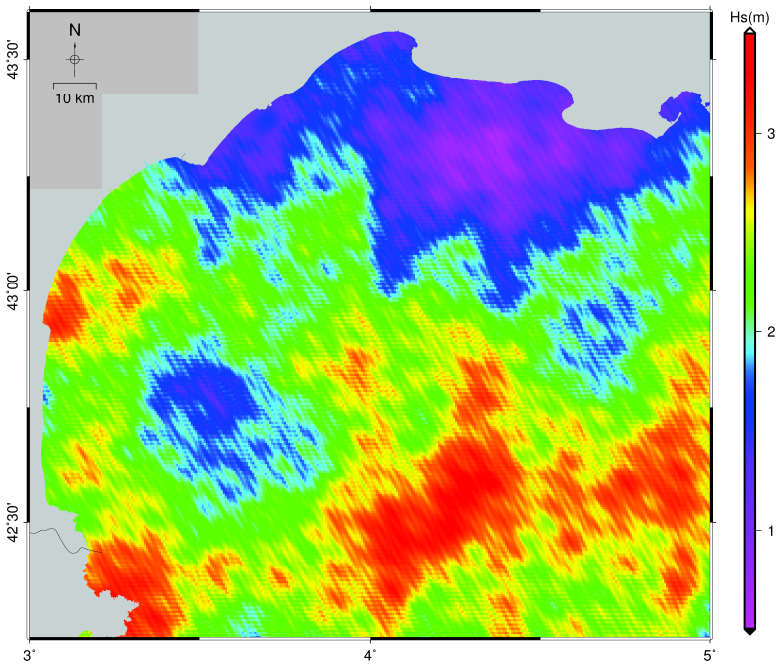


Figure 1.7: A random max-stable process simulated from a Schlather's max-stable model with an arbitrary dependence structure based on the exponential correlation function.

A simple measure to assess the dependence between random variables having extreme value distributions is the extremal coefficient (Smith, 1990).
Let us define a vector $(X_1, \ldots, X_N)$ of N dependent random variables that without loss of generality are assumed to be transformed to the unit Fréchet scale. If $(X_1, \ldots, X_N)$ has a MEV distribution,

$$P(X_1 \leq x, X_2 \leq x, \ldots, X_N \leq x) = \exp\left(-\frac{\theta}{x}\right), \tag{1.12}$$

where θ is the extremal coefficient of $(X_1, X_2, \ldots, X_N)$. Value of θ ranges between 1 and N the number of random variables into the multivariate vector. Limiting case $\theta = 1$ means that variables are strictly dependent, whereas $\theta = N$ represents the full independence.
For a max-stable process $Z(\cdot)$ we focus on the bivariate extremal coefficient function (Schlather and Tawn, 2003) $\theta(\cdot)$, which is given by

$$P(Z(s) \leq z, Z(s+h) \leq z) = \exp\left(-\frac{\theta(h)}{z}\right). \tag{1.13}$$

In the context of max-stable models, the bivariate representation of the extremal coefficient function is privileged because they can be simply derived from their (known) bivariate distribution. As recalled in Bacro and Gaetan (2012) we get for the former presented models:

— Gaussian extreme value process (Smith, 1990): $\theta(h) = 2\phi\left(\frac{\sqrt{h'\Sigma^{-1}h}}{2}\right)$, where

ϕ is the standard normal distribution and Σ the associated covariance matrix;

— Extremal Gaussian process (Schlather, 2002): $\theta(h) = 1 + \sqrt{(1-\rho(h))/2}$, where ρ is the correlation function of the stationary Gaussian process.

Estimation of Extremal Coefficient

Bacro and Gaetan (2012) states that for a single value $\theta(h)$ several estimators are defined (Smith, 1990; Cooley et al., 2006; Bel et al., 2008). Some of them, largely accepted in the literature, are introduced hereafter.
Let us consider K independent copies of the bivariate vector $Z_h = (Z(s), Z(s+h))'$ having unit Fréchet margins and denoted $Z_h^{(k)} = \left(Z_{h,1}^{(k)}, Z_{h,2}^{(k)}\right)'$, $k = 1, \ldots, K$. A first estimator comes from the introduction of a way to characterise the spatial bivariate structure of a process $Z(\cdot)$ by means of the madogram

$$\nu(h) = \frac{1}{2}\mathbb{E}|Z(s+h) - Z(s)|. \tag{1.14}$$

Cooley et al. (2006) considers the madogram for transformed max-stable process $F(Z(\cdot))$ with $F(\cdot)$ the unit Fréchet distribution, known as the Fmadogram. In that case, Cooley et al. (2006) shows that

$$\nu_{F(Z)}(h) = \frac{1}{2}\mathbb{E}|F(Z(s+h)) - F(Z(s))| = \frac{1}{2}\frac{\theta(h)-1}{\theta(h)+1}, \tag{1.15}$$

which leads to the estimator:

$$\hat{\theta}(h) = \frac{1 + 2\hat{\nu}_{F(Z)}(h)}{1 - 2\hat{\nu}_{F(Z)}(h)}, \tag{1.16}$$

where $\hat{\nu}$ is an empirical estimator.
Since $1/Z(s)$ has an exponential distribution and $\min(1/(Z(s), 1/Z(s+h))$ has an exponential distribution with rate $\theta(h)$, a second estimator of $\theta(h)$ is naturally given by Smith (1990) as

$$\hat{\theta}(h) = K \Bigg/ \sum_{k=1}^{K} \frac{1}{\max\left(Z_{h,1}^{(k)}, Z_{h,2}^{(k)}\right)}. \tag{1.17}$$

It appears that in the context of threshold-based extreme value methods, where realisations above a high threshold are considered as extreme ones, the availability and interpretation of the extremal coefficient function θ remains the same. Without loss of generality, let us consider IID random variables $Y^{(1)}, \ldots, Y^{(N)}$ defined with unit Fréchet distribution as before.
Let us then consider predetermined thresholds vectors $(u_j^{(1)}, \ldots, u_j^{(N)})$ and M IID random vectors $(Y_j^{(1)}, \ldots, Y_j^{(N)})$, $1 \leq j \leq M$, where each $Y_j^{(k)}$ is observed only if

$Y_j^{(k)} > u_j^{(k)}$; otherwise $Y_j^{(k)}$ is censored at $u_j^{(k)}$.
In this context, Smith in Caires et al. (2011) defines the following estimator of the extremal coefficient function θ as:

$$\hat{\theta} = m \Big/ \sum_{j=1}^{M} \frac{1}{\max(Y_j, u_j)} , \tag{1.18}$$

where Y_j and u_j are defined as $\max\left(Y_j^{(1)}, \ldots, Y_j^{(N)}\right)$ and $\max\left(u_j^{(1)}, \ldots, u_j^{(N)}\right)$, respectively; m is the number of excesses $Y_j > u_j$.

Simulation

The simulation of max-stable processes is divided in two categories: unconditional and conditional simulation (Bacro and Gaetan, 2012; Ribatet, 2013). Unconditional simulations stand for processes generated randomly, whereas conditional simulations stand for processes where some observed data condition the simulation. The later is particularly used when a quantity of interest (e.g., quantiles, joint-probabilities of exceedances, ...) is needed at other locations of the already known (observed) sites.

Unconditional Simulation From (1.11), the simulation of a max-stable process $Z(\cdot)$ would consist in computing the point-wise maxima over an infinite number of random processes. Fortunately, results show that in practice only a few numbers of realizations can be generated to represent $Z(\cdot)$, making the simulation of a max-stable process accessible.
Few methods exist to simulate max-stable processes in the literature. Among those, the one relying on the spectral representation (1.11) of max-stable processes and which is very efficient is described in the following algorithm 1 and illustrated in Figure 1.8

Algorithm 1: Unconditional simulation of max-stable processes having unit Fréchet margins.

Input : An upper bound $C > 0$, a stochastic stationary process $W(\cdot)$.
Output: One simulated max-stable process.

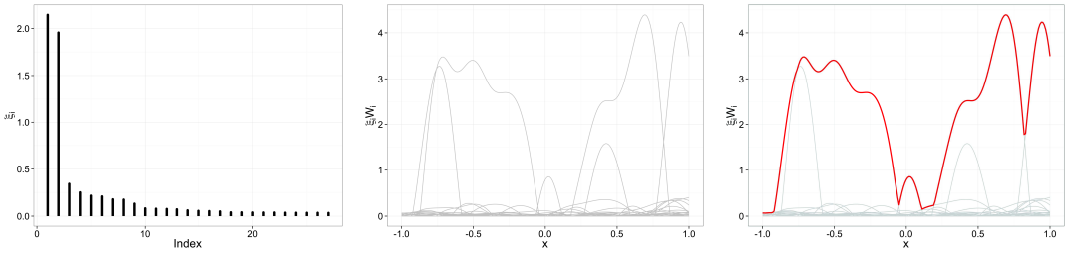
hasConverged $\leftarrow$ false, $k \leftarrow 0$, $T_0 \leftarrow 0$
while *!hasConverged* **do**
  $k \leftarrow k + 1$
  $E_k \sim \text{Exp}(1)$
  $T_k \leftarrow T_{k-1} + E_k$
  $\xi_k \leftarrow T_k^{-1}$
  $W_k(\cdot) \sim W(\cdot)$
  if $C\xi_k \leq \max_{1 \leq i \leq k} \xi_i W_i(s)$ **then**
    hasConverged $\leftarrow$ true
return $Z(s) = \max_{1 \leq i \leq k} \xi_i W_i(s)$

$(T_i)_{i \geq 1}$ is by construction a Poisson process on $(0, \infty)$ with intensity measure dt and $(\xi_i)_{i \geq 1}$ is a Poisson process on $(0, \infty)$ with the required intensity measure $x^{-2}dx$. Since ξ_k decreases to 0 as $k \to \infty$, the algorithm converges. Here we supposed that the stochastic process $W(\cdot)$ is uniformly bounded by a finite and positive constant C. An approximate algorithm has to be introduced if this condition is not fulfilled. In that case, choosing $P(W(s) > C)$ small enough does show good performances.
More recently Lantuéjoul et al. (2011) introduced the exact simulation of the Poisson storm process, exploiting specific properties of the random storms. However for some other max-stable models not introduced in previous section (e.g., Brown-Resnick), performing the simulation of processes is much more difficult. In that case, more sophisticated procedures have to be considered to simulate the processes (Oesting et al., 2012). Such a work is far from the main objectives of this thesis. Hence those processes will not be used in the applications.

Conditional Simulation Conditional simulation of a max-stable process is as useful as difficult to obtain.
If we consider $\mathbf{x} = (x_1, \ldots, x_K)$, $x_k \in \chi$, a vector of locations and $\mathbf{z} = (z_1, \ldots, z_K)$ the expected values, the aim is to sample from

$$Z(x) | \{Z(x_1) = z_1, \ldots, Z(x_K) = z_K\}, \quad x \in \chi, \tag{1.19}$$

where Z is as simple max-stable process on χ, i.e. non degenerate with unit Fréchet margins (de Haan and Pereira, 2006).
Only recently Wang and Stoev (2011) introduced a solution to construct such a conditional process for max-linear processes. This work was followed by Dombry et al. (2012); Dombry and Eyi-Minko (2013) who worked as well on a method to construct conditional process but in a less restrictive case. They showed that is



(a) Realisation of the Poisson process ξ.

(b) Simulation of bounded Gaussian processes amplified by the Poisson process: $\xi_i \times W_i$.

(c) Red line is the returned max-stable process .

Figure 1.8: Illustration of the procedure to simulate a one-dimensional Gaussian extreme value (max-stable) process.

possible to decompose the conditional max-stable process in two parts. The first one is a set of random functions, called extremal functions, contributing to the conditioning event $Z(\mathbf{x}) = \mathbf{z}$. The second part is a set of so-called sub-extremal functions, which are random functions that do not contribute at the conditioning points $\mathbf{x}$ but may contribute at other locations. The combination of simulations from those two sets of random functions makes a conditional max-stable process. Recently, Lantuéjoul and Bel (2014) introduced a new algorithm improving significantly the performances of the conditional simulations, allowing the simulation of max-stable processes conditioned up to hundred of points.

Applications

Even if the mathematical framework to deal with spatial extreme values have been discovered and justified years ago, only since few years such theory have been applied on several topics and in particular in environmental contexts as presented in Toulemonde et al. (2015).
Performances of such stochastic modelling have been shown in several applications over various topics. Among others, one reference in the extreme value community for the use of max-stable processes is Blanchet and Davison (2011) who study the heavy snow events in Alps. Gaume et al. (2013) work also on the same topic with max-stable processes in a very pedagogic way. Similarly, Davison and Gholamrezaee (2011) focus on the study of extreme heat-waves in Switzerland. Oesting et al. (2013) use conditional simulation of bi-variate max-stable processes to model extreme wind gusts in Germany by detecting the dependence structure between the forecast events and their observations. Also performing conditional simulation of max-stable processes from weather forecast models, Bechler et al. (2015) are interested in the capacity of downscaling extreme values for floods events in south of France.

More related to the topic of this thesis, Raillard et al. (2013) model extreme significant wave heights by fitting marginally a max-stable process along the time dimension. From a more technical point of view, Wadsworth and Tawn (2012) work notably on evaluating the limiting asymptotical dependence of the max-stable processes and propose alternatives validated on same kind of data-set.
Closely related to the max-stable processes through the use of extreme value theory, Weiss et al. (2014) develop an interesting approach relying on the formation of homogeneous region for regional frequency analysis of significant wave heights. Ewans and Jonathan (2014) develop conditional models to assess the extreme significant wave heights as well.
This short review is far from being exhaustive but demonstrates the global and recent efforts and interests of mathematical research aiming to modelling extreme events in those environmental contexts.

1.4.2 Generalised Pareto Process

Generalised Pareto process is the natural extension of the univariate and multivariate generalised Pareto distribution. For the multivariate contexts, Rootzén and Tajvidi (2006) showed that a generalised Pareto distribution is reached in limit when considering normalised peaks-over-threshold conditioned to the fact that at least one component is an exceedance (see Section 1.3.2). Ferreira and de Haan (2014) introduced the framework of generalised Pareto processes by extending this multivariate approach to infinite dimensional spaces. To do so they take into account the normalised and conditioned processes by its supremum over the space. In particular they showed that for a stochastic process $\{X(t)\}_{t\in T}$, if it exists continuous normalising functions $\{a_n \geq 0\}_{\{n\geq 1\}}$ and $\{b_n\}_{\{n\geq 1\}}$ such that

$$\frac{X - b_n}{a_n} \mid \left\{ \sup_{t\in T} \frac{X(t) - b_n(t)}{a_n(t)} > 0 \right\} \tag{1.20}$$

converges weakly in the space of continuous function, as $n \to \infty$, then the limit is a generalised Pareto process. Then Dombry and Ribatet (2013) leverage the approach by considering conditional events characterized through a continuous and homogeneous risk function $\ell(\cdot)$, leading to the characterisation of the ℓ-Pareto process. More recently Thibaud and Opitz (2015) work on a new class of generalised Pareto processes, the elliptical-ℓ-Pareto process. They present it in a more practical approach. Most notably, they provide efficient algorithms for the inference of their model and to realise conditional and unconditional simulations. To our knowledge, this is the most advanced practical demonstration of the generalised Pareto processes on real data set.
Regarding the global scope of the thesis that is focused on the development of a full methodology aiming at easing the assessment of coastal hazards, generalised Pareto processes are not directly considered in the following applications, but are somehow linked (see Section 5.3.4). Besides this remark, one may state that it

would be a valuable add to work with such processes, most notably in terms of physical interpretation of the simulated events, that is more natural. Such an application represents actually the purpose of a future work.

1.5 Space-time Modelling

Studying and simulating space-time extreme events is mandatory as soon as questionings considered are based on an event-scale, i.e. the evolution of extreme event through time of its realisation is of importance.
Either max-stable processes or GPD processes are defined on $\mathbb{R}^d$, with $d \geq 1$. If in the previous section $d = 2$, there is no theoretical restriction to rise the value of d and be able to model space-time extreme processes with such approaches. However, to deal with higher dimension problems implies (notably) to deal with much more complex and resource demanding inference processes.
The promising works of Davis et al. (2013a,b); Huser and Davison (2014); Nicolet et al. (2015) are ones of the very few and recent applications to the space-time context. Due to the scarcity of such applications in the literature, ones can still challenge the flexibility of these models to take into account complex space-time dependence structure. Another remark is that the interpretation of a simulated process from such models might be confused from a physical point of view.
Beside full-parametric models, alternatives exist to simulate such space-time process. For instance, the semi-parametric presented in Caires et al. (2011); Groeneweg et al. (2012). From our knowledge, such an implementation is highly promising as well.

1.6 Conclusion

Along this overview we have seen how to statistically model extreme events, from a univariate context to higher-dimensional context. Justified from asymptotical results, most of those methods becomes in practice widely used and accepted to tackle extreme value modelling.

Recalling the initial questionings, we are looking for information of extreme waves not only in a single location but along entire coastlines or within entire coastal regions. To deal with such a complex spatial physical phenomenon, a good practice is to take into account the underlying spatial dependence structure to limit the underestimation or overestimation of extreme realisations.
The presented approaches in Section 1.4 are particularly relevant to model them. However some improvements can still be realised. In particular models may represent much easily complex dependence structures, as the ones presenting asymptotical independence.
In case that the time evolution of an extreme event matters for the comprehension of the phenomenon, space-time approaches have to be considered. To deal

with such space-time dimension problem is one of the challenge of this thesis and is particularly discussed in Chapter 5.
Beyond the mathematical definition of models, any extreme values study – as generally in stochastic modelling – rests on the availability of a good quality data set. A good quality data set for such extreme modelling means a data set containing accurate and (regular) long time-series. Adding the spatial context on top of it, observations as to be spatially well represented.

This thesis is related to the assessment of coastal hazards in which extreme waves are highly implied. In this sense, we are likely to apply the presented Extreme Value Theory and its extensions to such an environmental phenomenon. Before that, let us review in the following chapter the bases of sea-waves.

Chapter 2

Waves, from Physics to Numerical Modelling.

Chapter Summary

This chapter focuses on the main data utilised in this thesis: the waves. What is the motivation of using such data? Waves can be assimilated to the main amount of energy in coastal areas. Since this PhD is likely to tackle questionings related to littoral hazards (in extreme conditions), waves are considered as the mandatory variable to consider.
In this chapter we overview the physic of waves and we discuss how such data can be observed. Observation methods are particularly important since we know from the previous chapter that accurate and long historical data sets are required to perform extreme value analyses. Among the different sources of data, the numerical modelling of waves is presented in detail.

2.1 Introduction to Waves

2.1.1 Generalities

Surfers are particularly familiar with waves: they ride them using their dynamic to move onto the water and enjoy the pleasure of water-skiing. A good surfer uses to ride waves ranging from 0.5 m to 5 m. When they fall while riding a 0.5 m, there is no consequence. On the other hand, falling from a 5 m wave can be severely damaging, if not mortal. Both the motion of a surfer and the

potential injuries issued by a fall are a slight overview of the embedded energy of a wave. When writing this thesis, G. McNamara has set the world record in 2013, riding a wave of about 30 meters, almost the highest recorded waves (Liu et al., 2008). This is straightforward to consider that such a wave is source of a tremendous energy, and its impact to any asset (coastline, petroleum platform, port, embankment) might be strongly damaging.
Naturally waves of all kind are of great interest and their study is focused on their **creation**, **propagation** and **dissipation**.
Waves can be classified in function of their **creation** factor source as illustrated in Figure 2.1 (Munk, 1950):

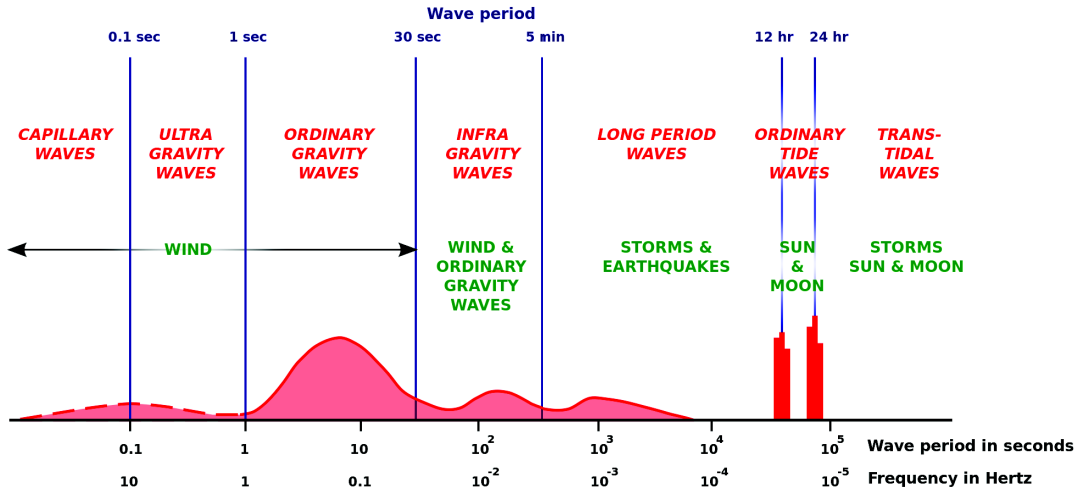


Figure 2.1: Classification of the spectrum of ocean waves according to wave period and factor source of creation adapted from Munk (1950).

From Figure 2.1 we can see the most observed waves – and waves that we will consider in the next chapters – are the ones directly or indirectly generated by the wind. Their period T ranges from 0.1 s to 30 s and are generally called gravity waves. This name stems from the characteristic of their **propagation**, mainly ruled by the gravity force. As soon as the curvature of the wave is too important, the surface tension (surface capillary force) governs the propagation. This tension is only important for the small period waves, also called capillary-waves. In the opposite side regarding wave-periods, waves generated by geophysical processes (e.g., earthquake) have periods T reaching several hours as it has unfortunately been observed recently in Indonesia (Lavigne et al., 2007) or Japan (Fujii et al., 2011). Tide waves that are induced from astrophysical processes (e.g., Moon) have even longer wave-periods.
When looking at the sea-surface of oceans, all waves are gathered. Fortunately it is relatively simple to filter such a signal and distinguish the long wave-period processes against the gravity waves. Capillarity waves are however harder to dissociate but it is artificially done to simplify computations.

Waves **propagation** is well known thanks to the work of Stokes, Airy, Rayleigh and Boussinesq in the 19th century. The emergence of new technologies and in particular numerical modelling (relying on the theories of cited scientists) has largely contributed to the recent evolution on the understanding and forecasting of the main characteristics of sea-states conditions. However the **generation** and **dissipation** of waves are still imperfectly understood.

2.1.2 Wave Analyses: Wave-wave versus Spectral

Until the end of the World-War II, waves were only observed relatively to their height. In particular sea-states were described by the maxima of the observed waves. After 1945, variability of waves has started to be considered in observations and forecasts. By now we distinguished two kind of waves analyses: wave-wave analysis and spectral analysis. Wave-wave analysis is particularly adapted for studies focusing on phenomena linked to celerity thresholds or surface-curve like the wave breaking. In the opposite the spectral approach is more adapted for wave forecasting.
Unlikely the spectral analysis, which is the approach intensively used in this document, the wave-wave analysis is shortly described here. The latter allows introducing the basis of waves statistics. Wave-wave analysis relies on the definition of individual waves, delimited by the time interval between two consecutive zero-down-crossing (i.e. the measure of the point crossing down the mean sea level).
With such a definition, wave parameters (here called variables from a statistical point of view) are defined. A wave has a height H, a period T, a direction θ, and so forth.
Ones have been interested in modelling the distribution of those random variables. In particular it has been shown that the N individual wave heights $H_1, H_2, \ldots, H_N$ of a certain time series follows a Rayleigh distribution (Figure 2.2), expressed here with its survival

$$P(H > h) = \mathrm{e}^{-(h/H_{\mathrm{rms}})^2}, \tag{2.1}$$

where $H_{\mathrm{rms}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} H_i^2}$ and the H_i denote the individual wave height in a certain time series.
The Rayleigh distribution is generally suitable for commonly observed waves. As soon as the waves are quite high, the distribution of Tayfun (1980) must be considered. Tayfun (1980)'s distribution is more realistic since most of the non-linear waves effects are taken into account. Indeed those non linear aspects are generally responsible of dependences between wave variables and induces a skewness of the distribution.

To summarise such a sea-state distribution, a common variable named the significant wave height H_s is determined. From wave-wave analysis, H_s is defined as $H_{1/3}$: the mean of the third highest waves of the time-series. It roughly corresponds to what an experimented seaman would report in a same sea-state stationary condition. Another variable used is $H_{\max}$ which is the maximum wave height observed, therefore highly depending on the length of the time series. Note that the famous rogue waves are the observations for which their heights $H > 2.1H_{1/3}$.

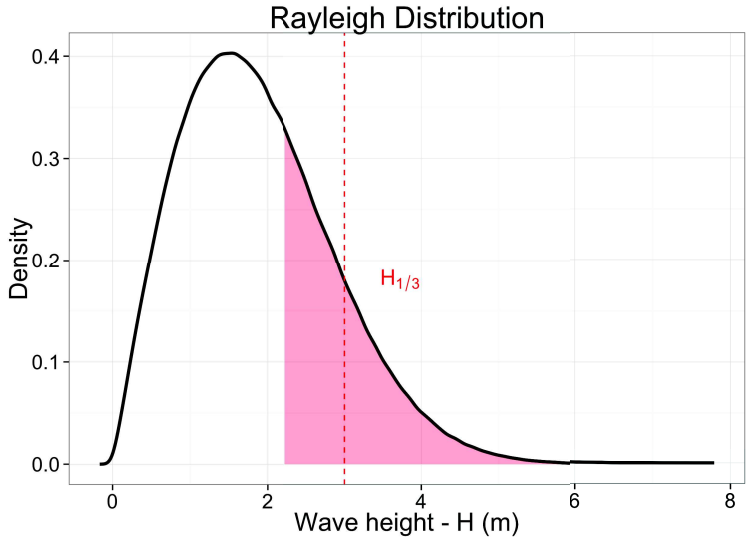


Figure 2.2: Rayleigh density function $f(H; 1.5)$. The red area is the third highest observed waves in a certain time-series. The dashed line represents $H_{1/3}$, the mean of the third highest observed waves also assimilated to the significant wave height H_s.

In this thesis we focus on extreme waves. Recalling that a return period T is associated to a return level x_T, a level for which the probability of being greater is $1/T$ (i.e. quantile $1 - 1/T$ of the max distribution law), we will be interested in extreme waves corresponding to long return periods. For instance the Netherlands Government requires to assess waves having a 10,000-year return period (Caires et al., 2011) in order to construct their coastal flood defences. With such quantities, the Rayleigh or even the Tayfun distribution (Tayfun, 1980) is not suitable anymore. In this context, extreme-value modelling introduced in Chapter 1 is the key component of any study.

2.2 Mathematical Description of Linear Waves

This section introduces the mathematical background of wave fluid motion analysis. This step is required to understand where the spectral wave analysis stems from. To go further from the briefly description of wave theory presented here, the reader may consult Dalrymple and Dean (1991); Ardhuin (2011), which

this introduction is inspired from.
As in any physical approach, the theory comes with some notations. In this chapter, waves are described by several quantities as

— H the wave height,
— T the wave period,
— L the wave length,
— θ the direction,
— $k = 2\pi/\mathrm{L}$ the wavenumber,
— a the wave amplitude,
— ka the wave slope,
— h the water depth,
— $\bar{\zeta}$ the mean free surface level,
— $D = h + \bar{\zeta}$ the local water depth.

Those quantities are going to be discussed in an Eulerian environment. In this environment, the position is given by the horizontal vector having two components $\mathbf{x} = (x, y)$ and the vertical position z. Celerities are their respective temporal derivatives denoted $\mathbf{u} = (u, v)$ and w.

In this mathematical introduction, waves are isolated and other phenomena like wind and current are not taken into account. This is to simplify the equations. Our goal here is to describe the motion of uniform (linear) waves, the one for which the wave slope is small ($ka \ll 1$) and the quantity a/D as well ($a/D \ll 1$). They are named small-amplitude waves.

2.2.1 Wave Motion

The scheme starts from the application of the Navier-stokes equations, assuming that the seawater is a perfect fluid:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + w\frac{\partial \mathbf{u}}{\partial z} = -\frac{1}{\rho_w}\nabla p + \nu\left(\nabla^2 \mathbf{u} + \frac{\partial^2 u}{\partial z^2}\right), \tag{2.2}$$

$$\frac{\partial w}{\partial t} + \mathbf{u} \cdot \nabla w + w\frac{\partial w}{\partial z} = -g - \frac{1}{\rho_w}\frac{\partial p}{\partial z} + \nu\left(\nabla^2 w + \frac{\partial^2 w}{\partial z^2}\right), \tag{2.3}$$

$$\nabla \cdot \mathbf{u} + \frac{\partial w}{\partial z} = 0, \tag{2.4}$$

where ρ_w is the water density, ∇ the horizontal gradient and ν the viscosity, g the gravity constant and p the pressure.

To simplify the computation, we are assuming that

1. The pressure is known and uniform;
2. The fluid flow is incompressible and non-viscous;
3. The density is constant;
4. The bottom is horizontal.

Moreover, the motion of the fluid is assumed irrotational. Hence celerity derives from a potential ϕ, such that $\mathbf{u} = \nabla\phi$ and $w = \frac{\partial\phi}{\partial z}$.

By replacing the potential ϕ instead of the celerities into (2.2) and (2.3), we obtain the Bernouilli equation

$$\frac{\partial\phi}{\partial t} = -\frac{1}{2}\left[|\nabla\phi|^2 + \left(\frac{\partial\phi}{\partial z}^2\right)\right] - \frac{p}{\rho_w} - gz + C(t), \tag{2.5}$$

where $C(t) = 0$ in the sequel.

We can state that the condition of continuous surface celerities is

$$w = \frac{\partial\phi}{\partial z} = \mathbf{u}\cdot\nabla\zeta + \frac{\partial\zeta}{\partial t} = \nabla\phi\cdot\nabla\zeta + \frac{\partial\zeta}{\partial t}, \quad \text{for} \quad z = \zeta, \tag{2.6}$$

and that (2.4) becomes equivalent to the Laplace equation for ϕ :

$$\nabla\cdot\mathbf{u} + \frac{\partial w}{\partial z} = \nabla^2\phi + \frac{\partial^2\phi}{\partial z^2} = 0, \quad \text{for} \quad -h \le z \le \zeta \tag{2.7}$$

and for an horizontal bottom the equation of continuous vertical velocity is

$$w = \frac{\partial\phi}{\partial z} = 0, \quad \text{for} \quad z = -h. \tag{2.8}$$

Additionally, after few manipulations on (2.2) and (2.3) we obtain the beginning of a wave equation :

$$\frac{\partial^2\phi}{\partial t^2} + g\frac{\partial\phi}{\partial z} = g\nabla\phi\cdot\nabla\zeta - \frac{1}{2}\frac{\partial\zeta}{\partial t}\frac{\partial^2\phi}{\partial z\partial t} - \left(\frac{\partial}{\partial t} + \frac{\partial\zeta}{\partial t}\frac{\partial}{\partial z}\right)\left[\nabla\phi\cdot\nabla\phi + \left(\frac{\partial\phi}{\partial z}^2\right)\right] + C(t), \text{ for } z = \zeta. \tag{2.9}$$

Equations (2.8) and (2.9) provide surface and bottom boundaries conditions for (2.7). Therefore they allow to find solutions. These equations, (2.7)-(2.9), form the so-called Euler equations. It is remarkable that the right term of the surface condition (2.9) is non-linear.

2.2.2 Towards Linear Waves

In the context of small-amplitude waves, the wave slope is small ($ka \ll 1$) and the quantity a/D as well ($a/D \ll 1$). Under this conditions, it has been proved that the non-linear term of (2.9) can be ignored and that from a first order Taylor development to get (2.9) for $z = 0$ instead of $z = \zeta$, the linearised wave equation

is

$$\frac{\partial^2 \phi}{\partial t^2} + g\frac{\partial \phi}{\partial z} = 0, \quad \text{for} \quad z = 0. \tag{2.10}$$

From a Fourier decomposition, then replacing the solution into the Laplace equation (2.7) and taking into account the bottom boundary condition, we obtain the relation

$$\frac{\partial^2 \Phi}{\partial t^2} + gk \tanh(kD)\Phi = 0, \tag{2.11}$$

having as solution

$$\Phi(t) = \mathcal{R}(\Phi_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\sigma t}), \tag{2.12}$$

with the dispersion relation given by Laplace (1776),

$$\sigma^2 = gk \tanh(kD). \tag{2.13}$$

As highlighted in Dalrymple and Dean (1991), noting that by definition a propagating wave travels a distance of one wave length L in one wave period T, and recalling that $\sigma = 2\pi/T$ and $k = 2\pi/L$, the speed of wave propagation C can be expressed from (2.13) as

$$\left(\frac{2\pi}{T}\right)^2 = g\frac{2\pi}{L} \tanh kh. \tag{2.14}$$

If elevation surface phase is defined as

$$\Theta = \mathbf{k} \cdot \mathbf{x} - \sigma t + \Theta_0, \tag{2.15}$$

with $0 \leq \Theta_0 \leq 2\pi$ and the amplitude

$$a = \mathrm{i}\frac{\sigma}{g}\Phi_{\mathbf{k}}, \tag{2.16}$$

the solutions of Airy (1841)[1] for the free surface elevation, the horizontal and vertical celerities and the pressure are

$$\zeta = a\cos\Theta, \tag{2.17}$$

$$\mathbf{u} = a\frac{\mathbf{k}}{k}\sigma\frac{\cosh(kz+kh)}{\sinh(kD)}\cos\Theta, \tag{2.18}$$

$$w = \sigma\frac{\cosh(kz+kh)}{\sinh(kD)}\sin\Theta, \tag{2.19}$$

$$p = \bar{p}^H + \rho_w g a\frac{\cosh(kz+kh)}{\cosh(kD)}\cos\Theta, \tag{2.20}$$

where the mean hydrostatic pressure $\bar{p}^H = -\rho_w g(z-\bar{\zeta}) + \bar{p}_a$ with p_a being the atmospheric pressure. These is the linear approach of the wave propagation.

In reality the waves are not fully linear but Stokes (1849) extended the Airy's solution to take into account the neglected non-linear terms in (2.9). Even if the latter improves the fit with actual observations of waves, the Airy's solution is a reliable approximation for deep-bottom waves propagation which are almost irrotational, without being so far from the reality to the coast on wave-breaking zones.

2.3 Fundamentals of Spectral Wave Analysis

2.3.1 Mathematical background

The analysis of the wave motion as described in the previous section becomes sophisticated as soon as the area of study is larger than the wave length L, implying the handling of irregularities and a consequent amount of variables for each time and locations of observations.
Alternatively, a statistical analysis named Spectral Analysis has been fostered since the 60's, concurrently to the popularizing of the computer science and the well-known Fast Fourier Transform (FFT) algorithm.
The water level (or sea free surface) elevation ζ is a highly irregular signal in any location. Indeed waves are in reality not monochromatic (i.e.single frequency).
Thanks to the Fourier series and since waves are assumed to satisfy the linear wave theory (locally), the water level signal can be decomposed into superposed sine waves with well-known characteristics.
This statement stems from the decomposition of the water level as an infinite series of sine and cosine functions oriented in all possible directions. The complex

1. Origins of this wave theory are well detailed in Craik (2004).

notation is of the form

$$\zeta(\mathbf{x},t)=\sum_{\mathbf{k},s} Z^s_{\mathbf{k}} \mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-(\mathbf{k}\cdot\mathbf{u}_A+s\sigma t)]}, \tag{2.21}$$

where $Z^s_{\mathbf{k}}$ denote Fourier's amplitudes and σ is given from $k=|\mathbf{k}|$ by the dispersion relation (2.13). From this decomposition and several mathematical manipulations, the wave spectrum corresponding to the variance of the free surface elevation ζ can be obtained. For instance,

$$E=\int_0^\infty \int_0^{2\pi} E(k,\theta)\mathrm{d}k\mathrm{d}\theta, \tag{2.22}$$

is the wavenumber-direction spectrum. Other expressions of the wave spectrum are also used. In particular, wave-lengths and wave frequencies are interrelated via the dispersion equation (2.13) and

$$E(k,\theta)\mathrm{d}k\mathrm{d}\theta = E(f,\theta)\mathrm{d}f\mathrm{d}\theta. \tag{2.23}$$

The **wave density spectrum** (i.e. right side of (2.23)) defines the repartition of the wave energy [2] along frequencies and direction. Unlike the signal of the free surface elevation, the density spectrum is relatively regular and allows compressing the information of the full signal. This is particularly suitable to numerical computing and forecasting. Illustrations of such spectra [3] are presented in Figure 2.3 and Figure 2.4. The former represents an omni-directional wave spectrum and the latter a directional-frequency one.

As recalled in Tolman (2014), without currents E is a conserved quantity. In case of currents, spectral component is no longer conserved due to the work done by current on the mean momentum transfer of waves. Hence, ones are also interested in using the wave action spectrum A as

$$A(k,\theta)=\frac{E(k,\theta)}{\sigma}, \tag{2.24}$$

which is conserved whatever the case (Whitham, 1965; Bretherton and Garrett, 1968).

2.3.2 Parameters Reconstruction

When needed, the signal is reconstructed from the spectrum definition by statistical procedures, since the phases of the original signal are not conserved. As

2. Actually E, the variance of the water level elevation, is not precisely the wave energy. The relation of the true wave energy is given as $\rho_w g E$, with ρ_w the water density and g the gravity. As it is generally the case in the ocean community, E is abusively called energy in the sequel.

3. Data were obtained from the HyMeX program, sponsored by Grants MISTRALS/HyMeX and Météo-France.

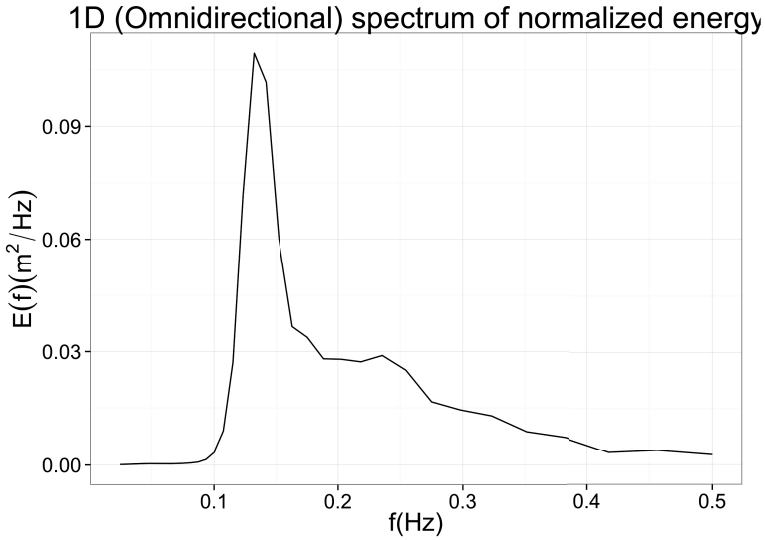


Figure 2.3: 1D (omnidirectional) spectrum of normalized wave energy observed from a drifting surface buoy, 2012/09/25 06:00 GMT, (N42.495205;E5.506991), with a range of 40 frequencies from 0.025 to 0.5 Hz, i.e. period from 2 to 40 s). The so-called wave energy corresponds to the integration of the variance of the free surface elevation. The area under the curve is a measure of the total energy in the wave field.

in the wave-wave analysis, the first quantity of interest is generally the significant wave height H_s. From the spectrum we define

$$H_{m0} = E^{1/2} = 4\left[\int_0^\infty \int_0^{2\pi} E(f,\theta)\mathrm{d}f\mathrm{d}\theta\right]^{1/2}. \tag{2.25}$$

In practice $H_{1/3} \simeq H_{m0}$ (Longuet-Higgins, 1952). H_{m0} is therefore the spectral representation of the significant wave height. The denotation $m0$ stands for the zero moment of power spectrum, which is more generally defined for the order p as

$$m_p = \int_0^\infty \int_0^{2\pi} f^p E(f,\theta)\mathrm{d}f\mathrm{d}\theta. \tag{2.26}$$

Several other quantities often used in ocean engineering derive from the spectrum. In particular, f_p is the peak frequency, with $E(f_p) = E_{\max}$ and the peak period $T_p = 1/f_p$. Other famous periods $T_{m0,1}$,$T_{m0,2}$ and $T_{m0,-1}$ stem from the period of order p defined as

$$T_{m0,p} = \left[\frac{\int_0^{f_{\max}} \int_0^{2\pi} f^p E(f,\theta)\mathrm{d}f\mathrm{d}\theta}{\int_0^{f_{\max}} E(f)\mathrm{d}f}\right]^{-1/p}, \tag{2.27}$$

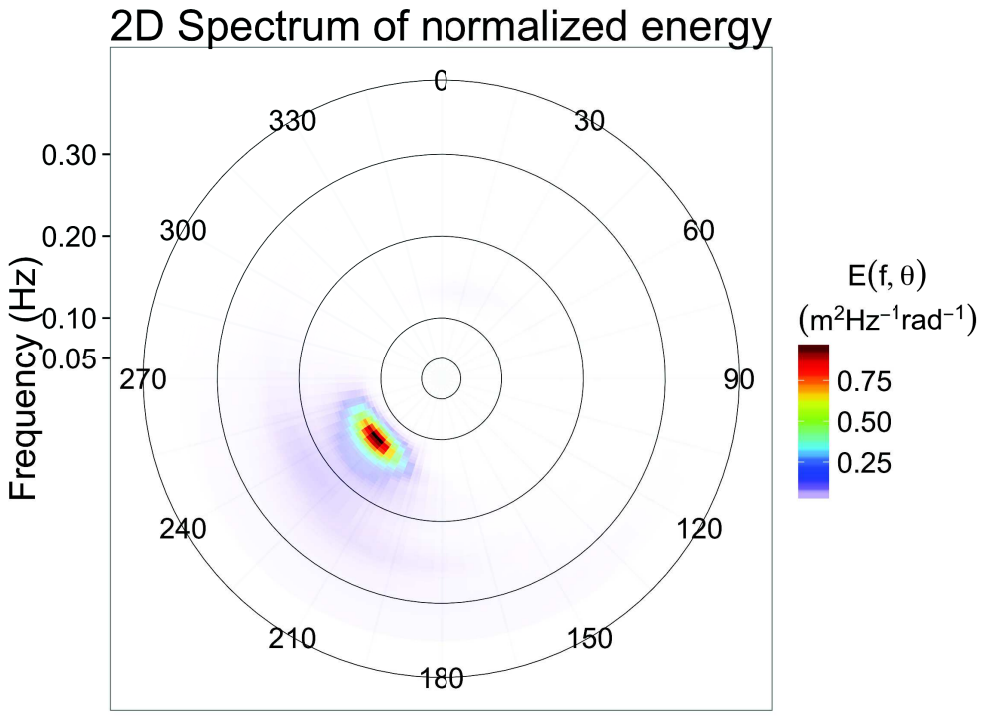


Figure 2.4: 2D spectrum of (normalized) wave energy from a drifting surface buoy, 2012/09/25 06:00 GMT, (N42.495205;E5.506991). The Energy is decomposed on directions (72) and frequencies (40 interpolated to 120 for graphical concerns, from 0.025 to 0.5 Hz, i.e. period form 2 to 40 s). The graph is zoomed on frequencies from 0.025 to 0.35 Hz.The colour is the wave spectral density. Lower frequencies ($f = 1/T$) are situated in the centre of the graph. Here the peak of the energy is observed for waves going towards (oceanographic convention) the south-west (230 degrees) for a period of $1/0.132 = 7.57$ (s).

with $f_{\max}$ the highest frequency observed. Finally, if we define

$$a_1(f) = \int_0^{2\pi} E(f,\theta)\cos\theta \mathrm{d}\theta \; / \; \int_0^{2\pi} E(f,\theta)\mathrm{d}\theta, \tag{2.28}$$

$$b_1(f) = \int_0^{2\pi} E(f,\theta)\sin\theta \mathrm{d}\theta \; / \; \int_0^{2\pi} E(f,\theta)\mathrm{d}\theta, \tag{2.29}$$

then the mean wave direction for the frequency f is

$$\theta_m(f) = \arctan\left(\frac{b_1(f)}{a_1(f)}\right). \tag{2.30}$$

In particular, $\theta_m(f_p)$ is the main wave direction (or peak wave direction). Ones are also interested in the mean wave direction θ_M defined by integrating over the direction as

$$\theta_M = \arctan\left(\frac{\int_0^{\infty} b_1(f)\mathrm{d}f}{\int_0^{\infty} a_1(f)\mathrm{d}f}\right). \tag{2.31}$$

To reconstruct the signal from a statistical approach as detailed here is valid in mostly all applications. However a wave-wave analysis would be preferred for applications when the phases of waves are of first interest, such as in the breaking zone.

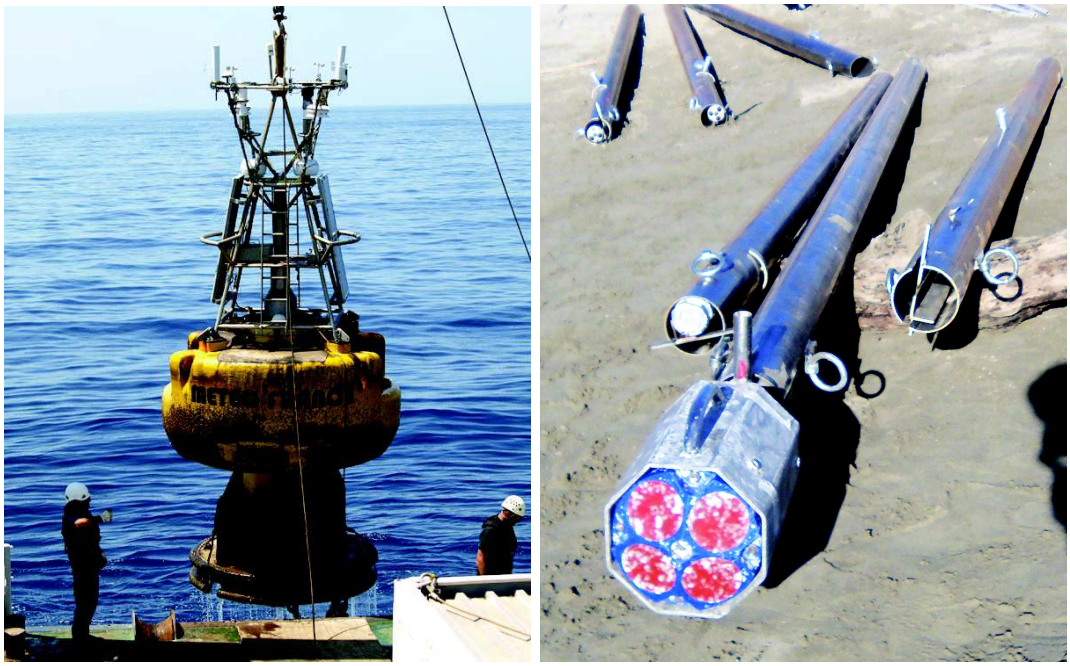
2.3.3 Observation of Waves

Many methods exist to observe waves. In this short overview we discuss five of the most used ones.

The first one is surface-buoys (Figure 2.5(a)). Surface buoys are in charge of monitoring the free surface elevation (e.g., Hamma and Goasguenb, 2004). Some of them (or combination of them) have the capacity to record the directional component of the waves propagation. Surface buoys are generally an accurate source of data. However, since the installation and maintenance of such tools are relatively expensive, the surface buoys network is quite sparse in the spatial dimension. This is the same statement for length of time-series. Only few surface-buoys have been installed on very long campaigns.

The first alternative considered here to observe the waves is the use of ADCPs (Figure 2.5(b)), standing for Acoustic Doppler Current Profiler(s). ADCPs are generally used within a network and are associated to the measure of pressure. From this measure a very accurate estimation of the wave spectrum (2.23) can be derived. Their (vertical and geographical) positions need to be perfectly known in order to be consistent. This constraint and their relative high cost (roughly tens of thousands of euros) make the ADCPs as an accurate tool but hard to handle. They use to be utilised for short measurement campaign on very local

areas.



(a) Replacement of the Lion's surface buoy (61002) (Photo Credits: SHOM).

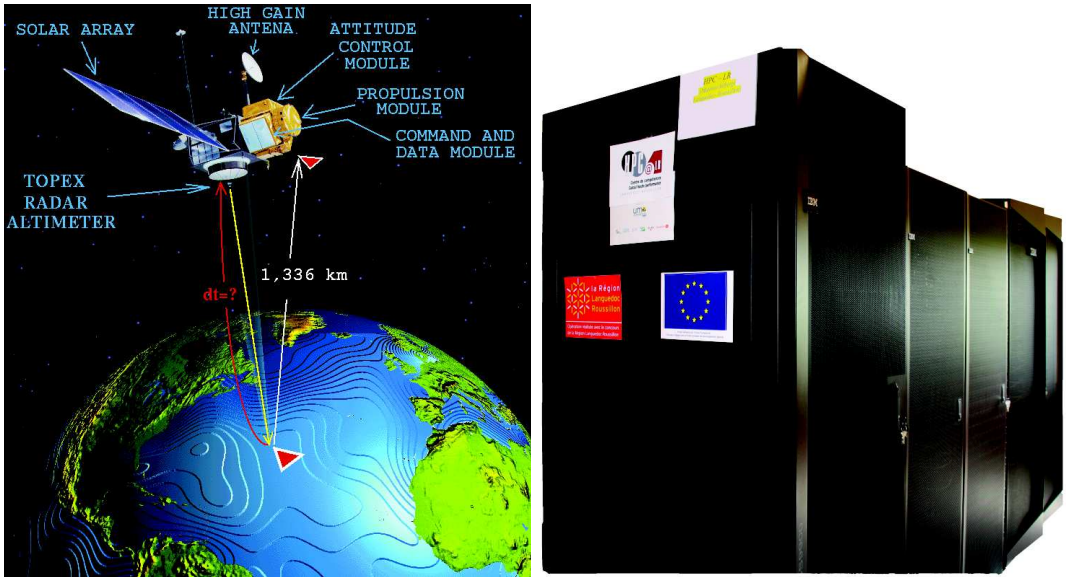
(b) ADCP deployed at a beach in the Saintes-Maries-de-la-Mer (France), for the campaign code-named Rousty1412.

Figure 2.5: Tools to (directly or indirectly) measure waves in situ.

Satellite altimeters (Figure 2.6(a)) bring another alternative to measure the sea surface (e.g., Fu, 1996). Measures provided by satellite altimeters are accurate as well and time-series begin to be quite long (few decades). However the use of such data sets implies two main drawbacks. Firstly, only the wave heights are observed. This might be useful for validation but is generally a limiting factor as soon as applications require to deal with other sea variables like the direction or period of waves. Secondly, satellites tracks are non-regular through time and space around the globe. It can therefore be a challenge to process such a data-set when applications are on a specific region of the globe and require to have regular time observation.

More recently, the use of Synthetic Aperture Radar (SAR) technology embedded into satellites (e.g., ENVISAT Frappart et al., 2006) is fostered to measure the sea-states. Such a measure is preferred to the altimeters since they can measure not only the wave height but the wave spectra. Those data sets are largely used to perform data assimilation in the waves forecast. In the very near future, even directional wave spectra will be available thanks to the enhancement of this

technology and new satellites campaigns [4].



(a) TOPEX/POSEIDON satellite altimeter and orbit placement (Photo Credits: NOAA).

(b) Octantio High Performance Cluster (HPC) of HPC@LR centre. This is the cluster used to perform waves numerical modelling, which is a resource demanding process.

Figure 2.6: Alternative tools to measure waves: remote measuring of sea-surface elevation from satellite altimeter and numerical modelling.

The very last alternative to observe waves presented here are the numerical modelling. They consist in a set of algorithms in charge of resolving the (coded) physical equations of the wave theory. Waves fields are therefore simulated from the numerical models. A simulation is used as a forecast, a hindcast or a reanalysis. The forecast concerns the prevision of the future sea-states. An hindcast is an historical data-set: the numerical model is used as in the forecast mode but the forcing fields are known (because they have been already observed) for the full period of the hindcast. A reanalysis is based on the same principle as the hindcast, but the simulation is conditioned to observation points of measure. In any case, such simulated fields are convenient because their spatial representation is controlled via the use of a computational grid serving the numerical resolution of the equations. Another remark is that unlikely the previous way of observation, (very) long historical data sets can be produced thanks to the climatic reanalysis forcing fields. Numerical models are approximations of the reality. Hence they are never perfect and sometimes the accuracy of such simulated data set can be challenged. Before using a time-series issued by a numerical model, a precise

4. For instance please consult https://cfosat.cnes.fr/fr.

validation step has to be performed, as detailed in Section 2.5.

Beside those solutions, other tools exist to measure ocean surface. We can briefly cite the wave staffs, the High Frequency Radards or even camera records as in Leckler et al. (2015), but they are out of scope of this short presentation.

2.4 A Step Forward in Wave Physic: in brief

Among other elements, we have introduced the theory of the linear waves propagation, under hypotheses. We have also seen that the signal of waves can be stored into a powerful mathematical tool: the wave density spectrum. Such a spectrum represents the sea-states energy from which we can derive other variables related to the waves (e.g., H_s, T_p, θ_M, and so forth). In this section, let us make a short introduction on the physical factors and properties that are responsible of the perturbation of the sea-states. In the sequel, high frequencies waves are set aside of those comments. Additional physic has to be taken into account to explain their properties. In particular the role of the surface tension, which is out of the scope from this introduction.

2.4.1 Spectral Balance

We have seen that (2.24) is a conserved quantity, even in the presence of currents. Thanks to that definition we can express the wave propagation as

$$\frac{DA}{Dt} = \frac{S}{\sigma}, \tag{2.32}$$

where A is the action wave spectrum, D/Dt is here the total derivative and S is the net effect of sources and sinks for the wave spectrum E. Since left part of (2.32) considers linear wave propagation as presented before, any perturbing effects are gathered in the expression of S. In the next subsection we identify physical phenomena in charge of the balance of (2.32).

2.4.2 Dominant and Limiting Factors

First of all, a perfectly steady sea surface requires a physical perturbation to produce waves. The dominant factor in charge of this creation is the wind.
Roughly speaking, when the stress of the wind over the sea surface is important enough, waves are created (wind-waves) and then propagate (swell). The action of wind is not the single one implied in the formation of waves. A second factor as important as the wind is the distance on which the wind impacts the sea surface. It is called the fetch distance (Figure 2.7). This is one of the reasons that for equivalent winds, lakes have smaller waves than seas or oceans. The fetch distance is generally linked to the fetch geometry area, also having a role in the

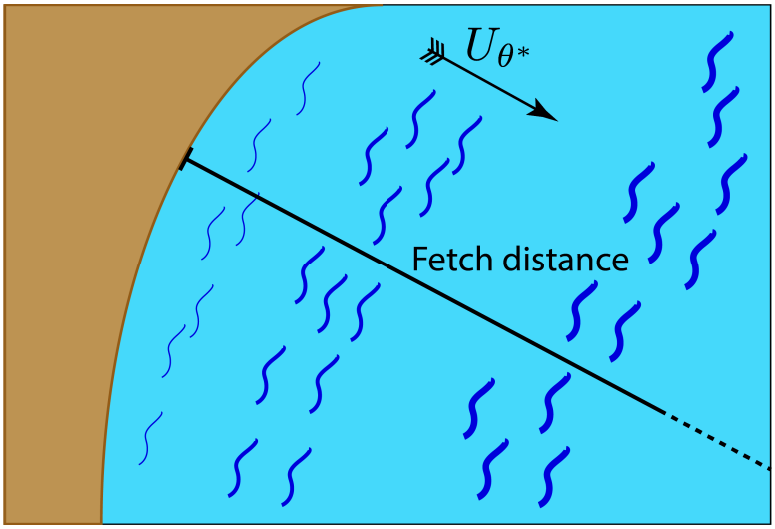


Figure 2.7: Fetch distance represented for a wind of strength U with direction θ^*. Without water depth consideration, the more the fetch distance the higher the waves.

creation of waves. They will behave differently if the wind blows over a straight or over a free open surface.
Not only the area on which the wind has an exchange of energy to the free water surface, but the duration of exposure to the wind is a limiting factor as well.
Water depth largely impacts the wave creation and propagation. We are going to detail its role in the next subsections. As a limiting factor of wave creation, we can cite for the moment the obvious reason concerning the amount of water available in shallow water.
Few others physical parameters may contribute or limit the waves creation, as for instance the temperature stability between the air and the water surfaces (sources of wind gustiness), the effect of high currents and even the rain.

2.4.3 Wind-Wave Interactions

Waves are generated from a transfer of energy from the wind to the sea surface. By creating pressure variation over the sea surface the wind might be a source $S > 1$ of energy for the waves if $U/(C\cos\theta_u - \theta) > 1$, where U and θ_u are respectively the celerity and direction of the wind, C and θ still the celerity and direction of the waves.
It can exist a damping effect due to the wind when waves are faster than the wind or with an opposed direction. Waves are therefore a source of energy to the wind when $U/(C\cos\theta_u - \theta) < 1$.
From the wave energy spectrum point of view, the source term S is generally of the form (e.g., Miles, 1957; Janssen, 1982; Ardhuin et al., 2010),

$$S = S_{\text{in}}(k,\theta) + S_{\text{out}}(k,\theta) = \sigma\beta E(k,\theta), \tag{2.33}$$

where β is known as non-dimensional growth parameter. S_{in}, the transfer of energy from wind to waves is to oppose to S_{out}.

2.4.4 Non-linearity

In reality, ocean waves are not exactly linear. Sources from non-linearity are numerous and will not be detailed here. However the main result of studies over the non-linearity of waves is that two waves processes can give birth to a third wave process at a different frequency. Exchanges of energy within the wave spectrum are therefore continuous and are named the wave-wave interactions. In shallow waters, application of this theory is not proven yet.
To balance the wave energy spectrum, a source term named S_{nl} is introduced.

2.4.5 Energy Dissipation in Infinite Depth

Several physical mechanisms dissipate the energy of waves. For instance, the conversion of the mechanic energy to heat is due to the viscosity of the fluid (Lamb, 1932). Water turbulences at every scale are also responsible of a part of the dissipation of the energy (Phillips, 1961). A considerable amount of energy is also dissipated during the wave breaking process (including white capping).
Terms named S_{dis} and S_{turb} represent these sources of dissipation of energy from the spectral point of view.

2.4.6 Littoral Physical Processes

Effects presented before assume that waves evolve in deep water. In case of shallow water, additional processes have to be considered. Most notably wave-bottom interactions S_{bot} (e.g., Shemdin et al., 1978) are responsible of a part of the dissipation of energy but also influence the mechanic of waves.

Indeed, physical mechanisms as the refraction, diffraction and reflection available in area of shallow water have also an impact on the wave propagation. Consequently they have to be included in the wave energy spectrum balance equation. In extremely shallow water, depth induced breaking and so-called triads wave-wave (three-waves interactions) are additional interactions impacting the equilibrium of the wave density spectrum.
For convenience, those last cited interactions use to be gathered in the non-linear source term denoted S_{nl}.

2.4.7 Summary

Finally, to better represent the real waves and therefore balance the spectral relation of the wave propagation (2.32), the total source term S is defined here

as

$$S = S_{\text{in}} + S_{\text{out}} + S_{\text{nl}} + S_{\text{dis}} + S_{\text{turb}} + S_{\text{bot}}. \tag{2.34}$$

2.5 Numerical Modelling

Numerical modelling is an alternative of the direct observation of waves. In order to obtain accurate time-series of sea-states, several modelling steps have to be performed. We review these key points in this section.

2.5.1 Numerical Modelling Families

There exist three main families of numerical wave models.

1. The first family of models include phase-resolving models. In these models the sea surface is resolved. They are out of scope when the area of study is to large since the (analytical) resolution of the equations are too costly. Despite this constraint, they are perfectly suitable for studies focusing on phenomena linked to celerity thresholds or surface-curve like the wave breaking. Most notably, some phase-resolving models implement Boussinesq's equations and are by definition dedicated to the accurate simulation of wave processes in shallow-waters (e.g., Filipot et al., 2013).
2. Then comes the family concerning the spectral wave models, also known as the phase-averaged models. As presented before, these models rely on the balance of the energy spectra of waves as (2.32). Since the very first spectral wave model issued by the French Weather Service in 1956 (Gelci et al., 1956), three generations have been released. First generation wave models did not consider non-linear wave interactions. They were included in the second generation. The third generation (actual) spectral wave models explicitly represent all the physics (e.g., non-linearity and dissipation) relevant for the development of the sea-states in a spatial dimension. The most-known are WaveWatchIII® (WW3) (Tolman, 1991), Simulating WAves Nearshore (SWAN) (Booij et al., 1999) or Tomawac (Benoit et al., 1996).
3. Recently, a third alternative of the previous methods has raised. This is the family of models based on particle analysis and named Smoothed Particle Hydrodynamics (SPH). For such an approach, the fluid is seen as a sum of particles that are modelled one by one. The modelling scheme is based on the mechanic interactions of these particles. The few applications realised so far are impressive in their capability to reproduce the reality (e.g., Larroude and Oudart, 2012; Lubin and Glockner, 2013). However SPH methods are tremendously resources demanding from a computational point of view. By now, it seems unreachable to model wide areas by SPH. However SPH modelling has to be taken into consideration to rise

the accuracy of the modelling of the fluid motion, especially in extremely shallow-waters, near the shore.

In the sequel we are only considering numerical wave model based on a spectral resolution.

2.5.2 Forcing Fields

Waves generation, propagation and dissipation rest on major physical factors as

— the wind,
— the surface currents,
— the fetch areas,
— the bathymetry.

Obviously, those forcing parameters need to be as accurate as possible in any initiative heading to accurately model waves. Most notably, for the realisation of an hindcast, atmospheric and ocean forcing fields issued from validated re-analyses are the choice to favour.

2.5.3 Physical Parameterisation

The physic embedded into third generation wave spectral models is considered as reliable for the modelling of sea-states. Some limitations exist due to the parameterisation. Source terms equations are mainly derived from empirical observations and are always parametric. The selection of those parameters for any equation (named parameterisation) is a main source of bias in wave modelling. Some of them have been validated for certain conditions or physical constraints, but are not valid for the simulation of all kind of wave all around the world in any situation. Those parameters may also be adapted in function of the forcing fields considered. The modeller has therefore a high responsibility in the parameterisation step, which in this sense is as important as the embedded equations themselves.

2.5.4 Numerical Aspects

In numerical modelling, the equations are transformed into algorithms in order to be resolved. Several approaches exist in the literature. In particular, the finite difference approach is one of the most used to resolve physical equations. For this approach based on the discretisation in time and space of the problem, two main schemes exist: explicit and implicit.
An explicit scheme is based on the computation of a time-step $t+\Delta t$ from information given at a time t. Δt need to be relatively short, implying long computational processes. A too high value of Δt results in instability of the model. Usually, numerical constraints are present to avoid the instability of the models.
In contrary, an implicit scheme is based on the resolution of a (generally complex)

equation involving both information at t and $t + \Delta t$. Such a scheme is stable and allows to work on higher value of time steps. However the equation might be extremely complex to resolve and numerical diffusivity might be introduced.

2.5.5 Spectral Discretisation

The main scheme of a spectral wave model is to resolve the spectral energy balance (2.32). Such relation implies integration of spectra over all frequencies $-\infty < f < +\infty$ and directions $0 \leq \theta < 2\pi$. In practice, any model can use neither an infinity of frequencies nor an infinity of directions. In this sense the modeller needs to constraint the physics into a certain range of these parameters. This is obviously a source of approximation of the reality.

2.5.6 Spatial Discretisation

To represent the space within a wave spectral model, we distinct three families of computational grids which are the support of equations resolution.
The first family concerns the **regular** grids, as illustrated in Figure 2.8. Such grids are easily created and ease the computations and models are generally stable enough. However, if waves need to be accurately computed on a littoral areas (e.g., coastline, islands), we need to nest low definition grids to high definition grids (e.g., Michaud et al., 2012). Indeed, the wave spectrum is much more variable and sensitive in littoral areas than offshore, and thus requires a high definition of points to better represent the local sea-states behaviour. Since even offshore hydrodynamics processes are impacting the littoral areas, such a high definition grid would be extended to the full area of interest. This would be out of reach for largely extended areas, from a computational point of view. Therefore nesting is used to bring to the nearshore the boundaries conditions of the smallest computational grid.
A second way to represent the space from a computational grid point of view is through the use of the so-called curvilinear (or polar) grids, as illustrated in Figure 2.9. One advantage is to avoid nesting grids but nevertheless obtain a high density of computational points close to the area of interest, which is placed near the pole of the grid. The gain of computational time is important since the model has to be run only a single time. However, as for the regular grids family, some offshore obstacle (like an island) might be totally masked by the low definition of the grid at this scale. This could impact the accuracy of the waves simulation, since without wind a swell can cross about 30,000 km.
Unstructured computational grids (also named meshes) like the one presented in Figure 2.10 are a solution to avoid this drawback. As the curvilinear grids, unstructured grids are refined to the area of interests. The main difference is that area may be numerous. The unstructured grid allows representing perfectly any obstacle all around the area of interest, but can be also coarse on offshore areas where spectra are likely to be less variable. The major drawback of such a grid

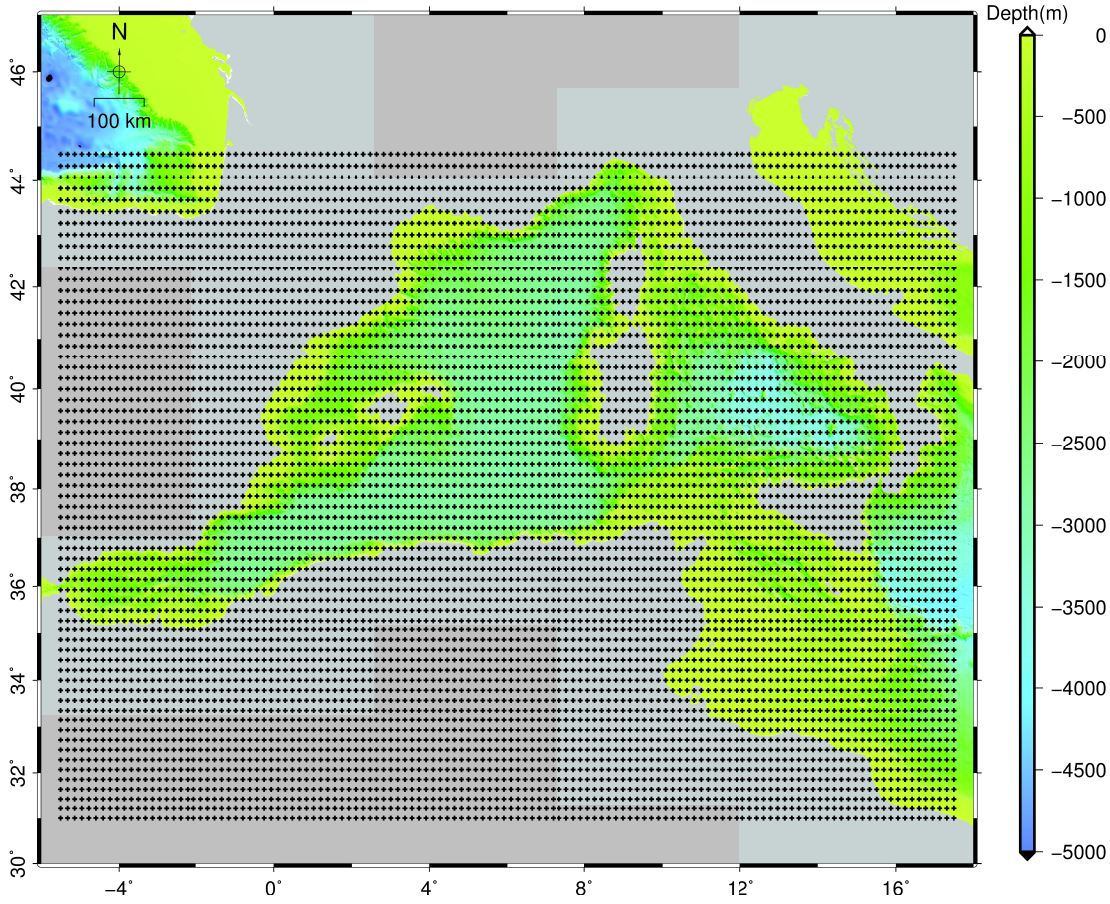


Figure 2.8: A regular grid (128 x 64 points) over the North Western Mediterranean sea. The spatial resolution is 0.1811 degrees on longitude and 0.2143 degrees in latitude. Bathymetry is plotted in colour.

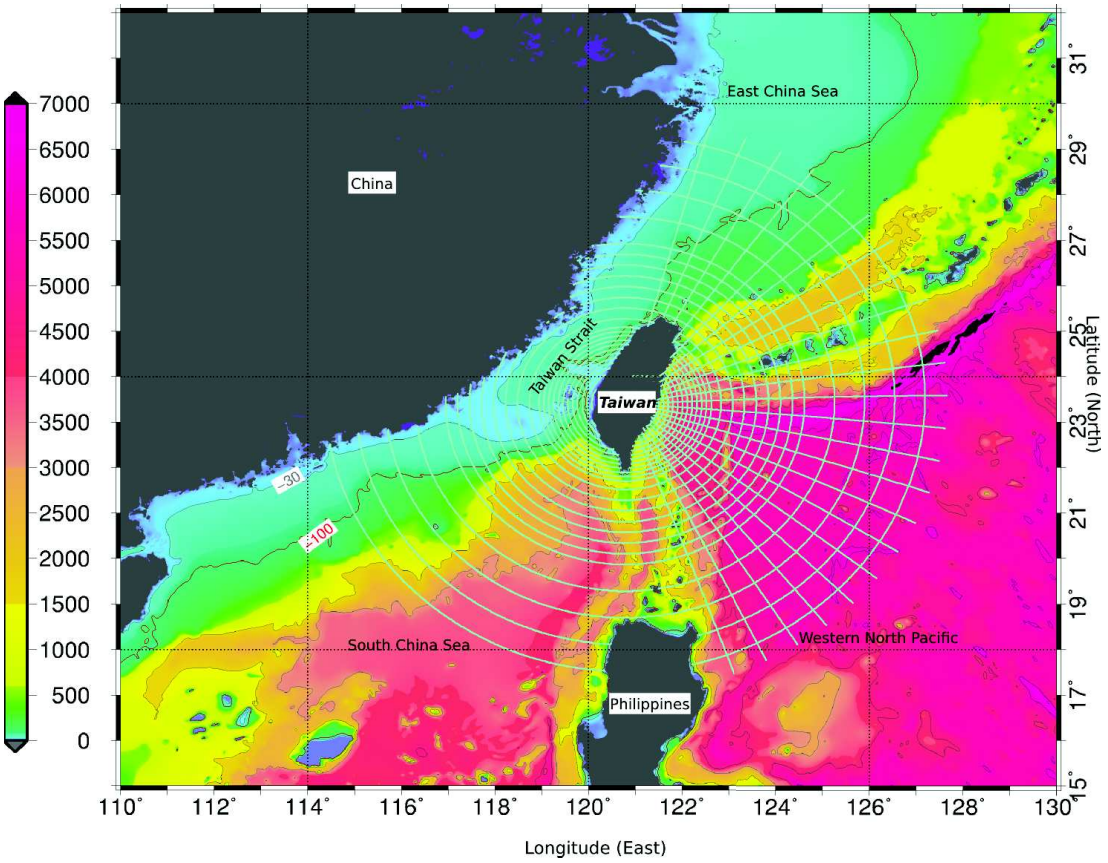


Figure 2.9: Curvilinear grid (822 x 322 points) for a regional modelling around Taiwan. Spatial resolution stretches from 400m at the shoreline to 5km offshore. Bathymetry is plotted in colour. This figure is extracted from Rétif (2015).

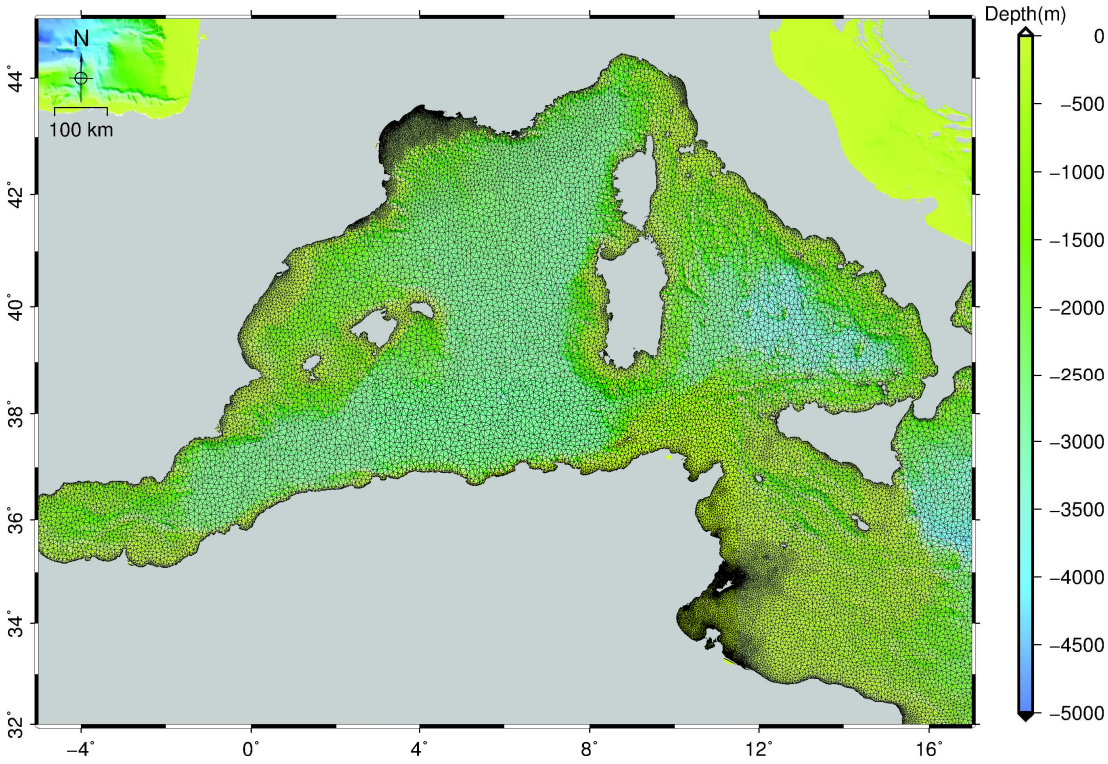


Figure 2.10: Unstructured triangular grid (47086 points). Spatial resolution stretches from 1000m at the shoreline to 12km offshore. Bathymetry is plotted in colour.

is that it can be difficult to generate it, even with assisted tools. Moreover the direct coupling (e.g., with circulation models) can be difficult as well. Curvilinear grids are preferred in this context.

2.5.7 Validation

Validation is an important step of the simulation procedure. Modellers have to validate the produced data sets in order to guarantee the accuracy of the created observations. This action requires having measure-points at the physical time of simulation and geographically situated within the studied area. For spectral model, the spectrum is rarely compared directly. Generally, waves variables derived from the spectrum (e.g., the significant wave H_s, the peak wave-period T_p or the mean wave direction) are the data compared. As introduced before in Section 2.3.3, sources of directly observed data might be diverse (e.g., surface buoys or altimeters).

At this step, the principal information about both Extreme Value Theory and sea-waves have been reviewed. In the following chapters we are going apply the former onto the latter in the goal of providing new methods to assess coastal hazards. Before that and in the next chapter, let us define the case study of this thesis and present the data on which we are working on.

Part II

Applications

Chapter 3

A 52-Year Wave Hindcast

Chapter Summary

As schematised in Figure 3.1, this chapter covers the production of a reliable data set of historical sea-states conditions over the case study area: the **Gulf of Lions** (GOL). This gulf is situated in the north-western Mediterranean sea where waves observations are poorly provided regarding both time and space dimensions. The challenge here is to get the most reliable and long wave features data set. In Oceanography, a common approach to deal with the scarcity of waves observations is the use of hindcast data sets. When the system studied lacks direct observations, hindcast data sets become the ideal candidates to apply extreme value modelling on (Chapter 1), in the goal of addressing coastal hazards.
In this chapter the area of interest is described before introducing the wave spectral model used to build the 52-year sea-states hindcast. Forcing fields and the model parameterisation is also discussed. Finally, results and their validation against surface-buoys records are discussed.

3.1 The Gulf of Lions and waves observations

The Gulf of Lions (GOL) is a semi-closed French coast area located in the north-western Mediterranean sea (NWM) as illustrated in Figure 3.2. This Gulf was named in reference of the meteorological conditions hitting it, which are considered as threatening as a lion Mistral (1979).

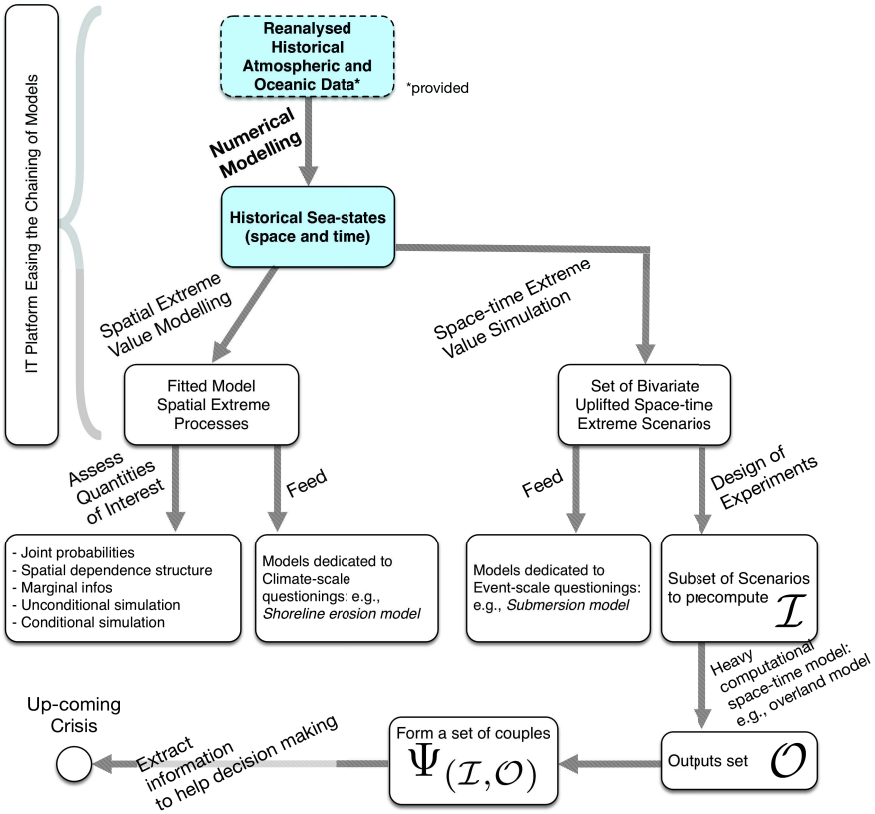


Figure 3.1: Regarding the schematic representation of the presented thesis, the current Chapter 3 covers the production of a reliable and historical waves data set.

In the GOL, weather conditions and sea-states use to change abruptly. This is due to several factors. Among them, Millot (1990) describes the simultaneous competition of many intense and variable phenomena (e.g., violent and surprising winds).
This (relatively) hostile area is therefore a natural laboratory reproducing diversified processes observable on several scales all around the world, and is subject to trigger coastal hazards. This make it a region of interest in many disciplines. For instance, Ferré et al. (2005); Leredde et al. (2007); Guizien (2009); Guerinel et al. (2012); Michaud et al. (2012) testify to the general interest of this area.
As them, and since the GOL is subject to coastal hazards, we set it as the region of interest of this Ph.D. Referring to Figure 2, we are likely to applied stochastic approaches to model extreme waves events in the GOL on the basis of the (simulated) data in that region.

We recall that the quality of statistical extreme value studies relies on the length and accuracy of the observed time-series. For spatial and space-time approaches, the modelling quality also relies on the spatial resolution of observation sites. Unfortunately only few surface-buoys are deployed and maintained in the GOL (see Figure 3.3).

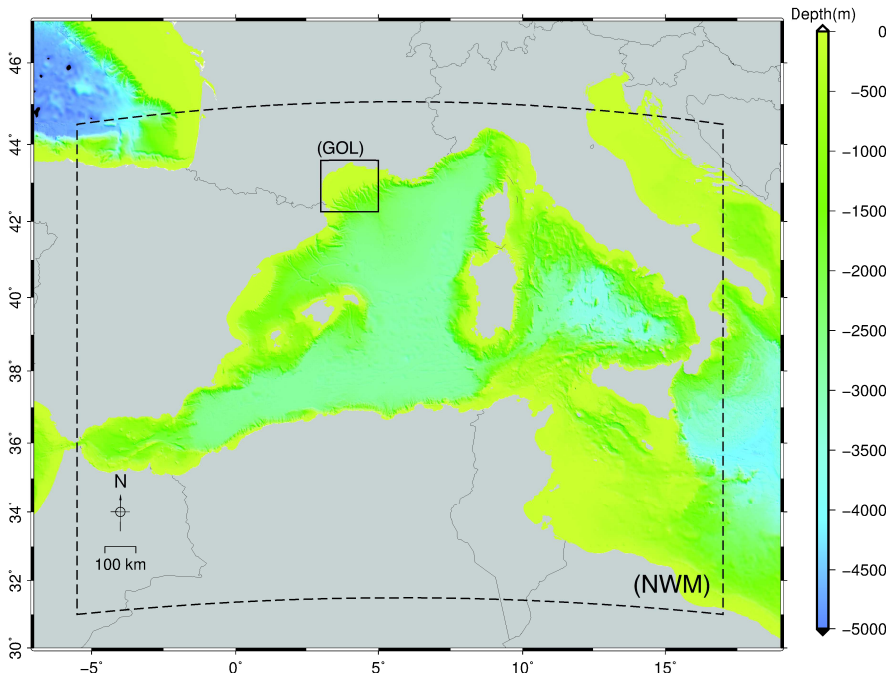


Figure 3.2: Spatial extension of the Gulf of Lions (straight box) within the North Western Mediterranean sea (dashed box). Bathymetry is plotted in colour.

Even if the data are accurately measured, time-series are too short and sparsely provided to accurately extrapolate information to long return periods and assess coastal hazards along an entire coastline.
In Chapter 2 satellite-altimeter data sets are introduced as an alternative to surface-buoys observations. However they only measure the significant wave heights. It is a real limit if studies are not only focused on the significant wave height variable but also on wave directions or wave periods. This issue is solved by the use of satellites embedding SAR radars but are very new, and observed time-series are short. Moreover since satellites tracks are non-regular through time and space around the globe, any extreme statistical analysis considering such data sets might be hard to handle, especially when the modelling concerns a space-time event in a fixed and relatively short area.
Naturally we choose a third alternative being the observations from numerical modelling, detailed hereafter. The requirements of the presented hindcast are to be reliable and to provide long time-series on a set of locations that well-cover the region of interest. In the literature, two wave hindcasts covering the GOL exist to the best of our knowledge: the hindcast in Morellato and Benoit (2010) covers the 1979-2008 period and its extension in Laugel et al. (2014) to the 1979-2010 period using CFSR (NCEP's Climate Forecast System Reanalysis) winds and unstructured meshes. Regarding them, we propose to realise a regional hindcast on a longer historical period (1961-2012), with a finest spatial resolution and using another cutting-edge wave model.

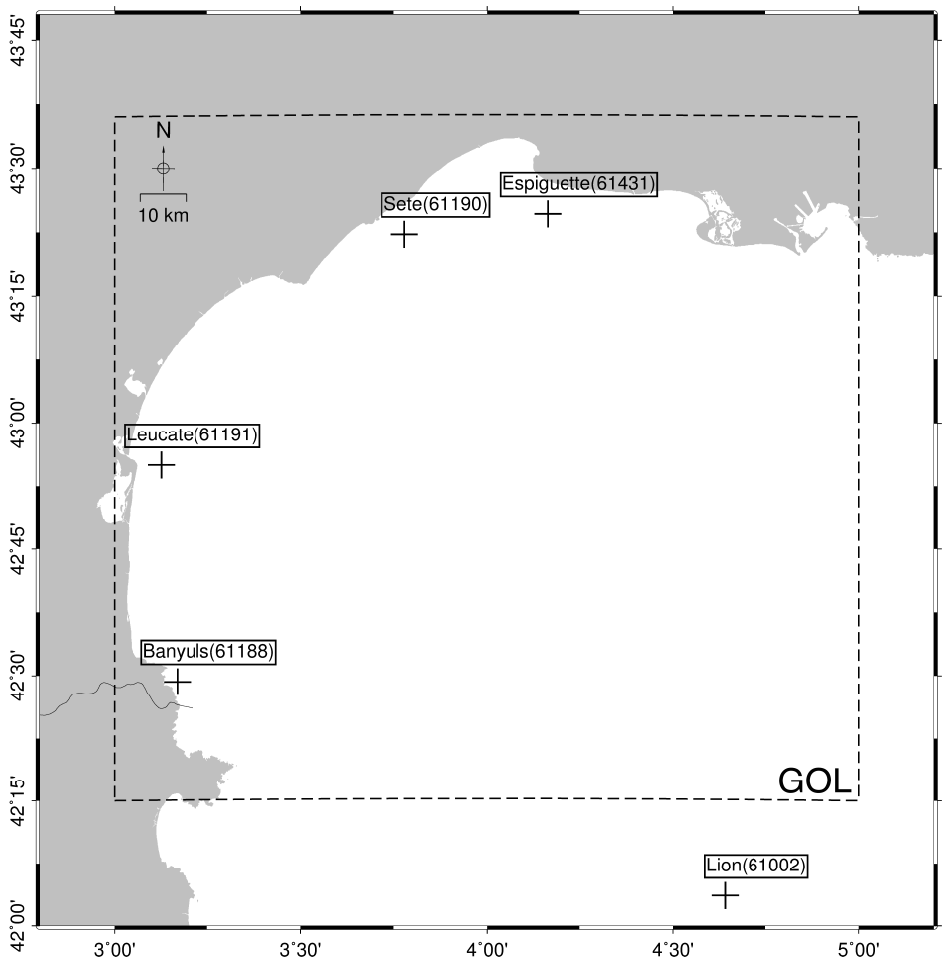


Figure 3.3: Locations of surface-buoys available in the GOL.

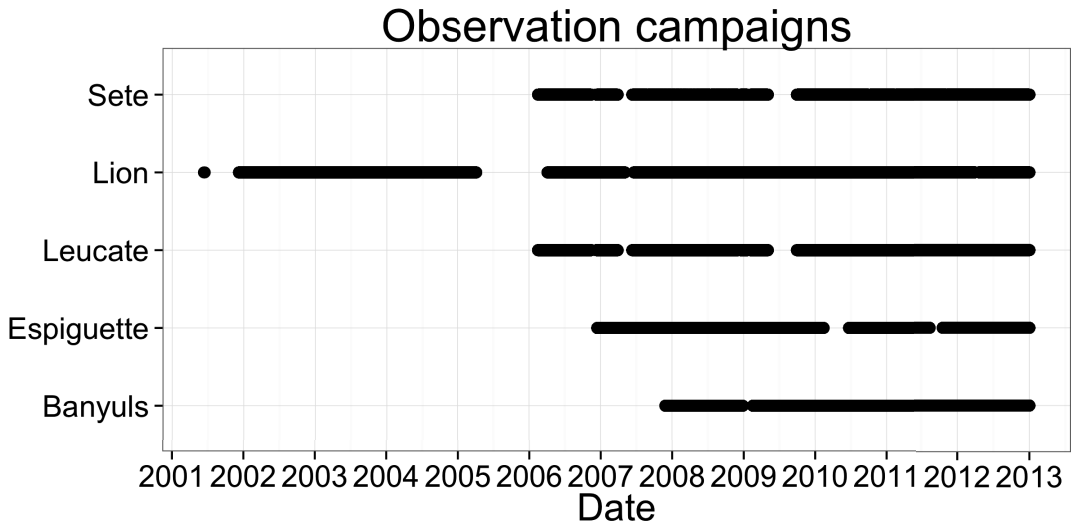


Figure 3.4: Recording campaigns of the surface-buoys located in the GOL.

From the literature, we can assume that any hydraulic process observed in the GOL results from a combination of those present in the NWM area only, i.e. the area extending from the Gibraltar's strait to the south of Italy (Figure 3.2). This statement provides the first parameter of this sea-states hindcast: the geographical envelope considered.

3.2 The Wave Model

To realise the hindcast, we use the WAVEWATCH III® (WW3) (Tolman, 1991, 2002, 2008, 2014) wave model. This is a third generation wave model solving the random phase spectral action density balance equation for wavenumber-direction spectrum (see (2.24) and (2.32)).
WW3 is supported by the NOAA/NWS/NCEP and known as one of the most reliable wave spectral model. It is used in production for daily forecasting but also for several hindcast and reanalyses. The model is suitable whether these simulations are *global* or *regional*.
In its earlier releases, WW3 was mainly dedicated to the computation of offshore waves. Since the version v3.14 (Tolman, 2008), the physic embedded into the model (e.g., wave-breaking in surf zone) allows to better approximate shallow water processes. The presented hindcast is produced with the very last version: v4.18 (Tolman, 2014).

3.3 Forcing Fields

As far as the goal of the presented hindcast is to be used to perform extreme values statistical studies, the longer the time series the better the extrapolation to large return period. Hence we are only considering data set with long period records. The geographical envelope considered (i.e. NWM) is relatively short. The hindcast cannot rely on state-of-the-art *global* reanalyses because of their incapacity of catching very local processes due to their weak resolution. In this context, we need to use *regional* Climate and Ocean reanalyses to force the wave model.

3.3.1 Atmospheric Fields

According to experts from CNRM [1], the best option to match our requirements while forcing the wave model is the data set from Herrmann and Somot (2008), namely the ARPERA reanalysis. This reanalysis is built from a tilted and stretched grid of the famous and widely used climate model named Arpege (Déqué, 2007).

1. French National Meteorological Centre.

Thanks to its zoom capacity, atmospheric data (e.g., winds) are given on a 50km-resolution grid illustrated in the Figure 3.5. The time step of the reanalysis is of 6 hours, from the 1st of January 1959 to the 31th December 2012. As far as WW3 (v4.18) does not handle non-regular grids as forcing fields, the data is interpolated to a regular grid [2]. The resolution of the interpolated grid is $1/8^\circ$ to fit the same resolution of the surface current data (presented in the next subsection). The interpolation is performed with a weighted-near-neighbour algorithm, which consists in making a weighted average of the closest data, up to a selected radius: $1/4^\circ$ [3]. This interpolation can be discussed, but it relies on the relative linear smoothness of wind fields over sea areas.

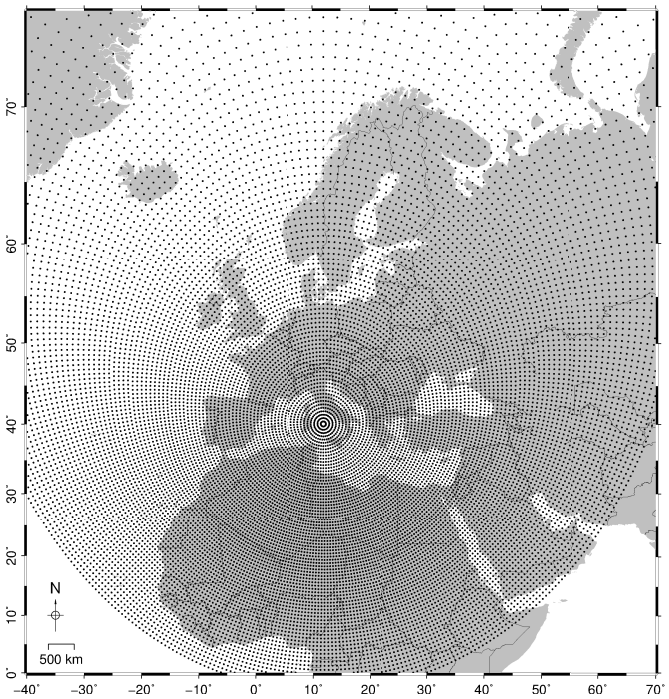
3.3.2 Ocean Fields

In the literature the Mediterranean sea is often described as an Ocean laboratory. The coastlines, islands, straits and shallow areas define its complexity. To accurately represent the ocean circulation in this context, a 10 km resolution model at least is needed. The reanalysis called NEMOMED-8-24 (Herrmann et al., 2010), stemming from the development of Beuvier et al. (2010) by the CNRM climate team, provides accurate surface ocean fields on the grid represented in Figure 3.6(a). This reanalysis has a spatial resolution stretching between 9 to 12 km and a time resolution of 24 hours. Available for a long historical period (1961-2012), NEMOMED-8-24 is forced by the atmospheric fields of ARPERA. The association of the both fields are therefore the ideal candidate for our regional hindcast.
Surface currents are extracted [4] from this grid with a small step of interpolation, like for the atmospheric fields but with a smaller radius for the weighted-near-neighbour algorithm.

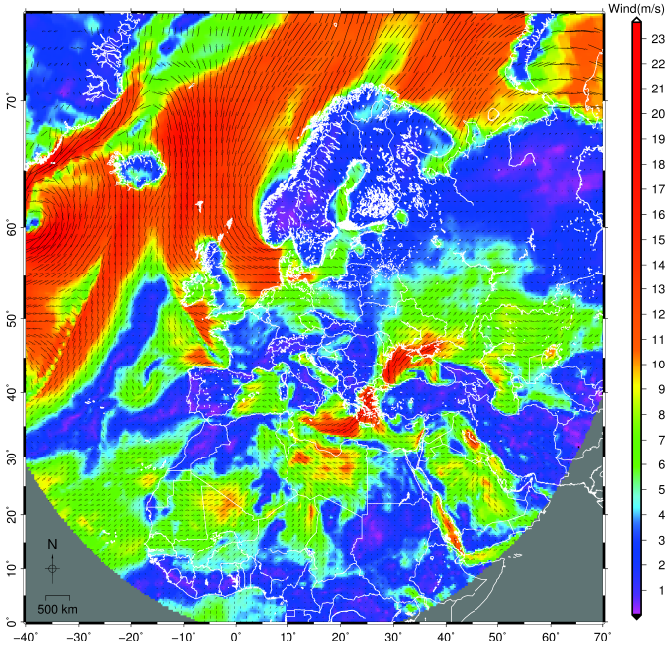
3.3.3 Bathymetry

A reliable bathymetry is a mandatory input for wave simulations. Most notably in littoral areas, where the bathymetry directly impacts the propagation of waves.
It is observable that the bathymetry in the NWM region is highly variable. For some places, the sea bottom is deep close to the shore (e.g., the French Rivera). However it exists places at the contrary, like the GOL. Indeed the GOL is composed of an inner-continental shelf observable in Figure 3.2. The dissipation of waves are completely different in these both regions. It is therefore important to rely on a reliable bathymetry.
For the production of the presented hindcast, we use the bathymetry constructed

2. See https://github.com/rc-34/mirmidon-toolbox/tree/master/scripts/convertMedAtmosFlux
3. This radius of $1/4^\circ$ is empirically chosen, in order to use enough data to avoid NaN but not to many to keep a relevant information and avoid smoothing gusts observations.
4. See https://github.com/rc-34/mirmidon-toolbox/tree/master/scripts/convertMedOceanFlux.

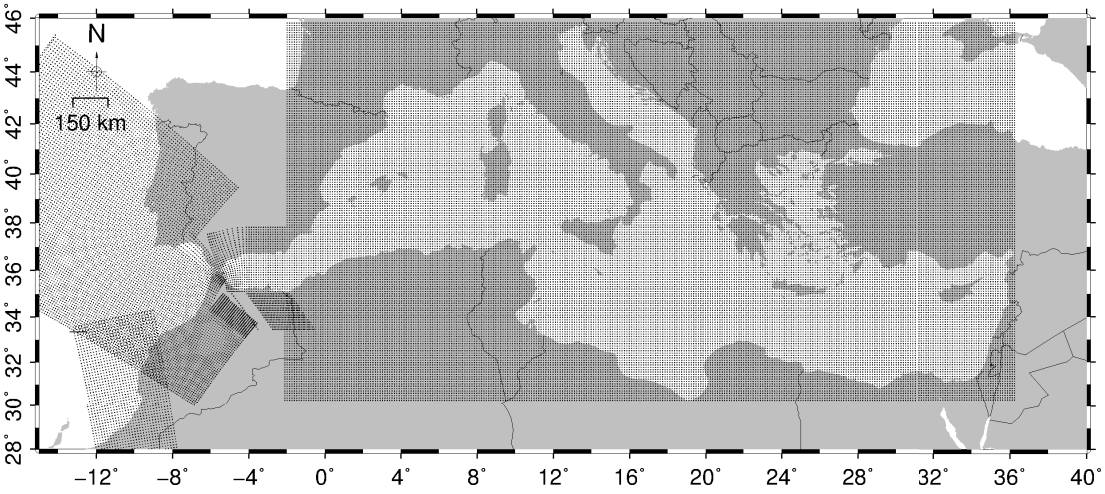


(a) The ARPERA computational (polar) grid, centred on Mediterranean sea. The spatial resolution of the grid is about 50km.

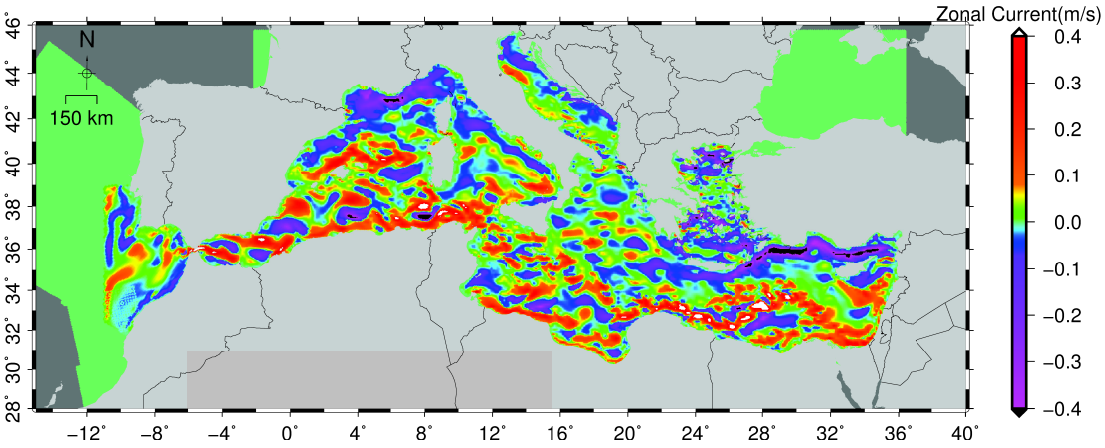


(b) One time-step of winds field from ARPERA (20-01-2012 18:00 UTC). ARPERA has a temporal resolution of 6 hours from 1959 to 2012.

Figure 3.5: ARPERA reanalysis.



(a) The NEMOMED-8-24 computational grid, encompassing the Mediterranean sea. The spatial resolution of the grid is from 9 to 12 km.



(b) One time-step of Zonal currents field from NEMOMED-8-24 (20-01-2012 18:00 UTC). NEMOMED-8-24 has a daily temporal resolution from 1961 to 2012.

Figure 3.6: NEMOMED-8-24 reanalysis.

by the SIROCCO TEAM [5], leveraging the famous GEBCO bathymetry with other measures campaigns. This bathymetry has a resolution of 0.00833 °, which is around 800 m in the GOL.

3.4 Computational Mesh

To better represent the irregularities of both the coastline and bathymetry, and also the several obstacles (e.g., islands) existing in the NWM, the model performs computations on an unstructured mesh. Figure 3.7 shows the finalised grid on the NWM area and provides a zoom on the refined GOL area as well. The computational mesh gets the following characteristics:

— The entire coastline and any islands in the area are segmented at a spatial resolution of 1km.
— The largest edge between two points – at offshore – is of 12 km.
— The mesh is refined between 1 and 1000 m depth, which explains the node density in such areas.
— A specific refinement appears in the GOL, with edges of the mesh up to 200 m long.

In total, the mesh is composed of 47086 computational nodes. They are 3944 for the GOL only.

3.5 Model Parameterisation

The wave spectral model WW3 uses a spatial and directional discretisation. For the presented hindcast, the frequency discretisation is realised by taking frequencies exponentially spaced from 0.0345 Hz to 0.5473 Hz at an increment of 10% (i.e. wave period from 1.8 s to 29 s). The spectrum is computed for 24 directions (15 ° increment). Several source terms can be activated to solve (2.32). The ones used for this hindcast are listed in Table 3.1. Source terms parameterisation stems from an adapted version of the test case T405 presented in Tolman (2014) (Table 2.6). For more details on those, please consult Tolman (2014).

3.6 Results

WW3 resolves (2.32) for each node of the computational mesh. Full spectra are not conserved in the outputs since it would represent too many data. Instead, derived sea-states (i.e. 20 variables like mean wave direction or peak wave period, Figure 3.8) are stored at an hourly time step for any computational node of the mesh and for the 52-year historical period (1961-2012).

5. http://sirocco.omp.obs-mip.fr/accueil/Accueil.htm.

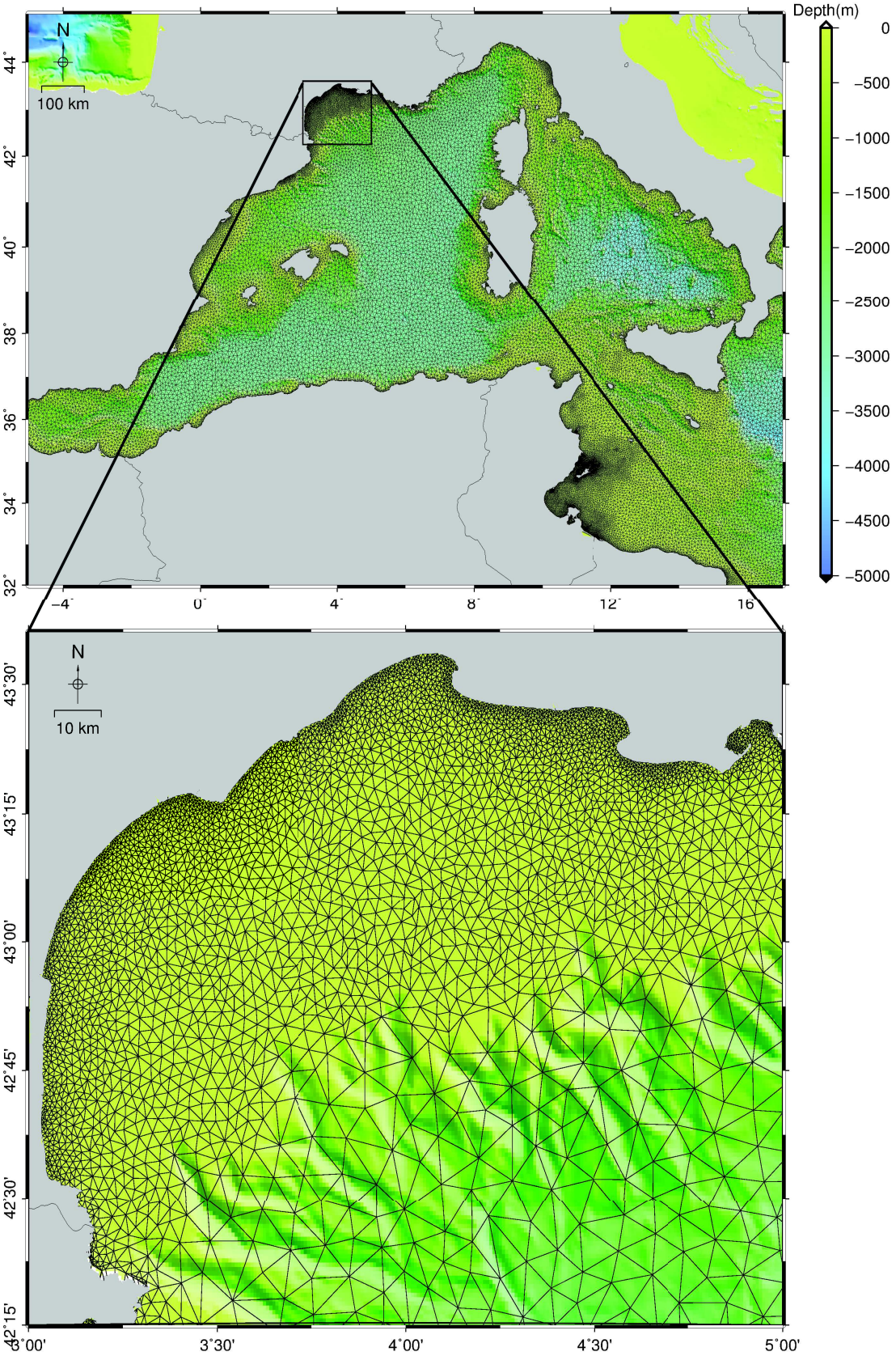


Figure 3.7: The computational unstructured mesh used for the hindcast modelling composed of 47086 computational nodes, with a zoom on the GOL (refined area).

Source term	Switch WW3	Comments
Propagation Scheme	PR3 + UQ	Higher order schemes with Tolman (2002) averaging technique and Third order propagation scheme
Linear input	SEED	Spectral seeding of linear input
Input and dissipation	ST4	Ardhuin et al. (2010) source term
Non linear interaction	NL1	Discrete interaction approximation (DIA) Hasselmann et al. (1985)
Bottom friction	BT4	SHOWEX bottom friction formulation
Depth induced breaking	DB1	Battjes and Janssen (1978)
Reflection	REF1	Enables reflection of shorelines
Shallow water	MLIM	Use Miche-style shallow water limiter
Triad interaction	TR0	No triad interaction used
Bottom scattering	BS0	No bottom scattering used

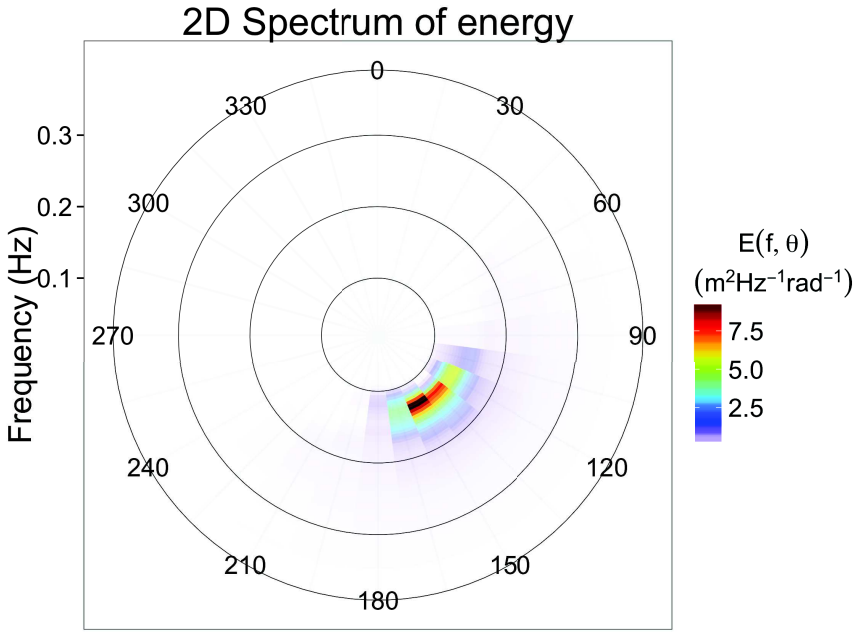
Table 3.1: Parameterisation of the WW3 model defined for the presented hindcast.

Beside the results available at computational nodes, hourly full wave spectra (Figure 3.8(a)) are stored for a list of 228 stations. These data are computed for validation and comparison purpose. They can also be used to define boundary conditions of smaller grids in that region.
To build the hindcast, WW3 is run in parallel on 30 nodes (240 CPUs) provided by the HPC@LR's HPC cluster [6]. The overall simulation takes 31 days in a row [7]. The produced fields form a 1.2 To data set stored in binary files (NetCDF4 format).

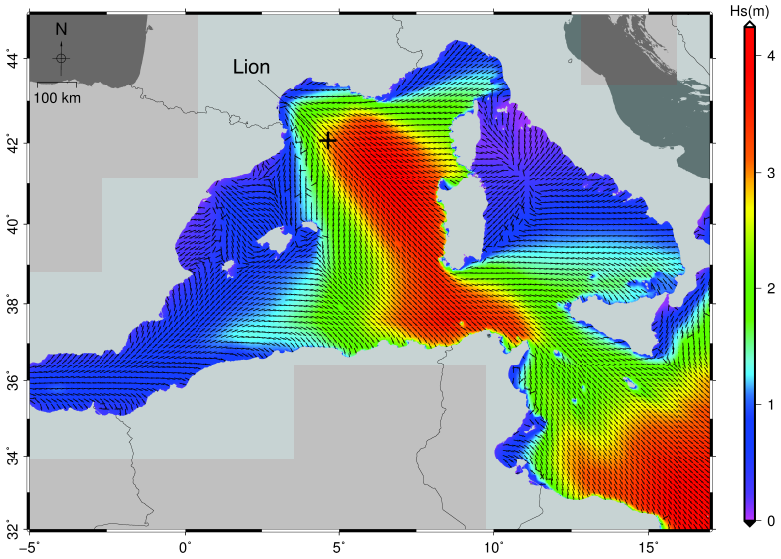
Data simulated are validated against the 5 observation sites illustrated in Figure 3.3. Four of them are near-shore surface-buoys and the remaining one is offshore (Lion). Full simulated spectra are available at those locations but we validate the variables of greatest interest for our applications: the significant wave height, the mean and peak wave directions and the peak wave period. Only the significant wave is investigated here.

To evaluate the performance of the hindcast, widely used statistical tools are introduced. Those tools allow to compare two time-series: $x_1, x_2, \ldots, x_n$ the reference time-series and $y_1, y_2, \ldots, y_n$ the simulated one. They are defined as follows.

6. Hindcast was performed in the Octantio cluster 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 University of Montpellier.
7. Scripts available at https://github.com/rc-34/mirmidon-toolbox/tree/master/scripts/megagol-autorun



(a) 2D spectrum of wave energy modelled by WW3 at 2012/01/01 00:00 GMT at the Lion's surface-buoy location (one of the 228 validation point). The simulated spectrum is decomposed onto 24 wave directions and frequencies exponentially spaced from 0.0345 Hz to 0.5473 Hz at an increment of 10%. The maximum of energy is at nearly 150 °, meaning a wave propagation towards the south-east (oceanographic convention).



(b) Extracted field of significant wave height. Vectors are the mean wave directions, interpolated to a 15km/15km grid to ease the reading of the map. The black cross is the location of the Lion's surface buoy.

Figure 3.8: Example of hindcast output.

Definition 1 (Correlation)

The Correlation coefficient (COR) expresses the linear dependency between the two time-series and is defined as

$$COR = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}, \tag{3.1}$$

where $0 \leq COR \leq 1$. The more COR approximate 1, the better the simulated series is linked to the reference (observations).

Definition 2 (Relative difference)

$$RDIFF = \frac{\bar{y} - \bar{x}}{\bar{x}}, \tag{3.2}$$

is the relative difference (RDIFF) between the two time-series. The more RDIFF value is close to 0, the better the simulated time-series matches the reference one, in mean. The divisor term allows to compare two RDIFF values at different locations to assess their performances.

Definition 3 (Root Mean Squared Error)

The root mean squared error (RMSE) is a value expressed in the unit of the time-series. In this sense it is easily interpretable and is a good measure of accuracy of the simulated time-series against the reference one. It represents the mean differences from one time-series (simulated) against the other (reference). It is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}}. \tag{3.3}$$

One may also normalise the RMSE to gives the NRMSE, defined as $NRMSE = \frac{RMSE}{y_{\max} - y_{\min}}$. The marginal NRMSE might be compared from each other to sea where the model is the more reliable.

Definition 4 (Scatter Index)

The scatter index (SI) is the relative value of the RMSE and is given by

$$SI = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{\sum_{i=1}^{N} x_i^2}}. \tag{3.4}$$

It allows to compare indexes from different group of time-series. One may compare the marginal SI to show where the model is the more reliable.

Station	COR	RDIFF	RMSE	SI	Period
Lion	0.942	-0.0857	0.428	0.223	2001-2012
Espiguette	0.908	-0.1540	0.275	0.311	2006-2012
Sete	0.908	-0.1810	0.289	0.351	2006-2012
Leucate	0.896	-0.0682	0.251	0.295	2006-2012
Banyuls	0.883	-0.0667	0.279	0.324	2007-2012

Table 3.2: Statistical validation parameters computed from the simulated time-series against the observations from the 5 surface-buoys. For each location, these parameters are computed from their respective campaign period.

These measures are computed to investigate the accuracy of the simulated significant wave heights and are presented in Table 3.2. It is remarkable that the significant wave height at the offshore location (Lion) is better represented than the four near-shore locations, regarding both its correlation 0.942 and scatter index 0.223 which are the best. This is in agreement with the complexity of modelling waves in littoral areas. However, the performances of the model on those four littoral locations are still suitable. For instance they present RMSE values inferior to 0.30 m and correlation coefficient around 0.90. The high value of the RMSE is partly explained by an additional investigation of the observed time-series (from buoy) commented at the end of this section.

Additionally to the computed indexes, another diagnostic of validation is the visual interpretation of the time-series as illustrated in Figures (3.9-3.13). Those are the plotted time-series of significant wave height (m), peak wave period (s), mean wave direction ($^\circ$) and peak wave direction ($^\circ$) for the 2012 year. These figures demonstrate a fit of quality.

Only the simulation of peak wave periods is sometimes exploding for few time-steps (< 1 % of observations). The main explanation comes from the frequency discretisation: when the frequencies are too high (capillarity waves), the model is no more able to simulate them under our parameterisation. Frequencies out of range of the parameterisation are arbitrary set by WW3 to 0 in the outputs and periods ($T = 1/f$) becomes infinite. For the four littoral surface buoys presented below, we post-process the time-series to set infinite periods values to NaN and therefore not reveal them in the graphs.

The presented indexes are representative for the assessment of the mean behaviour of the modelled time-series against the observed ones. As far as we are looking for assessing extreme waves, we need other diagnostics to show if the modelling is reliable for rare events as well. In this sense and to validate the hindcast, it is mandatory to have a look on the quantile-quantile plots of observed time-series against modelled ones. Figure 3.14 shows these plots for locations of the four nearshore surface-buoys of the GOL.

The plots emphasize a satisfactory fit quality for these near-shore locations. Even so, both Espiguette and Sete locations seems to be slightly underestimated by

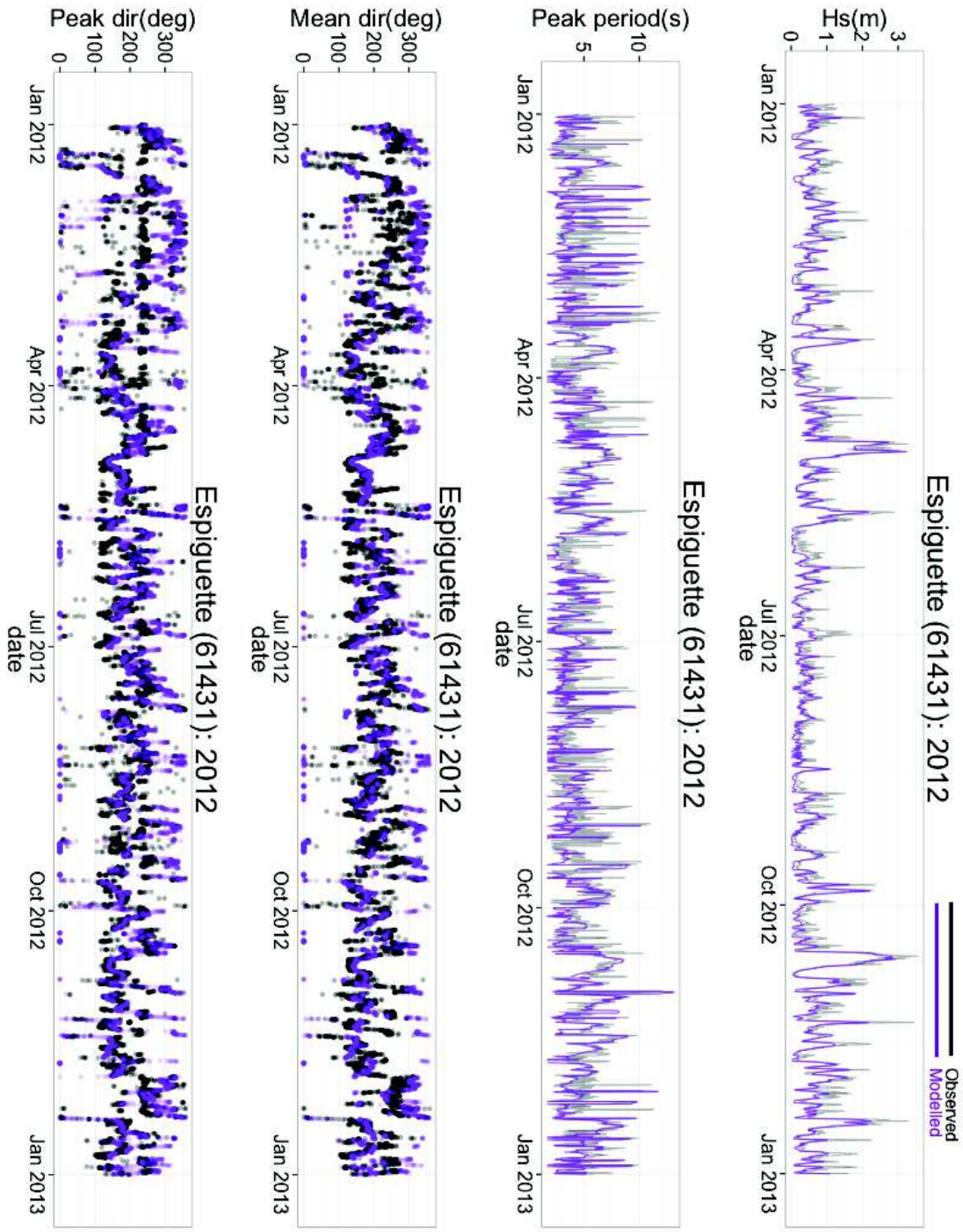


Figure 3.9: Validation of modelled time-series against observation at Espiguette for 2012.

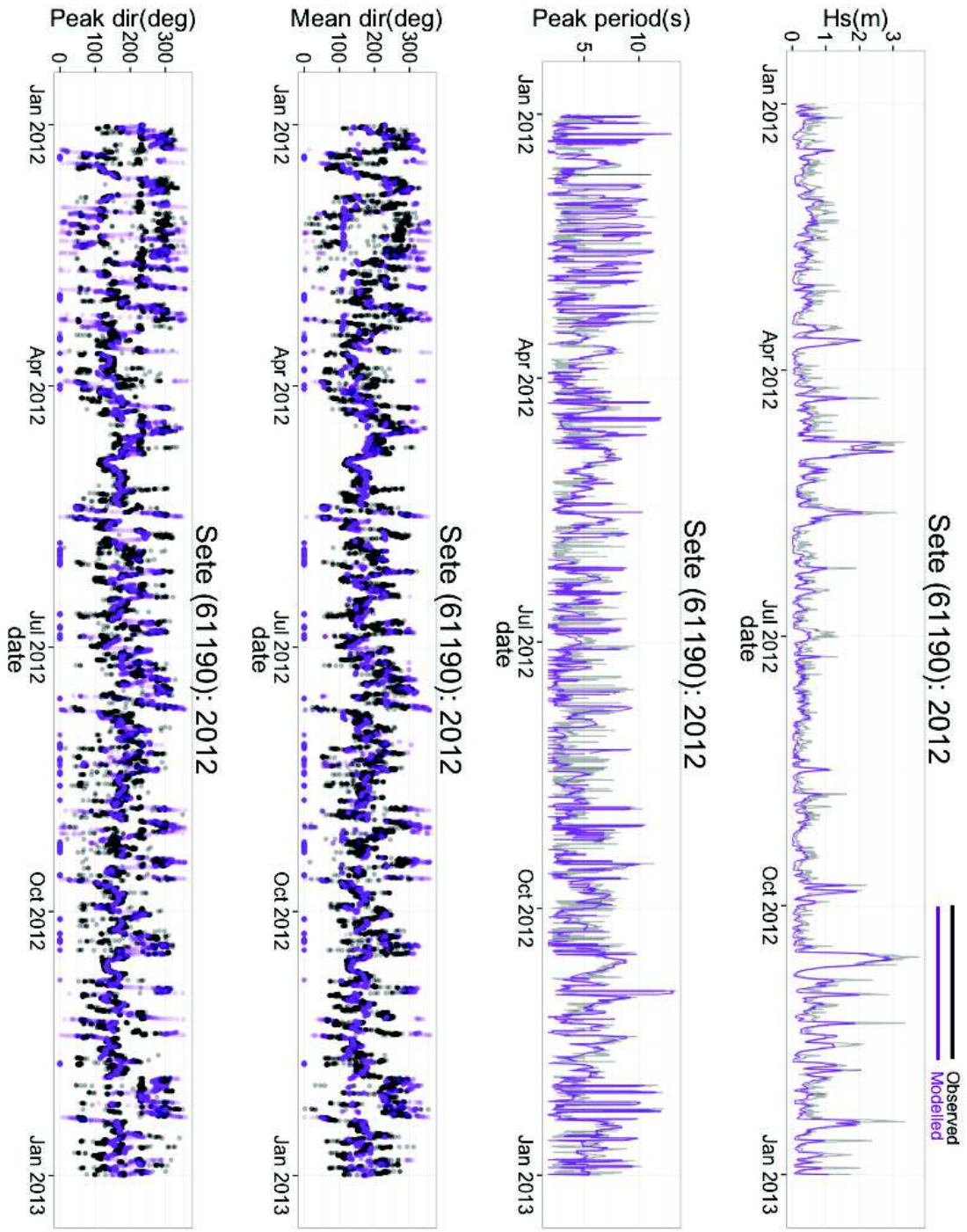


Figure 3.10: Validation of modelled time-series against observation at Sete for 2012.

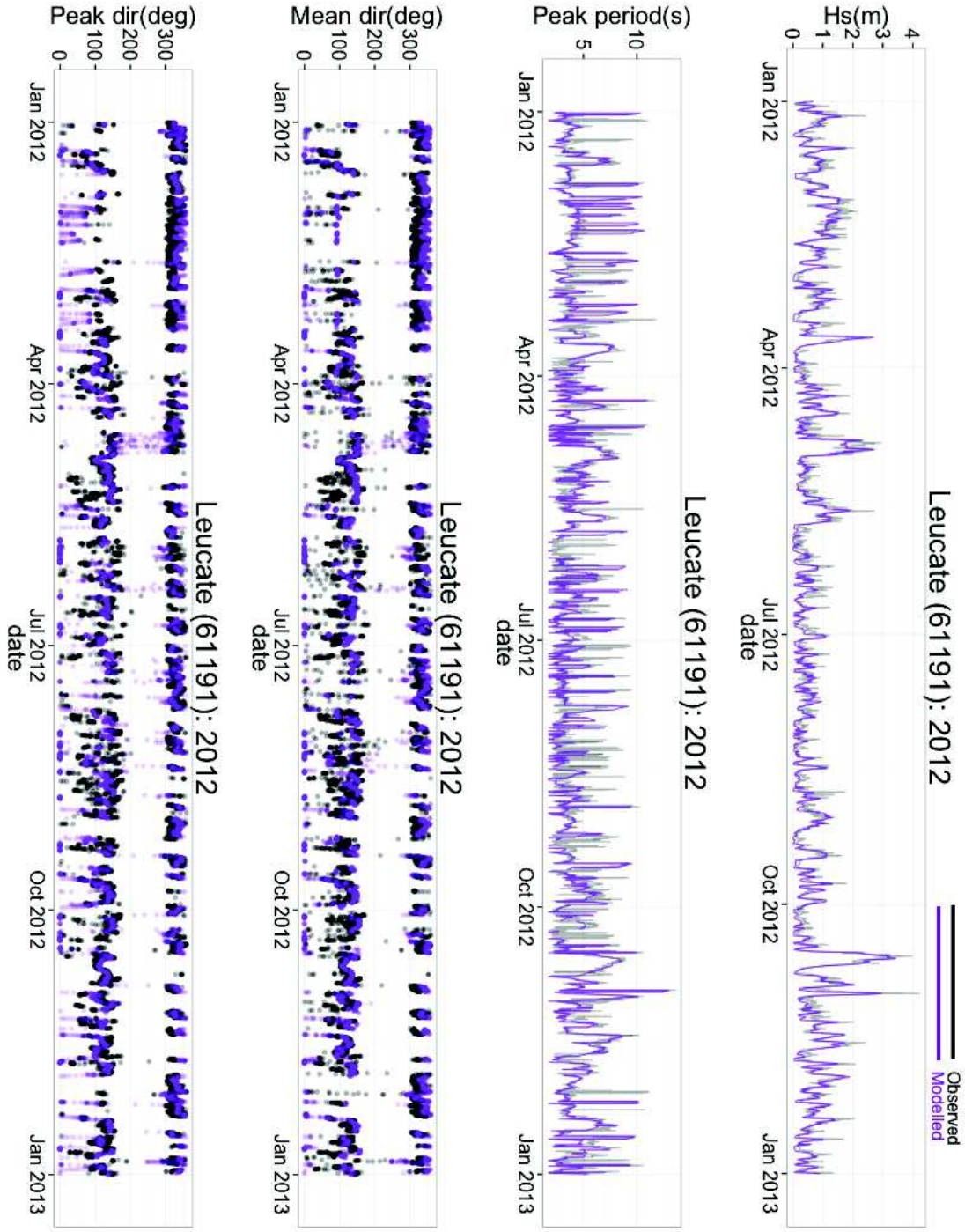


Figure 3.11: Validation of modelled time-series against observation at Leucate for 2012.

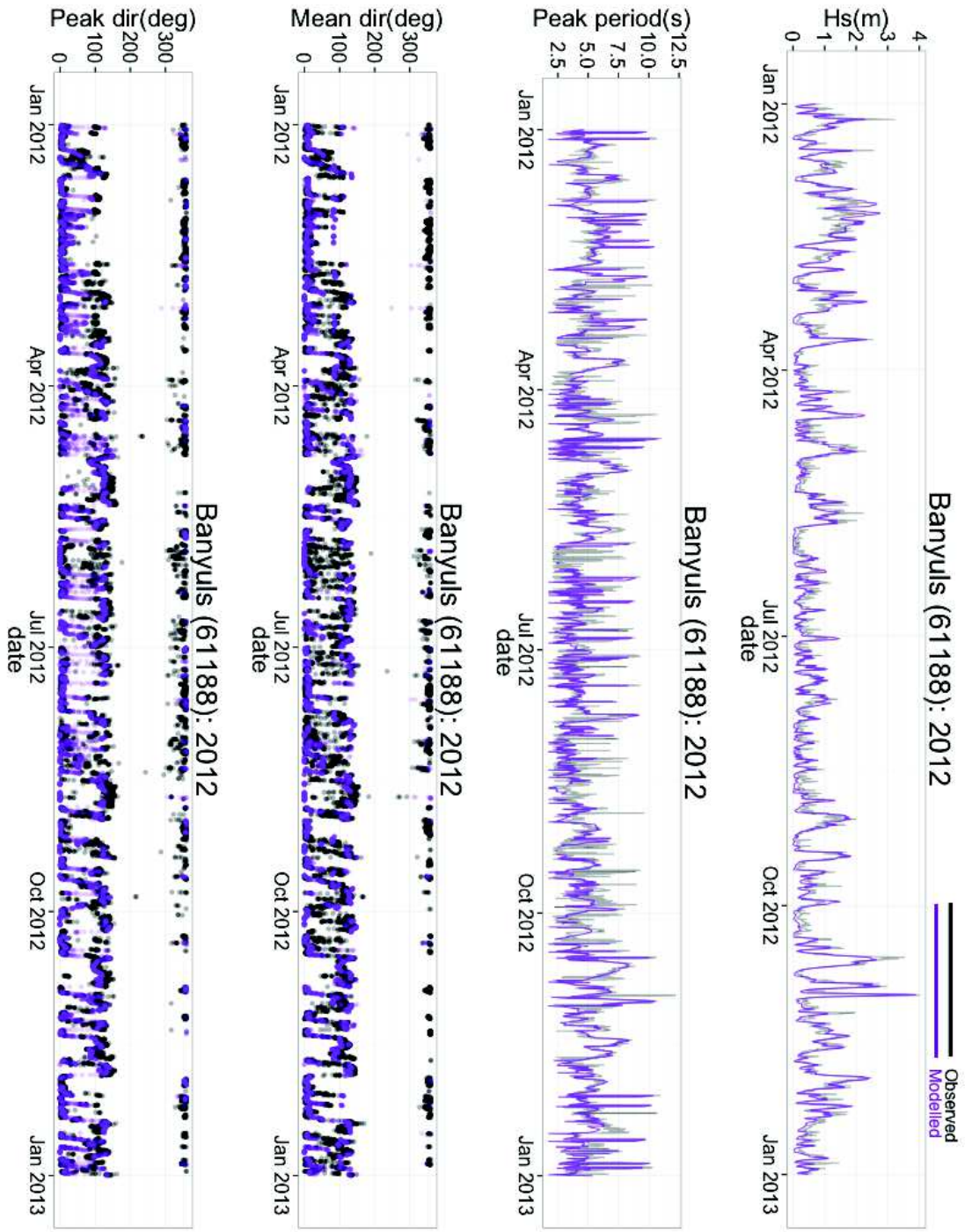


Figure 3.12: Validation of modelled time-series against observation at Banyuls for 2012.

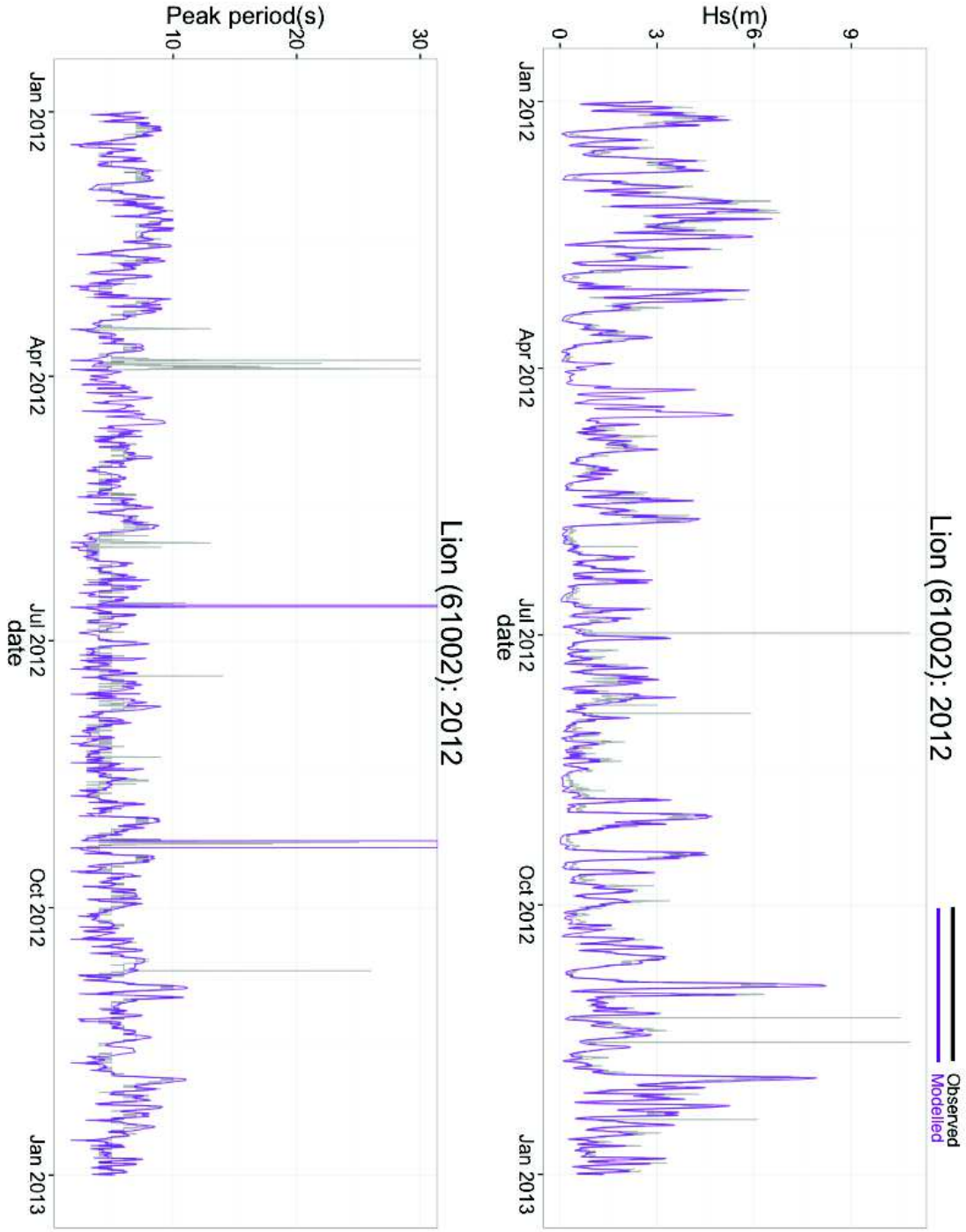


Figure 3.13: Validation of modelled time-series against observation at Lion for 2012.

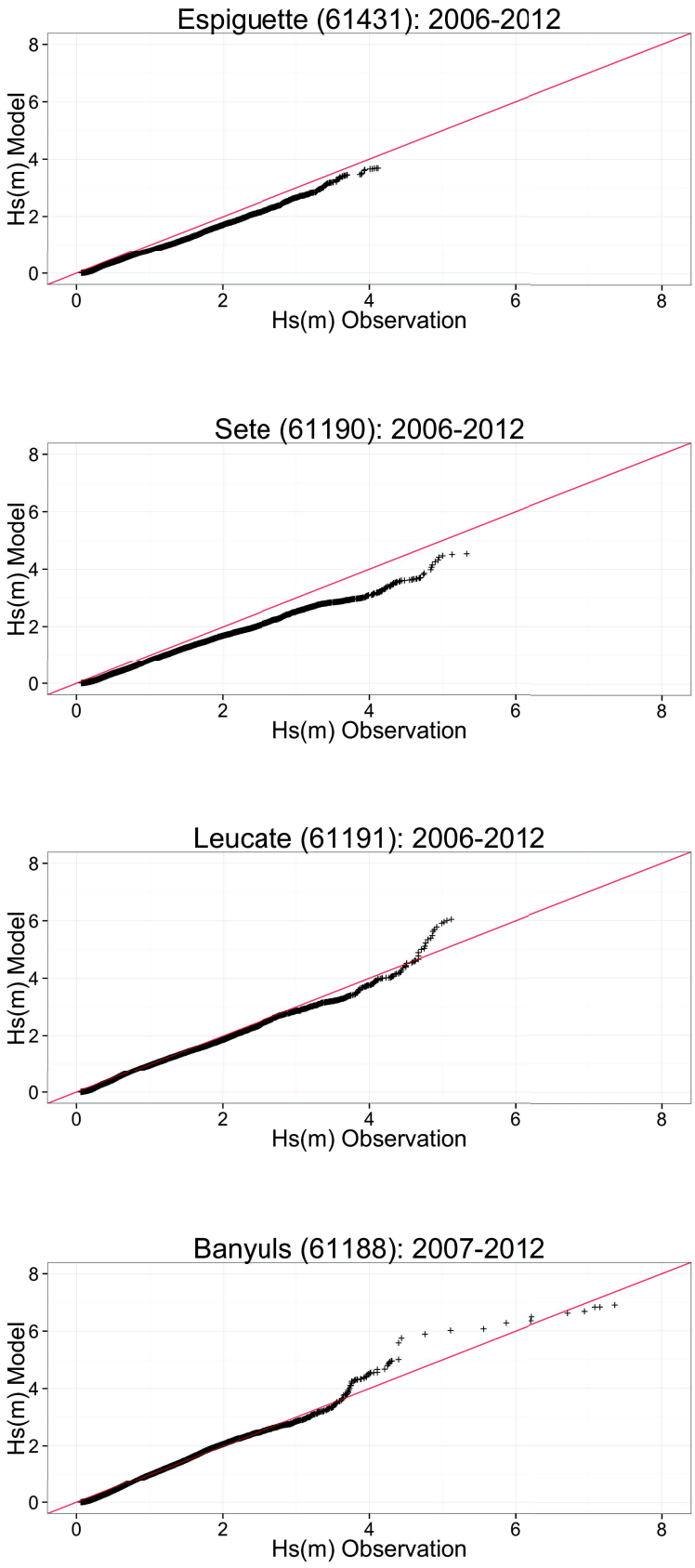


Figure 3.14: Quantile-quantile plot of observed time-series against modelled time-series for the four littoral locations.

the simulation for the high quantiles. In the opposite, data of Leucate station are rather overestimated for these high quantiles. At Banyuls, significant waves height > 3.8 m are overestimated but the fit is better for very high quantiles.

Regarding the Lion's surface-buoy, a preliminary step has to be conducted to remove outliers data but after computation of indexes from Table 3.2. Indeed a set of values are abruptly changing from low values to very high, as referenced in Table 3.3. From a physical point of view these data are not representative of (standard) evolution of sea-states and these data are considered as outliers. Many reason may explain those, like a boat passing by or a maintenance operation or even a sensor issue.
The pointed peaks referenced in Table 3.3 are removed from the time-series before displaying the quantile-quantile plot at Lion location in Figure 3.15. According to this process, the hindcast modelling reinforces the quality of the direct observations.
This last quantile-quantile plot shows again a slight underestimation for high quantile, but it is remarkable that the wave model is able to simulate very high waves, since the highest significant wave height modelled at Lion surface buoy location reaches 8.21 (m).

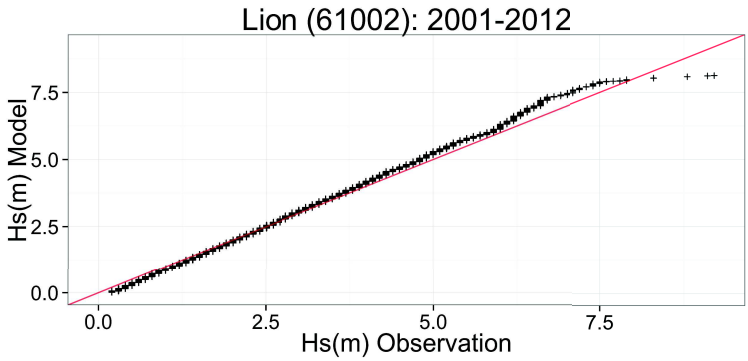


Figure 3.15: Quantile-quantile plot of observed time-series against modelled time-series for the Lion surface buoy.

3.7 Conclusion

Along this chapter the hindcast alternative has been explored to get a reliable data set over the GOL. A 52-year sea-states data set is built with cutting edge models and regional reanalyses forcing fields. The data presented here are generally a little bit underestimated or more rarely overestimated for high quantiles.
In the case of the Lion's surface buoy, some outliers data are identified. Presented assumptions of error measure might be verified in using another source of observation, like for instance validating the data with the use of satellite-altimeters.

Row	Date	Obs. Hs (m)	Obs. Ts (s)	Model Hs (m)	Model Tp (s)
356	2001-12-22 12:00:00	2.0	6	1.413306	5.560783
357	2001-12-22 13:00:00	11.9	6	1.379559	5.512919
358	2001-12-22 14:00:00	1.7	6	1.355907	5.420606
11418	2003-04-13 17:00:00	1.1	4	0.8300802	3.870950
11419	2003-04-13 18:00:00	11.1	4	0.8826900	3.981634
11420	2003-04-13 19:00:00	1.3	5	0.9372973	4.161441
15512	2003-10-18 19:00:00	5.0	10	4.054460	9.864575
15513	2003-10-18 20:00:00	9.4	9	3.893597	9.798878
15514	2003-10-18 21:00:00	4.2	9	3.727370	9.692098
24455	2004-11-03 05:00:00	1.2	5	1.130726	4.564687
24456	2004-11-03 06:00:00	11.1	5	1.117556	6.514265
24457	2004-11-03 07:00:00	1.1	5	1.090516	6.643392
27364	2005-03-07 12:00:00	5.0	8	4.544047	8.984739
27365	2005-03-07 13:00:00	9.8	8	4.498085	8.952805
27366	2005-03-07 14:00:00	4.0	8	4.492004	8.927857
29410	2006-06-01 04:00:00	3.7	7	3.636850	8.028714
29411	2006-06-01 05:00:00	13.7	7	3.616661	8.024514
29412	2006-06-01 06:00:00	3.9	7	3.593398	8.017008
51064	2009-02-24 21:00:00	2.7	6	1.638565	6.221772
51065	2009-02-25 00:00:00	13.5	19	1.500213	5.943340
51066	2009-02-25 03:00:00	2.0	6	1.325276	5.718832
82353	2012-11-16 21:00:00	0.8	4	0.6338978	4.833770
82354	2012-11-16 22:00:00	10.8	4	0.6431389	4.793915
82355	2012-11-16 23:00:00	0.9	4	0.6539590	4.754400
51056	2009-02-23 13:00:00	2.5	6	4.567213	8.933917
51057	2009-02-23 18:00:00	10.7	2	4.346717	8.590184
51058	2009-02-24 03:00:00	3.7	7	4.113022	8.409851
51064	2009-02-24 21:00:00	2.7	6	1.6385646	6.221772
51065	2009-02-25 00:00:00	13.5	19	1.5002128	5.943340
51066	2009-02-25 03:00:00	2.0	6	1.3252755	5.718832
51067	2009-02-25 06:00:00	10.3	10	1.0896575	5.537393
51068	2009-02-25 12:00:00	5.8	15	0.7681069	4.650116
72491	2011-08-30 02:00:00	0.3	5	0.2636432	2.395410
72492	2011-08-30 03:00:00	10.4	5	0.2596883	2.421481
72493	2011-08-30 04:00:00	0.4	5	0.2543055	2.434690
79158	2012-06-30 06:00:00	0.8	4	0.4882389	3.934016
79159	2012-06-30 07:00:00	10.8	4	0.4991300	3.926865
79160	2012-06-30 08:00:00	0.8	4	0.5181398	3.917089
82169	2012-11-08 12:00:00	0.5	4	0.2791750	2.431106
82170	2012-11-08 13:00:00	10.5	4	0.2732719	2.465889
82171	2012-11-08 14:00:00	0.5	4	0.2618464	2.508744

Table 3.3: Outlier measures (highlighted) at the Lion buoy.

Such a validation would reinforce the validation of the hindcast, on longer time-series than the one observed by surface buoys. However it has not been performed due to a lack of time.
Despite this remark, this hindcast data-set is considered in the sequel as satisfactory to pursue the presented work.

Chapter 4

Spatial Extreme Waves Modelling

Chapter Summary

This chapter presents a statistical extreme value study, aiming at addressing Climate-scale hazard questionings over an entire coastal zone (see Figure 4.1). This application relies on the use of the so-called max-stable models, particularly adapted to the study of extreme values in a spatial context.

Resting on the accurate historical data set of wave sea-states presented in the previous chapter, the case study is the Gulf of Lions (GOL). A preliminary study is first realised, particularly to diagnose the extremal dependence embedded in the observed extreme processes.

Then, we discuss the fit of several max-stable models on the data, and proceed to the selection of the best one. A risk analysis is finally performed in order to show the benefit of such spatial modelling in the assessment of extreme value quantities of interest. Most notably, joint probabilities of exceedances are computed on several locations, outperforming the theoretical results assuming the full dependence or the independence.

4.1 Introduction

We focus here on the development of a methodology to assess long-term coastal hazard. This chapter is therefore largely inspired from Chailan et al. (2014), which

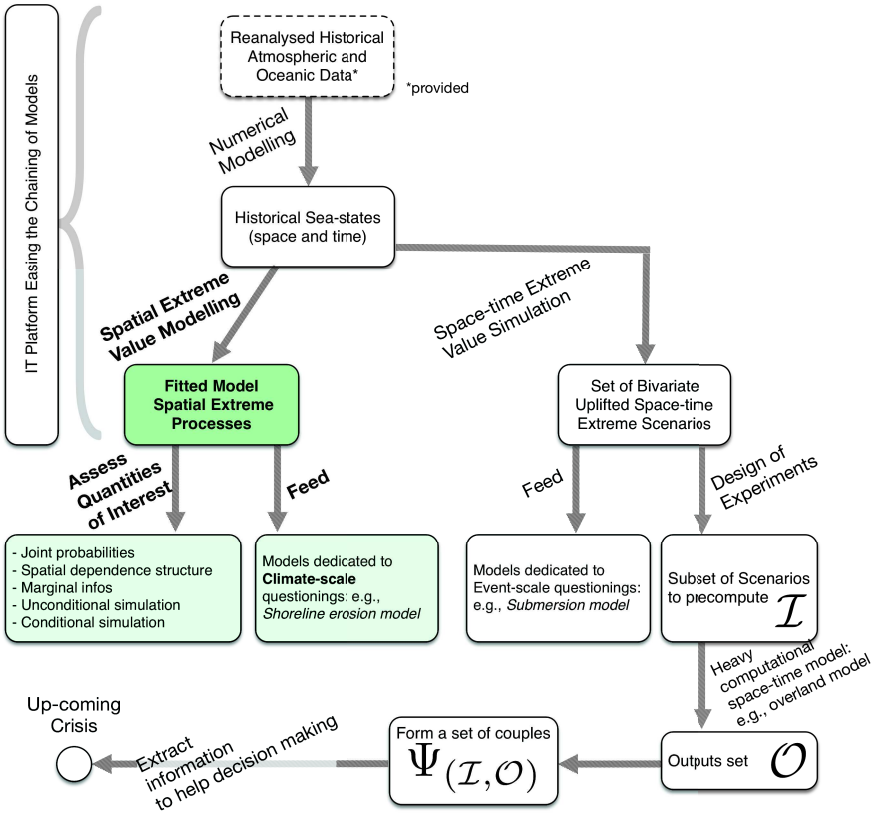


Figure 4.1: This Chapter 4 embraces a statistical spatial extreme value study, aiming at addressing Climate-scale coastal hazard questionings.

is the original work covering this topic.
Relying on the previously presented hindcast data set, we propose to apply the max-stable theoretical framework (Section 1.4.1) in order to study extreme waves of the Gulf of Lions, by modelling extreme significant wave heights (H_s), paying attention to the spatial dependence.
The domain S of this case study is the GOL (Figure 4.2). In the sequel, $\{Z(s), s \in S \subset \mathbb{R}^2\}$ denotes a spatial process of monthly maxima of significant wave heights over the GOL, taking its values in $\mathbb{R}$. Z is in fact observed at location $s \in \mathcal{M} \subset S$ where $\mathcal{M}$ is the set of locations of the computational mesh nodes.

4.2 Preliminary Analysis

Before entering in the modelling methodology, let us introduce some characteristics of the GOL.
Waves mainly originated from winds. In the GOL region, three prevailing winds are observed. First, the *Tramontane* is a strong and gusty wind having a North-West to South-East direction, which hits the west part of the GOL. Secondly, the *Mistral*, also a strong and gusty wind but coming from the North to the South,

hits the north part of the GOL. Finally, the last dominant wind is the *Marin*, which is a wind coming from the sea, i.e. East or South-East. Due to the limited fetch areas, the two first winds grow waves at off-shore locations only. In the opposite, since the GOL has a specific orientation, with a South-East open sea boundary with a large fetch area, the *Marin* is responsible of the creation of the most energetic waves impacting the GOL coastline. However, as observed in Guizien (2009), some locations may be protected thanks to the local coastline configuration. For instance south east waves would not directly impact the Banyuls' location because Cap de Creus provides a protection. In case of swells coming from the East, this cap has no longer effect.

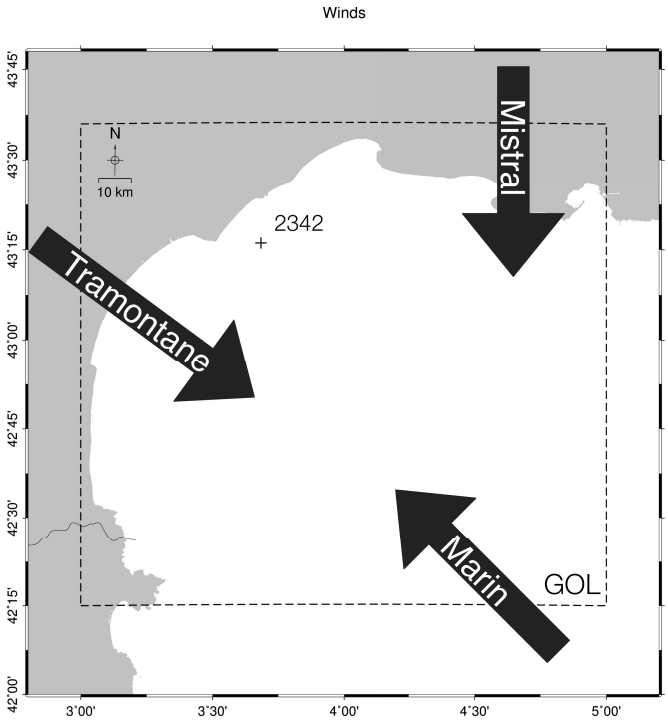


Figure 4.2: The Gulf of Lions, with an South-East open sea boundary and the observed prevailing winds in that region, namely Tramontane (north-westerly), Mistral (northerly), and Marin(south-easterly).

Another wave controlling parameter is the bathymetry. The GOL is composed of an inner continental shelf. Since the wave physic changes between deep and shallow water areas, the sea-bottom may be a factor revealing a spatial discrepancy of the observed wave heights.
Like many environmental phenomena, the significant wave heights observed in the GOL present a real seasonality (Figure 4.3), mostly due to the winds forcing: the previously cited dominants winds are likely to occur between September to April. Guizien (2009) talks also of a seasonality from October to March. As direct consequence, the most energetic waves have been observed in the same time interval since decades. In the sequel and to avoid this seasonality effect in the

modelling, only data through September to April are considered.

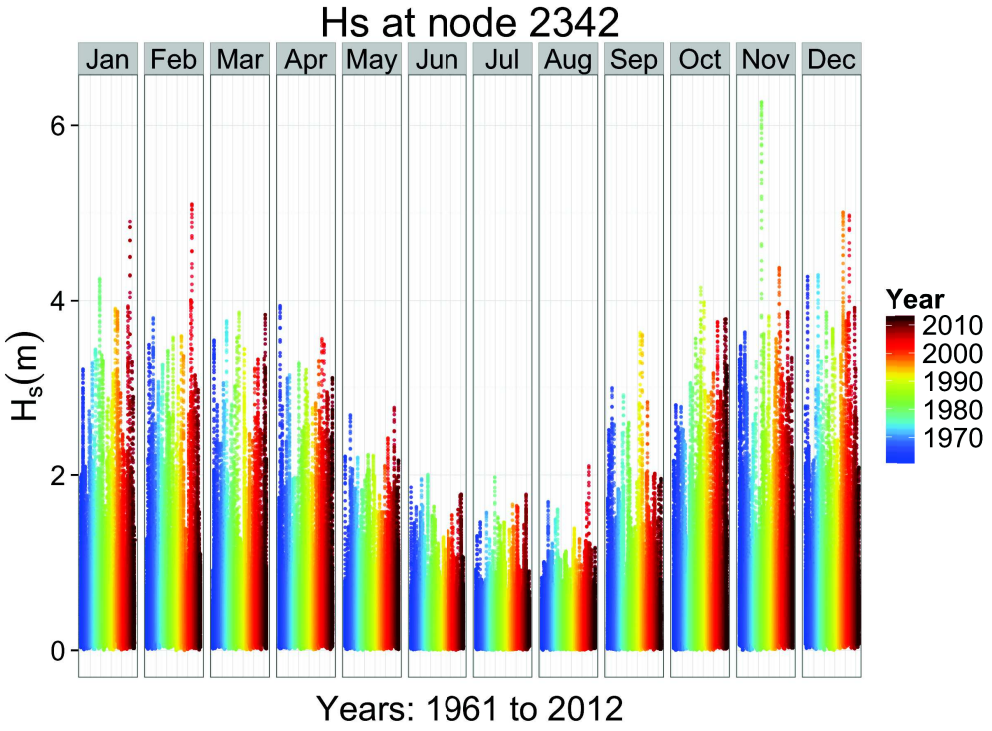


Figure 4.3: Time-series of the significant wave heights from the hindcast data set, at the computational node 2342 and grouped by month. A clear seasonality is observable.

In spatial extreme modelling applications, a good practice is to plot bivariate extremal dependence $\hat{\theta}(h)$ for any pair of locations available, in order to feel what kind of dependence structure rules the observed process (see Section 1.4.1). 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 $\hat{\theta}(h)$, with $h = (z_{ref} - z_j)$, $z_{ref} \neq z_j$, z_{ref} and $z_j \in \mathcal{M}$. A basic diagnostic is then to interpolate the result on the overall area of interest S, as presented in Figure 4.5.
To do so, we first select a subset $\chi \subset \mathcal{M}$ of locations. Borrowed from the domain of numerical modelling exploration, a Latin Hypercube Sampling (LHS) method (McKay et al., 1979) is used to randomly select 100 nodes from 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 4.4.

From the extraction of monthly maxima values of significant wave heights, the bivariate extremal dependence coefficient $\hat{\theta}(h)$ is computed between a set of reference sites and their associated pairs, using the estimator given in Equation 1.16 based on the Fmadogram measure. Then, results are interpolated over all

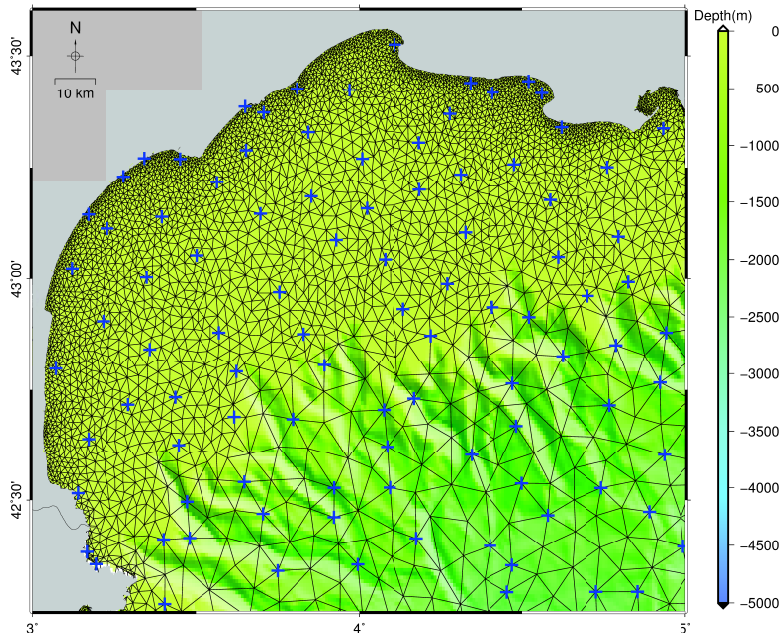


Figure 4.4: The Gulf of Lions, its bathymetry and the computational grid. Crosses are sites selected by an LHS algorithm to optimize the representation of the covered surface. This set denoted χ represents stations used for extreme modelling in the sequel.

the area. Such maps are given in example in Figure 4.5. From those maps several characteristics of the dependence structure can be highlighted.

Noticing that the particular orientation of the GOL allows swells from South-East and East directions to be more easily generated, the plotted estimations of extremal coefficient 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. The first explanation might be the fetch distance induced by two of the prevailing described winds, namely *Tramontane* and *Mistral*.

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 (Figures 4.5(a), 4.5(b), 4.5(e), 4.5(f)). When the reference location is taken in a way that it is exposed swells in all directions as in Figure 4.5(c), the dependence persists both for littoral and off-shore locations.

An other interpretation of the supposed anisotropy may include a more regional explication due to circulation pattern. Indeed the coastline at the extremities of the gulf use to block swells coming from South-West and North-East, which in those conditions impact only offshore sites. Following this circulation reasoning, some littoral areas (Grau du Roi, Beauduc) are far from dependent with some very close stations, because less subject to be impacted by the very energetic South-East swells.

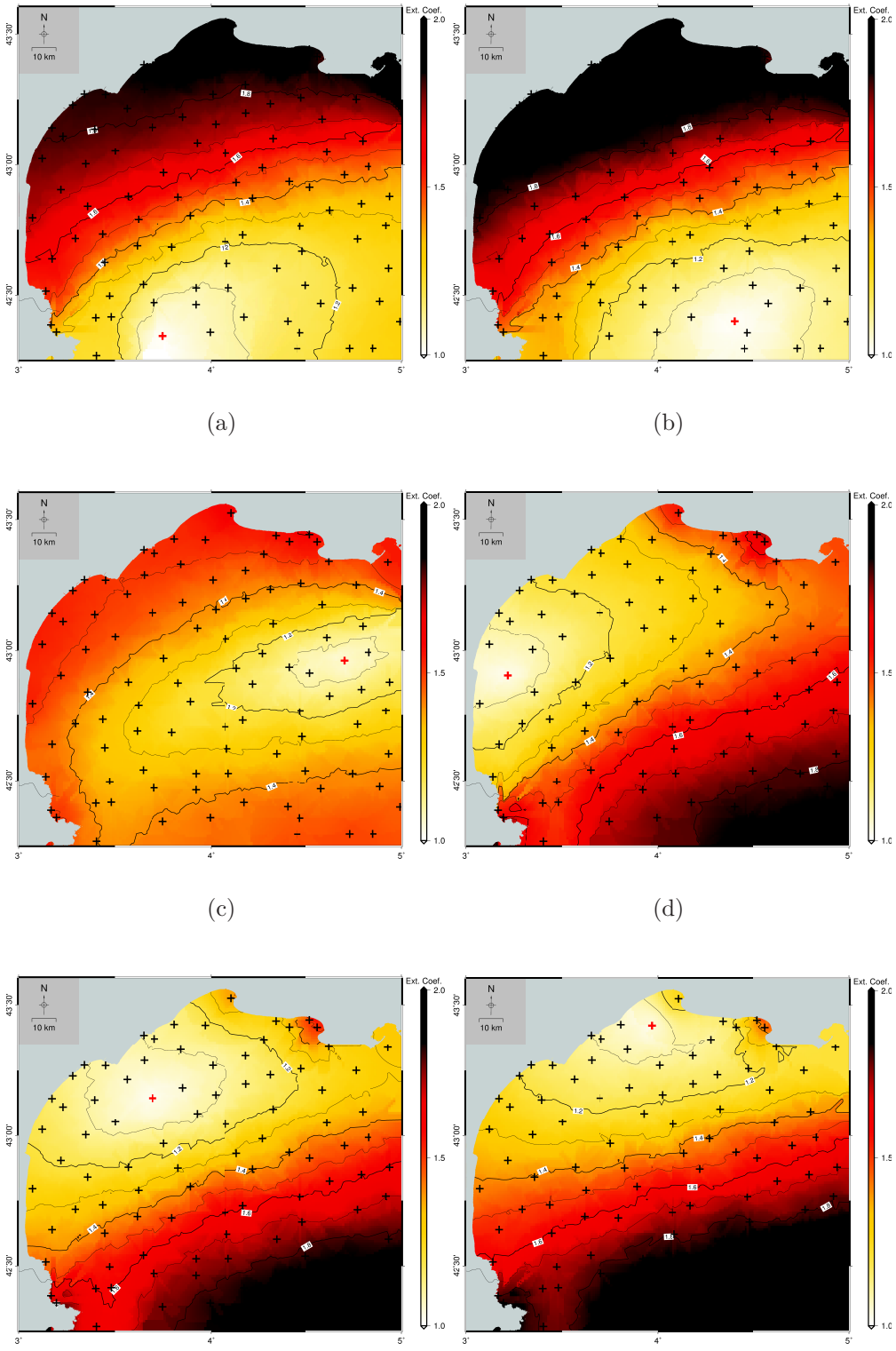


Figure 4.5: Interpolated bivariate extremal coefficient from observed hindcast data set $\hat{\theta}(h)$, with $h = z_{ref}, z_j$, between a reference site (red-cross) z_{ref} and all possible pairs z_j, with $z_{ref} \neq z_j$, z_{ref} and $z_j \in \chi$. Extremal coefficient are estimated by (1.16). Such map allows to feel the underlying dependence structure of extreme wave events of the Gulf of Lions.

4.3 Spatial Extreme Modelling

In this extreme analysis application, we propose to model extreme significant wave heights process, denoted $\{Z(s), s \in S\}$. This modelling has to pay attention to the underlying spatial dependence behaviour of such processes (Section 1.4.1). We propose to apply the max-stable theoretical framework (Section 1.4.1). The fitting procedure of max-stable models is here decomposed in two steps. The first one consists in finding an estimation of the GEV parameters in every locations of the space. This is achieved by using the definition of response surfaces, generally relying on co-variables as the longitude. If marginal GEV parameters are known in all locations of the space, it allows to transform the data back and forth to unit Fréchet margins. Such transformation gives access to the resolution of the second step, being the estimation of the dependence structure. The max-stable models used must be flexible enough to reach the behaviour of the observed dependence structure.
The inference of the parametric max-stable models is hard to afford as soon as the number of sites gets important. Therefore we do not use the entire set $\mathcal{M} \subset S$ of computational node observations, but have to decrease the number of sites considered. Therefore we restrict our methodology to the set $\chi \subset \mathcal{M}$ of 97 sites presented in the previous section.
One may investigate the sensitivity of the approach by increasing the number of sites considered or move their location or both. Such sensitivity analysis remains out of scope for this document.
To avoid seasonality effect, the former section outlines that summer months – *May, June, July, August* – are ignored. Our data set is therefore composed of $n = 52\,(\text{years}) \times 8\,(\text{months}) = 416$ marginal observations denoted $Z_i(x)$, with $i = \{1, \ldots, n\}$ and $x \in \chi$.

4.3.1 Marginal Transformation

Let $Z(s),\ s \in S$ be a random variable denoting the maximum of IID random variables (e.g., significant wave heights) at a site s. Thanks to the robustness of the univariate EVT results in the block-maxima approach applied on short-term dependent series (see Section 1.2.4), we expect that at any site $s \subset S$, $Z(s)$ will follow approximately the well-known and formerly presented Generalised Extreme Value distribution $GEV_{\mu_s,\sigma_s,\xi_s}$ (see Equation 1.1).
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. To simplify expressions and definitions of extreme mathematical objects or models, we use to transform the realisations to a standard extreme distribution. Since by definition an expression of the GEV distribution is given at any site $s \in S$, for all $s \in \chi \subset S$, we transform random variables $Z(s)$ to Fréchet distributed random variables $Z^*(s)$ – i.e. $P(Z^*(s) \leq z) = \exp(-1/z)$,

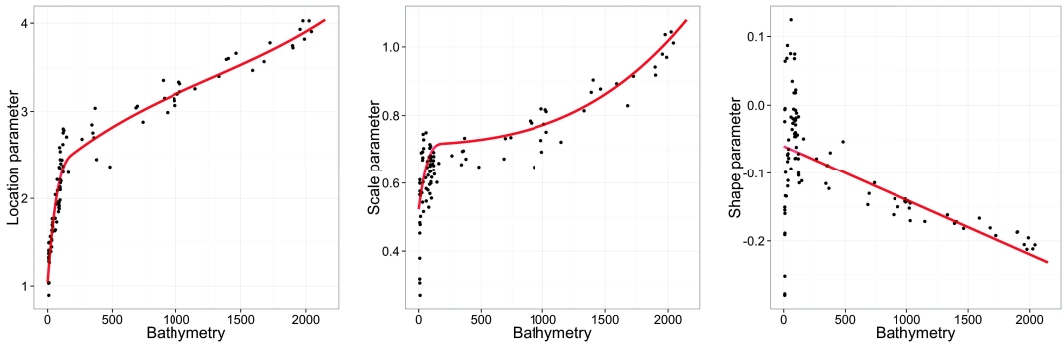
$z \geq 0$. To do so, GEV margins parameters have to be determined.
Several methods exist to define a continuous evolution of the GEV margins parameters through the area of interest.
In the one hand and because the computational grid of the hindcast is highly refined, marginal fits at grid points may lead to a good description of the GEV parameters spatial evolution. Even if this method could require a outliers identification routine due to the fit of models on numerous margins, the estimated parameters could be interpolated through the space of interest to have an expression of the GEV parameters at any point $s \in S$.
In the other hand, one can provide response surfaces to compute the GEV marginal parameters along potential covariates. Indeed while studying environmental phenomena, marginal fits use to show that $\mu(s)$, $\sigma(s)$ and $\xi(s)$ are spatially varying with covariates. 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

$$\mu(s) = \beta_0 + \beta_1 \text{bathy}(s) + \beta_2 \text{lon}(s) + \beta_3 \text{lat}(s). \tag{4.1}$$

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

$$f(s) = \beta_0 + \beta_1 s + \ldots + \beta_{m-1} s^{m-1} + \sum_{j=1}^{q} \beta_{m+j} \, |s_i - \nu_j|^{2m-1}, \tag{4.2}$$

with kernels ν_j of the associated radial basis function β_{m+j}.
Here, we do not compare those techniques but choose to work with response surfaces to reason in term of region instead of a site-by-site analysis. Hence we chose to fit 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 4.1. 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 4.2. Considering so, we end up with 300 combinations of those functions 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 4.6. It appears that the evolution against the bathymetry covariate is much more distinctive than along the longitude and latitude covariates. As described in the preliminary analysis, this can be explained by the physical wave process itself: the GOL presents an inner continental shelf, which directly controls wave formation and propagation. Also, we can notice that marginal distributions of littoral sites



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

Figure 4.6: 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.

have heavier tails than the distributions of offshore sites. The interpretation is much more ambiguous. This discrepancy may come from the fact that in specific (but unknown) configurations, moderate spectra are not transferred in the same way when waves propagate to the inner continental shelf, resulting in observing more energetic (extreme) spectra to the shore.

4.3.2 Model Inference

In this study, inference of the parametric models is performed thanks to the likelihood function. By definition it requires the joint density of any associated finite-dimensional multivariate distributions of the process.

Since the full likelihood is generally unreachable in this context, 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.

Let z_{ik} denotes a realisation of the process of maxima $Z^*(\cdot)$ for the i^{th}-period, at location k. Let us still 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

$$\mathcal{L}(\Psi) = \sum_{i=1}^{n} \sum_{k<l} \log f(z_{ik}, z_{il}; \Psi), \tag{4.3}$$

where $f(\cdot, \cdot)$ is the bivariate density of the max-stable process $Z^*(\cdot)$. 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

$$H(\Psi) = -\sum_{i=1}^{n}\sum_{k<l}\frac{\partial^2 \log f(z_{ik}, z_{il}; \Psi)}{\partial\Psi\partial\Psi^T} \quad \text{and} \tag{4.4}$$

$$J(\hat{\Psi}) = \sum_{i=1}^{n}\sum_{k<l}\frac{\partial \log f(z_{ik}, z_{il}; \Psi)}{\partial\Psi}\frac{\partial \log f(z_{ik}, z_{il}; \Psi)}{\partial\Psi^T}. \tag{4.5}$$

4.3.3 Model Selection

Once all models are fitted, the best one must be selected. Generally, the Akaike Information Criterion (AIC) is used to sort the models by balancing the goodness of their fit against their complexity. This standard use to detect parsimonious model cannot be used in this case because AIC relies on the determination of the full likelihood. An extension of the AIC working with the composite likelihood was introduced by Varin and Vidoni (2005), namely the Composite Likelihood Information Criterion (CLIC). The CLIC is defined by

$$CLIC = -2\mathcal{L}(\hat{\Psi}) + 2\text{tr}\left\{H(\hat{\Psi})^{-1}J(\hat{\Psi})\right\}, \tag{4.6}$$

where H and J are still the observed information and the squared score statistic matrices.
By definition, the lower the CLIC the better the model quality (goodness and parsimony).

4.4 Results

Having previously defined continuous functions to retrieve the GEV parameters over the area of interest (Section 4.3.1), 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 fitted models are reported in Table 4.1.

This table presents two sections. The first one concerns a model fitting an anisotropic dependence structure. The second one concerns models fitting an isotropic dependence structure.
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. Since the global objective of this study is not to compare all existing max-stable models but rather to provide a global methodology for the assessment of spatial extreme long-term coastal haz-

Model (Corr. function)	Isotropic	CLIC
Smith anisotropic	**NO**	**14975454**
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 4.1: CLIC of fitted models.

ards, we choose to not furnish additional efforts to test other anisotropic models. Thus we are going to use the Smith anisotropic fitted model for the following sections.

We are interested in modelling the underlying spatial dependence structure. In that case, one diagnostic of the fit of the model is to plot the bivariate extremal coefficient $\hat{\theta}(h)$ estimated from the empirical Fmadogram, but unlike in the preliminary analysis, the data used here are some realisations of max-stable processes simulated from the selected model. Such maps are represented in Figure 4.7 and must be compared to the reference Figure 4.5 from preliminary analysis. The first observation is that the presumed anisotropy revealed in Figure 4.5 is well reproduced by the model. As expected from the definition of a Smith max-stable model (since it cannot reach the asymptotic independence), the dependence is much more conserved even for long distances in 4.7. However the model has some difficulties to well reproduce the dependence in the previously spotted particular littoral areas (Grau du Roi, Beauduc). Generally, the performances of the model is satisfactory.

4.5 Max-Stable Model at Work

Beyond the description of the spatial dependence of extreme events, the fitted model can be used in several ways in order to assess coastal hazards of the studied region. Some of those are presented in this section.

4.5.1 Simulation of Spatial Extreme Processes

As reviewed in Section 1.4.1, one use of the modelling would be to stochastically simulate extreme processes. Simulations might be conditional or unconditional depending of the questionings, but they are always seamlessly performed at a low computational cost compared to the one of physical numerical modelling used for the hindcast. Figure 4.8 presents four unconditional simulated processes from the previously selected model. Each one represents a realisation of monthly

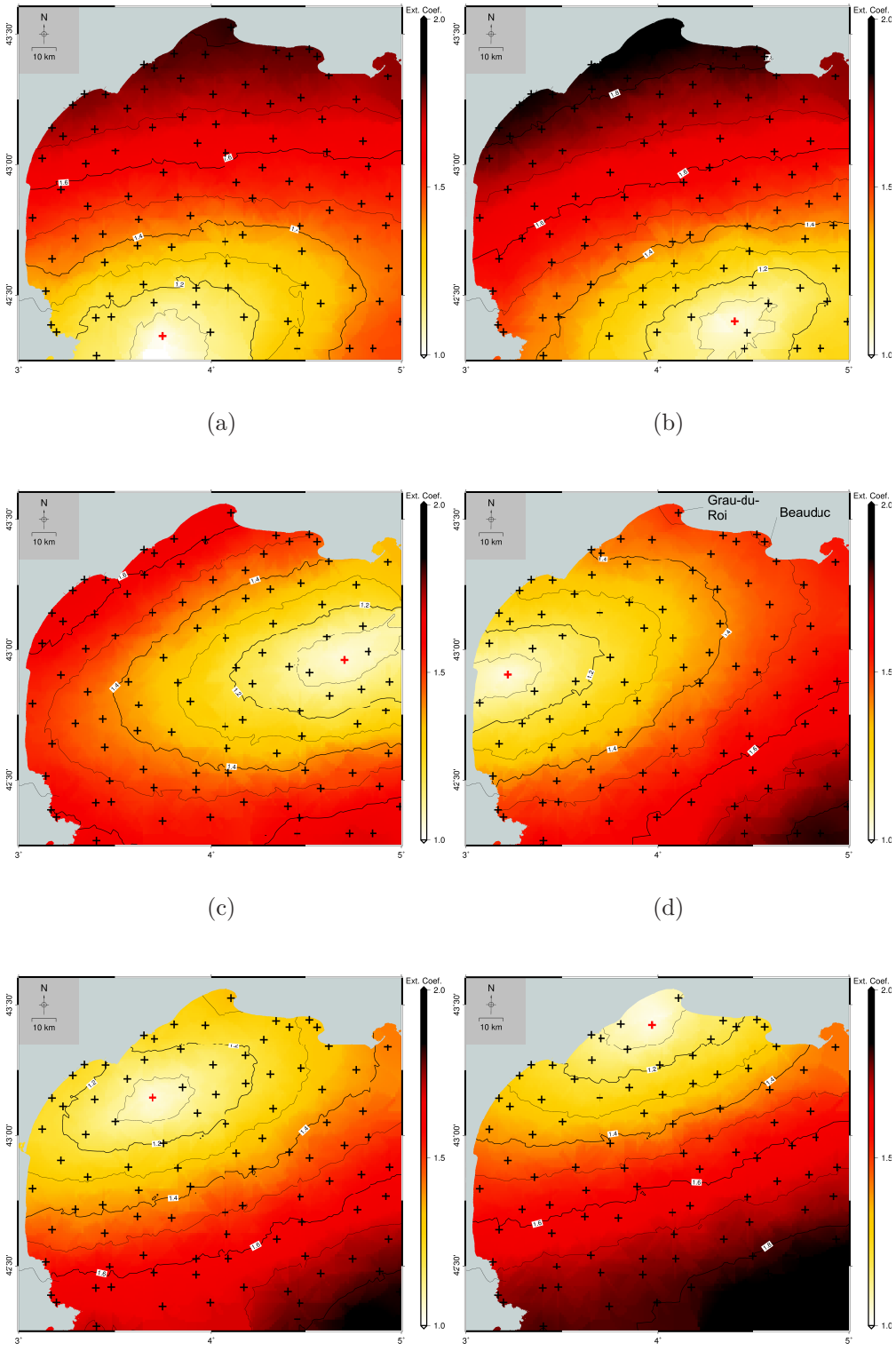


Figure 4.7: Interpolated bivariate extremal coefficient $\hat{\theta}(h)$ estimated from the empirical Fmadogram from 500 realisations of max-stable processes simulated from the best fitted model.

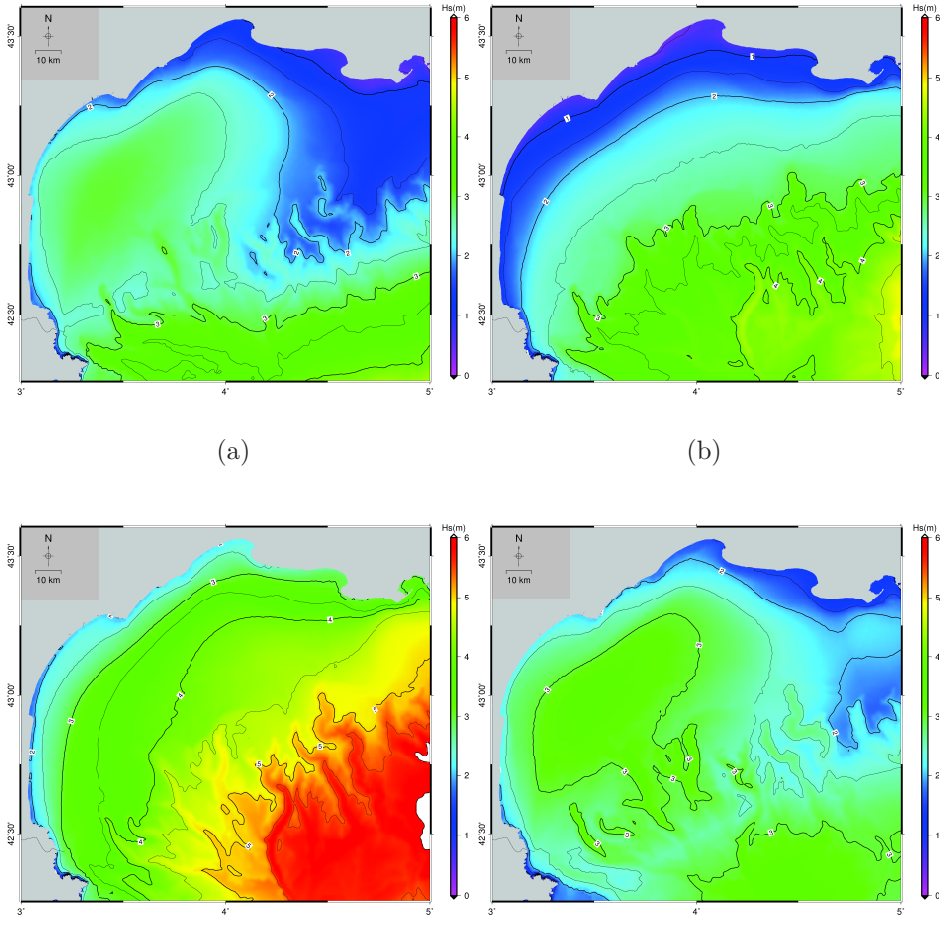
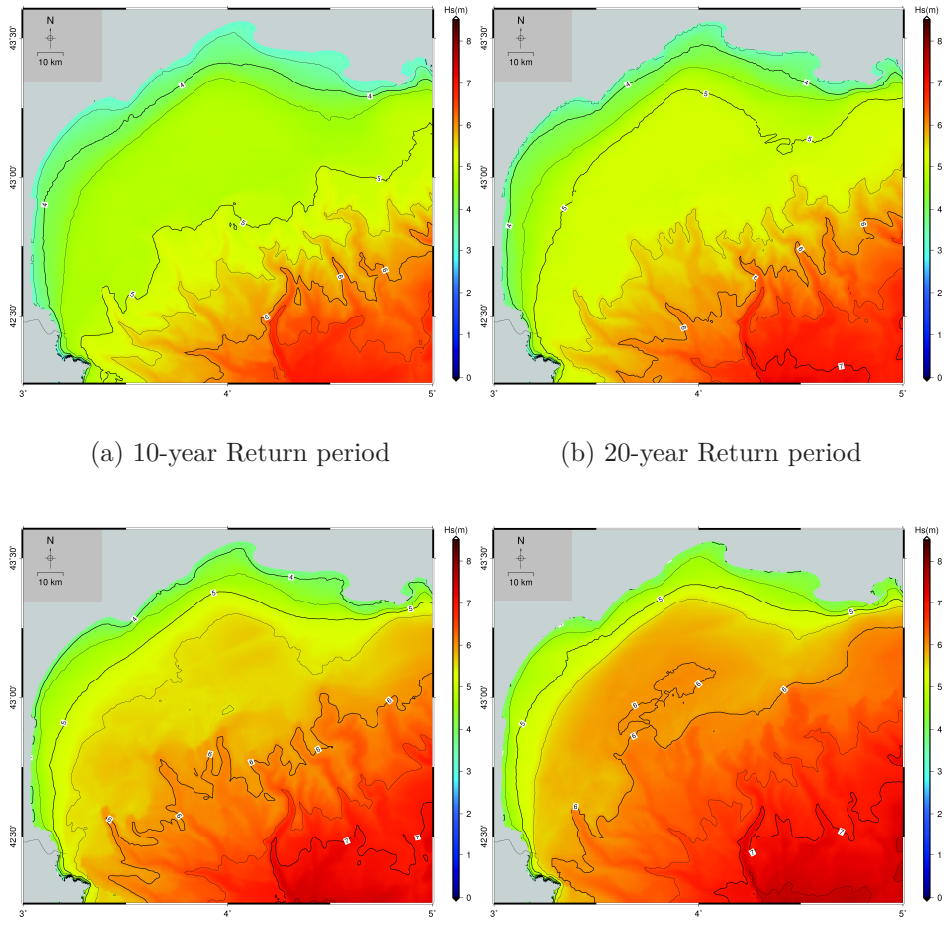


Figure 4.8: Four simulated processes from the best fitted max-stable model.

maxima significant wave heights over the GOL. These simulations takes into account the modelled underlying dependence structure of extremes.
One may feed physical littoral model responding to long-term questionings with such processes. For instance, since sediment transport is calculated from explicit formalisms that require waves features as main input parameters (e.g., Kamphuis, 1991), long-term shoreline change responding to extreme waves events can be forced with such processes. One may validate the simulated processes in terms of physics before.

4.5.2 Marginal Return Levels



(a) 10-year Return period

(b) 20-year Return period

(c) 50-year Return period

(d) 100-year Return period

Figure 4.9: Empirical marginal return levels computed from simulation of 5000 max-stable processes simulated from the best fitted model.

In coastal engineering, another paramount quantity in the dimensioning of structures or environmental studies are the return revels. From the simulated

processes implicitly taking into account the modelled dependence structure, the return levels are 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 simulate 5000 max-stable processes and then empirically derive the marginal return levels over the GOL. The results for 10, 20, 50 and 100 years are presented in Figure 4.9.

4.5.3 Risk Analysis: Joint Probabilities of Exceedances

In risk analysis, a quantity of great importance is the survival joint probabilities. Let us consider in this section that $Z^*(\cdot)$ denotes a process representing monthly maxima of H_s at the Fréchet scale. The survival joint probabilities are defined as $P\left(Z^*(\mathbf{s}) > r_t(\mathbf{s}),\ \mathbf{s} \in \chi^L \subset S^D\right)$ with $r_t(s_i)$ the return level of the t^{th}-period at site s_i, i.e. $P\left(Z^*(s_i) > r_t(s_i)\right) = 1/t$. Such a survival joint probability of exceedances may explain why the L selected sites are impacted (or not) by waves at the same scale during extreme conditions. Indeed, this probability determines whether those sites are more likely subject to observe exceedance of their marginal return level in a same period or not.
We recall that sediment transport computations are based on explicit formalisms requiring waves features as main input parameters. From joint probabilities of exceedances, one can identify some patterns to argue the behaviour of cross-shore or long-shore sediments transport responding to extremes. For instance, two close sites with a low joint probability of exceedances 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 exceedances may represent a more regional behaviour of waves. One can then discuss patterns of sediments transport at a regional and long-term scale.
If we do not rely on a spatial extreme analysis, one may assume the independence or full dependence of sites when computing any multivariate probability as the joint probability of exceedances. Assuming margins are transformed in unit Fréchet distribution, the theoretical joint probability of exceedances considering a full dependence of sites is given by

$$\begin{aligned} P(Z^*(\mathbf{s}) > r_t(\mathbf{s}),\ \mathbf{s} \in \chi^L) &= P(Z^*(\mathbf{s}) > r_t,\ \mathbf{s} \in \chi^L) \\ &= P(Z^*(s_1) > r_t,\ \ldots,\ Z^*(s_L) > r_t) \\ &= P(Z^*(s_k) > r_t), \quad s_k \in \chi \\ &= t^{-1}, \quad \text{by definition,} \end{aligned}$$

while for the total independence it is defined as

$$\begin{aligned} P(Z^*(\mathbf{s}) > r_t(\mathbf{s}), \mathbf{s} \in \chi^L) &= P(Z^*(s_1) > r_t) \times \ldots \times P(Z^*(s_L) > r_t) \\ &= P(Z^*(s) > r_t)^L \\ &= t^{-L}, \end{aligned}$$

with L the number of sites considered.

From several set of L sites, we compare the joint probabilities of exceedances $P(Z^*(\mathbf{s}) > r_t(\mathbf{s}), \mathbf{s} \in \chi^L)$ computed from 1) the theoretical full dependence case; 2) the theoretical total independent case; 3) the observations from the hindcast data set and 4) the best fitted max-stable model through the simulation of 5000 processes. The results are given in Figure 4.10.

Generally speaking the max-stable model outperforms the other theoretical cases for any set of sites and distances considered. When considering littoral sites altogether (Figure 4.10(a)), the model achieves to 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 (Figures 4.10(b), 4.10(c) and 4.10(d)), the model still seems to represent correctly the dependence of the observations. In that case the joint probabilities of exceedances lies in the very between of the dependent and independent case. Finally, when considering sites at any corner of the GOL and far away from each other (Figures 4.10(e) and 4.10(f)), joint probabilities of exceedances quickly drop to 0. It could be interpreted as the probability of having really extreme waves in each part of the GOL within the same period is weak.
In regards to these results against the theoretical independence and full dependence cases, we demonstrate the usefulness of modelling the underlying dependence structure in the assessment of joint quantities in the extremes.

4.6 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 in the goal of being used in risks analyses. For instance we introduce the possibility of stochastically model extreme events feeding models assessing long-terms questionings the capacity to compute accurate joint probabilities of exceedances.

The reliability of such stochastic modelling directly depends on the one of the data set of observed random variables. In our example, the sea-states hindcast must be representative and validated.

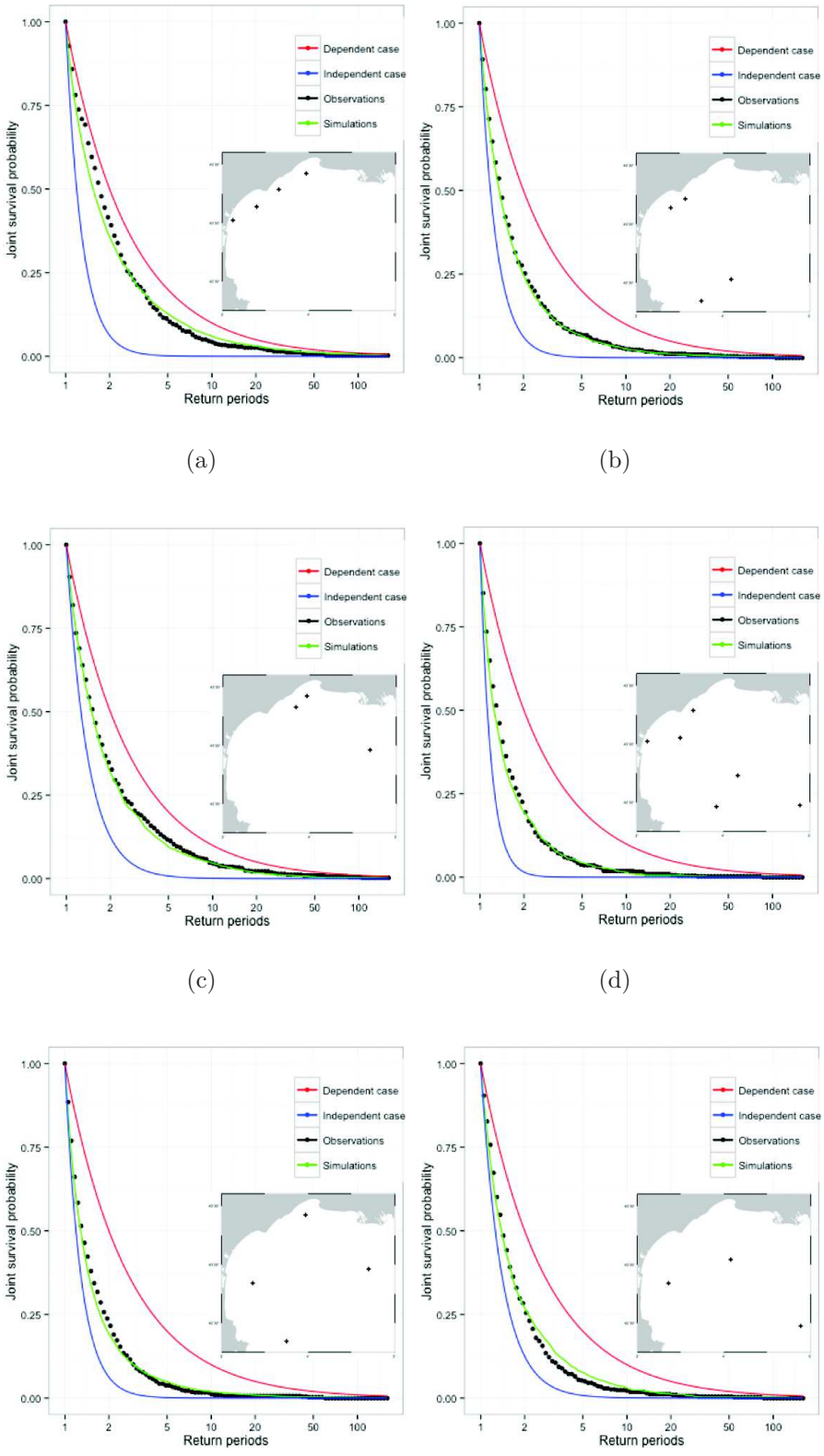


Figure 4.10: Joint probabilities of exceedance for different sites against return periods. Probabilities from observations are dot points and green curve is the probabilities computed from 5000 max-stable simulations of the best fitted model (Smith anisotropic). The two remaining lines correspond to theoretical independent and full dependent cases.

One other limit is that our approach is described in two steps: first the GEV margins parameters (with any kind of potential trends) and then the dependence structure parameters are estimated in different inference steps. Those can be estimated in a single step, having the advantage of getting a systematic way to compare the fitted models (CLIC) and preserve the parsimony of the one selected. Facing issues in the inference routine when dealing with all those parameters lead us to do not consider this approach.

We restrict our study to a limited number of max-stable models regarding the ones known in the literature. In particular we fit only one anisotropic model: the Smith model. To pursue investigations in fitting most complex max-stable models would be a valuable enhancement, most notably by using other models able to handle anisotropic cases as well. However, in regard to the main objective of the thesis, being the development of a methodology helping the assessment of coastal hazards, we preferred to deal this few numbers of models. This choice is reinforced because the selected model shows good performances, particularly in the assessment of the dependence structures.

In preliminary analysis we state that along the bathymetry, direction of waves origin and fetch distances are two controlling factors of extreme waves occurrence. It would be a valuable add to investigate such covariate in the modelling, whether by fixing it (i.e., one fit for one wave-direction sector) or by including it into the model. The latter is one of our perspective research.
In this study, extremes contained in the data are identified through maxima over block period. Like in the lower order cases, it would be worthwhile to use exceedances over threshold methods to catch more information from the extremes observed and obtain a better inference process.

If a part of ocean engineering questionings could or ought to be assessed at a long-term scale, some of them are not. Most notably, questionings where the time evolution of the extreme events is of concerns (e.g., submersion). The proposed method along this chapter is not suitable for such event-scale questioning due to the limited physical interpretation of the simulated processes.
However, these event-scale questionings require efficient (prompt and accurate) helps for decision making. In this sense and from an industrial point of view, methods to help the assessment of event-scales questionings are paramount.
Even if some improvements can be realised in the proposed methodology, we favour to move onto the space-time questionings challenge in the following.

Chapter 5

Space-time Extreme Waves Simulation

Chapter Summary

This chapter introduces a semi-parametric methodology to simulate space-time scenarios of extreme waves (see Figure 5.1). Providing a control on the extremeness of such simulated scenarios, they are intended to help the anticipation of coastal hazards at an event-scale. From the proposed methodology, assessment of hazards can include the time evolution of extreme events additionally to their spatial behaviour, which are a mandatory information in respect to those event-scale questionings.

In the following, the notion of extreme space-time processes defined from a threshold-based method is detailed. We also detail our motivations of simulation of those. To illustrate the benefits of the presented approach, a case study is given. It still concerns extreme wave processes of the GOL area. Observations used still stem from the proposed hindcast data set. On the basis of these data, a (second) preliminary analysis is performed to pay attention to the underlying dependence structure, but regarding the time dimension within the observed processes as well.

Then, the methodology about the simulation of extreme space-time processes is largely detailed and results of some simulated extreme scenarios are discussed.

To go further, a risk analysis is performed, relying on these simulated scenarios and implying a simple coastal physical model.

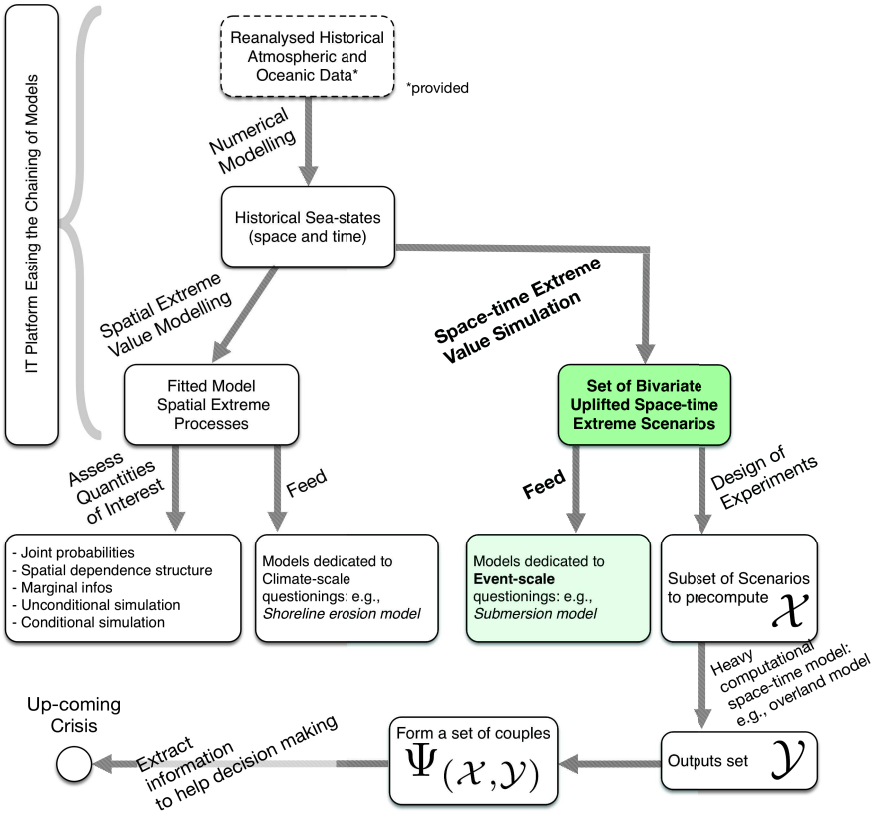


Figure 5.1: Chapter 5 presents a semi-parametric statistical methodology to simulate extreme space-time processes from observed intense ones. The final goal being to feed coastal physical model studying event-scale coastal hazard questionings, like the submersion phenomenon.

5.1 Introduction

Coastal hazards analyses are of prime importance in regard to the highly valuable stakes involved in their anticipation and management. In this chapter, largely inspired from our original work introduced in Chailan et al. (2015), we propose to address event-scale questionings, like the submersion phenomenon along an entire coastline.

Since they represent the main source of damaging energy, event-scale coastal hazards are likely to occur when sea waves conditions are extremes. Our proposition is therefore to obtain extreme wave processes which can be chained to other physical models as open boundary conditions, in order to study the selected event-scale coastal hazard questioning.

One may remark that hydraulic boundary conditions of event-scale coastal physical models are generally defined by three variables describing the sea-states conditions at an instant t: the mean wave direction $\psi(t)$, the significant wave height $H_s(t)$ and the peak wave period $T_p(t)$. Like in the previous chapter, we propose to assess coastal hazards considering the spatial dimension of the questionings. Therefore, since it is paramount to observe both the spatial behaviour and the

time evolution of the analysed process, we are likely to deal with space-time extreme wave processes (composed of H_s, T_p and ψ). Such a space-time process is multivariate, but in the sequel it is defined as extreme when there is at least one exceedance of site marginal threshold of H_s, meaning that a massive amount of (damaging) energy may impact the coastline.

In the case of observable space-time extreme wave events, direct observation methods suffer from the common drawbacks detailed in Chapter 2 (e.g., data scarcity in space and time), plus measuring issues due to the extremeness of the conditions (e.g., maintenance or sensors validity).
Such studies foster instead the use of wave numerical models since their reliability still holds for observable space-time extreme events. As soon as we consider very extreme events, the simulation from those models is generally unreachable. This is due to a lack of knowledge on boundary conditions (atmospheric and ocean) and also on the physical reliability of wave models for such extreme quantities. As an alternative we propose here to use statistical approaches.
Some statistical approaches have been presented to construct extreme scenarios of near-shore conditions like in Gouldby et al. (2014), but are generally not spatial. With max-stable processes, we used in Chapter 4 an approach aiming at addressing coastal long-terms questionings on an entire region or along a full coastline. The restriction to the long-term questionings stems from the physical interpretation of the fields simulated since those spatial extreme models deal with data aggregated along the time dimension (e.g., annual or monthly maxima). As stated in Section 1.5, ones introduced applications of max-stable processes to the space-time context as Huser and Davison (2014). However those applications are scarcely provided. Their capacity to model complex dependence structures can still be challenged. The physical interpretation of the simulated space-time processes issued by these models can be challenged as well.
In the proposed methodology we are focused on a semi-parametric approach stemming from parts of the original work of Caires et al. (2011); Groeneweg et al. (2012); Ferreira and de Haan (2014), summed up as follows.
Let $\{Z(s,t), s \in S,\ t \in \mathcal{T}_0\}$ be a space-time process considered as extreme, with $s \in S \subset \mathbb{R}^2$ the area of interest and $t \in \mathcal{T}_0 \subset \mathbb{R}^+$ the time dimension. For the sake of simplicity these space-time processes are called storms in the following sections.
The idea relies on three steps. First, Z has to be transformed from its original scale to a standard scale as $Z^* = T(Z)$ where T is a site marginal transformation. Then the process is uplifted by a coefficient denoted $\zeta > 1$, controlling the extremeness of the simulated process. Finally this uplifted process is transformed back to its original scale, making the process $T^{\leftarrow}(\zeta T(Z))$ a more extreme process.
This approach is mathematically justified. However it assumes that the space-time dependence structure is constant in the extremes, as it is the case in the context of max-stable process modelling. Caires et al. (2011); Groeneweg et al.

(2012) use this methodology to simulate space-time extreme processes.
We leverage this approach to perform a bivariate simulation of such processes. Indeed our main objective of having representative space-time extreme hydraulic boundary conditions requires to obtain processes of H_s and T_p at extreme levels, while assuming that site marginal mean wave directions are identical in the extreme. We also developed a distinct strategy of selection of storms, also named declustering. Finally and unlike those former studies, marginal distributions used for the standardisation of the data is based on the work of Thibaud and Opitz (2015).

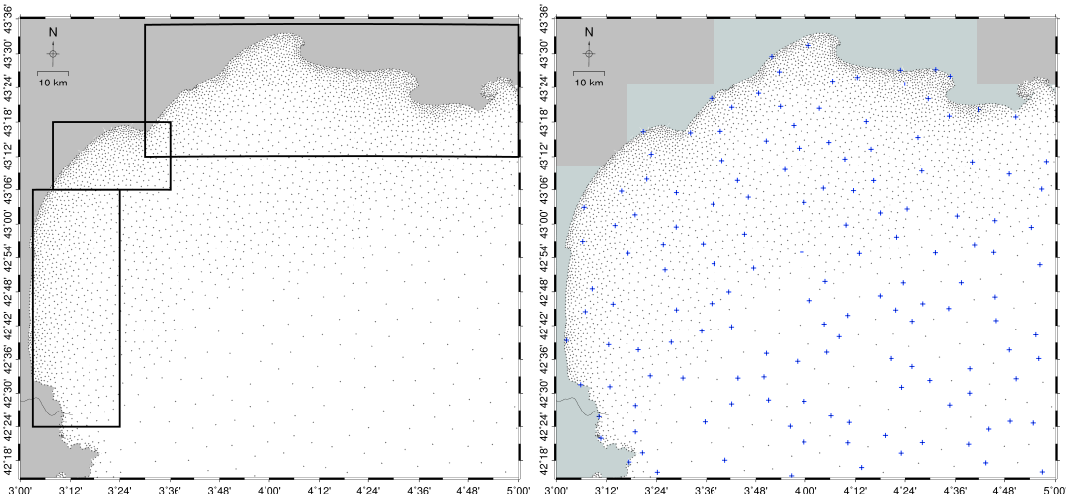
To demonstrate the usefulness of such approach, the behaviour of simulated space-time processes are discussed around a case-study: the quantification of the long-shore mass flux of energy in the GOL coastal area during extreme storms.
Since the presented methodology is applied on a large multidimensional volume of data, specific High Performance Analytics (HPA) algorithms are developed to work on the data, which brings forward an additional technical dimension.

5.2 Preliminary Analysis

The case study is still the Gulf of Lions. Sections 3.1 and 4.2 present the main physical characteristics of this region, where storm waves are likely to occur. In this chapter we have recourse to the same sea-states hindcast data set. However additional information need to be determined since we deal here with space-time events.
Indeed, we recall that we are interesting in working with space-time processes describing the mean wave direction ψ, the significant wave height H_s and the peak wave period T_p. From the hindcast data set, those processes derive from the computed wave energy spectra computed at each node of the mesh. For the GOL only, these three variables represent $3 \times 3944 \times 24 \times 365.25 \times 62 = 6\,430\,597\,344$ observations and are stored in a binary file of 19 GB.

From expert advices and due to the huge amount of energy transported by waves, storms relevant to study coastal hazards questionings are the one in which H_s reaches high values (or exceed a high threshold) inside a very littoral area denoted S^*. For the GOL, we choose the union of the determined areas (Figure 5.2(a)) that are set rectangular for technical constraints.
Like in the previous chapter and in order to observe the dependence structure both in time and space, we estimate pairwise extremal coefficients using an estimator provided by Smith in Caires et al. (2011) (see Equation 1.18), which is suitable in this specific context of threshold-based extreme values.

Relying on this estimator, two extremal coefficients $\hat{\theta}(k)$ and $\hat{\theta}(X, Y)$ are introduced and computed over the sea-states hindcast data set. They are computed



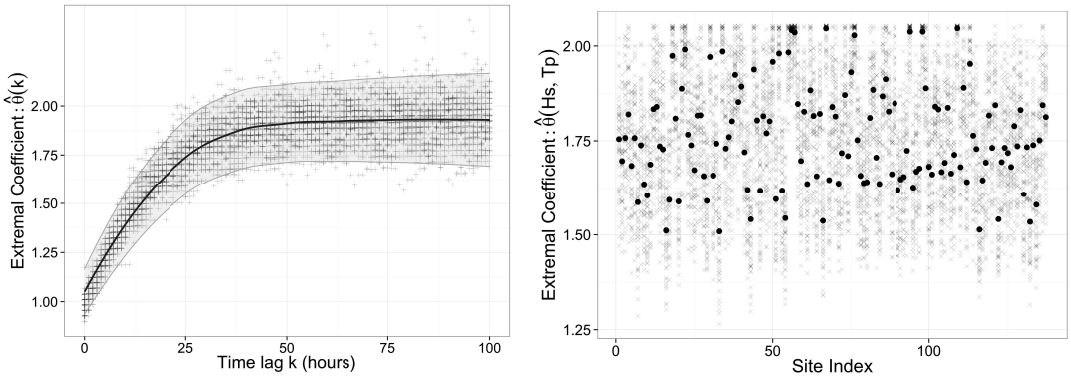
(a) Littoral area $S^* \subset S$ is the union of squared areas. From experts advice, if H_s is high in S^* the coastline is likely to be impacted. Wave data are available at the set of locations of the mesh nodes in this area, which is denoted $\mathcal{M}^* \subseteq S^*$.

(b) Crosses points form a subset χ of 140 sites selected from the locations of the computational mesh nodes. χ is constructed in manner of spatially representing all observations locations.

Figure 5.2: Spatial specification.

from a subset of 140 locations $\chi \subset S$ illustrated in Figure 5.2(b), selected in the same way as in Section 4.2.
The marginal dependence through time is measured by estimating the extremal coefficient $\hat{\theta}(k)$ between observations of a random variable at a single site but separated by a time lag k. The dependence between two random variables (X, Y) observed at a given site at the same time can also be assessed using the estimated extremal coefficient $\hat{\theta}(X, Y)$.

Figure 5.3(a) presents the estimation $\hat{\theta}(k)$ of $\theta(k)$ for pairs (Y_j, Y_{j+k}) where k is the time lag. In this case, $Y_j = \{\max(Y_j^{(s)}), s \in \mathcal{M}^* \subset S\}$ with $\mathcal{M}^*$ the very littoral presented before. The arbitrary choice of $\mathcal{M}^*$ is still related to the final goal of the document: quantifying coastal hazard. Therefore only storms impacting the shoreline area are considered in the measure. We can observe from Figure 5.3(a) that $\hat{\theta}(k)$ narrows 1.9 and becomes almost steady at $k = 50$. Hence, we state the dependence within a storm impacting the littoral will be considered as persistent only up to 50 hours.
Finally, Figure 5.3(b) is the estimation $\hat{\theta}(X, Y)$ between the two wave variables H_s and T_p at locations from the subset χ. From this illustration we can deduce that those two variables remain fairly dependent even within its extreme realisations.



(a) Extremal coefficient estimated between pairs $\{Y_t, Y_{t+k}\}$, with $Y_t = \max_{s\in\mathcal{M}^*}\{Y_{t,s}\}$ and k the lag in hour. The straight line and its shadow envelope are respectively a fitted polynomial regression model and its 95% predict interval.

(b) Extremal coefficient estimated between pairs $\{X_s, Y_s\}$, with X the significant wave height (H_s) and Y the peak wave period (T_p) at location $s \in \chi$. Dots are the median values from yearly estimated pairwise coefficients.

Figure 5.3: Estimations of coefficient extremal $\theta(\cdot)$ (see Equation 1.18) estimated for the full period (1961-2012) of the hindcast.

5.3 Method of Simulation

5.3.1 Extreme Space-Time Processes

In the sequel $\{X(s,t),\ s \in S,\ t \in \mathcal{T}\}$ denotes a random space-time process with S a compact subset of $\mathbb{R}^2$ and $\mathcal{T}$ a compact subset of $\mathbb{R}^+$. Such a random process represents a random variables collection indexed by both space and time which is in the space of continuous real functions on $S \times \mathcal{T}$ denoted $C(S \times \mathcal{T})$. We suppose that the stochastic process of interest is in the domain of attraction of a max-stable process. In other words, we suppose there exist continuous functions $a_n(s,t)$ positive and $b_n(s,t)$ such that the processes $\left\{\max_{1\leq i\leq n} \frac{X_i(s,t)-b_n(s,t)}{a_n(s,t)}\right\}_{(s,t)\in S\times\mathcal{T}}$ with $X_1, \ldots, X_n$ independent copies of X, converge in distribution to a max-stable process η in $C(S\times\mathcal{T})$. Since convergence of marginals and convergence of dependence structure can be split up, we consider in the sequel the standardised process $1 \,/\, \left(1 - G_{X(s,t)}(X(s,t))\right)$ where $G_{X(s,t)}$ corresponds to the distribution of $X(s,t)$. Such a process has marginal standard Pareto distributions and belongs to the domain of attraction of unit Fréchet. Following Thibaud and Opitz (2015), it is convenient to fix a high threshold function $u(s,t)$ and to assume that marginal distributions of this process satisfy

$$P(X(s,t) > x) = [1 + \xi(s,t)(x - \mu(s,t))/\sigma(s,t)]_+^{-1/\xi(s,t)}, \quad x > u(s,t) \qquad (5.1)$$

with real parameters $\mu(s,t) < u(s,t)$, $\sigma(s,t) > 0$ and $\gamma(s,t)$, such that the right-hand of (5.1) is less than unity.

As a consequence we can define more precisely the standardised process X^* we consider as follows

$$X^*(s,t) = T\left(X(s,t)\right) = \left[1 + \xi(s,t)(x - \mu(s,t))/\sigma(s,t)\right]^{1/\xi(s,t)}. \qquad (5.2)$$

5.3.2 Construction of Uplifted Storms

As presented in the introduction, the outline of the methodology lays in four steps. First, data are marginally transformed. It allows to manipulate the data at a standard scale. Here we use a transformation to reach the standard Pareto scale. Then we need to extract storms from the data-set. Once storms are extracted, the data are uplifted to higher values, with a control on the marginal amplification coefficient. Finally the data are transformed back to their original scale by inverting the transformation. In this subsection, details of this methodology are given.

The first step consists in standardising $X(s,t)$ to a standard Pareto scale according to (5.2). In practice, parameters are unknown and need to be estimated. We suppose the threshold and the parameters to be constant over time, depending only on space. One can alternatively use more sophisticated expressions of those quantities to deal with a potential non-stationarity or periodicity (or both) of the process. In each site, parameters estimations $\hat{\mu}(s)$, $\hat{\sigma}(s)$, $\hat{\xi}(s)$ are obtained by the maximum likelihood method using data above a high threshold $u(s)$. Since marginal data may have some short-term dependences they are therefore de-clustered before being used to estimate the parameters. This step allows to reach the independence condition assumed in the estimation procedure. Using such estimators in (5.2), let denote by $\left\{\tilde{X}^*(s,t),\ s \in S,\ t \in \mathcal{T}\right\}$ the considered standardised process.

The second step consists in extracting storms at a standardised scale from the data. Such a storm is a subset in the time dimension of $\left\{\tilde{X}^*(s,t),\ s \in S,\ t \in \mathcal{T}\right\}$, therefore defined as $\tilde{Z}^* = \left\{\tilde{X}^*(s,t),\ s \in S,\ t \in \mathcal{T}_0 \subset \mathcal{T}\right\}$.
Let $\left\{\tilde{Z}_i^*(s,t),\ i \in \{1,\ldots,p\}\right\}$ denotes a collection of such space-time processes and represent the p highest storms available in the – transformed – data-set. To detect those storms, the iterative scheme presented in Algorithm 2 is set up.

In this algorithm, δ is a time value allowing to set the size of a storm and ε is a time value to guarantee the independence of the storms. Those two values are generally defined from experts' advice or from preliminary analyses or both. For each iteration, the maximum value is searched over the subset of sites $\mathcal{M}^*$ which might be a single reference location, locations of the entire space S or locations

Algorithm 2: Storm selection

Input : $\left\{\tilde{X}^*(s,t),\ s \in S,\ t \in \mathcal{T}\right\}$, space-time observations at a standard scale.
p' the maximum of storms to select.

Output: $\{\tilde{Z}_i^*, i \in \{1,\dots,p\}\}$ with $p \leq p'$, a sorted collection of IID storms

begin
  $i = 1$, $\delta \leftarrow$ Cst, $\varepsilon \leftarrow$ Cst, $T \leftarrow \mathcal{T}$, $T' \leftarrow T$
  while $(i \leq p')$ and $(\max_{s\in\mathcal{M}^*, t\in T'} \tilde{X}^*(s,t) > 1)$ **do**
    $t_i \leftarrow \arg\max_t \left\{\tilde{X}^*(s,t)\right\}$ // $s \in \mathcal{M}^* \subseteq S$ and $t \in T'$.
    $\tilde{Z}_i^* \leftarrow \tilde{X}^*(\cdot\,, t)$ with $t \in T \cap [t_i - \delta,\ t_i + \delta]$
    $T' \leftarrow T' \setminus [t_i - \delta - \varepsilon,\ t_i + \delta + \varepsilon]$
    $i = i + 1$
  return $\{\tilde{Z}_1^*, \tilde{Z}_2^*, \dots, \tilde{Z}_p^*\}$

of some area in between.
One would notice that the stop condition of the algorithm implies that in each selected storm, there is at least one exceedance of the site marginal threshold. The algorithm would select storms until the required number of storms p' is reached or that the exceedance condition is no more satisfied.

The set $\left\{\tilde{Z}_i^*, i \in \{1,\dots,p\}\right\}$ forms the collection of storms in the data-set at a standardised scale. It is relevant to compare them from each other in term of their extremeness.
In the sequel, the definition of extremeness of a so-called storm $\{Z^*(s,t),\ s \in S,\ t \in \mathcal{T}_0 \subset \mathcal{T}\}$ relies on the level corresponding to the within-storm maxima $z_{max} = \max_{s,t}\{Z^*(s,t),\ s \in \mathcal{M}^* \subset S,\ t \in \mathcal{T}_0 \subset \mathcal{T}\}$. Consequently, a storm $\{Z_1^*\}$ is considered as more extreme than $\{Z_2^*\}$ if $z_{1,max} > z_{2,max}$.
In extreme value theory, a return period m is associated to a return level r_m of probability of exceedance for the distribution of the max of $1/m$. The level r_m is therefore reached once over the period m in mean. By definition this is no more than a quantile of high probability of the distribution of the max. We define the return period of a storm $Z^*(s,t)$ equals to the marginal return period associated to the within-storm maxima z_{max} observed at the location s_{max}. The location s_{max} is whether fixed as reference site or defined as equal to $\arg\max_{s\in\mathcal{M}^*}\left\{\tilde{Z}^*(s,t)\right\}$.
To obtain more severe storms (with longer return period), values of $\tilde{Z}_i^*, i \in \{1,\dots,p\}$ are multiplied by a coefficient factor superior to unity and denoted ζ_i. Hence $\zeta_i\tilde{Z}_i^*(s,t),\ \zeta_i > 1,\ i \in \{1,\dots,p\}$, is the collection of the uplifted storms at the standardised scale.
Finally, each uplifted storm is transformed back to its original scale by

$$\tilde{Z}_i(s,t) = \tilde{T}^{\leftarrow}\left(\zeta_i\tilde{Z}_i^*(s,t)\right), \quad i \in \{1,\dots,p\}, \tag{5.3}$$

where $\tilde{T}^{\leftarrow}(Y(s,t)) = \hat{\mu}(s) + \hat{\sigma}(s)\frac{Y^{\hat{\xi}(s)}-1}{\hat{\xi}(s)}$.
Therefore, through (5.2) to (5.3) we obtain a collection of heavier extreme storms from a set of observed extreme storms.

It is important to highlight that an observed extreme storm $Z_i^*(s,t)$ is defined if and only if

$$\max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1, \tag{5.4}$$

meaning that there is at least one exceedance of the site marginal threshold. This uplifting proposition relies on a mathematical justification given in the following section.
There is actually no limitation in uplifting bivariate processes $\{Z_{1,i}^*, Z_{2,i}^*\}$ conditioned to (5.4) is satisfied for one of the margin, as described in the following justification.

5.3.3 Justification

A mathematical justification of the storm uplift can be obtained through the following asymptotic equivalence for conditional distributions.
Indeed, following Caires et al. (2011),

$$P\left(\frac{T^{\leftarrow}(\zeta_i Z_i^*(s,t)) - b_{n\zeta_i}}{a_{n\zeta_i}} \in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1\right) \tag{5.5}$$

has the same limit (as $n\to\infty$) as

$$P\left(\frac{Z_i(s,t) - b_n}{a_n} \in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1\right), \tag{5.6}$$

where $Z_i^*(s,t) = [1+\xi(s,t)(Z_i(s,t)-b_n(s,t))/a_n(s,t)]^{1/\xi(s,t)}$ and $T^{\leftarrow}(y) = b_n + a_n\frac{y^{\xi}-1}{\xi}$.
Let us drop both i and (s,t) indexes for the sake of simplicity in the left part of the conditional probability. The former limit equivalence is valid since following

Ferreira and de Haan (2014),

$$P\left(\frac{T^{\leftarrow}(\zeta Z^*)-b_{n\zeta}}{a_{n\zeta}}\in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t)>1\right) \tag{5.7}$$

$$\begin{aligned}
&= P\left(\frac{a_n}{a_{n\zeta}}\frac{\left[\zeta^{\xi}\left(1+\xi\frac{Z-b_n}{a_n}\right)\right]-1}{\xi}-\frac{b_{n\zeta}-b_n}{a_{n\zeta}}\in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t)>1\right)\\
&= P\left(\frac{a_n\zeta^{\xi}}{a_{n\zeta}}\frac{1+\xi\frac{Z-b_n}{a_n}-\zeta^{-\xi}}{\xi}-\frac{b_{n\zeta}-b_n}{a_{n\zeta}}\in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t)>1\right)\\
&= P\left(\frac{a_n\zeta^{\xi}}{a_{n\zeta}}\left(\frac{Z-b_n}{a_n}-\zeta^{-\xi}\left[\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right]\right)\in A \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t)>1\right)\\
&= P\left(\frac{Z-b_n}{a_n}\in\frac{a_{n\zeta}\zeta^{-\xi}}{a_n}A+\zeta^{-\xi}\left(\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right) \mid \max_{s\in\mathcal{M}^*} Z_i^*(s,t)>1\right).
\end{aligned} \tag{5.8}$$

Since (de Haan and Ferreira, 2007)

$$\frac{a_{n\zeta}\zeta^{-\xi}}{a_n}\to 1 \text{ and } \zeta^{-\xi}\left(\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right)\to 0 \tag{5.9}$$

uniformly for $(s,t)\in\mathcal{S}\times\mathcal{T}$ as $n\to\infty$ the result follows.
There is no limitation to extend this reasoning to the bivariate context. Hence we can similarly show that

$$P\left(\frac{T_1^{\leftarrow}\left(\zeta_{1,i}Z_{1,i}^*(s,t)\right)-b_{1,n\zeta_i}}{a_{1,n\zeta_i}}\in A_1,\ \frac{T_2^{\leftarrow}\left(\zeta_{2,i}Z_{2,i}^*(s,t)\right)-b_{2,n\zeta_i}}{a_{2,n\zeta_i}}\in A_2 \mid \max_{s\in\mathcal{M}^*} Z_{1,i}^*(s,t)>1\right), \tag{5.10}$$

where $T_1^{\leftarrow}(y)=b_{1,n}+a_{1,n}\frac{y^{\xi_1}-1}{\xi_1}$ and $T_2^{\leftarrow}(y)=b_{2,n}+a_{2,n}\frac{y^{\xi_2}-1}{\xi_2}$, has the same limit (as $n\to\infty$) as

$$P\left(\frac{Z_{1,i}(s,t)-b_{1,n}}{a_{1,n}}\in A_1,\ \frac{Z_{2,i}(s,t)-b_{2,n}}{a_{2,n}}\in A_2 \mid \max_{s\in\mathcal{M}^*} Z_{1,i}^*(s,t)>1\right). \tag{5.11}$$

5.3.4 Remark and Directions

The storm uplift method we use comes from the asymptotically equivalence between conditional distribution presented in the previous section. It appears to be naturally linked to the GPD process framework (see Section 1.4.2).
We recall that Dombry and Ribatet (2013) generalise the framework of Ferreira and de Haan (2014) by considering conditional events characterized through a continuous and homogeneous risk function $\ell(\cdot)$. The case from Ferreira and de Haan (2014) corresponds to $\ell(f)=sup_{s\in\mathcal{S}}f(s)$ and the ℓ function we consider here corresponds to $\ell(f)=max_j f(s_j,t)$. In this sense, we state that the con-

ditional distribution we consider here corresponds to the distribution of a GPD process.
Other remarks can be done regarding the construction of the processes. First, note that in (5.3), the coefficient ζ_i can be determined in several way as far as it is superior of unity.
The first interpretation of the use of ζ_i is given by de Haan in Caires et al. (2011). According to him, using a coefficient ζ_i comes down to uplift the threshold of the peaks-over-threshold process Z_i. In particular, if the process is conditioned to $\max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1$, that means to $\max_{s\in\mathcal{M}^*} Z_i(s,t) > b_n$, the probability of Z_i to exceed b_n is $1/n$ and the probability that $\tilde{Z}_i$ exceeds $b_{n\zeta_i}$ is $1/(n\zeta_i)$.
We can also consider the special case $\zeta_i = \frac{T(z_m)}{T(z_{max})}$, where z_{max} is still the within-storm maxima and z_m is the return level corresponding to the m-year return period at location where z_{max} is observed. Implemented in Groeneweg et al. (2012), Smith in Caires et al. (2011) interprets such a transformation as an uplift from a storm with a given return period to a storm with a return period equal to m.
However, other choices for ζ_i could be proposed. For example, ζ_i, $i = 1,\dots,p$ could be proportional to independent realisations of standard Pareto distribution. In this specific configuration, our approach should be very similar to the constructive representation of the Pareto process proposed by Dombry and Ribatet (2013).

Aware of these remarks, we apply this uplifiting methodology on the hindcast data set. The results are given in the following section.

5.4 Results

We applied the method to our data set composed of the 52-year of sea-states conditions. To afford the computational demand of dealing with around 4000 sites locations, algorithms are implemented in a dedicated R code and is parallelised via a MPI interface. All computations are performed on a cluster composed of 96 cores.
From now on and for the sake of simplicity, the definition of storm embraces the multivariate space-time processes composed of H_s, T_p and directions ψ.
We worked on the 10 highest storms observed to uplift both H_s and T_p variables, resting on the proposed bivariate approach. In our case study, H_s is the variable that conditions the bivariate space-time processes selection. We justify this choice to guarantee to obtain highly energetic wave processes because at list one components in $\mathcal{M}^*$ exceed its threshold.

Since we are looking at modelling storms impacting the coastline only, we choose to set $\mathcal{M}^*$ from Algorithm 2 equals to the coastline-band area illustrated in Figure 5.2(a). This restriction in the area to detect storms prevents the selection of offshore storms that are not propagating to the coast. From the preliminary analysis in Section 5.2, we decide to consider that storms last about 49

hours. Consequently, the selected value of δ is equal to 24 (hours). To select only IID storms, the value of ε is equal to 49 (hours) as well. Both ζ_{i,H_s} and ζ_{i,T_s} are chosen to uplift original storms according to the two variables to the m-year return period, considering a reference site for s_{max} (node 2342 as illustrated in Figure 5.7(d)). We observe the variability of storms with m ranging from 25 to 150, by 25. Some illustrations of the available storms are given in the following Figures 5.4, 5.5 and 5.6.

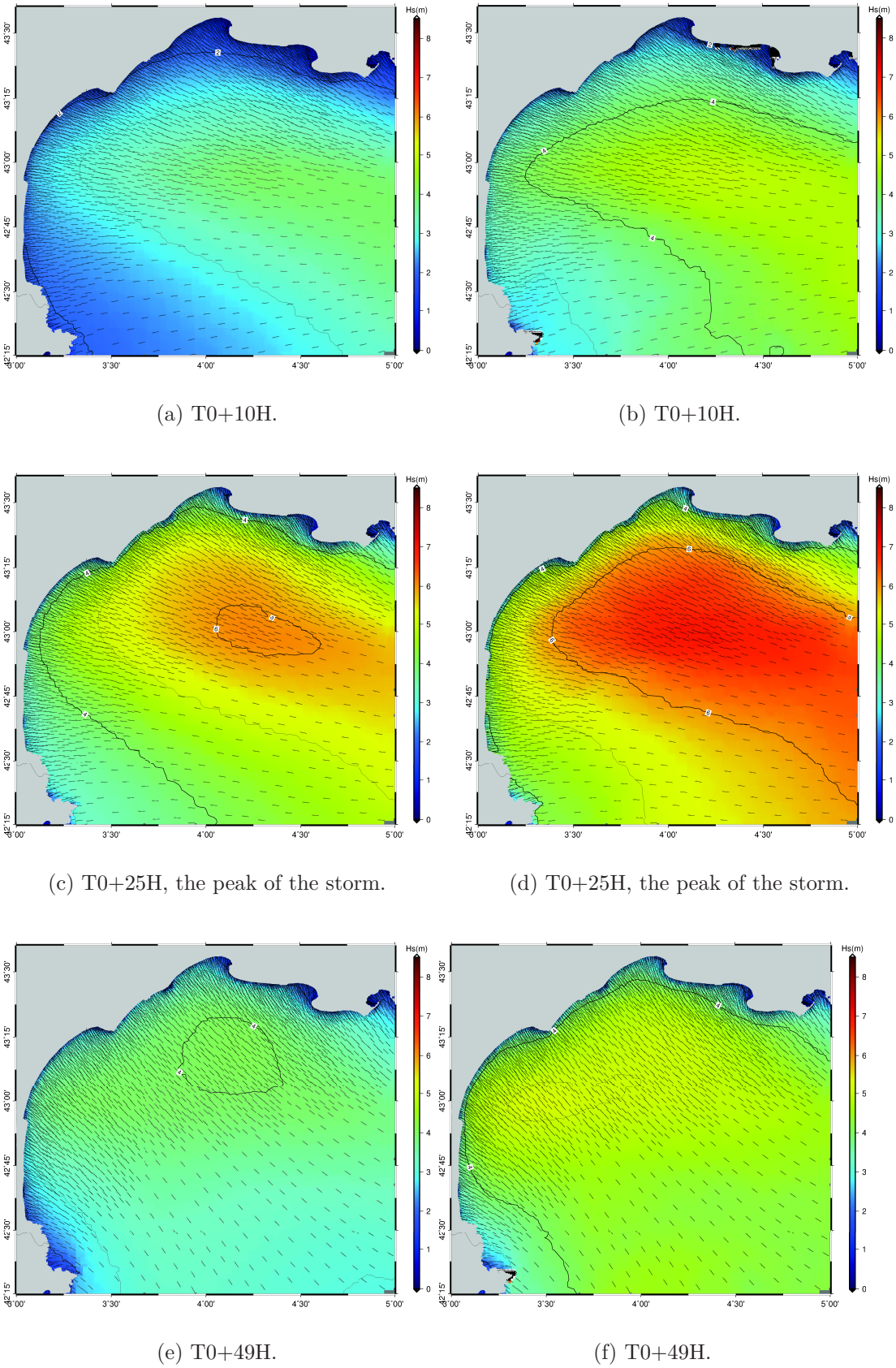
Figures 5.4 and 5.5 illustrate three time-steps of one of the selected storms and their corresponding uplifted processes towards the 100-year return period, respectively for H_s and T_p.
Mean wave directions are not transformed during the uplift method. This assume that the dependence structure for the mean wave direction within the extreme remains the same.
Among the set of 10 scenarios, the variability of the fields observed are quite large, but are non-surprisingly dominated by fluxes from South and SouthEast. This is a direct consequence choosing $\mathcal{M}^*$ as a very littoral area.

Figure 5.6 is the presentation of four instances of the same original storm uplifted towards different return levels. Only the significant wave heights are shown here. One may remark that the levels of Figure 5.6, given at the peak of the uplifted storm, are very similar to the return levels presented in the previous chapter in term of intensity of H_s. Physically speaking, the proposed space-time processes look like processes that are likely to be observed, in terms of both spatial and time-evolution dependences, which are actually conserved by the uplift procedure. They seem physically valid to feed littoral physical models.

5.5 A Risk Analysis

5.5.1 Mass Flux of Littoral Energy

Coastal hazards such as submersion, erosion or beach contamination are usually quantified from formulae that require the computation of the mass flux of energy towards the shoreline, given off the shoaling zone where waves do not interact significantly with the sea bottom. One usually discriminates cross-shore and long shore contributions, depending upon the goal of the application. For instance, the calculation of the alongshore-sand transport (Bagnold, 1966; CERC, 1984) requires the long shore mass flux of energy. In the following, we strictly consider the long shore impact ϕ of the deep water mass flux of energy $\mathcal{Q}$ to the shoreline, which is a relevant expression to tackle any analysis of shoreline dynamics. We model evolution of such a quantity during extreme wave storms.
For a given storm event $\mathcal{S}$, we compute the impact $\phi_{i,t}^{(\mathcal{S})}$ at a location $c_i \in \mathcal{C}$ and at a time t of the mass flux of energy $\mathcal{Q}_{i,t}$ coming from waves at a location $l_i \in \mathcal{L}$



(a) T0+10H.

(b) T0+10H.

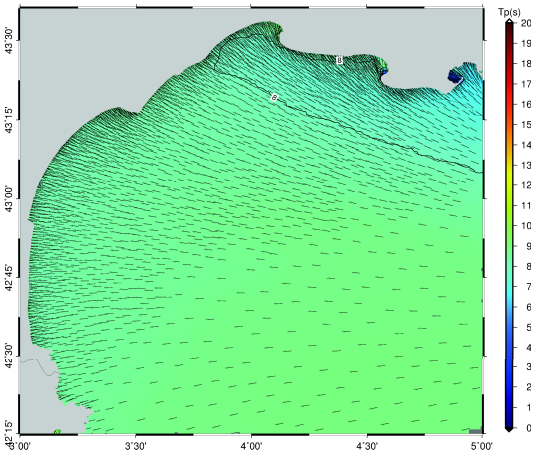
(c) T0+25H, the peak of the storm.

(d) T0+25H, the peak of the storm.

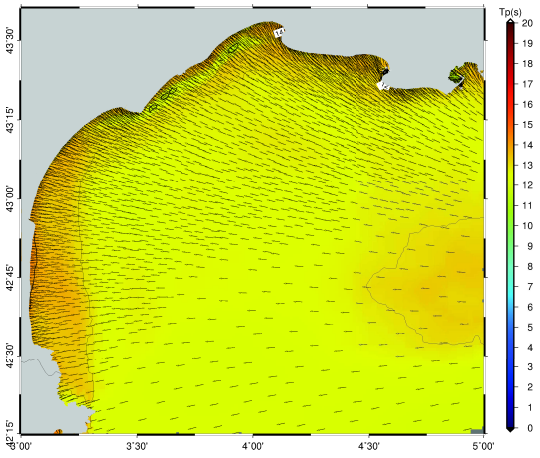
(e) T0+49H.

(f) T0+49H.

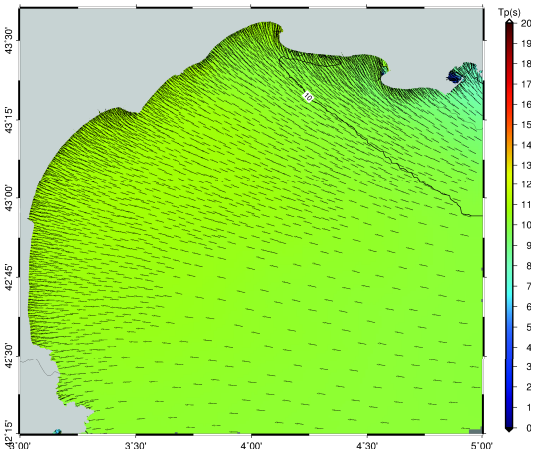
Figure 5.4: Illustration of some time-steps of a storm, for the H_s variable. The mean wave direction ψ are the vector. On left the original storm available from the hindcast data set. On right the corresponding storm uplifted at the 100-year return period.



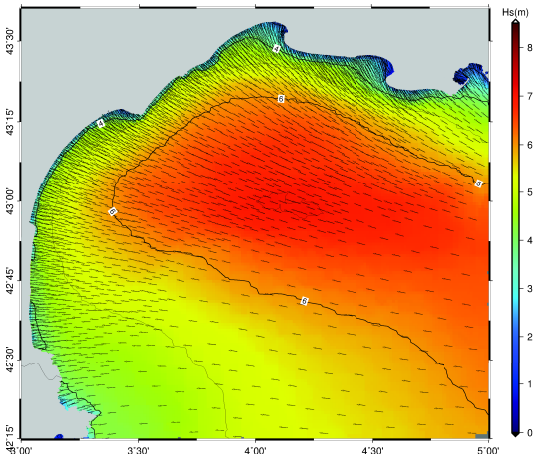
(a) T0+10H.



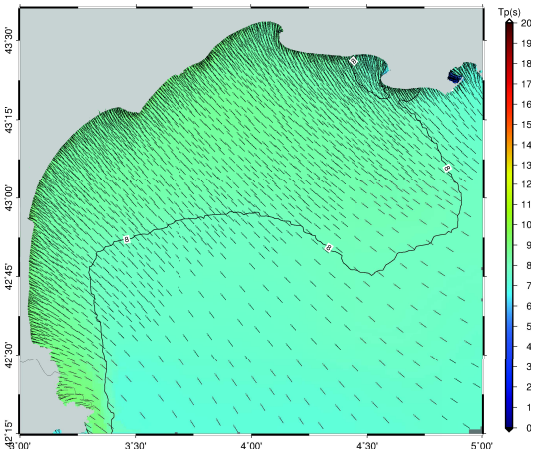
(b) T0+10H.



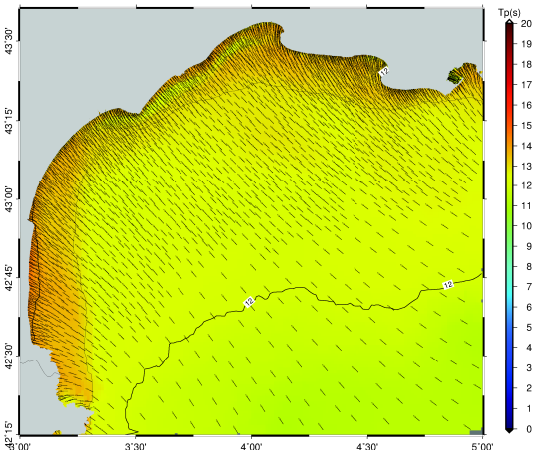
(c) T0+25H, the peak of the storm.



(d) T0+25H, the peak of the storm.

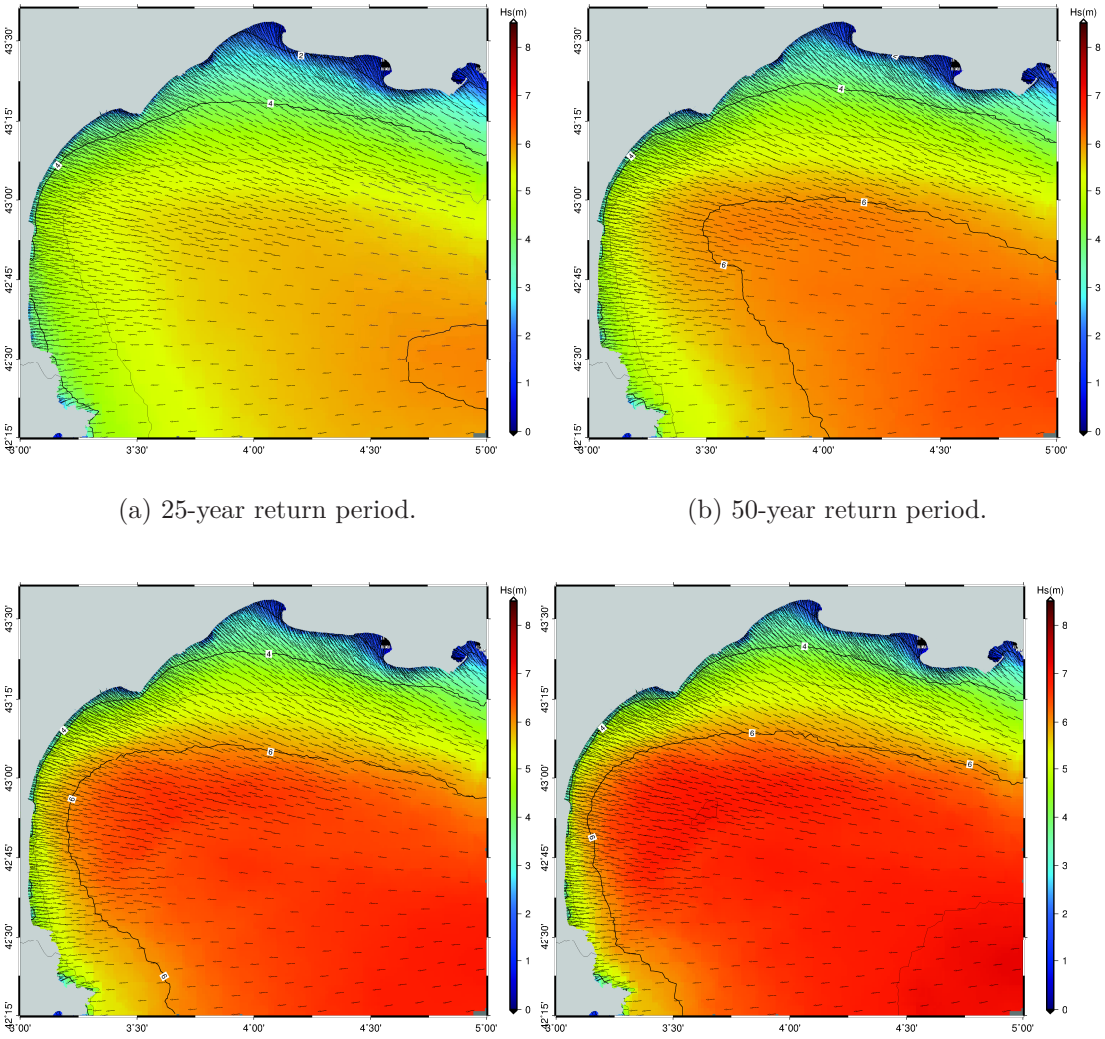


(e) T0+49H.



(f) T0+49H.

Figure 5.5: Illustration of some time-steps of a storm, for the T_p variable. The mean wave direction ψ are the vector. On left the original storm available from the hindcast data set. On right the corresponding uplifted storm of 100-year return period.



(a) 25-year return period.

(b) 50-year return period.

(c) 100-year return period.

(d) 150-year return period.

Figure 5.6: Illustration of the significant wave heights at the peak of the storm. It is the same storm uplifted four times at different return levels.

(see Figure 5.7(a). The longshore impact is calculated by :

$$\phi_{i,t}^{(\mathcal{S})} = \mathcal{Q}_{i,t} \sin(\omega_{i,t}) \cos(\omega_{i,t}) \tag{5.12}$$

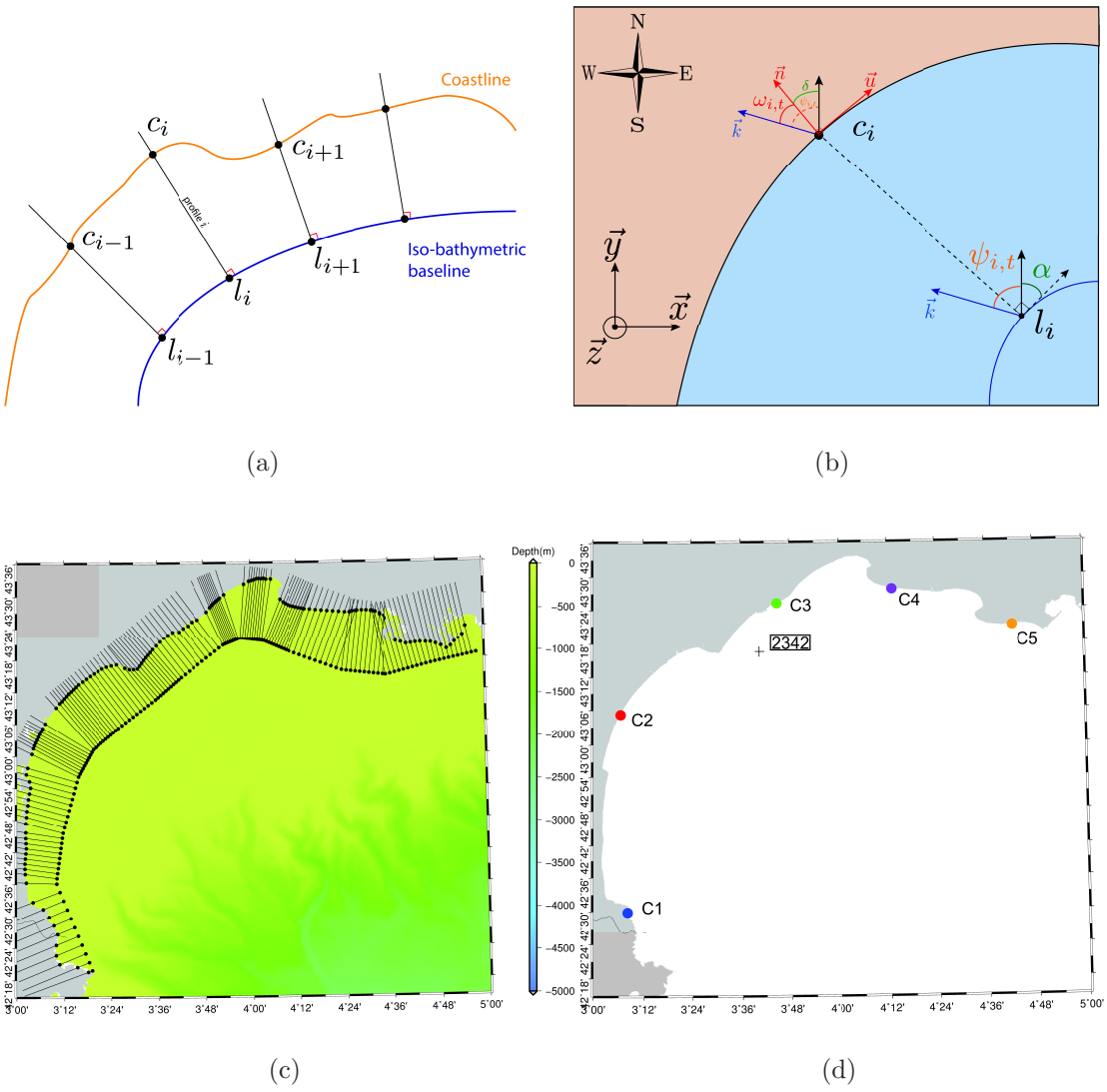
where $\omega_{i,t}$ represents the angle of the waves propagation at l_i at a time t as illustrated in Figure 5.7 and is function of the wave direction $\psi_{i,t}$.
Practically, $\mathcal{Q}$ is derived from the variables H_s, T_p characterising the sea-state conditions at various points along an iso-bathymetric baseline. Such a mass flux of energy is classically given by:

$$\mathcal{Q}_{i,t} = \frac{1}{8} \rho \, g \, H_{s_{i,t}}^2 \, T_{p_{i,t}} \tag{5.13}$$

where $\mathcal{Q}_{i,t}$ represents the mass flux energy at the location l_i and at a time t, ρ denotes the water volumetric mass density and g the gravity constant. This procedure can be performed both with the storms extracted from the hindcast data-set to monitor the impact of the past events, or with the uplifted storms to find out what would be the impact to the coast if the storms are more severe than the ones already observed, as presented in the following section.

5.5.2 Results

Some of the uplifted storms are used to compute the long-shore impact under these extreme conditions at any location c_i, from the definition of the global mass flux ϕ (see Equation 5.12). A set of 5 locations from the available c_i, coloured in Figure 5.7(d), have been peaked up as reference to discuss the assessment of the long-shore impact at the coastline of the GOL under extreme conditions.
Regarding the angles presented in 5.7, a positive value of ψ is interpreted as a long-shore contribution in the direction of $\vec{u}$. A negative value is therefore interpreted as a long-shore contribution in $-\vec{u}$.
Four figures are presented to give an overview of the various possibilities offered by the simulation of storms in the assessment of long-shore impact. Firstly, Figure 5.8 shows the response of the impact model at the 5 reference locations to the uplifted storm presented previously (see Figures 5.4-5.5), which is uplifted at the 100-year return level. Regarding this figure, it is very clear that in this configuration c_2, c_3 and c_4 are impacted towards the West and South West directions, revealing the presence of a eastern wave forcing. What is very interesting is that from such figure, the time evolution of the long-shore impact regarding the simulated extreme process can be explored.
The behaviour of the point c_1 seems to have a contribution towards the South East. This could be explained by the automatised procedure that computes angles at coastline, and which is not accurate when the shoreline is too irregular, as it is the case for the point c_1. It could be the explanation of this observed signal. However since this section remains an illustration, no more efforts have been dedicated to fix this issue.



(a) (b)

(c) (d)

Figure 5.7: (a) A schematic representation of the baseline and the creation of the n profiles. (b) Illustration of angles used to compute the impact of the wave energy flux at point l_i to its coupled coast point c_i. ωi_t denotes the angle of interest. It is the angle between the observed direction of the waves $\vec{k}$ at location l_i – at a time t – and the cross-shore direction at location c_i denoted $\vec{n_i}$. (c) The actual profiles construction over the GOL. Sea-states conditions are picked-up from a set $\mathcal{L} = \{l_1, \ldots, l_n\}$ of n points lying on an iso-bathymetric baseline. From those locations, n profiles normal to the baseline are created. Intersections of those profiles to the coastline form a set $\mathcal{C} = \{c_1, \ldots, c_n\}$ denoting the reference locations where mass flux energy are derived to. The number n is chosen to fit the resolution required along the shore. (d) The selected four locations analysed in the risk analyses and the reference location from where ζ_i coefficients are computed.

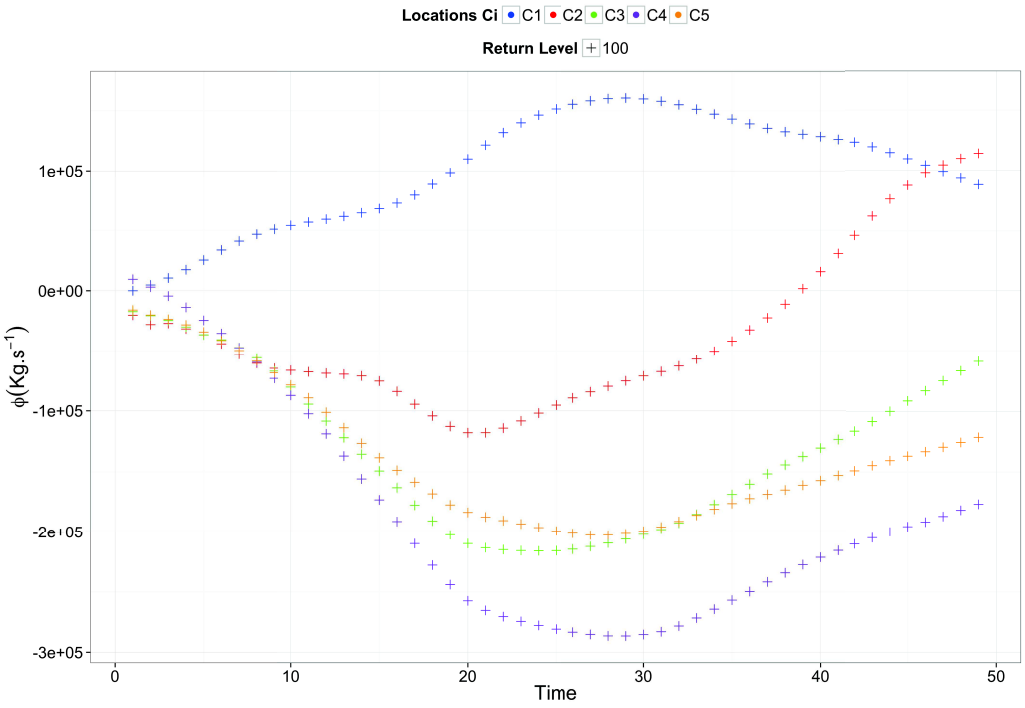


Figure 5.8: Evaluation of the long-shore impact ϕ at the 5 locations c_i for the 9^{th} highest observed storm, uplifted to the 100-year return period.

One may also look at the variability of the long-shore impact when storms are varying in extremeness, as defined before. Figure 5.9 represents what could be expected in terms of long-shore impact, at one location and for a given storm uplifted to various return levels.

An other interesting information in the assessment of long-shore impact is to look at the response ψ for several storms uplifted to the same return levels. This is illustrated in Figure 5.10 for the point c_5, which is situated at the very East of the GOL. From this figure we can state that the long-shore impact is likely to be towards the west (negative value of ψ), catching a consequent amount of energy from the storm coming from the open sea boundary of the GOL (i.e. from the East/ South East). This remark is in accordance with a physical observation that is identified when looking at the shoreline: the formation of sandy spits.
However and still in Figure 5.10, one of the storm selected has a positive impact during its realisation. This is not really surprising since as it is located at the edge of the GOL, this shoreline location is also subject to be hit by South and South-West storms, that are less frequent but even more damaging than the Eastern ones.

Finally Figure 5.11 is a mix of the possible combinations. It provides a simultaneously preview for various return levels of the storm and at the 5 locations of interest. Spatial patterns of long-shore impact regarding the intensity of a kind of storm might be determined from such a figure.
To summarise this risk analysis, and even if only few results of what it is ac-

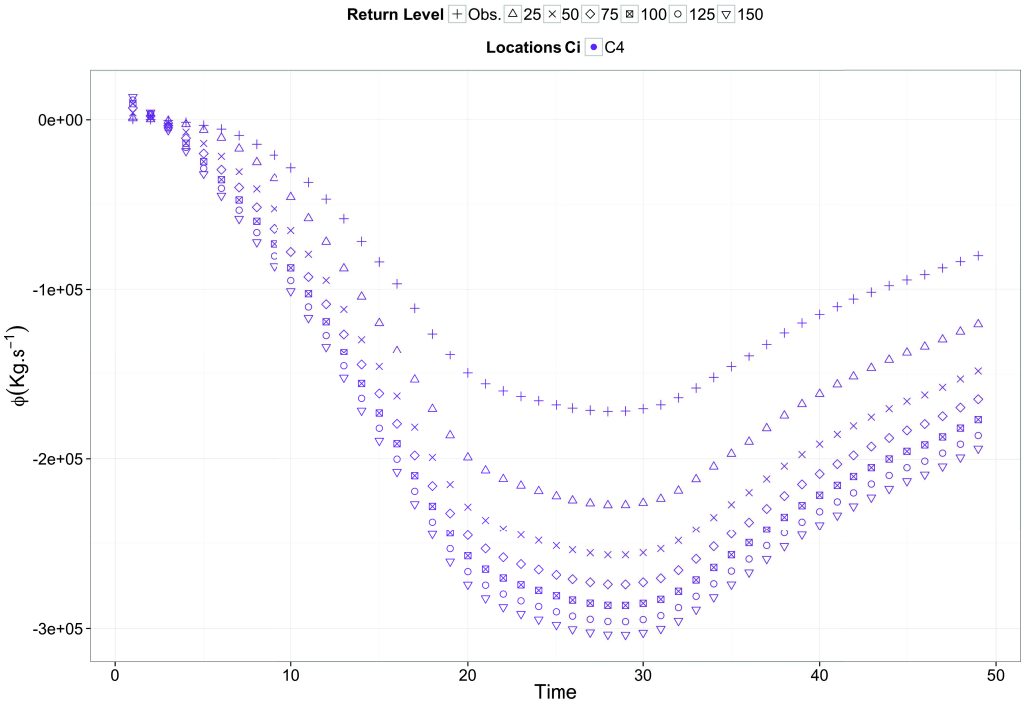


Figure 5.9: Evaluation of the long-shore impact ϕ at the location c_4, for uplifted storm to the 25,50,75,100,125,150-year return periods, from the 9^{th} highest observed storm. The impact computed from values of the observed storm are given as well for reference.

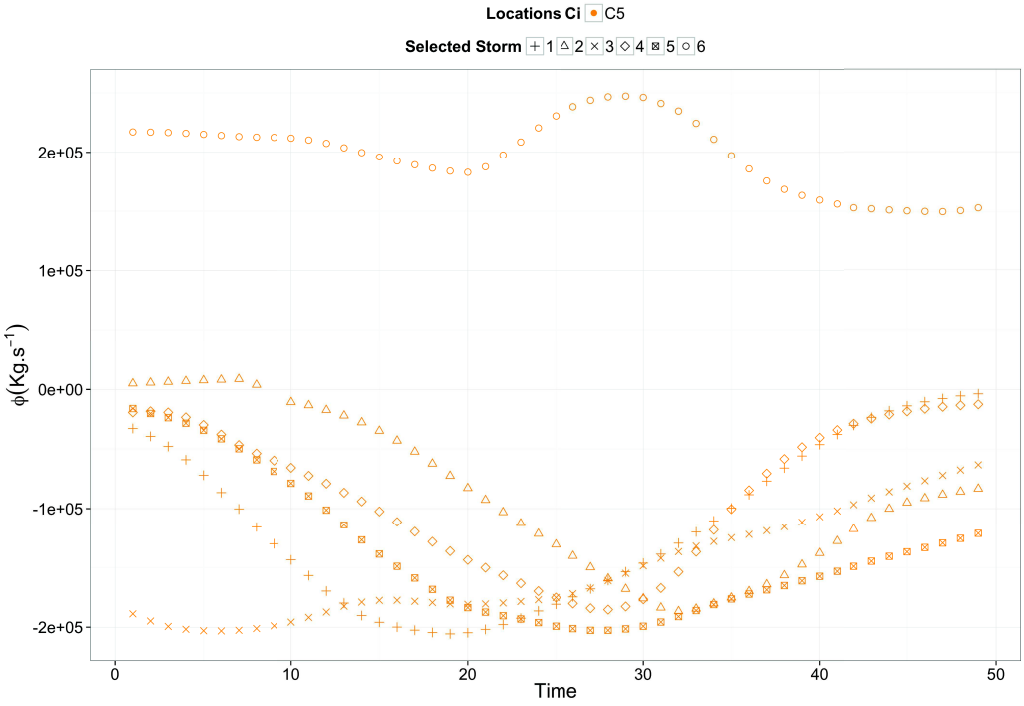


Figure 5.10: Evaluation of the long-shore impact ϕ at the location c_5 for a sample of observed storms, uplifted to the 100-year return period.

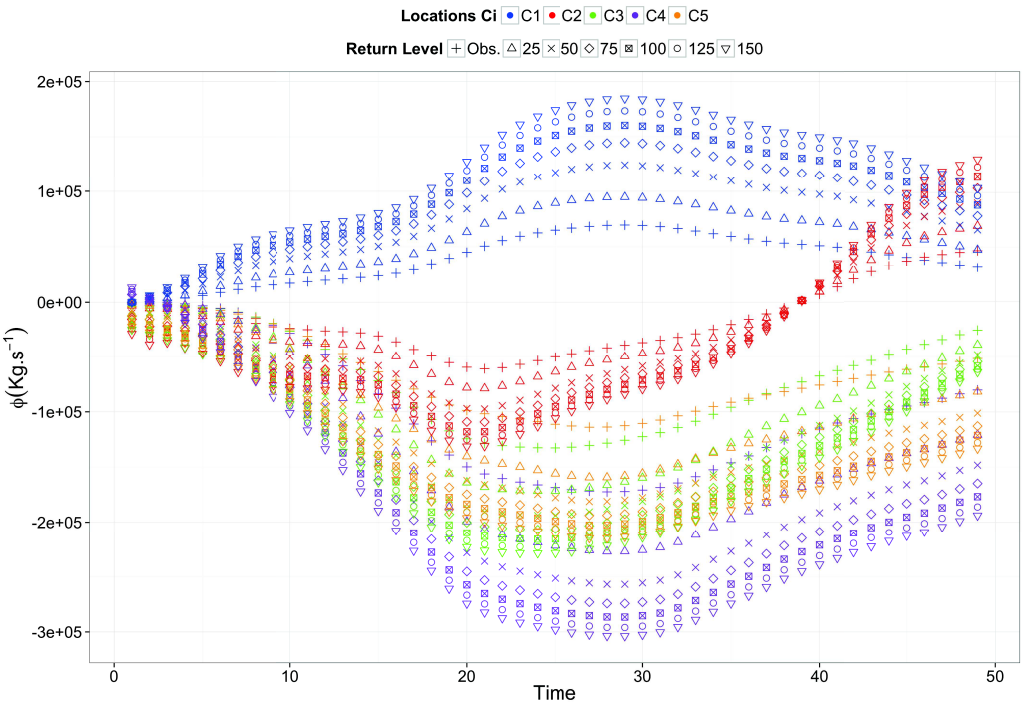


Figure 5.11: Evaluation of the long-shore impact ϕ at the 5 locations c_i for the 9^{th} highest observed storm, uplifted to the 25,50,75,100,125,150-year return periods. The impact computed from values of the observed storm are given as well for reference.

tually affordable to compute are presented here, we demonstrate the promising potential of the presented semi-parametric methodology towards the assessment of event-scale hazards.

5.6 Discussion

In this chapter we introduce a semi-parametric approach to simulate bi-variate extreme space-time wave processes. Such simulated storms are constructed to feed physical models assessing event-scale coastal hazards.
Like in the other chapters of this document, we apply the presented methodology on a case study over the GOL area.
The result of the simulated events is relevant in regards to the spatial information provided in the previous chapter, and what it is described in the literature about the storm waves episodes in the GOL.
To demonstrate the benefits of such a method, some simulated storms are used in a risk analysis. We show that thanks to the simulated processes on which a control of the extremeness is provided, the variability of the littoral long-shore impact can be assessed, both spatially and through the time evolution.

However some limits of the method and its implementation can be highlighted.

The first one is that the dependence structure of the variable are assumed as conserved in the extremes, like it is the case in the max-stable context. It could be a limit if this assumption is false and that very extreme scenarios are strictly different from the ones observed.
In a more practical aspect, one can argue that the storm size in the Algorithm 2 is fixed and symmetric around the peak value of the storm. This may not reflect the reality for all storms. Therefore replacing the current size by an adaptive one might be of interest to better represent those storms.
Other parameters of the algorithm can be argued as the littoral area $\mathcal{M}^*$. Even if its definition is paramount to assess littoral hazards, it could be interesting to test and find out the actual sensitivity of the storm detection regarding this space.

Beyond those few limits, we think that this method is promising and open many perspectives. One perspective of work would be the comparison between simulated storms issued by the presented semi-parametric approach and ones issued by other parametric approaches, and in particular the GPD processes as quickly introduced in Section 1.4.2. Such a comparison would be valuable since both approaches present similarities.
In the same time and after having performed a small risk analysis using some of the simulated extreme space-time waves events, one challenge is to use those storms to feed heavy computational physical models assessing other coastal hazards like flood overland models.
The use of such heavy models is motivated by the necessity of providing help towards the decision making in such crises. This last notion is further explored in the following chapter.

Part III

Industrial Implementation

Chapter 6

Towards Decision Tools

Chapter Summary

This chapter presents a subject of openness towards the creation of (IT) tools aiming at helping the decision in the anticipation and management of coastal hazards (see Figure 6.1). Such tools may rely on the previously introduced notions, and in particular the capacity of stochastically simulate extreme processes, seen as scenarios at controlled extremeness levels.
Hence, this chapter introduces a methodology based on a pre-computing principle, inspired from the subject of numerical model exploration and of case based reasoning algorithms, to provide information to the decision-maker without performing physical (and heavy) computation when an alert is raised due to the forecast of extreme conditions. We finally discuss the key points and the limits of this methodology in regards to its final goal of helping decision-making.

6.1 Introduction

This chapter, largely inspired from our former study in Chailan et al. (2012), introduces the proposed principle of pre-computation towards helping the decision-making in the anticipation and management of event-scaled coastal hazards (see Section 5.5.1).
When extreme conditions (e.g., storm waves) are forecast by weather agencies, alerts are raised to inform local decision-makers of threat. The consequences of such extreme conditions have to be accurately assessed, as quickly as possible.

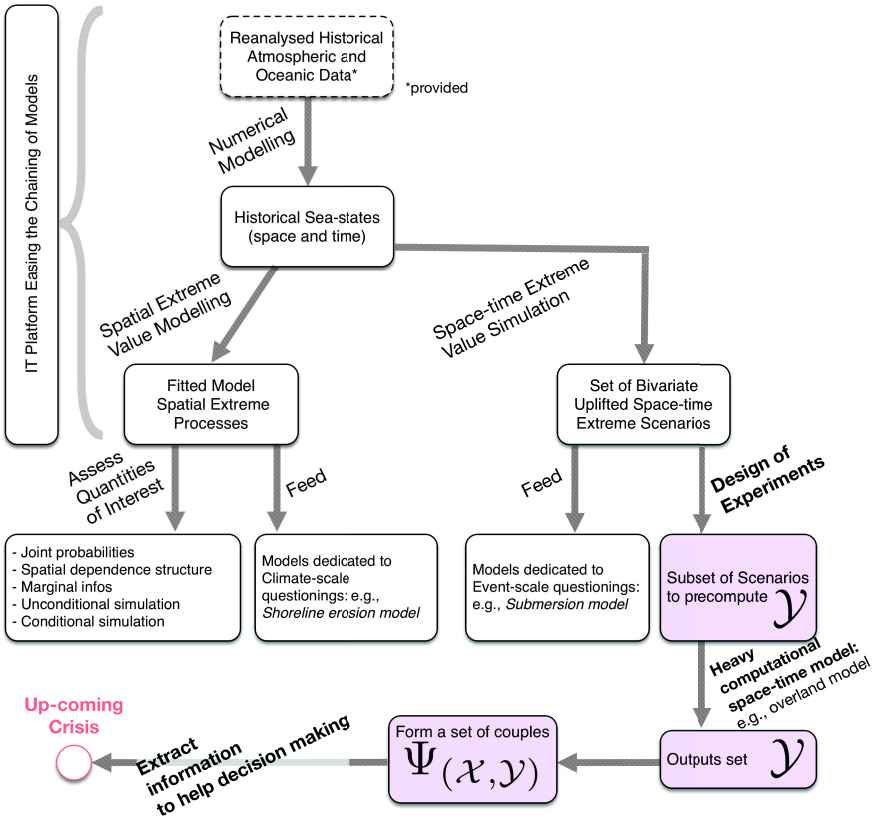


Figure 6.1: This Chapter 6 introduces the proposed principle of pre-computation to help decision making against coastal hazards. We propose an efficient way to select extreme scenarios to pre-compute, store the associated IO (Inputs and Outputs) couples and then query the system in order to provide an approximation off the on-coming crisis, which is likely to exist in the following few hours from the alert.

The notion of time is paramount. For instance in case of a potential submersion phenomenon, heavy physical models must be run to forecast whether the extreme conditions will involve or not the realisation of a devastating submersion. However, even if these computations are performed on High Performance Cluster (HPC) resources, they take time to be performed. This time can be longer than the realisation of the physical phenomenon itself. Due to the promptitude of these chained events and the stakes involved in their preventions, decision-makers must be provided with alternatives than the direct physical simulation. One of them is the presented pre-computation principle, inspired from Business Intelligence (BI) techniques such as the materialized view selection methods.

The pre-computation principle consists in anticipating the computation of a set of scenarios, in order to provide relevant information on a future crisis situation within a brief time. The information is provided whether by approximation of the result or by providing the exact result. The latter is possible if the actual crisis scenario matches one of the pre-computed scenarios.

This concept brings forward some scientific obstacles. Hereafter is the formalisation of those challenges.

Let f be the black box (heavy computational) model, taking a list of n inputs defined as $X = (x_1, x_2, \ldots, x_n) \in D_I = I_1 \times I_2 \times \ldots \times I_n$ and generates a list of m outputs $Y = (y_1, y_2, \ldots, y_m) \in D_O = O_1 \times O_2 \times \ldots \times O_m$, such that,

$$f : D_I \longrightarrow D_O \quad ; \quad X \longrightarrow Y = f(X). \tag{6.1}$$

Each deterministic simulation $Y = f(X)$ is very time and resource demanding. In the sequel, let us denote X as a scenario which is actually a realisation of the former random variable. The so-called pre-computing workflow, illustrated in Figure 6.2, consists in:

— Constructing a set of scenarios $\mathcal{X} = \{X_1, X_2, \ldots, \}$, from observed or stochastically simulated scenarios.
— For each scenario $X \in D_I$ available in $\mathcal{X}$, computing its result $Y \in D_O$ and store the couple. Forming *in fine* a set Ψ of IO (Inputs and Outputs) couples $\{(X_i, Y_i), i = 1, 2, \ldots\}$.
— Building an application $\bar{f}$ that given an input $X \in D_I$ and a positive integer $k \in \mathbb{N}^* \leq card\mathcal{X}$ returns the k closest scenarios found in $\mathcal{X}$ and thus forming a subset of $\mathcal{X}$ denoted $\bar{\mathcal{X}}_k$

$$\begin{aligned} &\bar{f}_k : D_I \longrightarrow D_I^k \\ &\quad X \longrightarrow \bar{f}_k(X) = \bar{\mathcal{X}}_k \subseteq \mathcal{X} \quad s.t. \quad card(\bar{\mathcal{X}}_k) = k \\ &\quad \text{and} \quad \forall Z_1 \in \bar{\mathcal{X}}_k, Z_2 \in \mathcal{X} \setminus \bar{\mathcal{X}}_k, d(Z_1, X) < d(Z_2, X) \end{aligned} \tag{6.2}$$

where $d : D_I^2 \longrightarrow \mathbb{R}$, D_I^2 defining $D_I \times D_I$, is a distance function to determine. The system finds the set of images pre-computed $\mathcal{Y}_k = \{Y_1, Y_2, \ldots, Y_k\}$, where $Y_i = f(X_i)$ and $X_i \in \bar{\mathcal{X}}_k$. Then it aggregates them to provide an approximation of the real simulation, denoted $\hat{Y} \approx f(X)$. $Y = f(X) = \hat{Y}$. In case of one scenario from $\mathcal{X}$ is equal to the queried scenario X.

$\bar{f}_k$ is evaluated very quickly and depends upon the number of scenarios of $\mathcal{X}$. We assume that the greater $card(\mathcal{X})$, the more accurate the results of $\bar{f}_k$, and consequently $\hat{Y}$.

Therefore the pre-computing approach is composed of four main challenges: to create a design of experiments to sample the scenarios to pre-compute, to efficiently perform pre-computations from a technical and storage point of view, to query the global system with a reference scenario X_{ref} in order to approximate its result with $\hat{Y}_{ref}$ via an aggregation function which is the last challenge. They are detailed in the following.

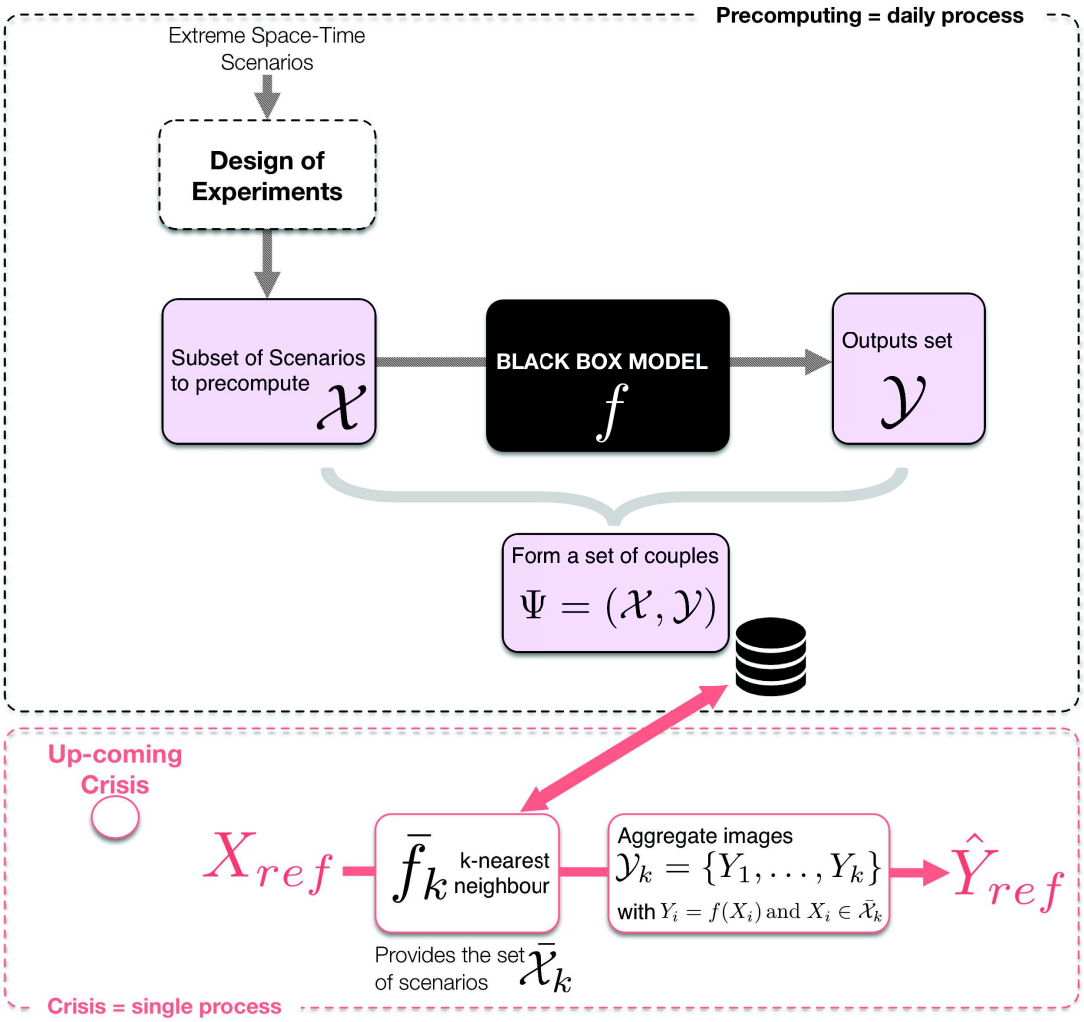


Figure 6.2: Schematic representation of the so-called pre-computing approach. It is composed of two workflows, running at different time-scale. As a daily work, a set $\mathcal{X}$ of space-time scenarios are selected (in a clever way) to be pre-computed. Their computations are performed with f the black box heavy computational model, and results are stored along their corresponding scenarios. They form a set Ψ of IO couples. When a crisis arises, a (real) scenario X_{ref} is provided by weather forecast agencies. On the basis of a distance function allowing to compare the scenarios, a function (here a k-nearest neighbour algorithm denoted $\bar{f}_k$) is used to gather a set of similar scenario outputs $\mathcal{Y}_k$ and estimate from it what is the output $f(X_{ref})$: the up-comping crisis.

6.2 Method

To illustrate these sections, we assume in the sequel that the application f is the one defined in Section 5.5.1, which is forced by a space-time extreme wave scenario X, also named storm according to the definition at the end of the previous chapter. Thus, the storm $X = X(s,t)$ is constructed from three random processes $H_s(s,t)$, $T_p(s,t)$ and $\psi(s,t)$ (respectively the significant wave height, peak wave period and mean wave direction).
To tackle the challenges identified above, the proposed pre-computing principle requires that any storm $X_i \in \mathcal{X}$ is indexed in an efficient way. Since each component of a storm X is a space-time process, its indexing is hard to handle. We reduce the dimension of the problem by indexing X by three characterising variables, being $\bar{H}_s \in D_{H_s}$, $\bar{T}_p \in D_{T_p}$ and $\bar{\psi} \in D_\psi$, with $D_{H_s} \subset \mathbb{R}^+$, $D_{T_p} \subset \mathbb{R}^+$ and $D_\psi = [0, 360[$. They are respectively the mean spatial values observed at the peak of the storm of the significant wave heights, the peak wave periods and the mean wave directions. However, even if they are indexed by those reduced-dimension variables, manipulated scenarios here still represent space-time processes, as defined in the previous chapter.
Let us also consider that the output $Y = f(X)$ is the vector whose components represent the mean long shore impact computed at each extracted coastal location c_i (see Figure 5.7) for the duration of the storm.
Finally and for illustration purpose, let us assume that f is a complex application and that one computation of $Y = f(X)$ requires huge computational and time resources.

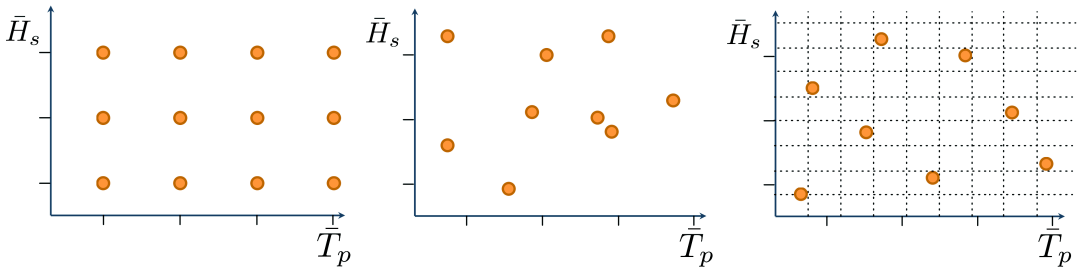
6.2.1 Design of Experiments

The first step of the method is the creation of the set $\mathcal{X}$ of space-time extreme scenarios, taken from observed or simulated events.
Considering the method of simulation of extreme space-time processes presented in the previous chapter and because $\bar{H}_s$ and $\bar{T}_p$ are continuously defined on $\mathbb{R}^+$, we have the possibility to obtain an infinity of scenarios by varying the choice of $\zeta_{i,Tp}$ and $\zeta_{i,Hs}$ but superior to unity. $\bar{\psi}$ is continuously defined on $[0, 360[$ but we have no control on its value in the previously presented semi-parametric approach.

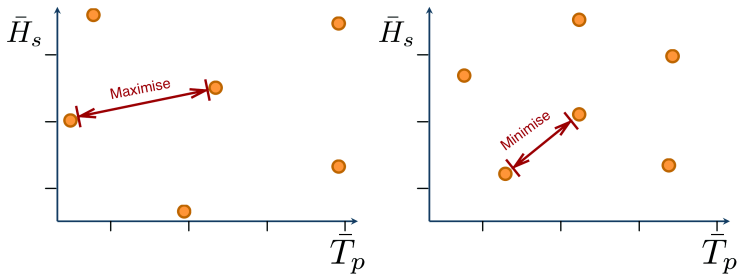
The goal of a design of experiments is to obtain a subset being representative of the space of definition D_I of X, that is the product of the domain of definition of each of its variable. In our example X has been resumed to $D_I^* = D_{H_s} \times D_{T_p} \times D_\psi = \mathbb{R}^+ \times \mathbb{R}^+ \times [0, 360[$. Most famous methods are reviewed in Faivre et al. (2013). Let us introduce some of them.
A first idea is to use a full factorial plan (e.g., Montgomery, 2008) that would consider all the combinations possible along the space of definition. This is clearly out of scope with such continuously defined random variables. One may therefore use a factorial plan by subsetting the domain of definition of each variable, making them discrete (see Figure 6.3(a)). However, criteria of subsetting may

still be argued.
An alternative to full factorial plan is the so called Monte Carlo sampling (e.g., Lemieux, 2009). In this approach illustrated by Figure 6.3(b), a matrix of experiments is constructed from the combination of random and independent samples along the domain of definition of each variable of X.

If there is an *a priori* on the behaviour of the model f, quasi Monte Carlo



(a) Factorial plan. (b) Monte-Carlo Sampling. (c) Latin HyperCube Sampling.



(d) Maximin. (e) Minimax.

Figure 6.3: (Top) Schematic (bi-variate) representation of the most used design of experiments algorithms. (Bottom) Examples of criteria to optimise an LHS sampling.

sampling (Lemieux, 2009) would be preferred in order to obtain from the same number of scenarios, a better performance on the representation of their outputs.

Latin Hypercube Sampling (LHS) (McKay et al., 1979; Stein, 1987) is a great alternative to better discretise the space of definition. To do so, each domain of definition of D_I is divided into a subset of N segments of probability equal to $1/N$. Then a value is randomly sampled to each of those segments. Once a value is placed into one of the segment, no more value can be sampled from this segment. For instance, a bivariate case with $N = 8$ is shown in Figure 6.3(c). Additionally to this sample techniques, a spatial recovering criterion can be utilised to optimised it. When using the *maximin* criterion, the sampling of the values within the Hypercube of the domain of definition is realised by maximising the minimum distance between each value (see Figure 6.3(d)). In the opposite, using

the *minimax* criterion would minimise the maximum distance between each value (see Figure 6.3(e)). They form respectively the families of LHS Maximin (Morris and Mitchell, 1995) and LHS low discrepancy (Minimax) (Jin et al., 2005).

One may also use sensitive analysis to diagnose for the scenarios X_i what are the input variables affecting the most their corresponding responses $Y_i = f(X_i)$ in term of variance. Such an analysis would allow to guide the design of experiments as well by reducing the considered domain of definition of the less influencing variables. For instance if the sensitivity analysis found out $\hat{T}p$ is not so much influencing the output, only a few discrete values of $\hat{T}p$ should be available in the design of experiments. Once selected the scenarios composing $\mathcal{X}$ have to be computed as described in the following section.

6.2.2 Pre-Computations and Storage

The second step of the the pre-computation principle is to compute and store the image Y_i of the selected scenarios X_i. Physical models dealing with coastal hazards are generally resources expensive. In that case, to compute any single output $Y = f(X)$ requires a dedicated HPC infrastructure. From this statement, there is no optimisation to define the arrangement of the sequence of scenarios to pre-compute, but the one from the job scheduler of the HPC cluster.

The pre-computed scenarios and their respective image via the model f represent a tremendous volume of data to store. The solution embedding a pre-computing module may prefer to handle meta-data of those couples instead of the couples themselves. Hence, one valuable asset of the proposed solution would be the decoupling of these meta-data management and the physical data sets.

6.2.3 Query System

In a crisis configuration, a decision-maker must have the possibility to query the system to obtain information on the upcoming crisis. The query system is inspired from the so-called Case-Based Reasoning approaches. Practically, providing an input (crisis) scenario X_{ref} which is a potential threat to the coast, the solution must be able to diagnose the potential risks by finding what would be the outcome $Y_{ref} = f(X_{ref})$. To do so, the stored set of IO couples Ψ must be queried to return the suitable information from pre-computed scenarios that are similar to X_{ref}.

For this purpose, we rely on the so-called "case-based" and "instance-based" reasoning. Several methods exist (Mitchell, 1997). In our work, we have considered the k-nearest neighbours method. In our context, the purpose is to find the set of the k-nearest pre-computed scenarios from X_{ref} denoted $\bar{\mathcal{X}}_k$ (see Figure 6.4).

The k-nearest neighbour algorithm is based on the definition of a distance func-

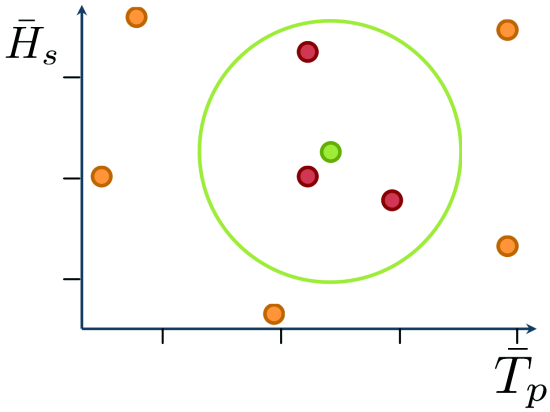


Figure 6.4: Schematic representation of a k-nearest neighbour algorithm, where $k = 3$ in this example. The green dot is the query X_{ref} and the red ones are the selected 3-closest values considering their distance d to the query, using here the euclidean distance.

tion d allowing comparing two scenarios. One valuable add would be to perform a sensitivity analysis over the variables of a scenario, in order to provide information and to weight the components of the distance function.
Then an estimation $\hat{Y}_{ref}$ of $f(X_{ref})$ has to be built from the set of pre-computed scenario images $\mathcal{Y}_k = \{Y_1, Y_2, \ldots, Y_k\}$, where $Y_i = f(X_i)$ and $X_i \in \bar{\mathcal{X}}_k$. One way to proceed is to aggregate the results Y_i of the k-closest scenarios pre-computed, weighted by their distance to the query X_{ref}. This aggregation can be tricky when the variables are of high dimensions. From our example, it can be computed as

$$\hat{Y}_{ref} = \sum_{i:Xi \in \bar{\mathcal{X}}_k} \left(Y_i \times \left\{ 1 - \frac{d(X_{ref}, X_i)}{\sum_{j=1}^{k} d(X_{ref}, X_j)} \right\} \right). \tag{6.3}$$

The weight of a pre-computed result $Y_i = f(X_i)$ of the ith closest scenario is inversely proportional to its distance to the query X_{ref}.
In our example, the interpretation of $\hat{Y}_{ref}$ is a vector of long shore impacts at locations c_i aggregated from the image Y_i of the k closest storms $X_i \in \mathcal{X}$.

One may propose to adapt the distance function in regards to both k and $\mathcal{X}$. Indeed, if those two values are close to each other, then the aggregation procedure would use scenarios that could be far from similar than the request X_{ref}.
An other remark is that using a mean function in the aggregation is subject to smoothing out the expected outputs. Especially in the context of extreme scenarios. Therefore one may use more sophisticated aggregation applications.

6.3 First Results

A first prototype is implemented in Chailan et al. (2012) to validate the pre-computing principle. More specifically, we show how the concept of pre-computed scenarios can be used for early-warning alert system tools. This prototype is complete enough to obtain good performances on a specific academic configuration. However, the full implementation to deal with a real case modelling as an event-scaled coastal hazard questioning, as presented in this chapter, is still on going.

6.4 Discussion

In this chapter we describe the pre-computing principle, aiming at easing the decision-making in the assessment of event-scaled coastal hazards, when promptitude is determining.
Any technical steps of this principle presented here have been implemented and validated in Chailan et al. (2012), but in simpler way. To go further in the industrialisation of the solution, this prototype had to be re-designed to be able to seamlessly perform simulations and store their results. Another point to address is to diagnose what are the best technical choices (software and hardware) to have an efficient IO resource, allowing to store and query multidimensional data sets in a clever way. These perspectives form the motivations of the next chapter.

Beyond the technical implementation, this pre-computing principle may be assisted by other statistical tools in the recommendation towards decision-making. In particular we can highlight the so-called meta-models. The role of such meta-model is to deliver a function $\tilde{f}$, which can mimic a black box model f by stochastically learning from IO couples $(X, Y = f(X))$. Many works have been proposed in the literature. However as soon as the domain of definitions of variables of the scenarios and associated images reach high dimensions, meta-models become hard to handle. Since we work on extreme scenarios, implementing such meta-model applications are a real challenge.
In terms of decision-making helping tools, we decide to limit the focus of this thesis on the pre-computation approach, although meta-modelling is a promising axis of development for an early warning system.

Chapter 7

Platform Prototype

Chapter Summary

Ruled by studies involving resources demanding and complex chains of physical et statistical (extremes) models, coastal hazards assessment activities require efficient IT platform to ease their realisations and help the decision making. This chapter is axed onto the demonstration of a platform prototype aiming at easing the chaining of the models (see Figure 7.1), and developed in the context of this thesis. From an industrial point of view, this platform is a technical base of what we consider the next generation of decision helping tools for coastal hazards.
After detailing our motivation to construct such a platform, its architecture and main components are presented. Then a short case study of numerical model chaining is given to demonstrate one valuable asset of the platform: models chaining.

7.1 Introduction

Ocean engineers and scientists assessing coastal hazards work with complex models which are generally resource consuming. For instance in previous chapters we have presented several steps of modelling, either considering chaining of physical numerical models or statistical ones.
Practically, such modelling process consists in 1) defining a model (f); 2) selecting and formatting input fields in the model's format ($\mathcal{I}_f$); 3) specifying parameters of the model ($\mathcal{P}_f$); 4) running – heavy – calculations on an HPC environment; 5)

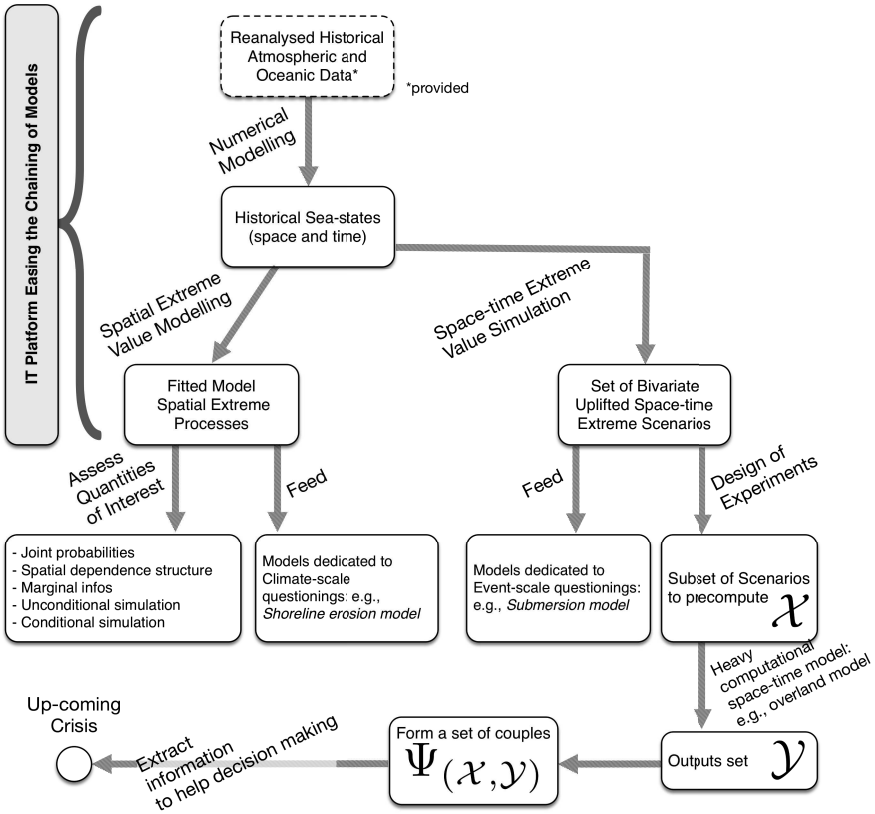


Figure 7.1: This Chapter focus on the description of the created platform prototype, aiming at easing in particular the chaining of hydrodynamic numerical model and (extreme) statistical ones.

analysing simulated outputs ($\mathcal{O}_f$).

Deelman et al. (2009) discusses the recent emergence of Scientific Workflow Management Systems (SWfMS). Such a system gathers computational tools enabling the composition and the automatic execution of complex modelling processes on distributed resources. Beyond their technical implementation, existing SWfMS (e.g., Callahan et al. (2006) or Deelman et al. (2007)) differ from each other on their capabilities (e.g. monitoring execution, dynamic human interference, data visualisation). We adapt and innovate parts of some existing SWfMS to build a platform responding to all requirements of hydrodynamics modelling. For instance, scientists needs to chain several models to compute water levels at a coastal area. Therefore like Deelman et al. (2007) the proposed solution is able to chain models without effort. This is based on a standardization of the IO data sets. Like Hunter et al. (2005) the proposed platform is interfaced to an HPC environment since computations of hydrodynamics models are time and resources demanding. In the same idea of what is possible with Callahan et al. (2006), users can explore the IO data sets stored into the remote environment.

Other features (e.g. monitoring workflow execution through web socket notifications) stem from other computational systems and nowadays technologies like

social networks. This chapter, largely inspired from our original work in Chailan and Rétif (2015), presents the architecture of the designed and developed solution along a discussion on its implementation and use.

7.2 Platform Architecture and Components

The proposed solution is web-oriented (see Figure 7.2). Its design relies on the decomposition of three layers. The Client Layer, the Business Layer and the Infrastructure Layer. Let us understand the Figure 7.2 from a top down approach. Communication between the Client layer and the Business layer is made through

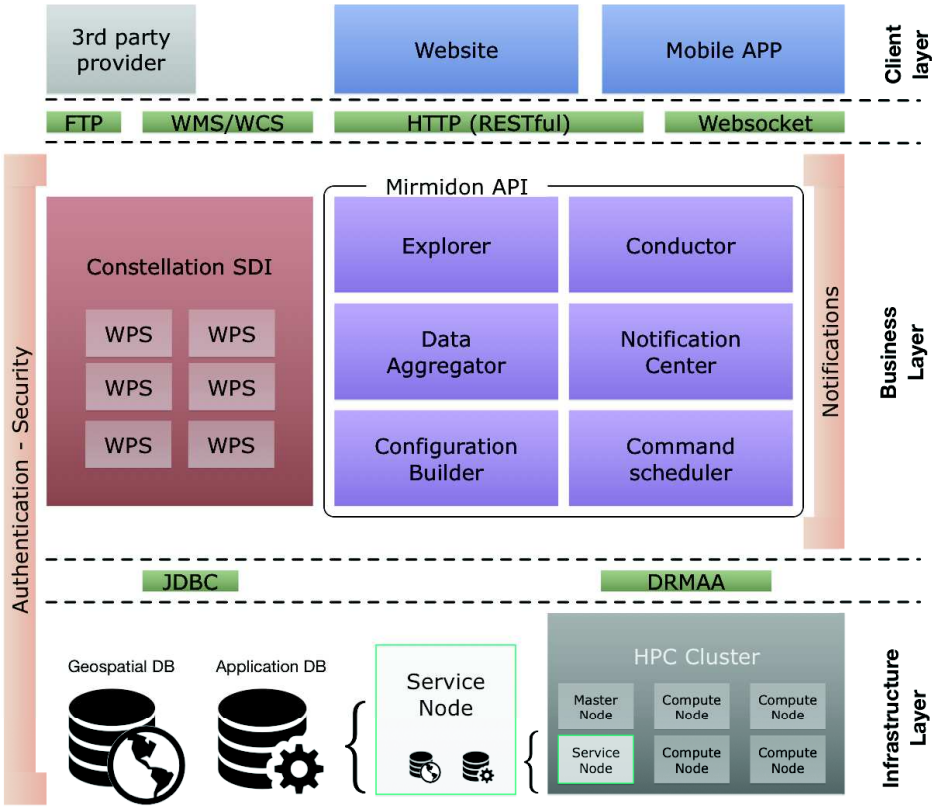


Figure 7.2: Overall architecture of the solution.

a classic HTTP RESTFul API and a FTP server. The FTP server provides a more suitable way to exchange large data files from and to the client side (Gigabyte(s) for input data and Terabyte(s) for output data). Since the platform aims to deal with geophysical model, it has the capacity to represent geospatial phenomena on a map. To do so, standardized web services of the Open Geospatial Consortium (OGC) are used, which supply standard HTTP requests to display data sets on a map such as Web Map Service (WMS). Such service allows to explore dynamically the input and output manipulated data sets.
The proposed solution is centred on the development of an API, so-called Mirmidon-

API. This is a RESTfull API implemented with Play! Java Framework (Reelsen, 2011). The main components of Mirmidon-API are:

1. The Data Aggregator, which provides functions to convert data ($\mathcal{I}_f$) from Proprietary Format to the Mirmidon Format and to insert converted data into a geo-spatial database (Figure 7.3);
2. The Explorer, which delivers functions to display the geo-spatial database records (in Mirmidon Format) on a map (Figure 7.4);
3. The Configuration builder, which gives access to functions to build a configuration (Figure 7.5). A configuration defines a model (f), its binary executable file, its input files ($\mathcal{I}_f$) (in Proprietary Format) and its configured input parameters files ($\mathcal{P}_f$) ready to be executed;
4. The Command scheduler, which provides functions to execute a configuration on HPC clusters;
5. The Conductor, which is in charge of the security, manages access of resources for a given user;
6. And the Notification Centre, which transmits notifications to the client side. Since tasks managed by Mirmidon-API are resources and time demanding, they are run asynchronously. Users have access to the notifications centre using web sockets (WS) to obtain information about the status of those asynchronous tasks (e.g. data transformation or HPC jobs). And all information is gathered into the user's dashboard (Figure).

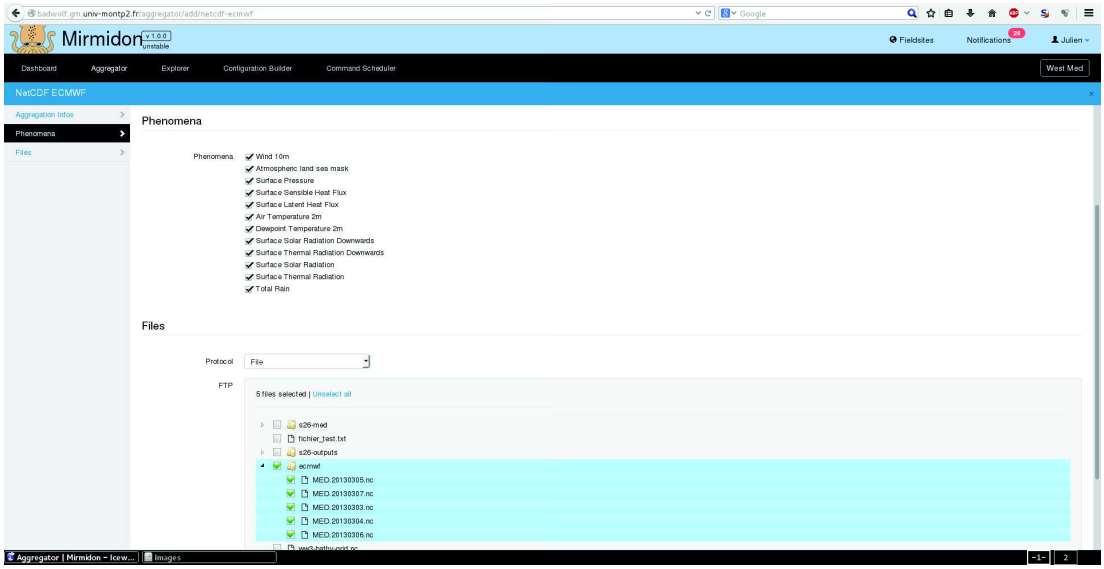


Figure 7.3: Screenshot of the Aggregator module, from where the user can aggregate any data set to the geo-spatial data base of the platform through the use of the Mirmidon Format.

Any data processing like file-format transformations, re-gridding or subtracting data sets are executed behind the use of Web Processing Services (WPS) respecting the OGC's standards. It allows deporting such resource-consuming

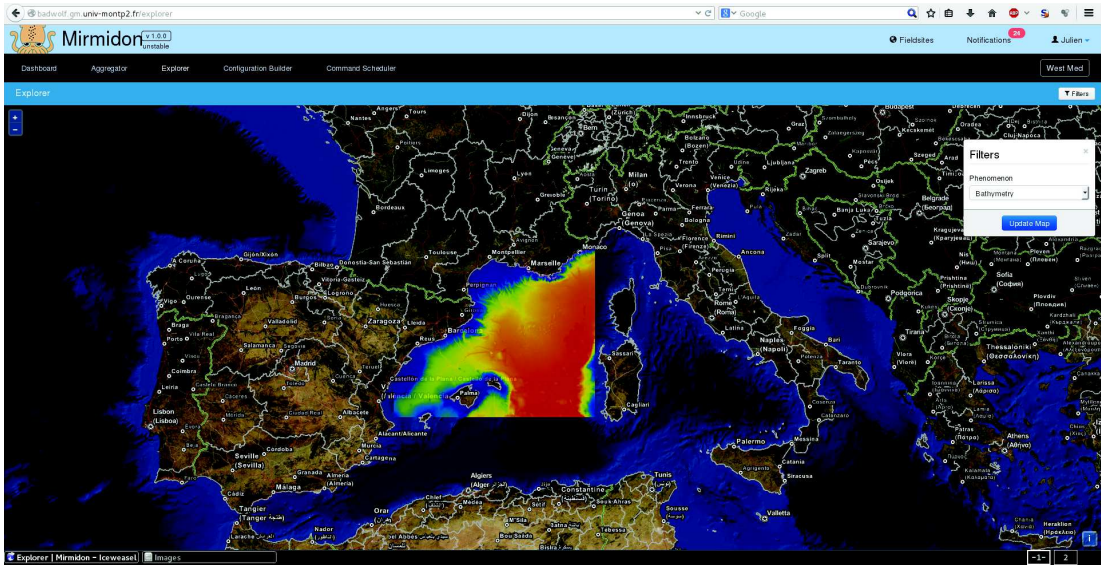


Figure 7.4: Screenshot of the Explorer module, from where the user can remotely consult the data set aggregated into the platform.

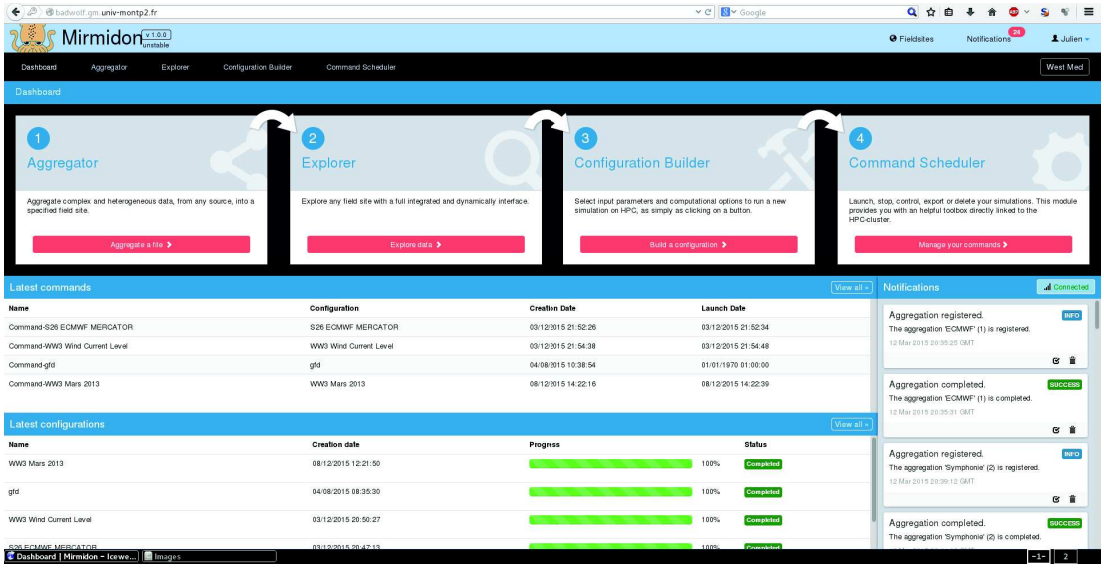


Figure 7.5: Screenshot of the Configuration-builder module, from where the user can create a new configuration (parametrised model and input files) ready to be submitted into the HPC cluster hosting the solution.

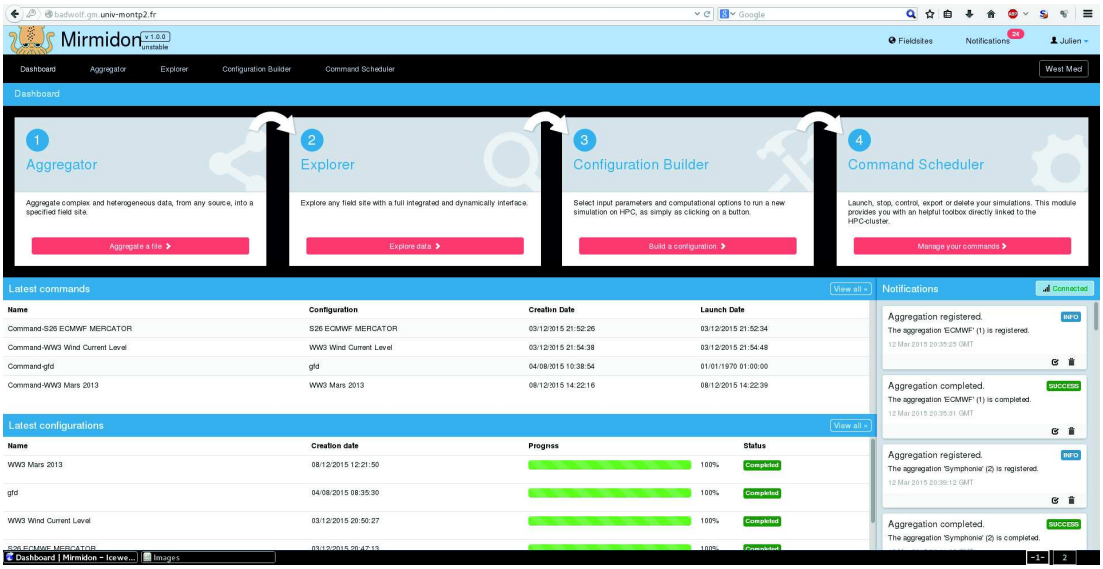


Figure 7.6: Screenshot of the Dashboard module. All information relative to actions of the user is gathered into this dashboard.

processes onto a dedicated server in the same LAN. Those services are deployed on a Spatial Data Infrastructure (SDI) server named Constellation SDI [1].

7.2.1 Data Storage

Two databases and a shared file system (FS) are used for data storage. The first database is a standard PostGreSQL database dedicated to store conventional applications information such as clients information and will not be discussed in this document. Beside this resource management database, a geo-spatial database named CoverageSQL is set up. CoverageSQL stems from the PostGIS extension of PostGreSQL and has been firstly introduced by Desruisseaux (2004). Any input and output data sets are converted in NetCDF binaries files (in Mirmidon Format) by the Data Aggregator module, and then are stored on the FS. Metadata of those files like geographical envelope, variable unity, time stepping and a link to the file is stored on the geo-spatial database. A major benefit of dealing with a geo-spatial database is the flexibility of queries. Whether a user looks after an input or an output data set, queries are composed of the variables (e.g., atmospheric pressure), the geographical envelop and the physical time range of interest. The database automatically retrieves data, gathers them and extracts a new standardized data set.
One remark is that a standard file (Mirmidon Format) follows the Climate and Forecast conventions (CF). This standardization is particularly suitable since data sets manipulated in hydrodynamics modelling (the one largely used in studies assessing coastal hazards) are generally clearly representable by the CF conventions.

1. See www.constellation-sdi.org

7.2.2 Workflow Control, Models Integration and Chaining

The solution uses a hybrid execution control model, which is monitored by the notification centre. Connections between data retrieving, data transformation and job submission are driven by a control-flow: a transfer of control from the preceding task to the one that follows. Inside the model submission itself, the control is data-driven: IO data represent the dependencies between each consecutive actions of the model (i.e. pre-processing, main processing, post-processing). Each model (f) integrated into the solution must be delivered with a toolkit composed of a Writer ($\mathcal{W}_f$) and a Reader ($\mathcal{R}_f$). The so-called Writer consists in converting a data set from the Mirmidon Format to the Proprietary Format of the given model. The converted data-set is denoted ($\mathcal{I}_f$). After computation by the numerical model (f), the output data-set ($\mathcal{O}_f$) is converted by the Reader from the Proprietary Format into the Mirmidon Format and then, according to the user, aggregated into the geo-spatial database. According to this workflow, scientists are able to chain models, in an infinite combination as described by the pseudo-code algorithm 3.

Algorithm 3: Chaining models.

Input :
A sorted list of model to chain: $\mathcal{F} = \{f_1, \ldots, f_n\}$,
I_1 the forcing fields of the first model.
Output: Output fields issue from the chaining of all models $\in \mathcal{F}$.

begin
  DB $\leftarrow I_1$ // Aggregate data for 1st model
  for $f_i \in \mathcal{F}$, $i \in 1, \ldots, n$ **do**
    if $i > 1$ **then** // If is a chaining model, last output becomes input
      $I_i \leftarrow \mathrm{DB}[O_{i-1}]$
    $I_{f_i} \leftarrow \mathcal{W}_{f_i}(I_i)$ // Convert from standardized data
    $O_{f_i} \leftarrow m_i^{(\mathcal{P}_i)}(I_{f_i})$ // Run model f_i
    $O_i \leftarrow \mathcal{R}_{f_i}(O_{f_i})$ // Convert to standardized data
    DB $\leftarrow O_i$ // Aggregate output data
  Serve_Data(O_n). // Serve data via WMS/WCS

7.2.3 Systematic Use of HPC Resources

Hydrodynamics modelling are resources and time demanding. To perform the calculations, users use HPC environments. In this sense an efficient solution must be connected to an HPC system.
Therefore the Business layer is instantiated on a service node of a targeted HPC

cluster. A service node is – like a compute node – a node of the cluster but is dedicated to offer services instead of computational resources. The core API (Mirmidon-API) uses the DRMAA library Troger et al. (2007) as an interface connection to submit calculation jobs on the HPC cluster. In this way, Mirmidon-API can obtain information on the state of jobs, can proceed to their cancellation or can pause it. DRMAA library provides a high level interface for several workload schedulers of the HPC market. We have successfully tested the compatibility of the presented solution with IBM Platform LSF, IBM Tivoli LoadLeveler and Sun Grid Engine.

7.3 Sea State Modelling: a Case Study

We propose to illustrate the capabilities of the platform by modelling sea-states [2] based on the chaining of a coastal circulation model (forced by atmospheric conditions and global currents) and a wave model (see Figure 7.7). This simulation is somehow similar to the one performed to create the hindcast data set presented in chapter 3. Additionally to the atmospheric conditions forcing the circulation model, surface currents and sea water level calculated by the circulation model are used as input of the wave model. We obtain the significant wave height. Assuming models are already implemented into the platform, the user

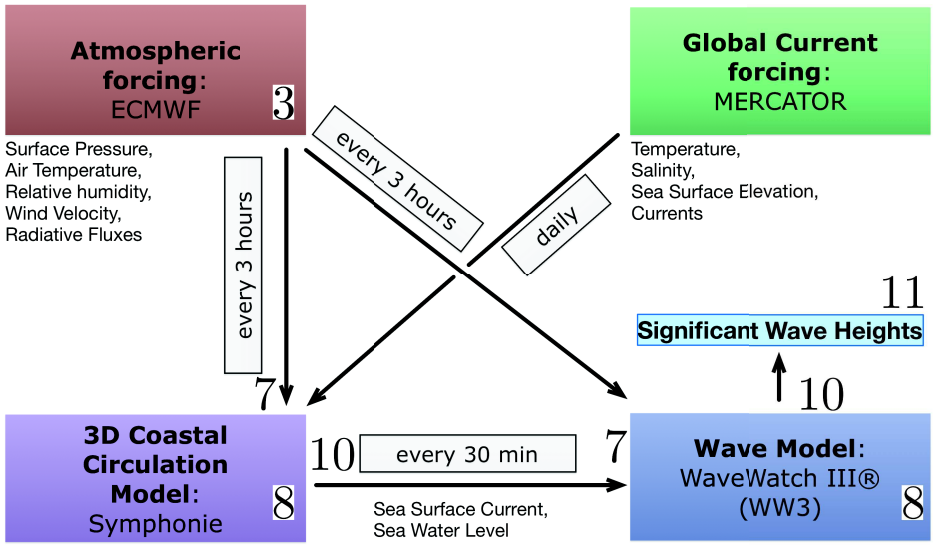


Figure 7.7: Demonstration workflow. Numbers reference the Algorithm 3.

can perform several simulation of significant wave heights without any concerns about having to manage the heavy data set produced and interacting with the HPC system. At the end of the simulation process, the user is notified and can

2. To go further, please consult the video recording of the platform demonstration at https://www.youtube.com/watch?v=st7MnO1QUec.

easily explore the data remotely through the web client interface as illustrated by the screen-shot in Figure 7.4.

7.4 Results

The detailed solution has been deployed on a private cluster during one year, within a research and development project named LittoCMS[3]. Two hydrodynamics models have been chained to produce the accurate sea-states conditions in extreme storms over the French Mediterranean coastline in a final goal of assessing coastal hazards.
As expected the solution can be seen as a dynamic workflow engines. A typical example is the procedure of validation of models chaining, which requires many runs with slightly changes in model parameters. This is made user-friendly and efficient for scientists by the capacity of switching from one workflow to another without any difficulty.
The notification centre and the control-flow of tasks execution are mandatories features in such solution. They alert on issues and allow making quicker decisions on daily use. For instance they allow cutting off a workflow in which tasks do not run correctly in terms or physics (e.g., divergence) or computing (e.g., computational resource crashing). Even if it is a well-known issue in SWfMS, we observe that a lot of time is spent in the IO data transfer from the user-side to the remote environment. Sometimes the transfer requires more time than the actual computations. In this sense, in situ analysis is a necessary feature. Hence the in-situ visualization through layered data access (e.g., WMS) is a real advantage because only satisfying data is downloaded by end-users.

7.5 Discussion

The presented web-oriented architecture aims at easing numerical modelling for coastal engineers and scientists. It is composed of decoupled and modular tools supplying both a transparent access to HPC environments and a user-friendly way of chaining models.
The solution is currently restricted in use to hydrodynamics numerical applications. Without high concerns, we can assume that this solution will perfectly behave for other modelling chains, including the previously presented statistical extreme value models.
On top of this solution and to leverage the approach presented in the previous chapter, both a Design of Experiment and a query-system modules would be implemented aside the presented prototype. They would allow to query the database storing couples of configurations and results. Once added, this additional materials would provide the users with what we consider the next generation of decision

3. Please consult the following website for further information on this industrial project: http://brli.brl.fr/17-projet-litto-cms-36.html.

helping tools for coastal hazards.

Because it has been spotted as a real bottleneck of the solution during the experiments steps, our perspective of work is to obtain better performance over the IO data transfer from and to the platform. From another technical point of view, a second perspective of work is to implement the current solution in a cloud infrastructure. The potential infinite scalability of such infrastructure offers a lot of technical and use-cases perspectives. For instance a crisis mode could be considered, in which the scientists may require the availability of huge computational resources for a very short time window.

Conclusion

We have seen that coastal hazards are grouped in two families that correspond to two different time-scales. They are respectively qualified of long-terms hazards or event-scale ones. In both cases, the most damaging hazards are generally due to extreme meteorological events.
The literature is scarcely provided with approaches to deal with such risk and at these two different time-scales. This is even more verified when the questionings cover not only a single location but an entire region. Taking place at the interface of three disciplines, the work of this thesis is highly motivated by the challenge of creating and exploring new methods that are likely to cover all these aspects, in the final goal of helping the decision-making towards the anticipation and management of coastal hazards.

Since sea-waves are the main source of energy into a littoral system, we have focused our study on this physical phenomenon. The first idea being to study the waves behaviour and thus to extrapolate useful information easing the decision-making on the anticipation and management of up-coming coastal hazards.

In order to quantify the hazards we have studied extreme events: events that are likely to occur only once in years or decades, but that are critically damaging for economical and ecological assets on the coast. Paradoxically, like in any study of extremes, we are interested in values on which the information is the scarcest. To extrapolate information and model extremes, we used a mathematical framework (and its extensions) that is widely accepted in such a case and known as the extreme value theory (EVT).

To apply EVT and extrapolate information to extreme quantities, the most representative data set of the studied phenomena have to be defined. Regarding our case study, namely the Gulf of Lions situated in the North Western Mediterranean sea, the available time series describing the sea-states conditions were scarce both in time and space dimensions. We therefore use an alternative to reproduce historical time series of sea states conditions over this region, using the cutting edge numerical wave models with their last parameterisations and several detailed forcing fields. The hindcast data set produced embraces an historical period of 52-year (1961-2012). Presenting a good performance in the validation step, we have also seen that such modelling may be used to consolidate the avail-

able time-series provided by surface-buoy campaigns.

Relying on this historical data set, we pursued our goal of creating methodologies to assess coastal hazards. Firstly we focused our work on providing long-terms coastal hazards assessment methods. We applied stochastic tools named max-stable processes to spatially model the extreme wave events present in the GOL and provide information on coastal hazards at a spatial and long-term scale. Our will of dealing with spatial processes stem from the definition of environmental phenomena that are generally spatially realised, as it is the case for the wave processes. Clear patterns of extreme waves were identified, most notably by discussing the modelling of the underlying dependence structure of those extreme values by the so-called max-stable models. Such a model was then used to tackle a risk analysis concerning the evaluation of joint probabilities of exceedances of high waves at several locations of the GOL. Their usefulness was demonstrated in this context, which opened perspectives of work that are discussed hereafter.
Secondly we focus our work on approaches able to deal with event-scale coastal hazards questionings. Those hazards are the ones for which the decision-making is generally the most critical. Therefore, any methods provided to ease the assessment of those is a valuable contribution.
We still considered the EVT and its extensions to spatially assess extreme events responsible of event-scale coastal hazards, but the time evolution of an event has to be handled as well in the methodology in order to tackle such questionings. To do so we worked with a threshold-exceedances based approach that allows to stochastically simulating space-time extreme processes of waves (significant wave heights, peak periods and mean directions), of a controlled extremeness. To demonstrate the usefulness of such an approach a second risk analysis is performed. It concerns the long-shore impact to the shoreline from offshore mass flux of energy, provided by the extreme waves processes. This study showed one of the potential application of the simulation of such space-time processes by studying the variability of the impact regarding the different storms and intensities used.

Whether to address long-terms or event-scale coastal hazards, one of the most important result of the stochastic applications is the capacity of simulating extreme events that are suitable to feed coastal physical models. Such a statement allowed us to define the principle of pre-computing, aiming at helping the decision making, especially towards the anticipation and management of event-scale hazards. The pre-computing principle directly uses the stochastically simulated space-time extreme processes. It is inspired from techniques of numerical model exploration and from case-based reasoning concepts.
A first IT platform prototype of the pre-computing principle was implemented. From it we were able to validate this principle on academic data. For real case studies and regarding to the complexity of the models used, of the necessity to chain them onto High Performance computational Clusters and of the massive multidimensional data set to handle, a second prototype have been designed and

developed to tackle these constraints.
The current implementation of this second IT platform prototype could be used in several ways. It was firstly used to ease the coupling of ocean and wave models. However the platform is not finished enough to be used as expected as an early warning system tool. Indeed, the development of the pre-computing module to embed in the platform is still in progress.

We have reviewed the main contributions of this thesis, which naturally lead us to at least as many perspectives of work.

The first result of the thesis is the hindcast data set of sea-states conditions over the North Western Mediterranean sea, motivated by the need of having long time-series but accurate as well. Indeed, we showed in this document that a representative data set is paramount to study the extreme values.
Along surface current and wind, bathymetry has a great role in waves propagation, especially in regions of shallow waters where dissipation and non-linear effects are accentuated. In the GOL region, several measurement campaigns of highly defined bathymetry have been realised in the last few years. Consequently a new and more refined bathymetry has been released after the creation of the presented hindcast. One perspective of work is to reinforce the accuracy of the hindcast by constructing a new computational mesh based on this bathymetry rather than on the former one. The computational mesh would be therefore refined as well, with paying attention to the global computational cost induced by such transformation. Thanks to this improvement, wave processes should be even better represented by the numerical model in very littoral areas.

To assess the long-terms coastal hazards we use some of max-stable processes available in the literature. It would be a valuable add to fit other known max-stable models on the data and compare their performances. In particular the ones that would allow to model an anisotropic underlying dependence structure or other models allowing the asymptotical independence or both.
Still regarding the application of max-stable models, the definition of GEV margins parameters might be improved by the use of other co-variables. As we use the bathymetry, both wave mean direction and fetch distance length are other co-variables that would potentially lead to a better fit of the model.
We have seen that the use of those max-stable models allows simulating conditionally or unconditionally extreme processes. To feed a physical model assessing a long-term coastal hazard questioning with such processes is the purpose of a future work.

Also, a promising perspective of work is to compare our presented results with ones that we would obtain in using the mathematical GPD process framework, being a threshold-based approach. By definition, the physical interpretation of simulated GPD processes is more natural than max-stable models and would

open new perspective in the chaining of models.

The semi-parametric approach is very promising to assess event-scale hazards, as it has been shown in the risk analysis about long-shore impact of the mass flux of energy transported by extreme waves processes. Further implementations of such simulated processes feeding coastal physical models (e.g., overland flood models) must be realised to leverage this approach. Also some physical criteria might be included in the uplift method to improve and control the suitability of the simulated processes. Such applications constitute a consequent perspective of work as well, and would lead us to the consolidation of methods aiming at easing the decision-making towards coastal hazards.

Finally, we have seen that the combination of the work presented in this thesis could be embedded in an IT platform. The basis of this platform have been developed. However some limitations are still existing and deserve to be improved. Most notably the development of the pre-computing module has to be finished. Once it is done, we view the challenge of creating meta-models in respect to the extreme quantities handled as an interesting perspective.
In any case, if it is made possible by the fact that the platform is very modular such an implementation lead us to new technical obstacles. We are likely to face them in the near future in order to provide decision-makers with a complete IT platform, wich would represent the next generation of coastal hazards alert system tools.

Bibliography

Airy, G. B. (1841), « Tides and waves ».

Ardhuin, F. (2011), *Etats de mer : hydrodynamique et applications*, tech. rep.

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., Van Der Westhuysen, A., Queffeulou, P., Lefevre, J.-M., Aouf, L., and Collard, F. (2010), « Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation », *Journal of Physical Oceanography*, 40(9), pp. 1917–1941.

Bacro, J.-N. and Gaetan, C. (2012), « A review on spatial extreme modelling », in: *Advances and Challenges in Space-time Modelling of Natural Events*, pp. 103–124.

Bacro, J.-N. and Gaetan, C. (2014), « Estimation of spatial max-stable models using threshold exceedances », *Statistics and Computing*, 24(4), pp. 651–662.

Bagnold, R. A. (1966), « An approach to the sediment transport problem », *General Physics Geological Survey, Prof. paper*.

Battjes, J. A. and Janssen, J. P. (1978), « Energy loss and set-up due to breaking of random waves », *Coastal Engineering Proceedings*, 1(16).

Bechler, A., Bel, L., and Vrac, M. (2015), « Conditional simulations of the extremal t process: application to fields of extreme precipitation », *Spatial Statistics*, 12, pp. 109–127.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004), *Statistics of extremes*, Theory and applications, With contributions from Daniel De Waal and Chris Ferro, John Wiley & Sons, Ltd., Chichester, pp. xiv+490.

Bel, L., Bacro, J.-N., and Lantuéjoul, C. (2008), « Assessing extremal dependence of environmental spatial fields », *Environmetrics*, 19(2), pp. 163–182.

Benoit, M., Marcos, F., and Becq, F. (1996), « Development of a third generation shallow-water wave model with unstructured spatial meshing », *Coastal Engineering Proceedings*, 1(25).

Beuvier, J., Sevault, F., Herrmann, M. J., Kontoyiannis, H., Ludwig, W., Rixen, M., Stanev, E., Béranger, K., and Somot, S. (2010), « Modeling the Mediterranean Sea interannual variability during 1961–2000: focus on the Eastern

Mediterranean transient », *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C8).

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

Booij, N., Ris, R. C., and Holthuijsen, L. H. (1999), « A third-generation wave model for coastal regions: 1. Model description and validation », *Journal of Geophysical Research: Oceans (1978–2012)*, 104(C4), pp. 7649–7666.

Bretherton, F. P. and Garrett, C. J. (1968), « Wavetrains in inhomogeneous moving media », in: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 302, 1471, The Royal Society, pp. 529–554.

Caires, S., de Haan, L., and Smith, R. L. (2011), *On the determination of the temporal and spatial evolution of extreme events*, tech. rep., report 1202120-001-HYE-004 (for Rijkswaterstaat, Centre for Water Management), Deltares.

Callahan, S. P., Freire, J., Santos, E., Scheidegger, C. E., Silva, C. T., and Vo, H. t. (2006), « VisTrails: visualization meets data management », in: *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, ACM, pp. 745–747.

CERC, U. A. (1984), *Shore protection manual*.

Chailan, R., Bouchette, F., Dumontier, C., Hess, O., Laurent, A., Lobry, O., Michaud, H., Nicoud, S., and Toulemonde, G. (2012), « High performance pre-computing: Prototype application to a coastal flooding decision tool », in: *Knowledge and Systems Engineering (KSE), 2012 Fourth International Conference on*, IEEE, pp. 195–202.

Chailan, R. and Rétif, F. (2015), « A generalised semi-parametric method to simulate extreme space-time event », in: *Applied Computing (AC 2015), Twelth International Conference on*, To appear.

Chailan, R., Toulemonde, G., Bouchette, F., Laurent, A., and Bacro, J.-N. (2015), « A generalised semi-parametric method to simulate extreme space-time event », Preprint.

Chailan, R., Toulemonde, G., Bouchette, F., Laurent, A., Sevault, F., and Michaud, H. (2014), « Spatial assessment of extreme significant waves heights in the Gulf of Lions », *Coastal Engineering Proceedings*, 1(34), management–17.

Coles, S. G. (2001), *An introduction to statistical modeling of extreme values*, Springer.

Coles, S. G. and Tawn, J. A. (1991), « Modelling extreme multivariate events », *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 377–392.

Coles, S. G. and Tawn, J. A. (1994), « Statistical methods for multivariate extremes: an application to structural design », *Applied Statistics*, pp. 1–48.

Cooley, D., Naveau, P., and Poncet, P. (2006), « Variograms for spatial max-stable random fields », in: *Dependence in probability and statistics*, pp. 373–390.

Craik, A. D. (2004), « The origins of water wave theory », *Annu. Rev. Fluid Mech.* 36, pp. 1–28.

Dalrymple, R. A. and Dean, R. G. (1991), *Water wave mechanics for engineers and scientists*, Prentice-Hall.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013a), « Max-stable processes for modeling extremes observed in space and time », *Journal of the Korean Statistical Society*, 42(3), pp. 399–414.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013b), « Statistical inference for max-stable processes in space and time », *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5), pp. 791–819.

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

de Haan, L. (1984), « A spectral representation for max-stable processes », *The Annals of Probability*, 12(4), pp. 1194–1204.

de Haan, L. and Ferreira, A. (2007), *Extreme value theory: an introduction*, Springer Science & Business Media.

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

de Laplace, P.-S. (1810), *Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et sur leur application aux probabilités*, Baudouin.

Deelman, E., Gannon, D., Shields, M., and Taylor, I. (2009), « Workflows and e-Science: An overview of workflow system features and capabilities », *Future Generation Computer Systems*, 25(5), pp. 528–540.

Deelman, E., Mehta, G., Singh, G., Su, M.-H., and Vahi, K. (2007), « Pegasus: mapping large-scale workflows to distributed resources », in: *Workflows for e-Science*, pp. 376–394.

Déqué, M. (2007), « Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values », *Global and Planetary Change*, 57(1), pp. 16–26.

Desruisseaux, M. (2004), « Pertinence des données altimétriques en halieutique appliquées à la pêche thonière », PhD thesis, Paris 6.

Dombry, C. and Eyi-Minko, F. (2013), « Regular conditional distributions of continuous max-infinitely divisible random fields », *Electron. J. Probab*, 18(7), pp. 1–21.

Dombry, C., Éyi-Minko, F., and Ribatet, M. (2012), « Conditional simulation of max-stable processes », *Biometrika*, ass067.

Dombry, C. and Ribatet, M. (2013), « Functional regular variations, pareto processes and peaks over threshold », *Soumis pour publication*, p. 48.

Ewans, K. and Jonathan, P. (2014), « Evaluating environmental joint extremes for the offshore industry using the conditional extremes model », *Journal of Marine Systems*, 130, pp. 124–130.

Faivre, R., Iooss, B., Mahévas, S., Makowski, D., and Monod, H. (2013), *Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement*, Editions Quae.

Ferré, B., Guizien, K., Durrieu De Madron, X., Palanques, A., Guillén, J., and Grémare, A. (2005), « Fine-grained sediment dynamics during a strong storm event in the inner-shelf of the Gulf of Lion (NW Mediterranean) », *Continental Shelf Research*, 25(19), pp. 2410–2427.

Ferreira, A. and de Haan, L. (2014), « The generalized Pareto process; with a view towards application and simulation », *Bernoulli*, 20(4), pp. 1717–1737.

Filipot, J.-F., Roeber, V., Boutet, M., Ody, C., Lathuiliere, C., Louazel, S., Schmitt, T., Ardhuin, F., Lusven, A., Outré, M., Suanez, S., and Hénaff, A. (2013), « Nearshore wave processes in the Iroise Sea: field measurements and modelling », in: *Coastal Dynamics 2013-7th International Conference on Coastal Dynamics*, http://www. coastaldynamics2013. fr/pdf_files/055_Filipot_Jean_Francois. pdf, p–605.

Fisher, R. A. and Tippett, L. H. C. (1928), « Limiting forms of the frequency distribution of the largest or smallest member of a sample », in: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, 02, Cambridge Univ Press, pp. 180–190.

Frappart, F., Calmant, S., Cauhopé, M., Seyler, F., and Cazenave, A. (2006), « Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin », *Remote sensing of Environment*, 100(2), pp. 252–264.

Fu, L.-L. (1996), « The circulation and its variability of the South Atlantic Ocean: first results from the TOPEX/POSEIDON mission », in: *The South Atlantic*, pp. 63–82.

Fujii, Y., Satake, K., Sakai, S., Shinohara, M., and Kanazawa, T. (2011), « Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake », *Earth, planets and space*, 63(7), pp. 815–820.

Galambos, J. (1975), « Order statistics of samples from multivariate distributions », *Journal of the American Statistical Association*, 70(351a), pp. 674–680.

Gaume, J., Eckert, N., Chambon, G., Naaim, M., and Bel, L. (2013), « Mapping extreme snowfalls in the French Alps using max-stable processes », *Water Resources Research*, 49(2), pp. 1079–1098.

Gelci, R., Cazalé, H., and Vassal, J. (1956), « Utilisation des diagrammes de propagation à la prévision énergétique de la houle », *Bull Inform Comité Central Océanogr Etude Côtes*, 8, pp. 170–187.

Gilli, M. and Këllezi, E. (2006), « An application of extreme value theory for measuring financial risk », *Computational Economics*, 27(2-3), pp. 207–228.

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

Gouldby, B., Méndez, F. J., Guanche, Y., Rueda, A., and Minguez, R. (2014), « A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis », *Coastal Engineering*, 88, pp. 15–26.

Groeneweg, J., Caires, S., and Roscoe, K. (2012), « Temporal and Spatial Evolution of Extreme Events », *Coastal Engineering Proceedings*, 1(33), management–9.

Guerinel, B., Bouchette, F., Lobry, O., Astruc, D., Azerad, P., Brambilla, E., Certain, R., Larroudé, P., Manna, M., Meulé, S., Rey, V., Robin, N., Sabatier, F., Sous, D., Martinie, D., and Arnaud, N. (2012), « Monitoring temps reel haute resolution d'un littoral: MAGOBS (Villeneuve-les-Maguelone, Golfe du Lion, France) », *Revue Paralia*, 12(Cherbourg, 12-14 juin 2012), pp. 595–602.

Guizien, K. (2009), « Spatial variability of wave conditions in the Gulf of Lions (NW Mediterranean Sea) », *Vie et milieu*, 59(3), p. 261.

Gumbel, E. J. (1960), « Distributions des valeurs extrêmes en plusieurs dimensions », *Publ. Inst. Statist. Univ. Paris*, 9, pp. 171–173.

Applications pratiques de la base de données CANDHIS de mesures d'états de mer in-situ (2004).

Hasselmann, S., Hasselmann, K., Allender, J. H., and Barnett, T. P. (1985), « Computations and parameterizations of the nonlinear energy transfer in a gravity-wave specturm. Part II: Parameterizations of the nonlinear energy transfer for application in wave models », *Journal of Physical Oceanography*, 15(11), pp. 1378–1391.

Herrmann, M. J., Sevault, F., Beuvier, J., and Somot, S. (2010), « What induced the exceptional 2005 convection event in the North Western Mediterranean basin? Answers from a modeling study », *Journal of Geophysical Research: Oceans (1978–2012)*, 115(C12).

Herrmann, M. J. and Somot, S. (2008), « Relevance of ERA40 dynamical downscaling for modeling deep convection in the Mediterranean Sea », *Geophysical Research Letters*, 35(4).

Hunter, A., Schibeci, D., Hiew, H. L., and Bellgard, M. (2005), « Grendel: A bioinformatics Web Service-based architecture for accessing HPC resources », in: *Proceedings of the 2005 Australasian workshop on Grid computing and e-research-Volume 44*, Australian Computer Society, Inc., pp. 29–32.

Huser, R. and Davison, A. C. (2014), « Space–time modelling of extreme events », *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2), pp. 439–461.

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

Janssen, P. A. (1982), « Quasilinear approximation for the spectrum of wind-generated water waves », *Journal of Fluid Mechanics*, 117, pp. 493–506.

Jin, R., Chen, W., and Sudjianto, A. (2005), « An efficient algorithm for constructing optimal design of computer experiments », *Journal of Statistical Planning and Inference*, 134(1), pp. 268–287.

Joe, H. (1990), « Families of min-stable multivariate exponential and multivariate extreme value distributions », *Statistics & probability letters*, 9(1), pp. 75–81.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009), « Stationary max-stable fields associated to negative definite functions », *The Annals of Probability*, pp. 2042–2065.

Kamphuis, J. W. (1991), « Alongshore sediment transport rate distribution », in: *Coastal Sediments*, ASCE, pp. 170–183.

Lamb, H. (1932), *Hydrodynamics*, Cambridge university press.

Lantuéjoul, C., Bacro, J.-N., and Bel, L. (2011), « Storm processes and stochastic geometry », *Extremes*, 14(4), pp. 413–428.

Lantuéjoul, C. and Bel, L. (2014), « Simulation conditionnelle du processus de Schlather », *46èmes Journées de Statistique de la SFdS*.

Larroude, P. and Oudart, T. (2012), « Sph model to simulate movement of grass meadow of posidonia under waves », *Coastal Engineering Proceedings*, 1(33), p. 56.

Laugel, A., Tiberi-Wadier, A.-L., Benoit, M., and Mattarolo, G. (2014), « ANEMOC-2 Atlantique et Méditerranée: calibration et validation de deux nouvelles bases d'états de mer construites par simulations numériques rétrospectives sur 1979–2010 », in: *XIII e Conférence Journées Nationales Génie Côtier Génie Civil, Dunkirk, France (in French)*.

Lavigne, F., Gomez, C., Giffo, M., Wassmer, P., Hoebreck, C., Mardiatno, D., Prioyono, J., and Paris, R. (2007), « Field observations of the 17 July 2006 Tsunami in Java », *Natural Hazards and Earth System Science*, 7(1), pp. 177–183.

Leadbetter, M. R. (1983), « Extremes and local dependence in stationary sequences », *Probability Theory and Related Fields*, 65(2), pp. 291–306.

Leckler, F., Ardhuin, F., Peureux, C., Benetazzo, A., Bergamasco, F., and Dulov, V. (2015), « Analysis and interpretation of frequency-wavenumber spectra of young wind waves », *Journal of Physical Oceanography*, 45.

Lemieux, C. (2009), *Monte carlo and quasi-monte carlo sampling*, Springer Science & Business Media.

Leredde, Y., Denamiel, C., Brambilla, E., Lauer-Leredde, C., Bouchette, F., and Marsaleix, P. (2007), « Hydrodynamics in the Gulf of Aigues-Mortes, NW Mediterranean Sea: In situ and modelling data », *Continental Shelf Research*, 27(18), pp. 2389–2406.

Li, L., Li, P., and Liu, Y. (2013), « Structural Reliability Based Design and Assessment Acceptance Criteria Development for Fixed Offshore Platforms in South China Sea Under Extreme Storm Conditions », in: *ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering*, American Society of Mechanical Engineers, V02BT02A044–V02BT02A044.

Liu, P. C., Chen, H. S., Doong, D.-J., Kao, C. C., and Hsu, Y.-J. G. (2008), « Monstrous ocean waves during typhoon Krosa », in: *Annales Geophysicae*, vol. 26, 6, Copernicus GmbH, pp. 1327–1329.

Longuet-Higgins, M. S. (1952), « On the statistical distributions of sea waves », *J. mar. Res.* 11(3), pp. 245–265.

Lubin, P. and Glockner, S. (2013), « Detailed numerical investigation of the three-dimensional flow structures under breaking waves », in: *Proc. 7th International Conference on Coastal Dynamics Conference*, pp. 1127–1136.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), « Comparison of three methods for selecting values of input variables in the analysis of output from a computer code », *Technometrics*, 21(2), pp. 239–245.

Michaud, H., Marsaleix, P., Leredde, Y., Estournel, C., Bourrin, F., Lyard, F., Mayet, C., and Ardhuin, F. (2012), « Threedimensional modelling of wave-induced current from the surf zone to the inner shelf », *Ocean Science*, 8, pp. 657–681.

Miles, J. W. (1957), « On the generation of surface waves by shear flows », *Journal of Fluid Mechanics*, 3(02), pp. 185–204.

Millot, C. (1990), « The gulf of Lions' hydrodynamics », *Continental Shelf Research*, 10(9), pp. 885–894.

Mistral, F. (1979), *Lou tresor dóu frelibrige ou Dictionnaire provençal-français*.

Mitchell, T. M. (1997), *Machine Learning*, 1st ed., McGraw-Hill, Inc.

Montgomery, D. C. (2008), *Design and analysis of experiments*, John Wiley & Sons.

Morellato, D. and Benoit, M. (2010), « Constitution d'une base de données d'états de mer le long des côtes françaises méditerranéennes par simulations rétrospectives couvrant la période 1979-2008 », *Revue Paralia*, 3, pp. 5–1.

Morris, M. D. and Mitchell, T. J. (1995), « Exploratory designs for computational experiments », *Journal of statistical planning and inference*, 43(3), pp. 381–402.

Munk, W. H. (1950), « Origin and generation of waves », *Coastal Engineering Proceedings*, 1(1), p. 1.

Nicolet, G., Eckert, N., Morin, S., and Blanchet, J. (2015), « Inferring Spatio-temporal Patterns in Extreme Snowfall in the French Alps Using Max-stable Processes », *Procedia Environmental Sciences*, 26, pp. 24–31.

Oesting, M., Kabluchko, Z., and Schlather, M. (2012), « Simulation of Brown–Resnick processes », *Extremes*, 15(1), pp. 89–107.

Oesting, M., Schlather, M., and Friederichs, P. (2013), « Conditional modelling of extreme wind gusts by bivariate Brown-Resnick processes », *arXiv preprint arXiv:1312.4584*.

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

Penrose, M. D. (1992), « Semi-min-stable processes », *The Annals of Probability*, pp. 1450–1463.

Phillips, O. M. (1961), « A note on the turbulence generated by gravity waves », *Journal of Geophysical Research*, 66(9), pp. 2889–2893.

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

Pickands, J. (1975), « Statistical inference using extreme order statistics », *the Annals of Statistics*, pp. 119–131.

Raillard, N., Ailliot, P., and Yao, J. (2013), « Modeling extreme values of processes observed at irregular time steps: application to significant wave height ».

Reelsen, A. (2011), *Play Framework Cookbook*, Packt Publishing Ltd.

Resnick, S. (1987), *Extreme Values, Regular Variation, and Point Processes (1987)*.

Rétif, F. (2015), « Modélisation du niveau instantané de la mer en conditions paroxysmales : Caractérisation des contributions à différentes échelles de temps et d'espace », PhD thesis, Université de Montpellier.

Ribatet, M. (2013), « Spatial extremes: Max-stable processes at work », *Journal de la Société Française de Statistique*, 154(2), pp. 156–177.

Rootzén, H. and Tajvidi, N. (2006), « Multivariate generalized Pareto distributions », *Bernoulli*, pp. 917–930.

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

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

Shemdin, O., Hasselmann, K., Hsiao, S. V., and Herterich, K. (1978), « Nonlinear and linear bottom interaction effects in shallow water », in: *Turbulent fluxes through the sea surface, wave dynamics, and prediction*, pp. 347–372.

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

Smith, R. L., Tawn, J. A., and Yuen, H. (1990), « Statistics of multivariate extremes », *International Statistical Review/Revue Internationale de Statistique*, pp. 47–58.

Stein, M. (1987), « Large sample properties of simulations using Latin hypercube sampling », *Technometrics*, 29(2), pp. 143–151.

Stokes, G. G. (1849), « On the theory of oscillatory waves », 8, pp. 441–455.

Tawn, J. A. (1988), « Bivariate extreme value theory: models and estimation », *Biometrika*, 75(3), pp. 397–415.

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

Tayfun, M. A. (1980), « Narrow-band nonlinear sea waves », *Journal of Geophysical Research: Oceans (1978–2012)*, 85(C3), pp. 1548–1552.

Thibaud, E. and Opitz, T. (2015), « Efficient inference and simulation for elliptical Pareto processes », *arXiv preprint arXiv:1401.0168*.

Tolman, H. L. (1991), « A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents », *Journal of Physical Oceanography*, 21(6), pp. 782–797.

Tolman, H. L. (2002), « Alleviating the garden sprinkler effect in wind wave models », *Ocean Modelling*, 4(3), pp. 269–289.

Tolman, H. L. (2008), « A mosaic approach to wind wave modeling », *Ocean Modelling*, 25(1), pp. 35–47.

Tolman, H. L. (2014), *User Manual and System Documentation of WAVEWATCH III® version 4.18*, tech. rep. 316.

Toulemonde, G., Ribereau, P., and Naveau, P. (2015), « Applications of Extreme Value Theory to Environmental Data Analysis », in: *Extreme Events: Observations, Modeling, and Economics (Geophysical Monograph Series)*, in press.

Troger, P., Rajic, H., Haas, A., and Domagalski, P. (2007), « Standardization of an API for distributed resource management systems », in: *Cluster Computing and the Grid, 2007. CCGRID 2007. Seventh IEEE International Symposium on*, IEEE, pp. 619–626.

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

Wadsworth, J. L. and Tawn, J. A. (2012), « Dependence modelling for spatial extremes », *Biometrika*, 99(2), pp. 253–272.

Wang, Y. and Stoev, S. A. (2011), « Conditional sampling for spectrally discrete max-stable random fields », *Advances in Applied Probability*, pp. 461–483.

Weiss, J., Bernardara, P., and Benoit, M. (2014), « Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights », *Journal of Geophysical Research: Oceans*, 119(5), pp. 2906–2922.

Whitham, G. B. (1965), « A general approach to linear and non-linear dispersive waves using a Lagrangian », *Journal of Fluid Mechanics*, 22(02), pp. 273–283.