Statistical learning for coastal risks assessment

Aurélien Callens

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR
ÉCOLE DOCTORALE 211

Laboratoire de Mathématiques et de leurs Applications de Pau (LMAP)

Statistical learning for coastal risks assessment

Aurélien Callens

A thesis submitted for the degree of
Doctor of Philosophy in Mathematics

Thesis defense planned for the 16th September 2021 in front of the thesis committee composed by:

Benoit Liquet	Professor	UPPA	Supervisor
Denis Morichon	Associate Professor	UPPA	Supervisor
Deborah Idier	Researcher	BRGM	Reviewer
Pierre Lafaye de Micheaux	Associate Professor	UM	Reviewer
Bruno Castelle	Director of Research	CNRS	Examiner
David Callaghan	Senior Lecturer	UQ	Examiner
Kerrie Mengersen	Professor	QUT	Examiner
Tom Baldock	Professor	UQ	Examiner

Abstract

Over the last decades, the quantity of data related to coastal risk has greatly increased with the installation of numerous monitoring networks. In this era of big data, the use of statistical learning methods (SLM) in the development of local predictive models becomes more legitimate and justified. The objective of this thesis is to demonstrate how SLM can contribute to the improvement of coastal risk assessment tools and to the development of an early warning system which aims to reduce coastal flooding risk.

Three methodologies have been developed and tested on real study sites. The first methodology aims to improve the local wave forecast made by spectral wave model with machine learning methods and data from monitoring networks. We showed that data assimilation with machine learning methods improve significantly the forecast of wave parameters especially the wave height and period. The second methodology concerns the creation of storm impact databases. Even though these databases are essential for the disaster risk reduction process they are rare and sparse. We therefore proposed a methodology based on a deep learning method (convolutional neural networks) to generate automatically qualitative storm impact data from images provided by video monitoring stations installed on the coast. The last methodology is about the development of a storm impact model with a statistical method (bayesian network) based exclusively on data acquired with diverse monitoring networks. With this methodology we were able to predict qualitatively the storm impact on our study site, the Grande Plage of Biarritz.

Keywords: Coastal flooding; Coastal risk; Deep learning; Early warning system; Machine learning; Monitoring network; Statistical learning methods.

Résumé

Au cours des dernières décennies, la quantité de données relatives aux risques côtiers a fortement augmenté avec l'installation de nombreux réseaux de surveillance. Dans cette ère de big data, l'utilisation de méthodes d'apprentissage statistique dans le développement de modèles prédictifs locaux devient de plus en plus légitime et justifiée. L'objectif de cette thèse est de démontrer comment les méthodes d'apprentissage statistique peuvent contribuer à l'amélioration des outils d'évaluation des risques côtiers et au développement d'un système d'alerte précoce qui vise à réduire le risque d'inondation côtière.

Trois méthodologies ont été développées et testées sur différent sites d'étude. La première méthodologie vise à améliorer les prévisions locales de vagues faites par un modèle spectral de vagues avec des méthodes d'apprentissage automatique et des données provenant de réseaux de surveillance. Nous avons montré que l'assimilation de données avec des méthodes d'apprentissage automatique améliore de manière significative la prévision des paramètres des vagues, en particulier la hauteur et la période des vagues. La deuxième méthodologie concerne la création de bases de données sur l'impact des tempêtes. Bien que ces bases de données soient essentielles dans le processus de réduction des risques de catastrophes, elles sont rares et peu nombreuses. Nous avons donc proposé une méthodologie basée sur une méthode d'apprentissage profond (réseaux de neurones convolutifs) pour générer automatiquement des données qualitatives sur l'impact des tempêtes à partir d'images fournies par des stations de surveillance vidéo installées sur les côtes. La dernière méthodologie concerne le développement d'un modèle d'impact de tempêtes avec une méthode statistique (réseau bayésien) basée exclusivement sur des données acquises avec divers réseaux de surveillance. Grâce à cette méthodologie, nous avons pu prédire de manière qualitative l'impact de tempêtes sur notre site d'étude, la Grande Plage de Biarritz.

Mots-clés: Apprentissage automatique; Apprentissage profond; Inondations côtières; Méthodes d'apprentissage statistique; Réseau de surveillance; Risque côtier; Système d'alerte précoce.

Acknowledgements

First and foremost, I would like to express my sincere thanks and gratitude to my PhD supervisors, Benoit Liquet and Denis Morichon, for giving me their trust and proposing me numerous opportunities over the past years. I would like to thank Benoit in particular for his continuous cheerfulness and broad knowledge about statistics that made our weekly meetings always rich and more than pleasant. I would also like to thank Denis who shared with me his deep knowledge about coastal engineering which was not my area of expertise and who helped me improve significantly my academic writing with his remarks and numerous proof-readings. Their guidance is without a doubt a key ingredient of this thesis.

Beside my advisors, I would like to thank Deborah Idier and Pierre Lafaye de Micheaux for their constructive remarks and suggestions about my work and for taking the time to read and review this thesis.

I would like to express my gratitude to the persons who collaborated with me during this PhD thesis and who made it all the more interesting: Stéphane Abadie, Matthias Delpey, Bruno Castelle, Agnès Petrau, Maialen Sagarduy, Pedro Liria, Irati Epelde.

I would like to give a special thank to all the personal of the doctoral school and the Laboratory of Mathematics and their applications of Pau, who helped me in any way for the administrative tasks and for the inscription of training courses.

I also thank my PhD and postdoc colleagues: Fania, Claire, Sebastien, Floren, Bastien, Téo for the stimulating discussions we had and the good time we shared during our meetings or during training courses.

From a more personal point of view, I am extremely grateful to my parents, my brother and my grandparents for their love and continuous support throughout my life and especially during this important period. I am also grateful to my friends Arthur, Charly for the gaming sessions spent together and their humor that always cheered me up. Finally, I am forever thankful to Mylène, my love and partner in

life, for reassuring and supporting me with so much kindness and patience at every step of this project.

Contents

Acronyms

ANN Artificial neural networks.

BN Bayesian networks.

CNN Convolutional neural networks.

DBN Dynamic bayesian networks.

DRR Disaster risk reduction.

EWS Early warning system.

SLM Statistical learning methods.

SNN Shallow neural networks.

SOM Self organizing maps.

1. Introduction

1.1 General context

Coastal areas have always been attractive habitats for human communities because of the abundant resources and valuable services provided by the diverse ecosystems associated with these areas (Barbier et al., 2011; Mehvar et al., 2018) and their strategic positioning at the interface between land and sea which offers access points to marine trade and transport. Over the last decades, the coastal zones have known an active urbanization (Seto et al., 2011) and population growth. Nowadays, the population density in coastal areas is significantly higher than in inland areas (Neumann et al., 2015). At world level, the coastal areas represent only 4% of the earth's total land area (Barbier, 2013), whereas they host 30% of the World's population (MEa, 2005). The same phenomenon is observed for Europe (Figure 1.1) where it is estimated that one third of the population live within 500 meters of the European seas or Oceans (European Commission, 2012).

The coastal zones also play a major role in the economy of human communities. As evoked earlier, their strategic positioning enables the construction of ports which are essential for international trade and transport. However, most of the economic value of these areas comes from the resources and services provided by the ecosystems present in these areas (Barbier et al., 2011). The main valuable services are tourism, recreation and storm protection services (Mehvar et al., 2018). In Europe, one third of the gross domestic product is produced within 500 meters of European seas and the economic value of these areas has been estimated between 500 and 1,000 billion Euros (European Commission, 2012).

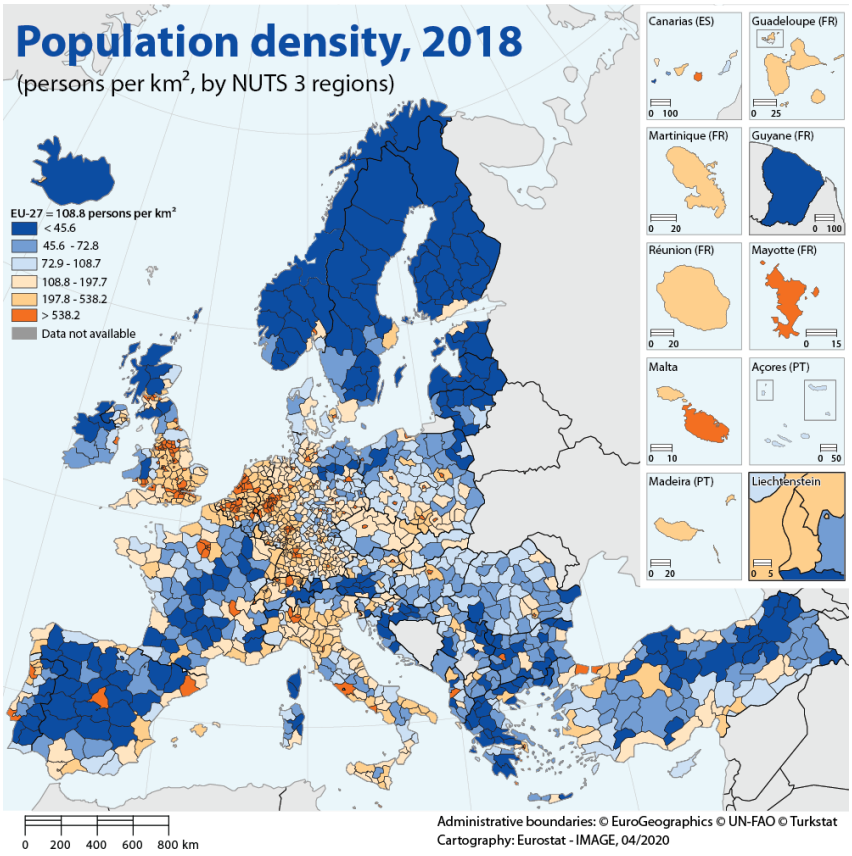


Figure 1.1: Map representing the population density for Europe in 2018. This map was reprinted from `ec.europa.eu/eurostat`.

1.2 Coastal risk

Coastal areas are exposed to different types of natural hazards such as storm waves, flood and erosion that can pose significant risks to assets and communities installed in these areas. According to the terminology of the United Nations relating to disaster risk reduction (UNGA, 2016), a risk is referred to the potential loss of life, injury, or destroyed or damaged assets which could occur to a system, society or a community in a specific period of time, determined probabilistically as a function of hazard, exposure, vulnerability (Figure

1.2).

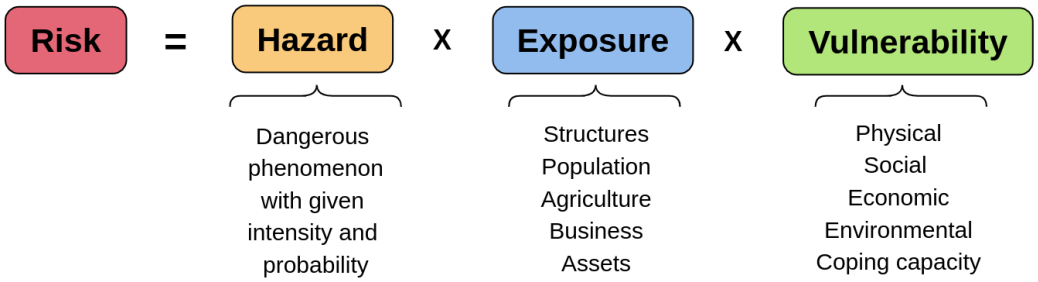


Figure 1.2: The concept of risk.

In environmental studies, **hazard** usually refers to a phenomenon with given intensity and frequency that may cause loss of life, property damage or economic disruption (non exhaustive list). In the case of coastal areas, the hazards are mainly natural hazards such as storm waves, coastal flooding or erosion. Two categories of hazards can be distinguished: rapid onset hazard that occurs at a time scale of days to weeks and slow onset hazard that occurs at a time scale of decades to centuries (Hill et al., 2020). **Exposure** alludes to the communities, the properties or other human assets which are subject to potential losses because of their location in hazard-prone areas (UNGA, 2016). Finally, **vulnerability** corresponds to the physical, social, economic and environmental factors associated with these exposed communities, properties or other human assets which increase their susceptibility to the impact of hazards.

1.2.1 Coastal flooding

In this thesis, a particular attention is paid to coastal flooding occurring on highly urbanized cities. This type of hazard falls in the class of rapid onset hazards. A flood is a general and temporary inundation of normally dry land areas. It is considered as coastal flooding when coastal processes such as tide, waves and storm surge are involved (Yang and Liu, 2020). This hazard

occurs during extreme water-level events when the total water level exceeds the elevation of a defense infrastructure. It can cause severe damages on assets or buildings located behind the defense infrastructures (Figure 1.3).

Figure 1.3: Coastal flooding on the Grande Plage de Biarritz during the Christine storm (04/03/2014). Image reproduced from www.rtl.fr.

The total water level is a combination of several components, namely the astronomical tide, the storm surge and the wave runup (Figure 1.4).

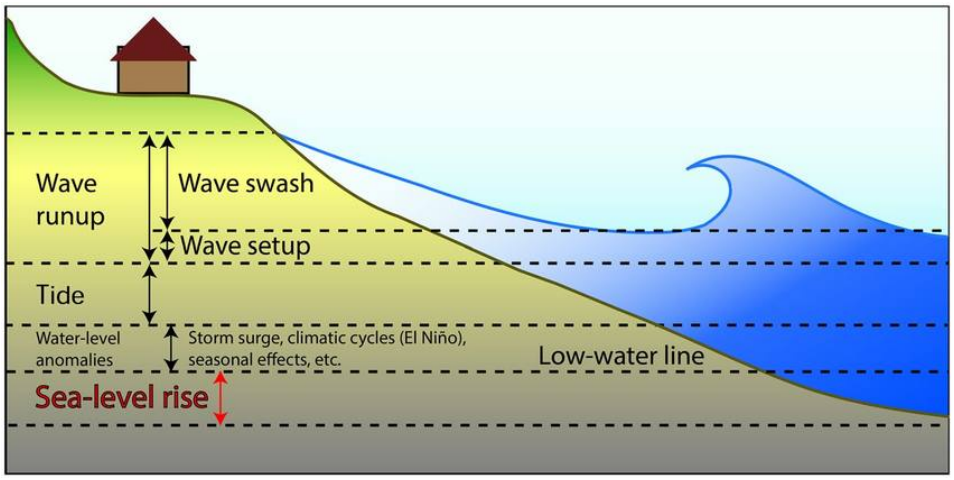


Figure 1.4: Different components of total water level. Reprinted from Vitousek et al. (2017).

Astronomical tide

The astronomical tide is one of the major components of the total water level. Tides are the rising and falling of the surface of the ocean caused by the combined gravitational attraction of the sun and the moon. In some regions, the tide level will determine whether or not there will be a flooding. For instance, it has been shown that coastal floodings occurred most of the times during high tides in the central part of the bay of Biscay (Breilh et al., 2014). Spring tides can also be an aggravating factor in coastal flooding. Twice a lunar month, the Earth, sun, and moon are nearly in alignment (full moon and new moon) which results in the so-called spring tides that are tides with slightly larger tidal range. A storm event occurring during a high spring tide is more likely to cause severe floods.

Storm surge

A sea surge corresponds to the difference between the observed sea level and the astronomical tidal level (Pirazzoli, 2000). This difference is the result of the interactions between the sea and several forcings such as the atmospheric

pressure, the winds or the waves. Atmospheric pressure has a significant effect on the surge level. This phenomenon is known as inverse barometric effect. It has been estimated that a decrease of 1 hPa in atmospheric pressure leads to an elevation of 1 cm in total water level (Harris, 1963). Winds can also move water masses toward the coast and lead to increased surge level. In shallow coastal areas, winds have generally stronger effect on sea surge than atmospheric pressure (Arnaud and Bertin, 2014). Finally, it has been proven that sea state can influence significantly the surge level (Bertin et al., 2015). During a storm, the extreme conditions (low pressure, high winds and large waves) lead to a significant surge which is qualified as storm surge.

Wind waves induced components

By breaking on the shore, waves have a direct effect on the total water level. This effect is named "wave runup" and is defined as the maximum vertical extent of wave uprush on a beach or structure above the still water level. The wave runup is the sum of two components: "wave setup" which is a nearly static component and "swash" which is a dynamic component at the wave scale. Wave setup is the increase in mean water level above the stillwater level due to momentum transfer to the water column by waves that are breaking or otherwise dissipating their energy. The swash corresponds to the vertical elevation of the lens of water that washes up on the beach after a wave has broken. The swash dynamics is controlled by incident waves (period smaller than 20s) and infragravity waves (period ranging between 30 and 300s) (Stockdon et al., 2006). The contribution of each component varies with the beach type in such a way that swash in dissipative beaches is dominated by infragravity waves, and in reflective beaches by incident waves (Plant and Stockdon, 2015).

During a storm event, wave height and period are generally larger than average. This can cause a significant elevation of the total water level because wave height and period are positively correlated with wave runup (Diwedar, 2016). It has been shown that the contribution of waves to the total water

level is more important than the contribution of winds in zones where the continental shelf is reduced (Kennedy et al., 2012). Waves can also interact with other components of the total water level. Several authors have demonstrated that waves can influence the storm surge (Nicolle et al., 2009; Bertin et al., 2012, 2015). Indeed, the sea state can enhance the wind friction, inducing the movement of larger bodies of water toward the coast which results in larger storm surge.

1.2.2 Coastal flooding and global change

The risk related to coastal flooding is expected to increase significantly in the future. On one hand, global warming will have an impact on all the components of the total water level which are responsible for coastal flooding. Since the beginning of the last century, the global mean sea level has been rising at an accelerated rate (Chen et al., 2017) and is expected to keep rising during the next decades, reaching an increase of 50 cm (RCP 4.5) to 80 cm (RCP 8.5) by the end of the century (Kopp et al., 2014; Mengel et al., 2016). The sea level rise and the global warming will induce changes in the amplitudes and phases of tides (Idier et al., 2017; Pickering et al., 2017). Even though an increase in storminess in the north Atlantic for the future decades is still debatable (Bengtsson et al., 2009; Feser et al., 2015), wind waves (Perez et al., 2015) and storm surges (Marcos et al., 2011; Little et al., 2015) will be affected by global warming. These changes, especially in mean sea level, will result in more frequent and intense coastal floodings (Vitousek et al., 2017; Vousdoukas et al., 2018b; Taherkhani et al., 2020).

On the other hand, the stakes on coastal areas will keep growing. More people will be exposed to coastal flooding as the populations in these zones are expected to increase significantly in the future decades (Neumann et al., 2015). Urban areas will keep developing in coastal zones to welcome the growing populations resulting in more assets, buildings and infrastructures to protect from coastal flooding. In the study of Vousdoukas et al. (2018a), it has been esti-

mated that the current expected annual number of people exposed to coastal flooding of 102 000 will reach 1.52-3.65 million by the end of the century for Europe depending on the scenario considered. Concerning the current expected annual damage for Europe, the authors found an increase by two to three orders of magnitude by the end of the century if nothing is done (from 1.25 billion to 93-961 billion euros depending on the scenario used).

1.3 Reducing coastal risk

Given the current flooding risk and its future increase in coastal areas, efficient disaster management is needed to identify, assess or reduce the risks of disaster. Generally, the process of Disaster risk reduction (DRR) is represented as a cycle (Figure 1.5) with 4 steps: response, recovery, prevention, preparedness.

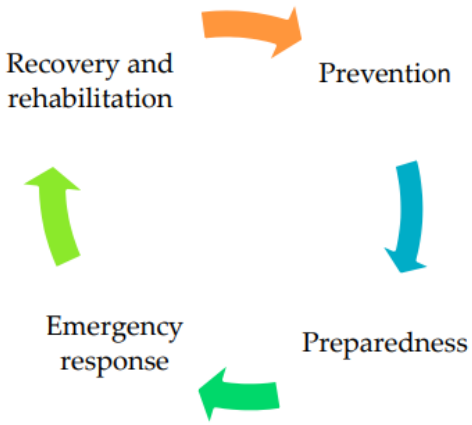


Figure 1.5: Disaster reduction cycle. Reprinted from Yang and Liu (2020)

The response corresponds to the actions and measures taken directly before, during or immediately after a disaster in order to save lives, reduce health impacts, ensure public safety (UNGA, 2016). The recovery is the step occurring in the post-disaster phase, and corresponds to the restoring of the economic, social, environmental assets (non exhaustive list) affected by this hazard. Prevention and preparedness steps occur during the pre-disaster phase.

The prevention step corresponds to all the activities and measures taken to avoid existing and new disaster risks. Finally, the preparedness step represents the knowledge and capacities developed by governments, communities to effectively anticipate, respond to the impacts of likely, imminent or current disasters (UNGA, 2016).

Among all the measures employed during the disaster management process, we can distinguish two families with different aims: mitigation or preparedness. The measures aiming to mitigate directly the impact of a coastal hazard are often structural measures that correspond to engineering constructions. The measures aiming to prepare and prevent a disaster are often nonstructural measures such as actions, legislation or warning systems (Meyer et al., 2012). Our interest lies in these measures as they are more proactive, relatively cost effective and require a short time to implement.

1.3.1 Preparedness

The measures aiming for preparedness focus on two key factors which are the knowledge about the hazards and the ability to predict in advance such events to give an early warning to the communities.

Coastal monitoring networks

The most efficient way to build knowledge about hazards is to install permanent monitoring systems. Over the last decades, numerous coastal monitoring systems have been installed along the coasts in order to study the processes that are responsible for coastal flooding. These systems are generally regrouped into networks to provide continuous and sustainable data. Along the french coasts, we find the RONIN network operated by SHOM that regroups tidal gauges which are measuring the total water level, the CANDHIS network operated by CEREMA constituted by numerous offshore directional wave buoys which record the wave characteristics and a network of weather stations operated by Metéo-France. There are also remote sensing systems such as video monitoring station or satellites. Video monitoring

systems are installed on the beach to study coastal processes such as beach morphology changes, wave runup, and coastal currents (Splinter et al., 2018; Buscombe and Carini, 2019). Examples of networks for video monitoring systems are euskoos (https://www.kostasystem.com/fr/) for the Basque coast, WebCAT (https://secoora.org/web-cameras/) for the east coast of USA or the coastal imaging network of the water research laboratory of New South Wales (http://ci.wrl.unsw.edu.au/) for australian coasts. Satellites are employed to perform various tasks such as mapping ocean currents (Klemas, 2012), estimating wave parameters (Shao et al., 2016; He et al., 2006) and also monitoring meteorological conditions (Kidd et al., 2009). During disaster or post-disaster phase, remote systems can assess the impact of the hazard which can be precious for building Early warning system (EWS) (Klemas, 2009). Data collected from monitoring systems are essential for the preparedness as they are used as input for statistical learning models or EWS (Valchev et al., 2014; Van Dongeren et al., 2018). They are also used as ground truth to calibrate and validate numerical wind wave models (Lavidas and Venugopal, 2018).

Predictive models and Early warning systems

To complement coastal monitoring systems, the development of accurate predictive models and Early warning system (EWS) are crucial measures for the preparedness step. According to UNGA (2016), an EWS designate a system including four interrelated key elements: (1) disaster risk knowledge based on the systematic collection of data and disaster risk assessments; (2) monitoring, analysis and forecasting of the hazards and possible consequences; (3) dissemination and communication of timely accurate warnings and associated information on likelihood and impact; and (4) preparedness at all levels to respond to the warnings received. The development of EWS is considered as the most cost-effective measure, as EWS save both lives and properties (Rogers and Tsirkunov, 2011).

Various methods can be used inside an EWS to forecast the characteris-

tics of a hazard or predict its impact. These methods can be divided into 2 main groups: physics-based methods (numerical modeling) and data-driven methods (statistical learning). Physic-based models are representations of physical processes in mathematical terms. They are usually composed by one or more governing equations that are based on theory and fundamental knowledge. Because these equations are often impossible to solve or because initial or boundary conditions are not known (Larson, 2005), numerical modeling techniques are usually employed to approximate solutions. On the opposite, data-driven approaches are based on the analysis of the data about a specific system. They aim to find relationships between the system variables (input and output) without explicit knowledge of the physical behavior of the system (Solomatine and Ostfeld, 2008).

This thesis focuses on the data-driven methods which are getting more attention over the last decade because of the increasing amount of data collected by the monitoring networks and the rapid development of data science.

1.4 Statistical learning methods

Statistical learning is one of the data-driven approaches. It is a field combining machine learning and statistics, it regroups tools and methods for modeling, predicting and understanding complex data. SLM aim to solve different type of learning problems such as unsupervised learning which aims to find the structure of the data or supervised learning which aims to map data to a desired output. The latter is our main focus because it has applications in predictive modeling. Over the past year, the field of statistical learning has gained increasing interest and attention as can be seen from the Figure 1.6.

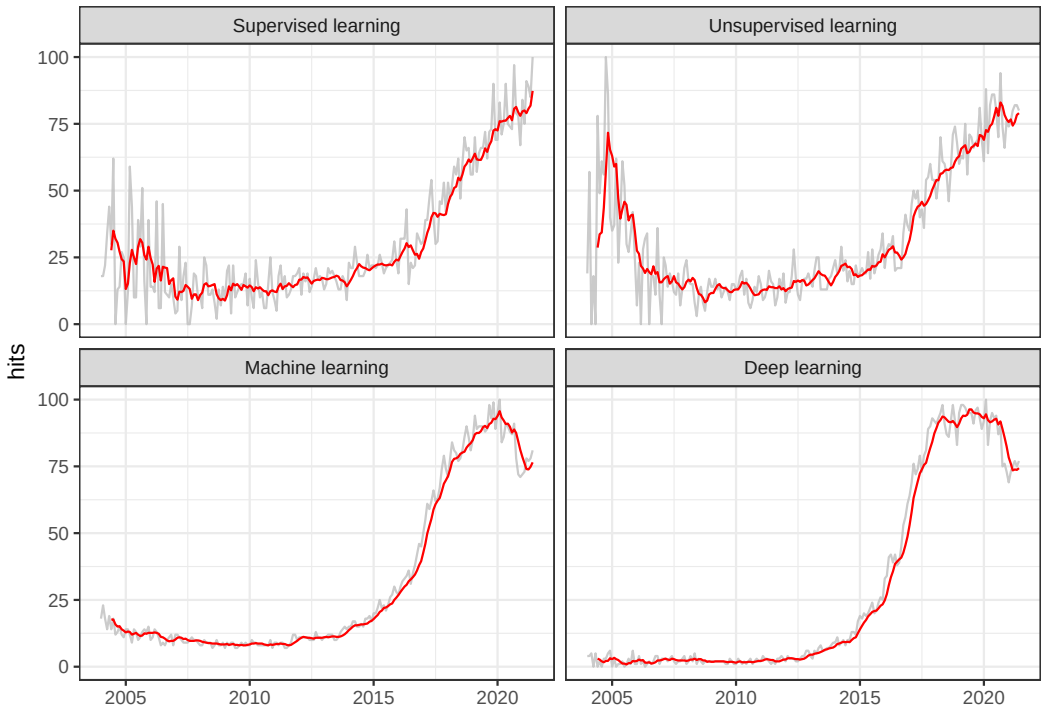


Figure 1.6: Normalized measure of the interest made by Google Trends for different terms related to statistical learning. Measures are represented in grey lines and moving average over 6 months are represented in red lines. Calculation of the normalized measure is detailed here: `https://support.google.com/trends/answer/4365533?hl=en`.

Our interest will lie upon the use of three categories of SLM namely machine learning, deep learning and statistical methods in coastal risk analysis. Machine learning methods are algorithms that learn pattern from pre-defined data features in order to make the most accurate predictions for new data. The features extraction in machine learning is a manual process that requires domain knowledge. Although they are considered as black box models, they are widely employed due to their prediction accuracy (Burkart and Huber, 2021). Deep learning is sub-class of machine learning methods based on deep artificial neural networks which are algorithms inspired by the structure and

function of the brain. Deep learning methods aim to discover the mapping from data features to the target variable but also discover which features to extract from data contrary to machine learning methods. Deep learning methods generally require a large amount of data and computational power to train. Finally, statistical methods are methods based on statistic, they can be used as predictive models but they are more oriented toward the understanding of the data and the study of relationships between variables.

1.4.1 Machine learning

In coastal risk analysis, machine learning methods are commonly employed due to their predictive ability. They are employed to perform short-term forecast of coastal processes responsible for coastal flooding or to improve forecast of the same processes made by process-based numerical models.

Concerning these two applications, we observed in the literature a preference for Shallow neural networks (SNN) as a machine learning algorithm. Indeed, they are used as off-the-self method in numerous works. SNN designate a class of feed-forward Artificial neural networks (ANN). They are generally simple ANN with three layers of nodes: an input layer, one hidden layer and an output layer. They have been employed to obtain short-term forecast of wave parameters, storm surge and tide level (Table 1.1). The explanatory variables (input) are mainly measurements obtained from monitoring networks (buoy, weather stations, tide gauge) or predictions made by numerical wave models.

They have also been employed to improve the predictions made by numerical wind wave models with a data assimilation procedure called error prediction method (Babovic et al., 2001; Makarynskyy et al., 2005; Moeini et al., 2012; Deshmukh et al., 2016; Londhe et al., 2016). In this case, the explanatory variables are wave parameters simulated by the wind wave model and meteorological variables and the response variable are the deviations of the wind wave model. In Londhe et al. (2016), this data assimilation method with SNN led to a significant improvement of significant wave height forecast, with the

correction of the underestimation (no remaining bias after the assimilation) and RMSE in average 50 % lower.

Table 1.1: Scientific works using SNN in the predictive modeling of tide, waves and storm surge.

Modeled parameter	References
Tide	Lee (2004), Makarynskyy et al. (2004), Makarynska and Makarynskyy (2008), Granata and Di Nunno (2021)
Wave height	Deo et al. (2001), Tsai et al. (2002), Makarynskyy (2005), Mandal and Prabaharan (2006), Browne et al. (2007)
Storm surge level	Lee (2006), Tseng et al. (2007), You and Seo (2009), Hashemi et al. (2016), Bezuglov et al. (2016), Kim et al. (2018), Lee et al. (2018), Quintana et al. (2021)

Despite the popularity of SNN, other machine learning methods have been employed to predict coastal processes responsible for coastal flooding. In the literature, significant wave height has been predicted by support vector machines (Elgohary et al., 2017; Berbić et al., 2017), regression trees (Etemad-Shahidi and Mahjoobi, 2009) and random forest (Mafi and Amirinia, 2017). The storm surge has been also predicted by different algorithms such as support vector machines (Hashemi et al., 2016), regression trees and random forest (Granata and Di Nunno, 2021). Over the last decades, an increasing number of works have been comparing the SNN with other machine learning methods proving that this method is not the best for all problems. In Granata and Di Nunno (2021), they showed that regression trees outperforms SNN in most

of the case for the tide forecast. Same observation was made on the prediction of significant wave height (Etemad-Shahidi and Mahjoobi, 2009).

1.4.2 Deep learning

With the increasing amount of data and computational power, deep learning methods such as recurrent neural networks have been increasingly used over the last years in the predictive modeling of coastal processes. Recurrent neural networks are deep neural networks specialized in the sequences and time series prediction (Bengio et al., 2017). This type of networks have been used successfully to accurately predict time series of significant wave height (Mandal and Prabaharan, 2006; Sadeghifar et al., 2017; Savitha et al., 2017; Alqushaibi et al., 2021) or to predict storm surge level (Di Nunno et al., 2021; Quintana et al., 2021). More recently, they have been used to improve the predictions of a numerical wind wave model by integrating both the local data and the temporality in the errors of the numerical wind wave model (Zhang et al., 2021).

Convolutional neural networks (CNN) which are deep neural networks specialized in images analysis are also more and more employed to analyze the large quantity of images created by the video monitoring stations located on the shore. However, there are only few applications of CNN based on the images of the video monitoring stations such as the estimation of the nearshore bathymetry (Benshila et al., 2020) or wave-tracking (Kim et al., 2020). So far, they do not have a direct application for the analysis of coastal risk.

1.4.3 Statistical methods

In the study of coastal risks, statistical methods are used for their ability to bring understanding about the data and to study the relationships between variables. Among the statistical methods, Bayesian networks (BN), a class of probabilistic graphical models, are appreciated methods in coastal risk modeling. Their low computational cost, their ability to represent complex systems

by integrating different sources of data and their intuitive representation are non negligible advantages compared to other modeling approaches such as numerical process based models.

Most of the applications of BN in coastal engineering domain concern the translation of forcing variables (e.g. wave, weather and tide conditions) into impact and damages on the shore during storm events. For instance, they have been employed to predict coastal cliff erosion (Hapke and Plant, 2010), shoreline retreat (Beuzen et al., 2018), dune retreat and erosion (Palmsten et al., 2014; den Heijer et al., 2012) and barrier island response (Plant and Stockdon, 2012; Wilson et al., 2015) resulting from coastal storms. They are also used as surrogates of process-based models. In Poelhekke et al. (2016) and Plomaritis et al. (2018), BN are trained using output data from many pre-computed process-based model simulations. Once trained, the networks can be conditioned with forecast of the hydraulic boundary conditions to obtain instantaneously forecast of onshore hazards. By avoiding the computation time associated with process based models, bayesian networks are a great assets for operational EWS.

1.4.4 Advantages and limitations of SLM

The main advantage of SLM over process-based models is the computation time. The training time for such methods is usually in the order of minutes (or hours for deep learning) and the predictions made by these methods are instantaneous compared to numerical process-based models. In addition, SLM are relatively simple to use and with enough data they can achieve similar or even higher performances than numerical wind wave models for site-specific wave parameters (Browne et al., 2007; Savitha et al., 2017).

The main limitation of these methods is that their performances depend directly on the quantity and quality of the data. Deployment of these models are only possible in sites equipped with monitoring systems. In the future, data availability will not be as much as a problem. Indeed, monitoring networks

are expected to be more dense as new technological advances will improve sensors and their connectivity, and as the cost of the components will decrease (Marcelli et al., 2021). In this future context, the use of statistical methods adapted to big data will be necessary (Xu et al., 2019).

1.5 Problematic and objectives of the thesis

Considering that coastal flooding risk is expected to increase in the next decades and that SLM offers a promising potential in an era of big data, this thesis aims to answer the following question:

How SLM can contribute to the improvement of coastal risk assessment tools and to the development of an EWS which aims to reduce coastal flooding risk ?

This problematic is broken down into 3 sub-questions that will be investigated in this thesis:

- How to improve, with SLM, the local forecast of spectral wave models that are known to underestimate wave parameters during extreme events?
- Can we employ SLM to constitute storm impact databases that are rare and not routinely collected but are essential in the disaster risk reduction process?
- Can we fully develop an accurate predictive model linking offshore hydraulic boundary conditions into onshore hazards based on data collected from monitoring network and SLM?

1.6 Outline of the thesis

The presentation of the research carried out in this thesis is organized in 4 chapters each dedicated to a specific scientific issue linked to the application

of SLM to the study of coastal flooding and its prediction (Figure 1.7).

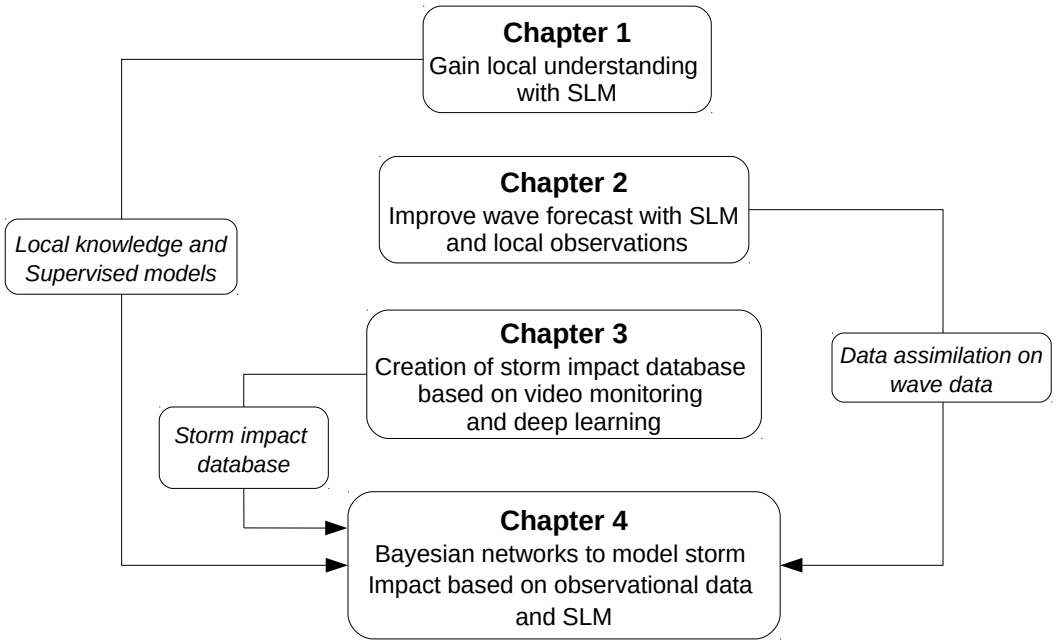


Figure 1.7: Connection between the chapters of this thesis.

The first chapter acts as an introduction on SLM and their applications in the study of coastal processes. The objective of this chapter is to demonstrate the interest of using these methods by highlighting the fact that they can provide knowledge about a coastal process at a local scale. Local knowledge is essential in the development of an EWS. This statement is especially true if the EWS is built with a statistical learning method such as bayesian networks, which is the case in this thesis. This chapter is organized in three parts, each presenting a different application of SLM in the study of coastal processes. Two of them concern supervised learning with the modeling of two coastal processes related to coastal flooding: storm surge level and wave runup on the beach. The last application of SLM concerns unsupervised learning and shows how to characterize local wave climate with clustering method.

In the second chapter of this thesis, we demonstrate the ability of SLM to improve local wave predictions made by numerical wind wave models. The

accuracy of the data used in the development of the EWS is crucial to have reliable warnings. However, it has been shown that numerical wind wave models have a tendency to underestimate certain wave parameters during stormy conditions. We present alternatives SLM (random forest and gradient boosting trees) to neural networks which are traditionally used in the literature to perform data assimilation. We demonstrate that these methods, that have never been used for this task, leads to better improvements than neural networks in addition to providing explicability with the predictive power of the variables. This chapter is constituted by an article published in *Applied Ocean Research*.

The third chapter demonstrates that SLM, especially deep learning methods, can be used to create automatically a storm impact database with images from video monitoring networks. Building a bayesian network that can be employed in EWS to predict coastal flooding requires a lot of data. It needs data about the coastal processes that are responsible for coastal flooding (tide, waves, storm surge, meteorological conditions) but it also needs data about the impact of coastal flooding. Unlike data about the coastal processes, data about the impact of coastal flooding are rare and sparse. Convolutional neural networks, which are deep learning methods, are used to classify the video monitoring images into three storm impact regimes which are categories of coastal flooding risk. Once trained, these networks can predict the storm impact regimes of newly created timestacks, generating an incremental storm impact database. This chapter is composed by an article published in *Remote Sensing*.

The fourth chapter shows that SLM can be employed to predict coastal flooding risk. This chapter regroups the models and data acquired through the previous chapters and assemble them in a Bayesian network that can be employed in a EWS to predict coastal flooding risk (Figure 1.7). In this chapter, we evaluate the predictive performance of a BN exclusively based on observational data from diverse monitoring networks. In addition, we propose a methodology based on SLM to extend the storm impact and atmospheric

surge data that are limited for the study site of Biarritz. A second BN is trained on the extended database and its performances are compared with the first BN to see the gain of performances associated with the extension of the database. This chapter is constituted by an article in preparation for a submission to *Natural Hazards*.

Finally, we conclude this thesis by answering the main research question, summarizing the scientific contributions associated with this thesis and discussing the potential perspectives.

2. Gain local understanding with SLM

2.1 Introduction

The objective of this chapter is to demonstrate the interest of SLM in the modeling of coastal processes by showing that they can provide knowledge about these processes or the study site. This chapter presents three examples where SLM were employed to model two different processes which are storm surge, wave runup on the beach and to characterize wave climate.

In the first part of the chapter, we show how to model storm surge level with several environmental variables related to the tide, the winds and the waves. Different methods are tested to find the best predictive model and variable importance analysis is performed on the best model to learn which variable is the most predictive for the study site. The insights brought by the variable importance analysis are discussed with the characteristics of the site and what is known in the literature.

The second part of this chapter is about the modeling of wave runup on the beach depending on the characteristics of the offshore waves. This part is quite similar to the first part of chapter one in terms of methodology, however, an additional statistical method is tested which is the generalized additive model. This method has the advantage to better illustrate the relationships between the variable of interest and the covariates (variables used to make the predictions).

The last part of this chapter aims to characterize the wave climate at Anglet's buoy location by studying the relationships between the local wave

characteristics observed at the buoy and the synoptic weather (sea level pressure) over the North Atlantic ocean. Contrary to the two other parts of this chapter, the characterization of the wave climate is an unsupervised task and other statistical learning methods must be employed such as dimensionality reduction (PCA) or clustering methods (Self Organizing Maps).

2.2 Sea surge modeling

Storm surges play a major role during coastal floodings, it is therefore essential for coastal researchers and stakeholders to have predictive models for this phenomenon. Numerical modeling is commonly employed for storm surge modeling. For a general review on this subject the reader is referred to the review of Flather (2000) and for examples of operational models to the works of Lionello et al. (2006) or Souza et al. (2013). Over the last few years, statistical learning models (especially shallow neural networks) have been employed to model the storm surge (Bezuglov et al., 2016; Hashemi et al., 2016; Lee et al., 2018). These models have the advantage to be less computationally expensive than process-based numerical models but they requires large observational dataset to be trained.

In this first part, we aim to train a statistical learning model to predict precisely sea surge level at Socoa tide gauge (South west of France), while learning about the local phenomenon by studying variable importance. Variable importance is a measure of the predictive power of a variable in a model. It can be used to sort variables from most to least predictive, allowing one to have more insights on the problem and to perform feature selection when there are too many variables. This model can be potentially used to predict operationally local storm surge and the insights brought by the variable importance analysis can be included in the development of an EWS based on Bayesian networks (BN).

2.2.1 Data

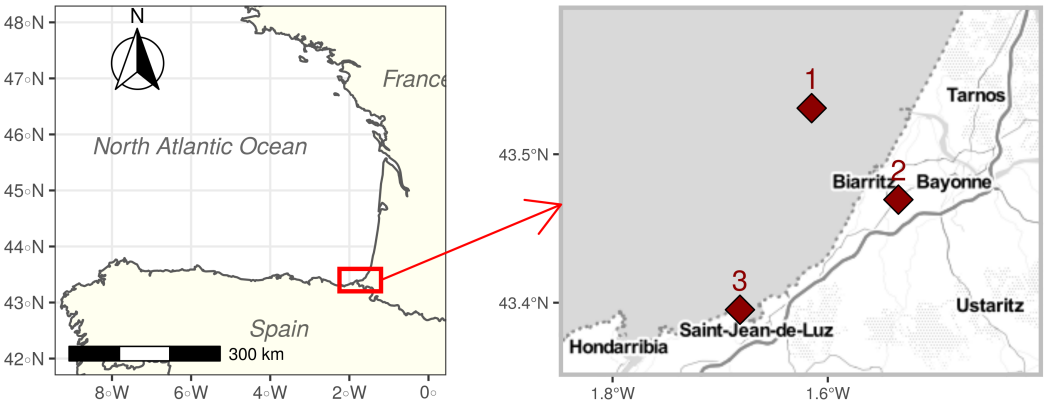


Figure 2.1: Map showing the locations of the directional wave buoy (1), the meteorological station of Biarritz (2) and the tide gauge of Socoa (3).

Storm surge data are obtained by subtracting the astronomical tide modeled by harmonic analysis from the total water level collected by the tide gauge of Socoa (Figure 2.1). This method is fully described in Appendix A. The data range from 2011 to nowadays with a hourly time step (Figure 2.2) and contain missing data for the period 2016-2019 due to dysfunctions of the Socoa tide gauge.

Meteorological data, including average wind speed above 10 meters, wind direction and atmospheric pressure are furnished by the French national meteorological service MeteoFrance. The data were collected hourly by the meteorological station of the Biarritz airport, located only a few kilometers from the study site (Figure 2.1). It covers the period ranging from 2013-01-01 to 2020-03-23.

Direct measurements of wave parameters are furnished by the National Center for Archiving Swell Measurements (L'her et al., 1999). They were made by a directional wave rider buoy (DWR MKIII) operated by the Center for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (CEREMA) and the University of Pau and Pays de l'Adour

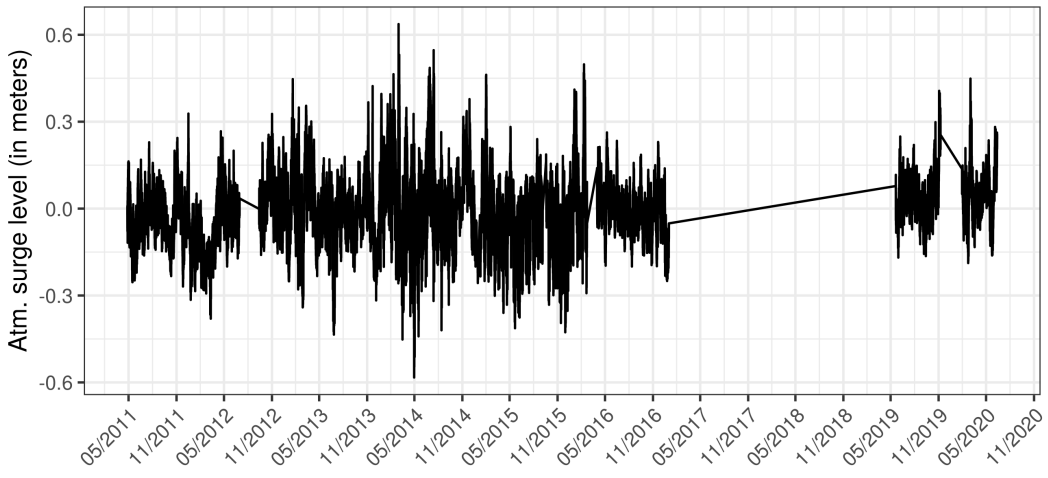


Figure 2.2: Atmospheric surge level extracted from the water level of Socoa tide gauge.

(UPPA). The buoy is located a few miles off the Basque Coast (Figure 2.1) at 50 meters water depth. Since its deployment in 2009, this buoy has been recording significant wave height H_s, peak period T_p, direction at peak θ_p and other wave parameters every 30 minutes.

By assembling the storm surge, the meteorological data and the wave measurements we obtain a dataset of 37401 hourly observations ranging from 2013-01-01 to 2017-01-01. Because temporal effects are not taken into account in the modeling of storm surge, the data are divided randomly into two sets : the training set containing 75% of the data and the testing set containing the remaining 25%. The explanatory variables are normalized (mean of 0 and standard deviation of 1) to facilitate the training of the algorithms.

2.2.2 Methodology

Supervised learning methods

Over the last few years, the use of Shallow neural networks (SNN) has been democratized for the modeling of water level related to the atmospheric surge,

few examples are presented in Table 2.1. The explanatory variables (Input) are mostly local meteorological conditions and storm characteristics (strength, location, etc...).

Table 2.1: Examples of atmospheric surge modeling with SNN in the literature.

Authors	**Year**	**Type of neural networks**	**Input**	**Output**
Lee	2006	SNN	For one location at time t: -Pressure -Wind velocity -Wind direction -Harmonic Analysis tidal level	Sea level for one location at time t
Tseng et al.	2007	SNN	-Typhon characteristics -Distance to station -Local wind direction -Local pressure -Local wind speed -Local astronomical tide Sometimes they used these variables at time t or t -1h .	Storm surge deviation (in meters) at time t+1h
Lee	2008	SNN	For one location at time t: -Pressure -Wind velocity -Wind direction -Harmonic Analysis tidal level	Sea level for one location at time t+1h or t+3h or t+6h.
Kim et al.	2018	SNN	-Longitude -Latitude -Central pressure -Moving speed of the storm -heading direction -Radius of exponential scale pressure	Normalized surge level
Lee et al.	2018	Generalized regression neural network	-Difference in pressure -Maximum wind speed	Storm surge level

Most authors take into account the temporal effect for the modeling of the atmospheric surge level and therefore use SNN as the "off-the-shelf" method. In this work, we do not take into account the temporal effects. It is why we are comparing several methods to find the most appropriate one for our problem. We tested usual supervised learning algorithms:

- Linear model
- Shallow neural networks

- Random forests
- Gradient boosting trees

The reader is referred to Hastie et al. (2009) for details on neural networks, random forest and gradient boosting trees. The algorithms all have the same explanatory variables to predict the storm surge level: the astronomical tide level (harmonic analysis), the local meteorological conditions (average wind speed and direction above 10 meters, atmospheric pressure) and the sea state characteristics (significant wave height, peak wave period and wave direction). The best statistical learning model is the one which obtains the best performances (lowest root mean squared errors) on the test set.

Bayesian optimization for hyperparameter tuning

For complex learning algorithms such as random forest, gradient boosting trees and neural networks, hyperparameter values must be optimized. Hyperparameters are parameters whose values are specified by the user before the training process begins, they affect the structure of a learning algorithm and how well it trains. They have a non negligible impact on the final results.

In this work, Bayesian optimization is employed to select the optimal hyperparameter values. Bayesian optimization is an iterative algorithm that aims to minimize an objective function, in our case the root mean squared errors (RMSE). First, it builds a probability model (Gaussian process) of the objective function. Then it uses this surrogate model to select the most promising values of hyperparameters to evaluate. Once the promising combination of values have been evaluated, the probability model is updated and searched again for the most promising combination. This process is repeated several times. This method is employed because it is very efficient for tuning hyperparameter values and it usually requires less iterations than other methods such as grid or random search (Bergstra et al., 2011). In depth details of this method are given in the works of Snoek et al. (2012); Marchant and Ramos (2012) and Shahriari et al. (2015).

To tune the hyperparameter values of our models, bayesian optimization is coupled with a 5-fold cross validation on the the training set. The objective function to minimize for the Bayesian optimization method is the average out-of-sample RMSE. The Bayesian optimization for our data is performed using the R package **Tidymodels**. This package allows for the training and tuning of several models simultaneously. To find optimal hyperparameters values for the different models (random forest, gradient boosting trees and neural networks), random combinations of hyperparameter values are first evaluated to serve as search base for the bayesian method (5 in this study), then an acquisition function (upper confidence bound) is used to find the next combination values to evaluate (this step is repeated 15 times).

Permutation Feature Importance

There is a lot of methods to compute variable importance depending on the type of model used. A complete review about these methods is presented in the work of (Wei et al., 2015). Permutation Feature Importance is one of the methods used to measure variable importance and is the one used in this work. This method was first described for random forest by Breiman (2001) and was later adapted to other models by Fisher et al. (2019).

This method consists in shuffling the values of a variable (process called permutation) and measuring the increase in the model's prediction errors associated with these permutations. A variable is considered as important if the permutations cause an increase in the model errors. This increase of model errors means that the model relied strongly on this given variable. On the contrary, a variable is considered as an unimportant variable if the permutations do not increase the model errors.

Permutation feature importance is implemented in numerous R package. This method is usually implemented by default in the packages specialized for random forest and gradient boosting trees. The model agnostic version of Fisher et al. (2019) used in this work comes from the R package **vip**.

2.2.3 Results

The table 2.2 shows the RMSE and MAE computed on the test data for the different models. From this table we can see that the performances of the different learning algorithms are quite close. Random forest shows slightly lower RMSE and MAE. The results obtained for our site are consistent with the literature. For example, Lee (2006), who performed a similar analysis with neural networks on several Taiwanese stations, found RMSE values between 5.39 and 9 cm.

Table 2.2: MAE and RMSE (in cm) computed on the test data for the different models.

	Linear model	**Random Forest**	**Gradient boosting trees**	**Neural Networks**
RMSE	6.61	5.73	6.54	6.56
MAE	4.86	4.09	4.78	4.81

The importance of the variables for random forest is displayed in figure 2.3. The variable with highest predictive power is the atmospheric pressure. This is expected as the atmospheric pressure influences significantly the surge level (Harris, 1963). The second, third and fourth most important variables are related to wave characteristics. In coastal areas with a reduced continental shelf (such as our site), it has been proven that the wind effect is limited (hence its low predictive power) and the contribution of both the atmospheric pressure and the waves is higher (Kennedy et al., 2012; Chaumillon et al., 2017).

2.2.4 Discussion

The main limitation of the SLM is the quantity of data. Training models with more data results in models that are able to generalize better from a higher amount of information and therefore results in better performances on unseen data. This is especially true as we are modeling sea surge which can become

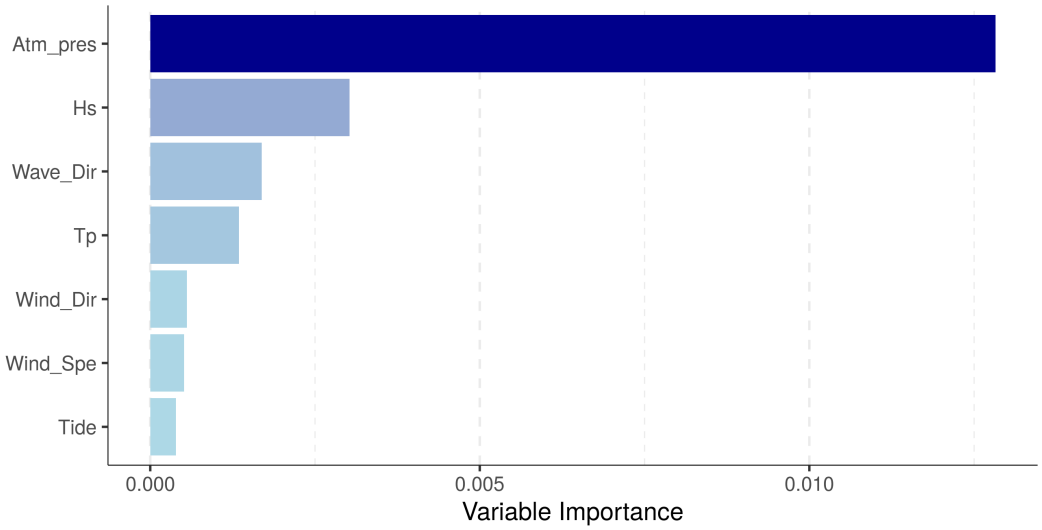


Figure 2.3: Permutation importance of variables used in Random Forest to predict storm surge level

extreme events (storm surges). Due to the extreme nature of certain events, the SLM need even more data to model correctly both the average sea surges and the storm surges.

With its promising performances, this model could be used in operational settings to forecast sea surge by using wave and meteorological forecast as explanatory variables. However, this is not recommended as this model has been trained with observational data and not on forecast data that might contain potential bias. The best solution for an operational forecast model should be to train the model with wave and meteorological forecast data in order to take into account the potential bias of the forecast models.

In the future, this work could be extended by adding explanatory variables containing temporal aspect. This could be the values of atmospheric pressure at previous time steps. In this framework, artificial neural networks could perform better as they are known to handle efficiently time series. Other input variables could be also used to improve the modeling such as the meteorological or wave data at different locations near the site.

2.2.5 Conclusion

We trained statistical learning models to predict the sea surge for the site of Socoa using several explanatory variables including meteorological conditions, sea state characteristics and tide level. By comparing the performances of these models on the test set, we defined random forest as the best model for this site with RMSE value of 5.73 cm.

The importance of the variables for random forest were computed with the permutation method. The variable importance analysis gave us insights about the local sea surge and the order of importance corresponded to the characteristics of the site and to what was found in the literature in similar sites. The knowledge acquired during this analysis can be helpful in the development of an EWS based on BN.

2.3 Runup elevation on the beach

Coastal floodings originate from the interaction of several processes which all influence the total water level. Among these processes, there are the tide, the meteorological conditions and also the waves. The contribution of the waves to the total water level is called wave runup and corresponds to the maximum onshore elevation reached by a wave, relative to the wave-averaged shoreline position (Flanders Marine Institute, 2021). Wave characteristics are available through observations (buoy) or numerical modeling. However, the wave characteristics only gives a broad idea of the runup on the beach. To transform offshore wave characteristics into runup elevation on the beach, two modeling approaches can be employed: numerical modeling or data-driven approaches.

There are various numerical models that solve hydrodynamic equations to estimate runup elevation on the beach. More details on these different models can be found in the work of Fiedler et al. (2018). Runup on the beach can also be modeled using empirical formula or statistical models estimated with local data. Many empirical formulas have been developed over the year, a complete review is given in the work of da Silva et al. (2020). A non negligible advantage for the data-driven methods is in the computational aspect, where the computational effort and time are lower than those of numerical modeling methods.

Measurements of the wave runup were performed on the study site of Biarritz (South west of France). By comparing the estimations obtained by the formula of Stockdon et al. (2006) (the most used empirical formula) with the measurements, large estimation errors were found, highlighting the fact that this empirical formula was not adapted to this site. An empirical formula based on a linear model was developed by the SIAME researchers. This second formula yielded runup estimation much closer to the observations (Figure 2.4).

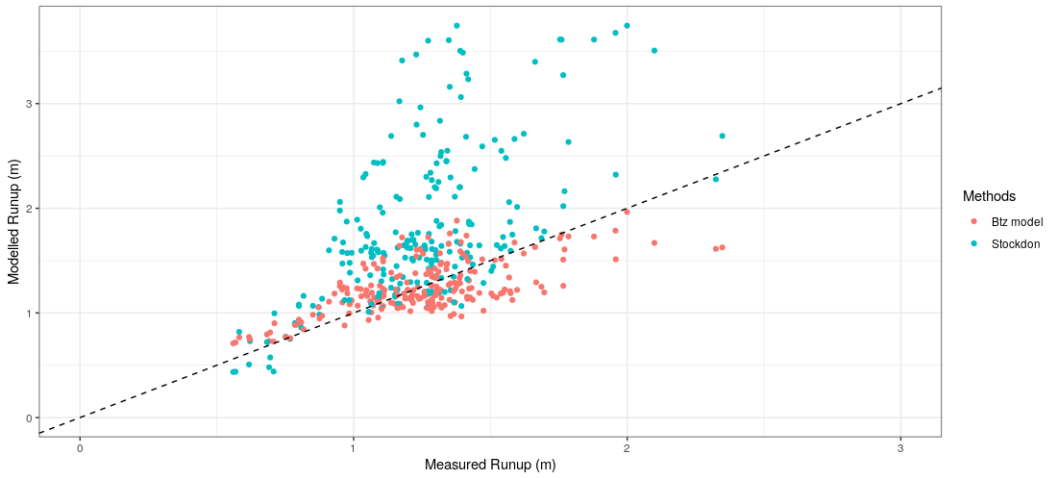


Figure 2.4: Measured runup on the beach versus the runup predicted by Stockdon formula and the local formula computed for the GPB.

In this second part of the first chapter, we aim to extend the work made the Grande Plage of Biarritz. We want to train a statistical learning model that yields better runup predictions than the empirical formula developed on the Grande Plage de Biarritz, while learning about the local phenomenon.

2.3.1 Study site

The Grande Plage de Biarritz (GPB) is an urbanized beach located in the South Ouest of France. This beach has a high socio-economic value due to its location near the city center, its historical heritage and its tourist appeal (Morichon et al., 2018). Stakes on this beach are high: infrastructures are located behind a sea promenade in the upper part of the beach. This embayed beach is 1.2 kilometers long and is delimited by two rocky outcrops. It is a intermediate-reflective with a steep slope (8 to 9%) in the upper part of the beach and a slight slope (1,5%) in the lower part of the beach. Finally, it is a mesotidal beach with a spring tidal range of 4.5 meters (Morichon et al., 2018).

2.3.2 Data

Runup measurements

In the framework of the MAREA project, runup was measured on the Grande Plage de Biarritz by the SIAME (Applied sciences in mechanics and electrical engineering) laboratory. The measurement of wave runup were obtained using the video monitoring station installed on the Grande Plage de Biarritz. This video monitoring station consists in 4 cameras, controlled by the open source software SIRENA (Nieto et al., 2010). The pixel intensities were sampled along two transects (Figure 2.5) at 1Hz during 14 min to produce timestack images.

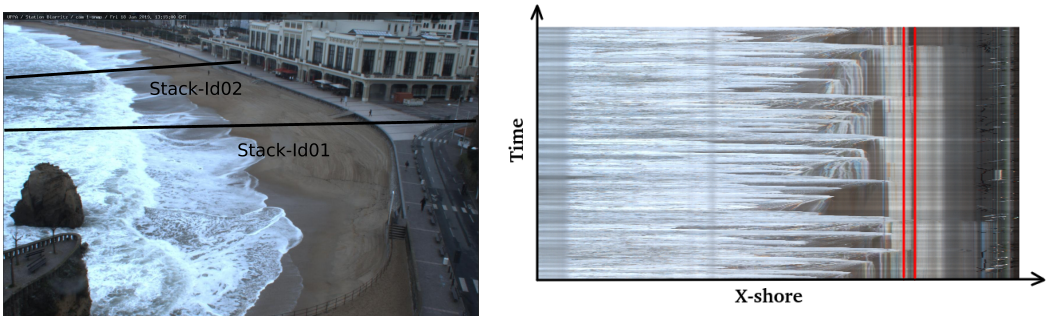


Figure 2.5: Transects monitored on the Grande Plage de Biarritz and example of associated timestack. The red line in the timestack represents the position of the seawall.

Standard transformations from image to world coordinates (image orthorectification) were employed. Then, Otsu's thresholding method was applied to the collected timestack images to obtain time series of the cross-shore position of the waterline. This semi-automatic extraction method has already been employed to extract waterline in timestack images (Vousdoukas et al., 2012) and aims to separate pixels into two classes based on their intensities. The obtained cross-shore coordinates were transformed into elevations using the topographic information collected with a NRTK GNSS at low tide during the

14 days of field surveys. These 14 days of fields survey are spread between the 4th December 2017 and the 24th January 2019. During these 14 days, runup R and the slope of the beach (β) were measured every 15 minutes.

Wave parameters

The wave parameters such as significant wave height H_s, peak period T_p, wave directional spreading ($disp$) and direction at peak θ_p are recorded by a directional wave rider buoy (DWR MKIII) operated by the Center for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (CEREMA) and the University of Pau and Pays de l'Adour (UPPA). The buoy is located a few miles off GBP at 50 meters water depth. Since its deployment in 2009, this buoy has been recording wave parameters every 30 minutes. By joining the runup measurements and the waves characteristics we obtain a dataset of 220 observations. Due to the small number of observations and the temporal discontinuity of the data, we do not consider the temporal aspect in the modeling step presented in the next section.

2.3.3 Methodology

Statistical learning methods

We compare several statistical learning methods with different characteristics: linear models, generalized additive models (GAM) and random forest. Linear models are the simplest method and will serve to determine baseline performance. Random forest are ensemble learning methods which rely on weak learners (classification and regression trees) to make a prediction. For more details concerning random forest, the reader is referred to the works of Breiman (2001) and Friedman (2001).

Generalized additive models are generalized linear models where the response variable (Y_i) depends linearly on a sum of smooth functions of covariates (Hastie and Tibshirani, 1990). The general notation is given below:

$$g(\mu_i) = A_i\theta + \sum_{1}^{j} f_j(x_{ji})$$

where $g(.)$ is a link function, $\mu_i = E(Y_i)$ and Y_i $EF(\mu_i, \phi)$ with $EF(\mu_i, \phi)$ denoting an exponential family with mean μ_i and scale parameter ϕ. A_i is a row of the model matrix for any parametric model component (intercept) and θ the corresponding parameter vector. Finally, f_j are smooth functions of the covariates x_j. By using smooth functions to represent the dependence between covariates and response variable, this model has a great flexibility. However, choices must be made especially about the type of smooth functions used and their associated smoothness. An extended review on the smooth functions can be found in Wood (2017). The visualization of the smooth functions associated with each covariates is really informative and can bring insight on the response variable. In the same spirit of the previous section, the permutation feature importance can be computed with the **vip** package.

Fitting the models

The different models tested are presented in Table 2.3. The linear models are fitted with the `lm` function in base R, GAM models are fitted with the **mgcv** package and the random forest with the **ranger** package.

Table 2.3: Characteristics of the different statistical learning models

Method used	Response variable	Covariates	Hyperparameters
Linear models	Runup	H_s, T_p, $disp$, θ and β	None
Linear models	Runup	Log transformed H_s, T_p, $disp$, θ and β	None
GAM	Runup	Splines for H_s, T_p, $disp$, θ but not for β	Default smooth functions (thin plates) Additional penalty term in the smoothness selection procedure with select = TRUE in mgcv package.
Random forest	Runup	H_s, T_p, $disp$, θ and β	Default values (number of trees = 500)

To assess in a objective manner the performances of the formulas, GAMs and random forest we perform a repeated (10 repetitions) of a 10 fold cross-

validation. The RMSE reported in the results section are the average of the out-of-sample RMSE.

2.3.4 Results

Comparison of the models

The averages of the out-of-sample RMSE are presented in the Figure 2.6. The model with the highest RMSE is the simplest model: the linear model with no transformation of the covariates. It is followed by the linear model with the log transformed variable, then the GAM. The lowest RMSE is obtained by the random forest. This observed order was expected as more complex algorithms are more able to learn complex patterns and therefore yield better results.

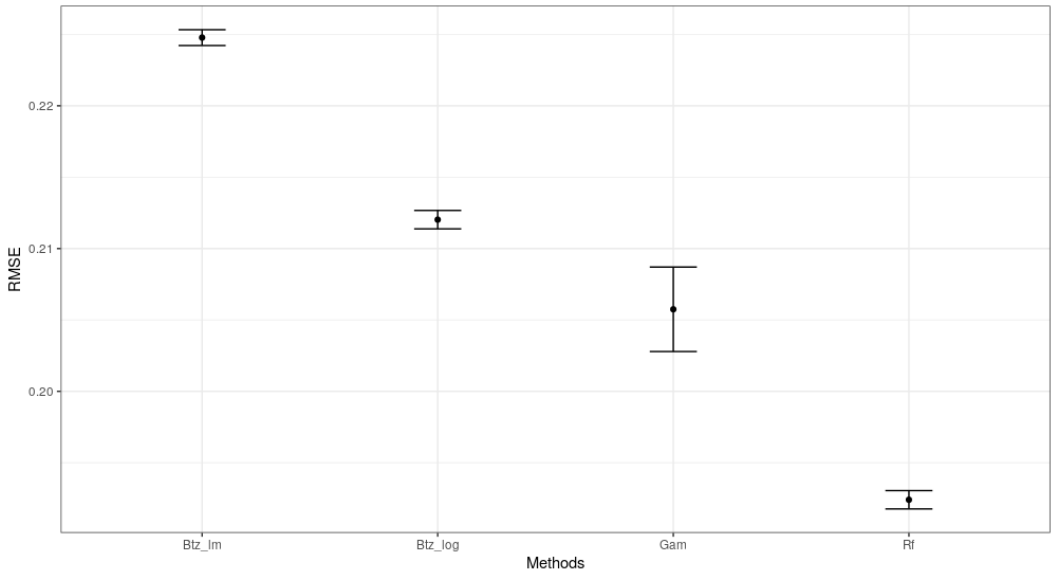


Figure 2.6: RMSE and associated standard error obtained for the 10 repeated 10-fold cross validation for different statistical learning methods.

The permutation feature importance is computed for the best algorithm (random forest) fitted on all data and is shown in Figure 2.7. It is clear that the significant wave height has the most predictive power for this model. It is

followed by the wave period, the slope of the beach and the wave directional spreading. The least predictive variable is the wave direction.

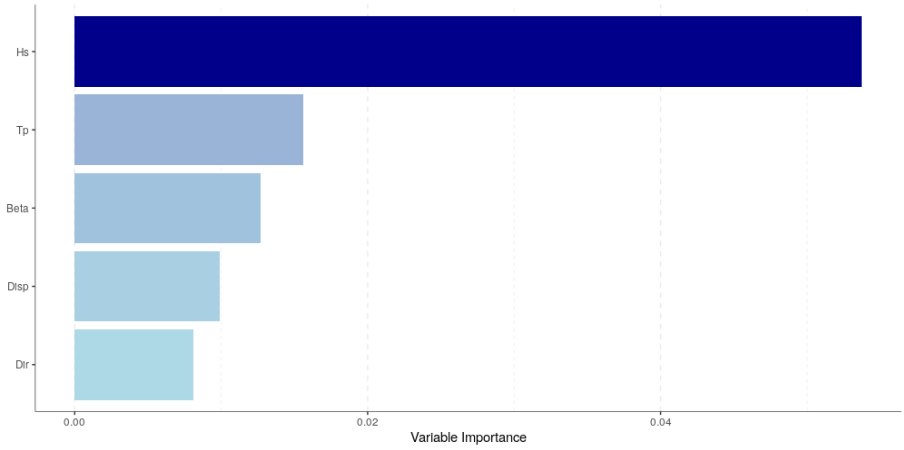


Figure 2.7: Permutation importance computed for the random forest model.

GAM

Even though GAM is not the best model, we decide to present a detailed analysis of this method because the visualization of the smooth functions associated with each covariates can bring valuable insights on the response variable. The detailed analysis is based on the GAM model fitted on all the data.

Basic model checking plots are presented in Figure 2.8 and show nothing problematic. The QQ-Plot is very close to a straight line which suggest that the distributional assumption is reasonable. The histogram of the residuals seems approximately close to normality. The residuals vs the linear predictor plot suggest a very slight increase of the variance as the mean increase.

Figure 2.9 shows the smoothing functions for the GAM model. For the significant wave height H_s the smoothing function is not monotonic. It increases as the value of H_s increases, however above 3.8 meters, the function starts to decrease. This observation is not logical as higher wave should lead to larger runup elevation on the beach. This could be due to the lack of observations during these conditions (high waves).

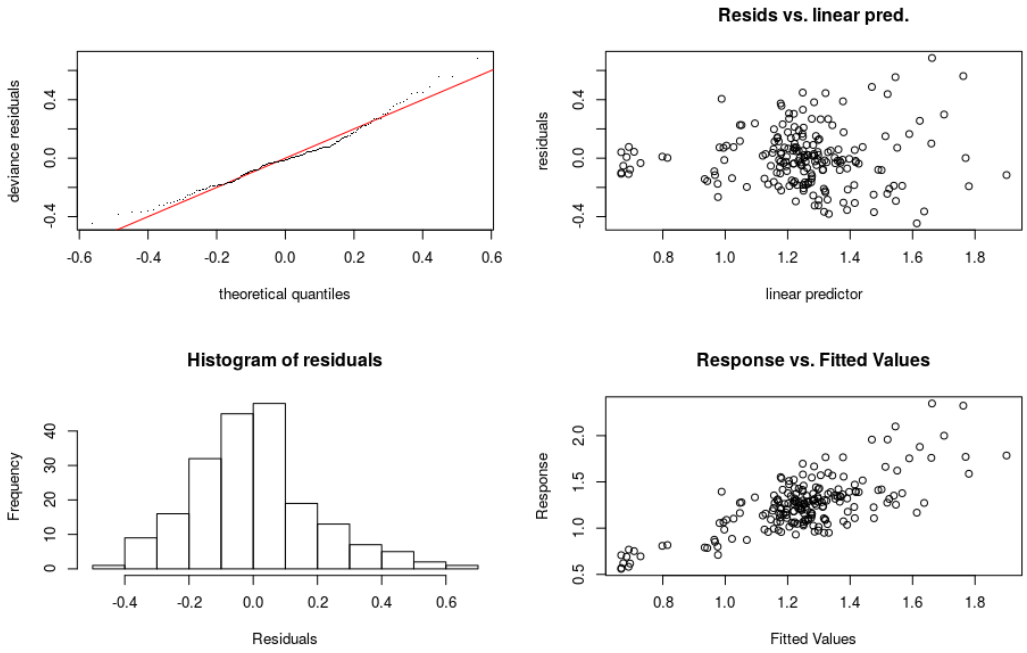


Figure 2.8: Basic model checking plots for GAM fitted on the data. The upper left panel represent a QQ-plot of the residuals while the upper right panel represent the residuals vs the linear predictor. The lower left panel is an histogram of the residuals and the lower right is a plot of the fitted values against the residuals.

The smoothing function of T_p is monotonic and positive. This is expected as waves with larger period lead to increased runup on the beach. On the contrary, the smoothing function of the wave directional dispersion is monotonic negative meaning that waves with lower directional dispersion lead to larger value of Runup. Finally, the slope of the beach seems to have a slight positive influence on the Runup elevation and the wave direction does not seem to be a good predictor as it was excluded by the regularization.

The variable importance computed by the feature permutation method is shown in Figure 2.10. Such as the random forest, the most predictive variable is H_s. The second most predictive variable is wave directional dispersion, it is

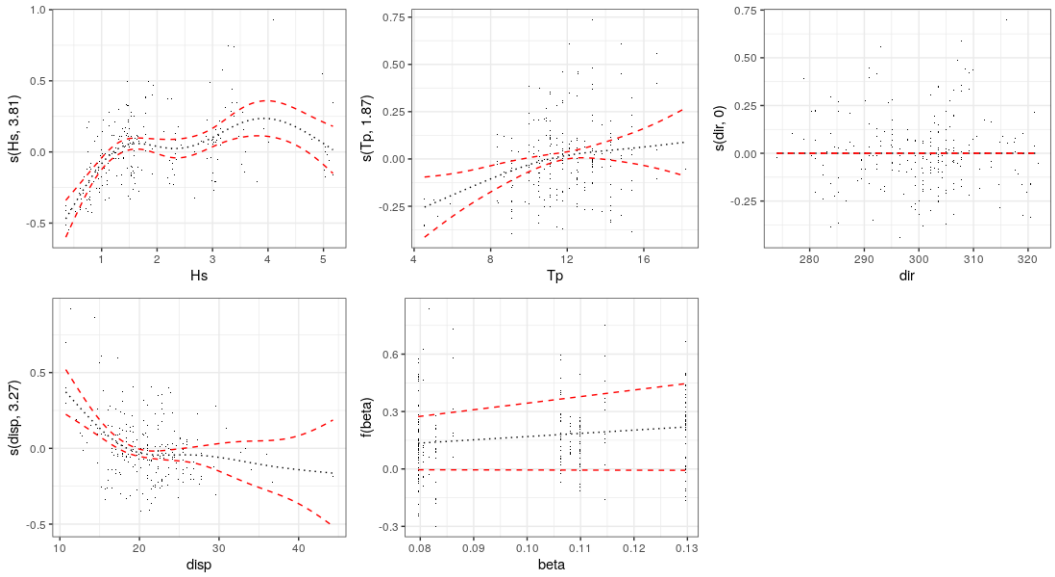


Figure 2.9: Spline functions estimated by the GAM model.

followed by the wave period at peak and finally the slope at the beach.

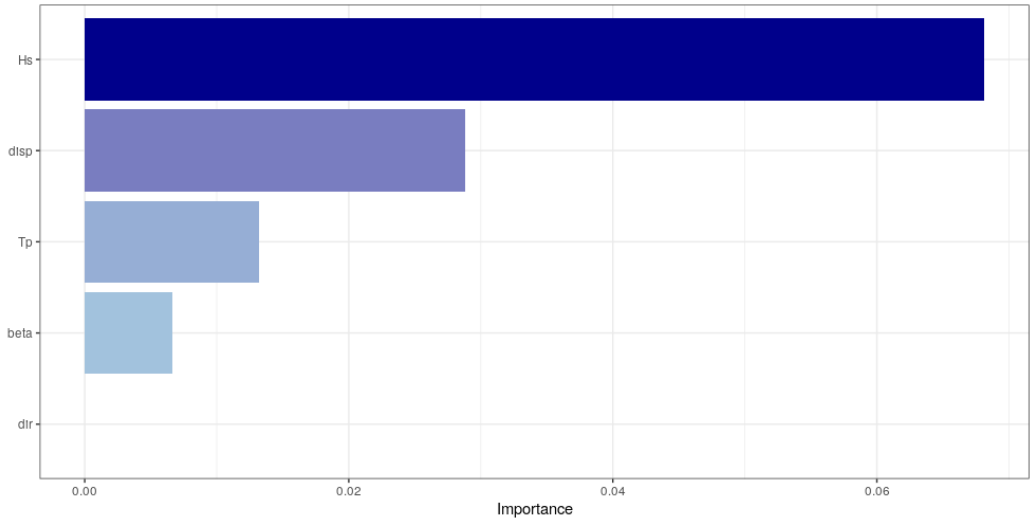


Figure 2.10: Permutation importance computed for the GAM model.

Predictions on buoy data

To get an idea of the predictions of our models (linear model with log transformation, GAM and random forest), we decided to predict the runup on the beach for each observation at the wave buoy. Because the slope of the beach is not available for each observation it has been removed from the models. The predictions of the models (fitted on the whole dataset of the measurement campaign) plotted against H_s are presented in Figure 2.11.

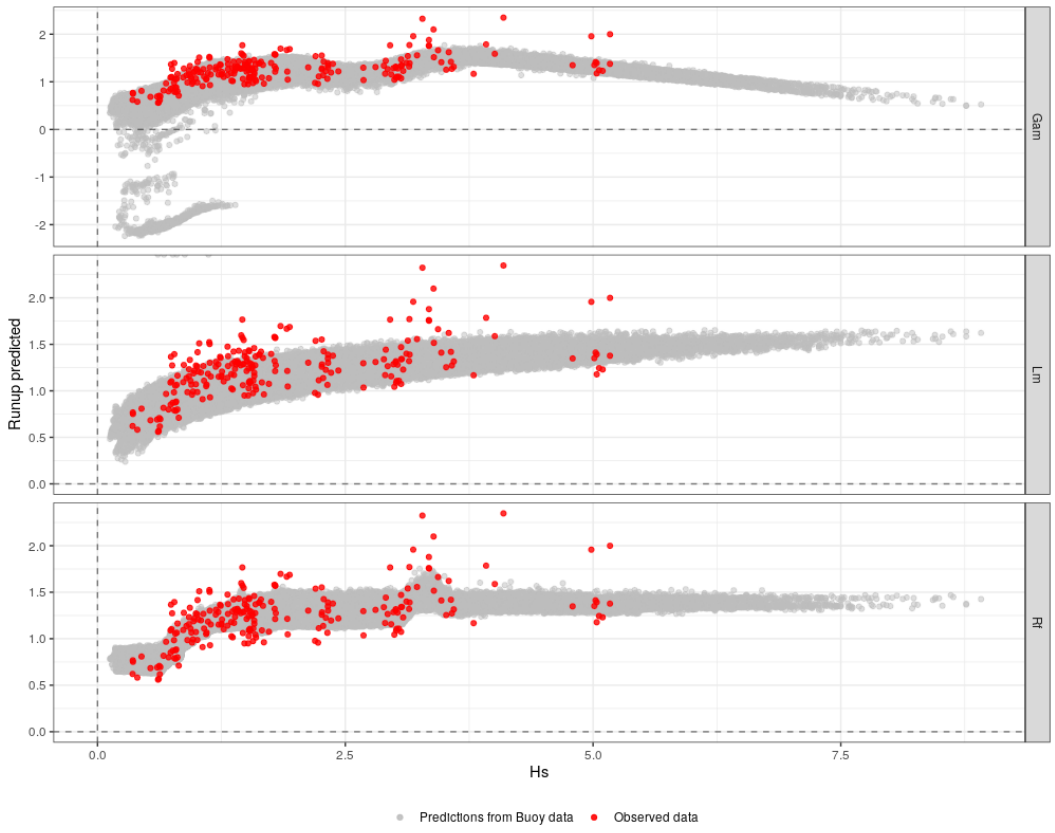


Figure 2.11: Plot of the runup predictions made with the buoy data against H_s. The red dots are the data collected during the measurement campaign and served as training data.

From this figure we can see that GAM does not seem to generalize well

for unseen data. Indeed, some predictions of the runup are negative and the runup elevation seems to decrease as the significant wave height increases. The predictions seems more reasonable for the linear model or the random forest.

2.3.5 Discussion

In this section we will discuss the advantages and downsides of statistical learning methods in the modeling of the runup.

Statistical models are a lot less computationally demanding than process based model such as Xbeach for run modeling. In addition, the statistical models are also easier to train and calibrate than the process based models and they provide tailored prediction for one site. They can bring valuable insights by analyzing the predictive power of each covariates used in the model. By using the GAM, the influence of each covariate can be visualized and analyzed.

Unlike process based methods, the statistical methods do not rely on physical processes. Consequently, they need a large amount of data to take into account such processes and this constitutes their main limitations. In predictive usage, a SLM trained on a short dataset will encounter numerous events outside the distribution the training data and the associated predictions will be erroneous. This is the case for the GAM model where it predicts negative runup during unseen conditions.

One solution could be to collect more data, however, in the case of runup modeling on the GPB it does not appear as the best solution. Indeed, to improve the models presented in this work, data collected during energetic events (large H_s and T_p) are needed. Such events could be dangerous for the operators measuring data. The cost of measurements campaigns can also be a hindrance.

A potential solution for the lack of data could be to switch to Bayesian framework in order to integrate prior knowledge of the physical process and to estimate uncertainties. When there are unseen conditions, the model will rely more on the prior knowledge than in data leading to more parsimonious

predictions.

2.3.6 Conclusion

We compared the performances of linear models, random forest and generalized additive model to find the best predictive model for the runup elevation on the Grande Plage of Biarritz. This comparison was made with a repeated 10-fold cross validation to assess objectively the performance of these methods on unseen data despite the low number of observations.

The results and variable importance were analyzed for two models: random forest and GAM. Random forest yielded the best predictions with a RMSE of 19.3 cm. The variable importance analysis showed that the variables with the most predictive power were the significant wave height and wave period at peak. For the GAM model, the RMSE was slightly larger (20.5 cm). Even though it was not the best predictive model, GAM gave valuable insights with the visualization of the effect of each covariate on the prediction (Smooth functions). In addition,the importance variable analysis for this model was slightly different as the two most predictive variables were the significant wave height and the wave directional spreading.

For the wave runup prediction, SLM are a great alternatives to empirical formulas or numerical modeling. By learning directly from the observations they provide "tailored" predictions for one site. In addition, they can provide valuable insights by studying the predictive power of the covariates or by looking at the relationships (smooth functions) between the covariates and the response variable in case of generalized additive models.

2.4 Wave Climate Characterization with statistical downscaling

Waves play a major role in many coastal processes ranging from sediment transport process to coastal flooding process. The influence of the waves on sediment transport and erosion processes is discussed in the work of Munk and Traylor (1947) or Masselink et al. (2014). Waves also have an impact on the coastal flooding process. This has been demonstrated in the first part of this chapter where wave characteristics showed a great predictive power for a storm surge model and also in the work of Bertin et al. (2015) where it has been proven that the waves increased significantly the storm surge level during Xynthia and Joachim storm events.

Deep knowledge about sea state characteristics and its seasonality is therefore essential not only for coastal management but also for safety at sea and on the coast. The wave climate can be characterized by analyzing wave data which can originate from two sources: observations (wave buoy, satellites) or modeled data (numerical models).

In the framework of the MICROPOLIT project, a preliminary work has been done to characterize the wave climate on the study site of Anglet (Callens, 2017)[1]. The wave data at the location of the buoy of Anglet come from a simulation covering the period from 1949 to 2014 with the spectral wave model WWII (Roland et al., 2012). Temporal decomposition was performed to extract the trend and seasonality of H_s and extreme value analysis was performed to investigate the potential value of extreme H_s and their return period. For H_s a slight increase was observed over the years and a strong seasonality was highlighted with larger waves during the winter.

In this last part of chapter 1, we aim to characterize more precisely the wave climate at the Anglet's buoy location by studying the relationships between

[1]Interactive web application presenting the work: `https://aureliencallens.shinyapps.io/application/`

the wave characteristics observed at the buoy and the synoptic weather over the North Atlantic ocean. To reach this objective, we employ the statistical downscaling method which tries to find statistical relationship between a local predictand (wave characteristics observed at the buoy) and a regional scale predictor (SLP field) which has been categorized into different weather types to facilitate the analysis.

2.4.1 Data

Sea level pressure

Sea level pressure data were provided by the research data archive of Colorado [2]. They come from two reanalysis:

- CFSR (Climate Forecast System Reanalysis) from 1979 to 2010 with 6 hour time-step with 0.5 ° resolution (Saha et al., 2010).
- CFSv2 (Climate Forecast System Version 2) from 2010 to 2020 with 6 hour time-step with 1 ° resolution (Saha et al., 2011).

The spatial points of the first reanalysis have been sampled to correspond to the spatial points of the second reanalysis which has a coarser resolution. The spatial domain of the data was restricted to North Atlantic: from 20 °N to 60°N and from 70 °E to 20°W with a spatial resolution of 1°. In land points (red points in Figure 2.12) are not considered for this work as they show a greater variability in atmospheric pressure (Camus et al., 2014).

Wave characteristics

The wave parameters such as significant wave height H_s, peak period T_p, direction at peak θ_p are recorded by a directional wave rider buoy (DWR MKIII) operated by the Center for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (CEREMA) and the University of

[2] website `https://rda.ucar.edu`

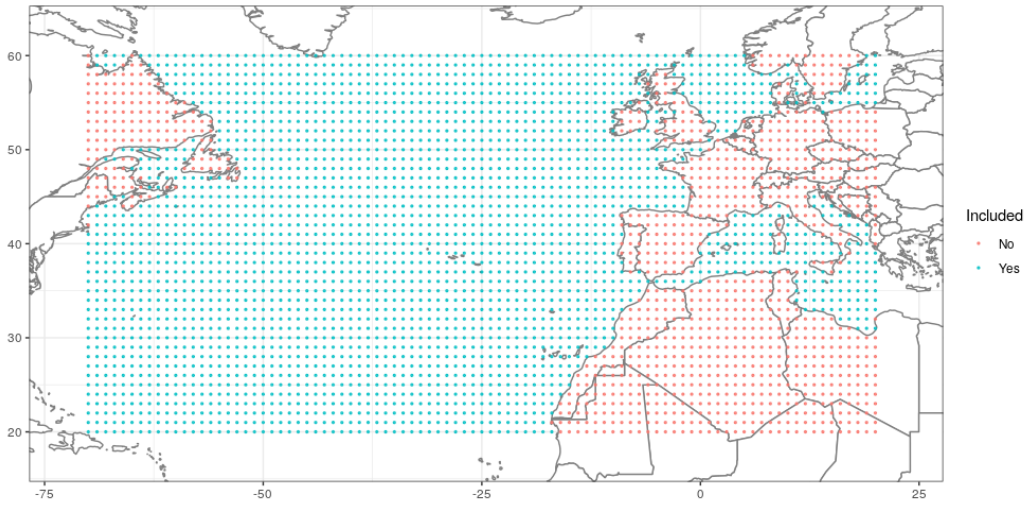


Figure 2.12: Spatial distribution of seal level pressure data with resolution of 1 °

Pau and Pays de l'Adour (UPPA). The buoy is located a few miles off the Basque Coast at 50 meters water depth. Since its deployment in 2009, this buoy has been recording wave parameters every 30 minutes.

2.4.2 Statistical downscaling

The statistical downscaling method aims to find statistical relationship between local predictands (wave characteristics) and regional-scale predictors such as sea level pressure and sea level pressure gradient (Camus et al., 2014). This method relies on a weather pattern approach: it uses several data mining techniques such as dimensionality reduction and clustering to simplify the atmospheric pressure data into weather types, then it finds statistical relationship between wave characteristics and each weather type. This method allows for the study of seasonal and inter-annual variability of the sea state (Camus et al., 2014). The simplified methodology is as follows:

1. Processing of the SLP data

- Computation of the squared SLP gradient
- PCA to reduce dimensionality of the SLP and SLP gradient data
- Clustering into different weather types (WT)

2. Analysis of statistical relationships between weather type and sea state
3. Forecast of the sea state by using the statistical relationships found previously

Computation of the SLP gradient

The first step of this method is to prepare the regional scale predictors: sea level pressure (SLP) and the squared gradient of SLP (SLPG). The gradient of SLP is important for this analysis as it is a better representation of the winds responsible for the wave generating process (Wang and Swail, 2006). SLP and SLPG are preferred over wind variables because global models are known to better represent sea level pressure than winds variables (Caires et al., 2006).

The squared gradient is computed for each point from the values of the four nearest grid points using the weights proportional to the inverse of the distance (Espejo et al., 2014). It is the sum of the squared zonal and squared meridional SLP gradients, which is proportional to wind energy (Wang and Swail, 2006).

Temporal averaging

The temporal range and the region of the predictor are really important when predicting local predictand. The generation time of the wave in the Atlantic Ocean must be taken into account. For north Atlantic Ocean, it is less than 5 days (Hegermiller et al., 2017). The ESTELA method (Pérez et al., 2014) can be used to determine the number of days to take into account. This method allows one to know the region of wave generation and the temporal range needed for the predictor. In this work, we choose to average the SLP and

SPLG data over 3 days (current day and 2 previous days) such as the work of Camus et al. (2014); Espejo et al. (2014).

Dimensionality reduction

The SLP and SPLG data for each point averaged over 3 days for the period 1970-2020 are stored in a 15065×5026 matrix. The dimensionality of this dataset must be reduced for the next step which is the clustering with Self organizing maps (SOM). A PCA is applied on this data to reduce the dimensionality while conserving a large proportion of the variability of the data. Only the n first components of the data representing 99% of the variability of the data is kept.

Clustering into weather regimes with SOM

The clustering with Self organizing maps (SOM) is done on the data projected on the n first components. SOM are a type of neural networks that rely on competitive learning. This algorithm aims to create a map of the data based on an iterative process involving neurons connected together by a grid. For one observation, the closest neuron is designated as the best match unit (BMU). The coordinates of the BMU are then shifted toward the observation point at a designated rate (learning rate). The neighboring neurons of the BMU are also slightly shifted toward the observation. This process is repeated several times for all the observations of the training set. After the training process, the grid of neurons are approximating the data distribution and new observations can be assigned to a neuron (best matching unit) which can act as a clustering.

To cluster the projected data, other methods can be employed such as k-means algorithm (KMA) or maximum dissimilarity algorithm (MDA). A complete comparison of the three algorithm are presented in the work of (Camus et al., 2011). In this work, SOM was chosen over the two other algorithms as it is the best technique to graphically characterize the multidimensional wave climate (Camus et al., 2011). By projecting the classification on

a two bi-dimensional lattice with proximity information (similar groups are closer), this method allows the visualization of patterns in multidimensional data which simplify the analysis of such data.

The clustering with SOM algorithm is performed with the R package **kohonen**.

2.4.3 Results

Weather types with SOM clustering

PCA was performed to reduce the dimensionality of the data before the clustering. Among all the components, the first 26 were kept for the clustering step. These 26 components explains 99% of the total variance of the original data. The data projected on the first 26 components of the PCA are mapped on a 10×10 rectangular grid by the SOM algorithm. A sensitivity analysis based on the total within-cluster sum of square (wss) was performed to select the number of cluster. The number of 100 clusters (10×10) was chosen as it showed great performance in term of wss and as this number was large enough to capture a wide range of atmospheric situations.

The mean SLP fields of the 100 weather types obtained by the SOM algorithm are displayed in Figure 2.13. From this figure, we can see that similar weather types are closer on the bidimensional lattice. On the upper-left corner, the majority of weather types display high-pressure systems in the Atlantic Ocean. On the contrary, on the lower-left corner the weather types mostly display intense low-pressure systems. These weather types can be identified as the two mains circulation pattern in the North Atlantic Ocean: North Atlantic oscillation (NAO) and the east Atlantic oscillation (EA). The former circulation pattern (NAO) is characterized by the presence of both a low-pressure system located over Iceland and a high-pressure pattern located around the Azores area (Mellado-Cano et al., 2019). This circulation pattern is represented in the upper part of the lower-left corner in the lattice. The latter circulation pattern (EA) is characterized by the presence of a low-pressure system in the

south of Iceland and west of Ireland (Mellado-Cano et al., 2019). The weather types presenting this pattern are located in the lower part of the lower-left corner of the lattice. On the right side, the weather types in the middle display high-pressure system while weathers type on the upper and lower corner show low-pressure systems (not as intense as the left side).

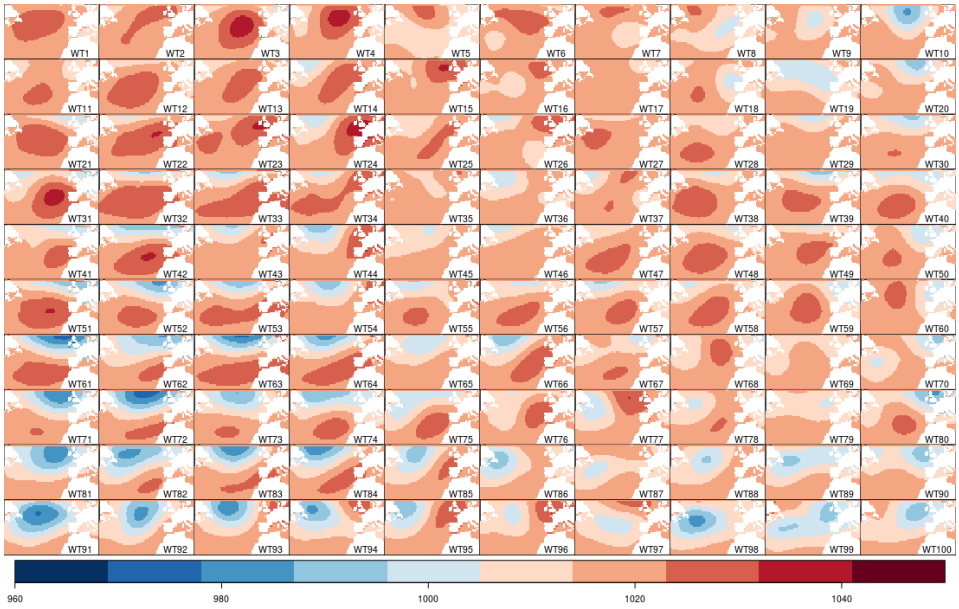


Figure 2.13: Weather types obtained by the self organizing map method. Blue color represents zones of low pressure and red color zone of high pressure.

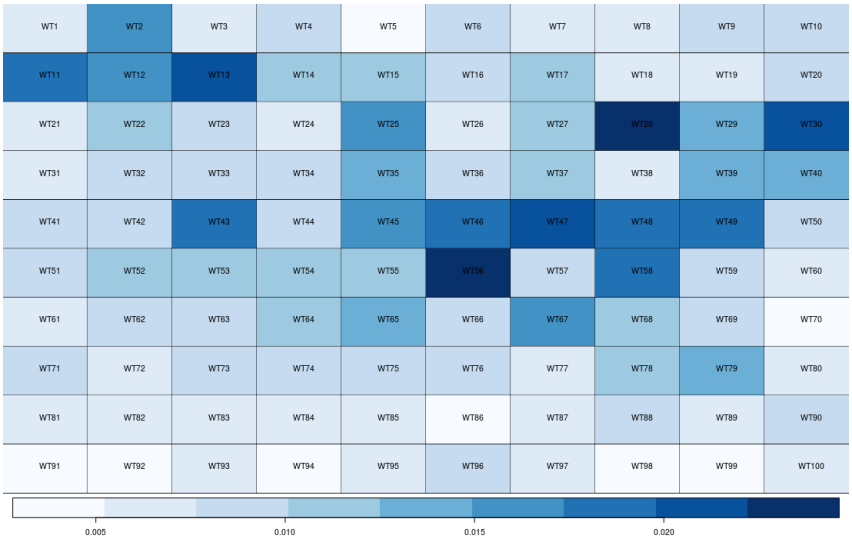
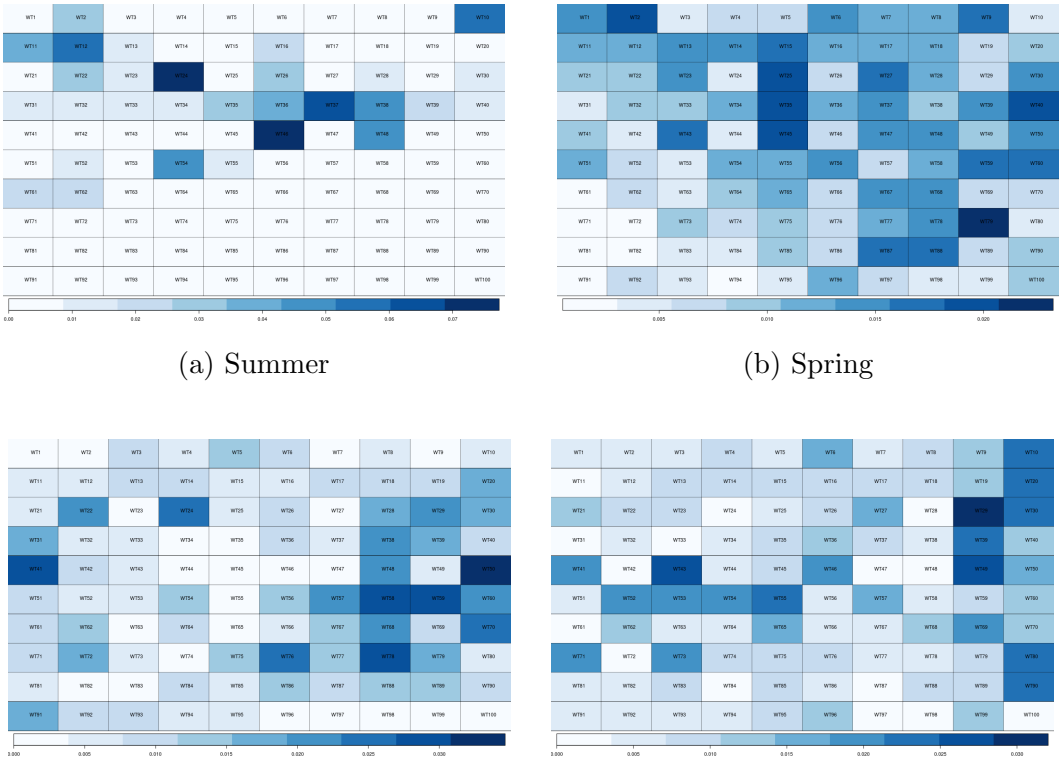


Figure 2.14: Observed frequencies of the different weather types for the period 1969-2019

The observed frequencies of each weather type (WT) computed for all the period is presented in Figure 2.14. The most represented weather types for the whole period are the weather types 28 and 56. Both of these weather types present a high-pressure system on the Azores islands. This high-pressure system is commonly known as Azores high and is semi permanent. It influences significantly the weather and climatic patterns of the north Atlantic Ocean.

The seasonal variability can also be analyzed using the weather types. The observed frequencies computed by season are presented in Figure 2.15. Summer season shows the less variability with the majority of WT located in the upper part of the lattice corresponding to WT with high pressure systems or with homogeneous pressure field with average pressure value of $1013hPa$. On the contrary, the other seasons show large variability. In winter and autumn seasons, the WT with intense low pressure system are much more frequent comparing to spring season.



(a) Summer (b) Spring

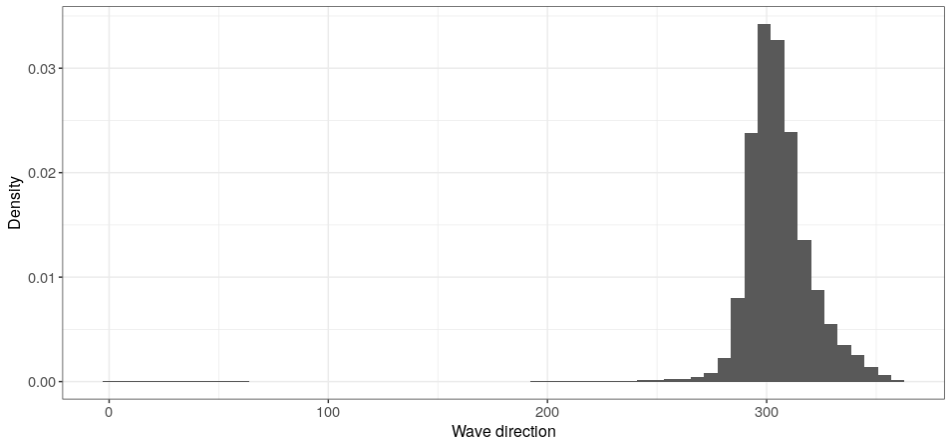
(c) Winter (d) Autumn

Figure 2.15: Observed frequencies of the different weather types depending on the seasons.

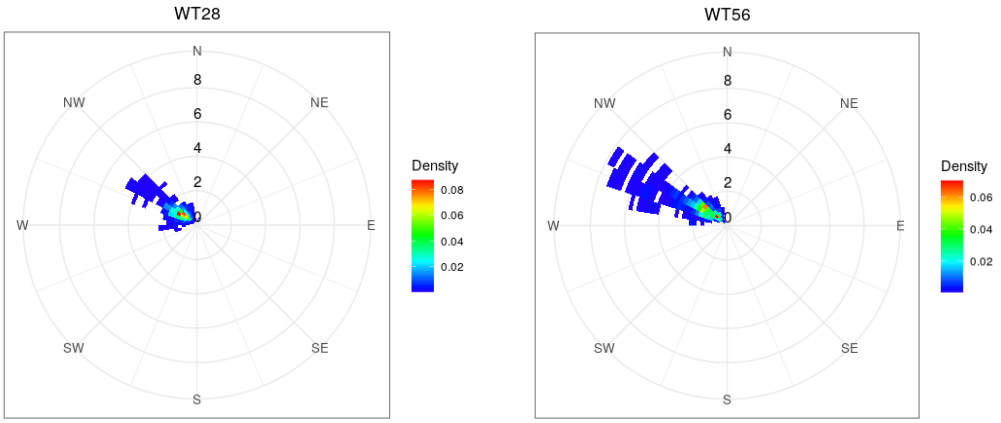
Relating weather types to sea state

The wave climate can be characterized by looking at the distribution of the wave parameters (H_s, T_p, θ) depending on the WT. Figure 2.16 shows the bivariate graphs of H_s and T_p for each WT. The distribution of H_s and T_p change significantly between WT. The highest values for both parameters are commonly found in WT displaying low-pressure system whereas the lowest values are found in system presenting strong high-pressure system or homogeneous pressure field. Unlike the significant wave height, the wave direction (θ) do not vary between weather types. For all the observations, this variable do not vary a lot: 75% of the observations have a direction contained in the

following interval: $[296.6, 312.6]$. The Figure 2.17 shows the histogram of θ and bivariate graphs of H_s depending on θ for the most represented weather types (WT28 and 56).



(a)



(b)

Figure 2.17: (a) Histogram of the wave direction for all the observations and (b) Bivariate graphs of significant wave height (Hs) and wave direction (θ) for the most common weather types.

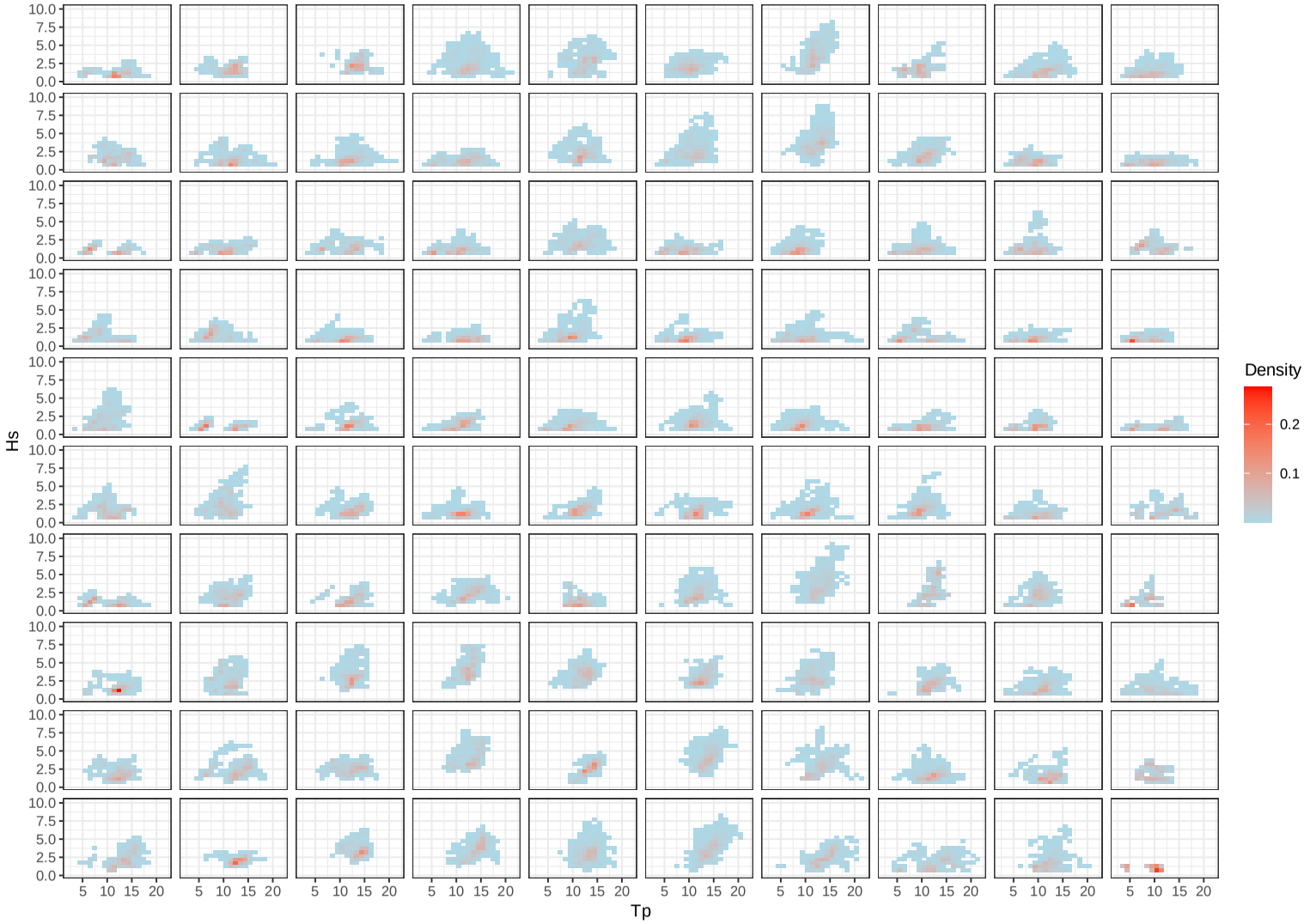


Figure 2.16: Bivariate graphs of significant wave height (Hs) and peak wave period (Tp) depending on the different weather types

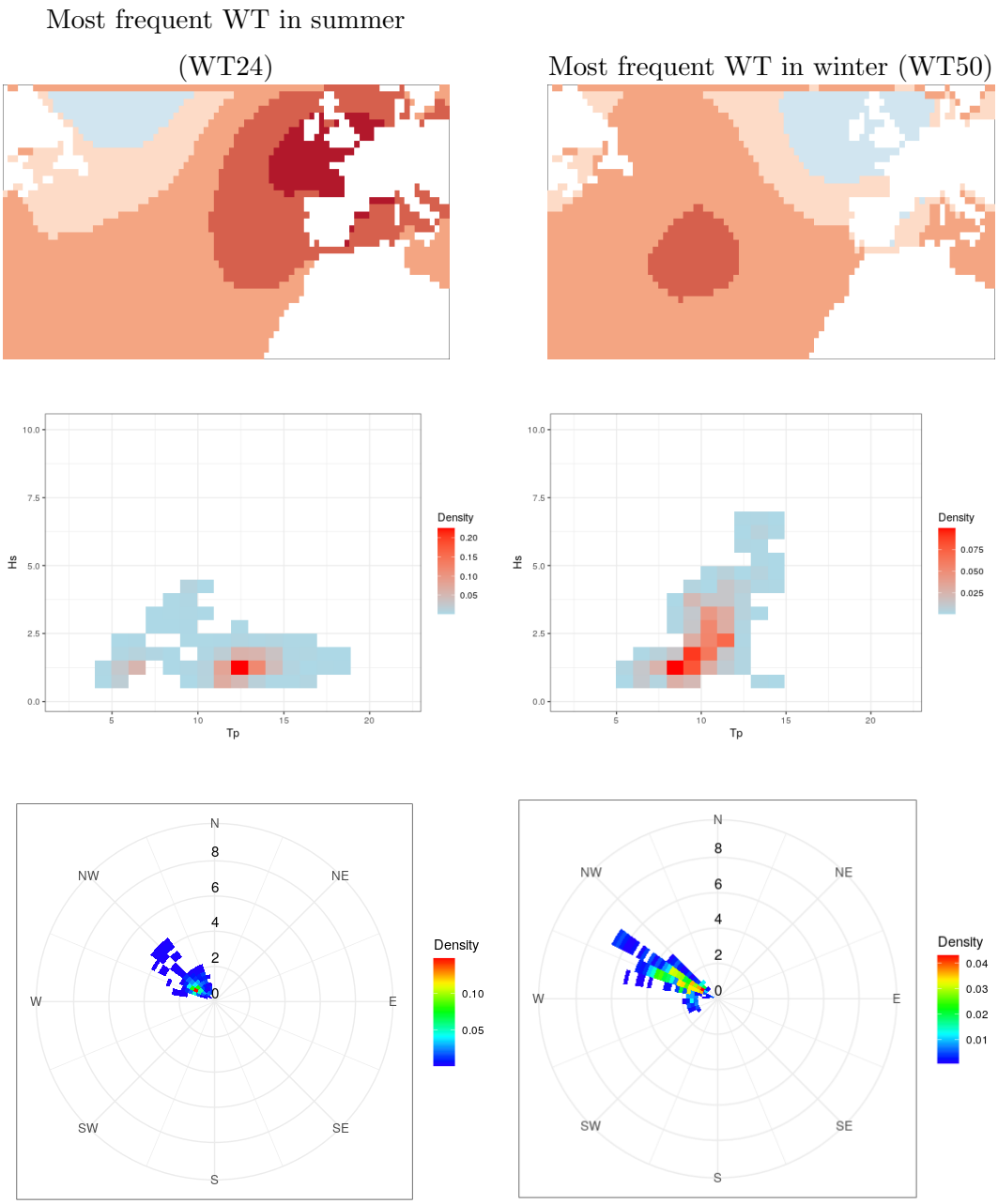


Figure 2.18: Comparison of the most frequent weather types for summer and winter

Figure 2.18 compares the most frequent weather types for summer and winter in terms of wave characteristics. The WT 24 display a high-pressure

system located on Europe while the WT 50 display a low pressure system located on England. There are significant differences in wave characteristics between these two WT. Indeed, the WT 50 (most frequent in winter season) display higher values and variability for significant wave heights. On the contrary, the WT 24 display higher values and variability for peak wave period. There is also a small difference in wave direction between these two WT: the mean wave direction is northwest for WT 24 and west-northwest for WT 50.

2.4.4 Discussion

The work presented in this short article is mainly exploratory and was aimed to characterize the wave climate and its seasonality at the Anglet's buoy location. However, the characterization of wave climate at one location is only one of the possibilities offered by the statistical downscaling method. Many extensions of this work can be considered as this method can be employed to:

- reconstitute the wave spectrum daily (Espejo et al., 2014) or monthly (Camus et al., 2014)
- perform wave climate projections by using SLP fields from different climate change scenarios (Camus et al., 2014; Perez et al., 2015)
- perform extreme value analysis for wave parameter depending on each weather type (Camus et al., 2016; Rueda et al., 2016)

When we use this method, we make 3 important hypothesis (Camus et al., 2014):

- Variability of the local variable should be explained by the statistical connection
- Changes in the mean climate should lie within the range of its natural variability

- The relationships should be stationary.

A main limitation of statistical downscaling method lies in the fact that these hypothesis might not hold true in a context of global change (Camus et al., 2014). Indeed, extreme events that did not occur in the past may occur in the future and such events would lie outside the range of the past variability. Moreover, global change may influence the statistical relationships observed in the past. It is worth noting that data availability is also a limitation for the statistical downscaling method. This method need a lot of data to see as much different events as possible. A site with low data availability can display large inhomogeneities when using this method (Rueda et al., 2019).

2.4.5 Conclusion

The statistical downscaling method based on weather types allowed us to better understand the relationships between the synoptic weather in the North Atlantic Ocean and the wave characteristics observed at Anglet's buoy.

We showed that among the 100 weather types, some of them are representative of the mains circulation patterns in the North Atlantic Ocean such as the North Atlantic Oscillation (NAO) or the East Atlantic Oscillation (EA). By studying the seasonal probabilities of occurrence of each WT, we highlighted large differences between seasons in terms of WT and variability. We also demonstrated that the distributions of H_s and T_p differ significantly between weather types. We observed that WT with low-pressure systems, more frequent in winter, are generally characterized by waves with larger H_s and T_p in comparison with WT with high-pressure systems or homogeneous SLP fields which are more frequent in summer.

For the disaster risk reduction (DRR) process, statistical downscaling method is an interesting tool for coastal researchers or stakeholders because it gives a better understanding of the wave characteristics and their variability at one location. As stated in the discussion section, the statistical downscaling method in this work is used only to characterize the wave climate. However, it can

be extended to perform wave climate projections or extreme value analysis which can bring insights respectively on the evolution of the future risk and on climate variability of extreme events.

2.5 Conclusion

This chapter showed the interest of using SLM to model processes involved in coastal flooding. Indeed, in the first two applications of this chapter, we showed that they can predict new observations accurately when the new data is in the range of the training data (interpolation) and that they can provide knowledge about these processes or the study site with variable importance analysis. In the last application of this chapter, we demonstrated that knowledge can also be acquired with unsupervised learning methods (PCA, SOM). These methods can be employed to detect patterns from big data and to gather similar observations into groups which is particularly useful in wave climate characterization.

Among the three examples of statistical modeling presented in this chapter, only the storm surge modeling can be considered for operational use in an EWS. Indeed, the statistical models trained to predict wave runup were not reliable for energetic conditions (storm events) due to a lack of data. Statistical downscaling could be used to predict wave characteristics depending on the synoptic weather type. However the wave characteristics can be more easily and accurately simulated with a spectral wave model.

3. Improve wave forecast at a specific location with ensemble methods and local observations

3.1 Introduction

In this chapter, we show that SLM, especially ensemble methods (random forest and gradient boosting trees) can be employed to improve significantly the spectral wave model predictions at a specific location. In the literature, numerous methods combining statistical models and process-based models have already been proposed to improve the numerical wave model predictions (Babovic et al., 2001; Makarynskyy et al., 2005; Moeini et al., 2012; Deshmukh et al., 2016; Londhe et al., 2016). However, they all have in common to use shallow neural networks as off-the-shelves method to perform data assimilation. Unlike previous works on this subject, we compare the performances of neural networks with ensemble methods that have never been used for this task. In addition, a special attention is given to hyperparameter tuning. Hyperparameters are parameters associated to machine learning algorithms that must be chosen before the training. These parameters influence greatly the final performance as they control the structure or the learning of the algorithm. In the literature, not a lot of importance is given to hyperparameter tuning and usually only a few combinations of hyperparameters are tested with grid search method. In this work, bayesian optimization, an automatic method, is employed to find the optimal hyperparameters. Finally, the potential benefits of data assimilation are investigated in a real case scenario: the computation

of wave run-up on the Grande Plage of Biarritz.

Scientific output:

This work resulted in the publication of an article (presented below) and two communications.

Publication:

- Callens, A., Morichon, D., Abadie, S., Delpey, M., & Liquet, B. (2020). Using Random forest and Gradient boosting trees to improve wave forecast at a specific location. Applied Ocean Research, 104, 102339.

Communications:

- "Using Random forest and Gradient boosting trees to improve wave forecast at a specific location", JDS 2021: 52èmes Journées de Statistique de la Société Française de Statistique (SFdS), Online conference (June 2021).

- "Improving numerical wave models with machine learning algorithms", Statistical Modelling in Ecology and Environmental Data Workshop, Anglet (September 2019).

3.2 Article: Using Random forest and Gradient boosting trees to improve wave forecast at a specific location

Using Random forest and Gradient boosting trees to improve wave forecast at a specific location

Aurélien Callens[a,*], Denis Morichon[b], Stéphane Abadie[b], Matthias Delpey[c], Benoit Liquet[a,d]

[a]*Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Pau, France*
[b]*Université de Pau et des Pays de l'Adour, E2S UPPA, SIAME, Anglet, France*
[c]*Centre Rivages Pro Tech SUEZ EAU FRANCE, Bidart, France*
[d]*Department of Mathematics and Statistics, Macquarie University, Sydney, Australia*

Abstract

The main objective is to present alternative algorithms to neural networks when improving sea state forecast by numerical models considering main spectral bulk parameters at a specific location, namely significant wave height, peak wave period and peak wave direction. The two alternatives are random forest and gradient boosting trees. To our knowledge, they have never been used for error prediction method. Therefore, their performances are compared with the performances of the usual choice in the literature: neural networks. We showed that the RMSE of the variables updated with gradient boosting trees and random forest are respectively 20 and 10% lower than the RMSE obtained with neural networks. A secondary objective is to show how to tune the hyperparameter values of machine learning algorithms with Bayesian Optimization. This step is essential when using machine learning algorithms and can improve the results significantly. Indeed, after a fine hyperparameter tuning with Bayesian optimization, gradient boosting trees yielded RMSE values in average 8% to 11% lower for the correction of significant wave height and peak wave period. Lastly, the potential benefits of such corrections in real life application are investigated by computing the extreme wave run-up ($R_{2\%}$) at the study site (Biarritz, France) using the data corrected by the different algorithms. Here again, the corrections made by random forest and gradient boosting trees provide better results than the corrections made by neural networks.

Keywords: Artificial neural networks, Data assimilation, Error prediction, Gradient boosting trees, Random forest, Wave forecasting.

1. Introduction

Nowadays, numerical wave models are routinely used to forecast wind generated waves. Although they provide satisfactory predictions at a regional scale and during mean wave conditions, it has been shown that they are less accurate for forecasting at a specific location (Londhe et al., 2016) and have a tendency to underestimate wave height during energetic wave conditions. This underestimation has been observed in different state-of-the-art wave models: see the work of Arnoux et al. (2018) for Wind Wave Model II (WWMII) and WAVEWATCH–III (WW3) ; Rakha et al. (2007) for WAve Model (WAM) and Moeini et al. (2012) for the Simulating WAves Nearshore model (SWAN). The errors in wave predictions are mainly due to inaccuracies in the wind input that forces the model. The winds used as forcing are numerically simulated and are known to underestimate high wind speeds (Moeini et al., 2012). This results in the underestimation of wave parameters by numerical wave models. Simplifying assumptions, approximations employed in the modeling process, discretization of the domain and a potentially wrong parametrization of the model can also be sources of inaccuracies in wave model predictions (Babovic et al., 2001, 2005).

When observation data are available, data assimilation can be used to improve the predictions made by numerical models. There are 4 main categories of data assimilation procedures (Refsgaard, 1997; Babovic et al., 2001): updating the input parameters, updating the state variables, updating the model parameters and finally updating the output parameters. The last procedure is called "Error prediction" method and is the most suitable approach to improve model predictions of different output variables at a specific location (Babovic et al., 2005). This procedure presents several advantages comparing to the other data assimilation procedures. First, it covers inaccuracies coming from all sources because it improves directly output variables. In addition, it can use a combination of external variables such as meteorological or wind data to increase the accuracy of the predictions. Lastly, it is easy to implement because it consists in only three steps and does not require multiple runs of numerical wave model. First, the deviations between the modeled values and measured values are computed. Then, machine learning algorithms are used to forecast these deviations. Finally the deviations predicted by the algorithms are incorporated to the predictions of the numerical model for the next time steps, resulting in a more accurate wave forecast.

This method has been successfully applied on hindcast data (Makarynskyy

*Corresponding author

et al., 2005; Deshmukh et al., 2016) and has even been implemented in real time setting in the works of Babovic et al. (2001) and Londhe et al. (2016). To our knowledge, only artificial neural networks have been tested to forecast the errors in the data assimilation. However, according to the so-called "No Free Lunch" theorem, there is no single model that works best for all problems (Wolpert, 2002). It is therefore necessary to try multiple models and find the one that works best for our particular problem. The performance of artificial neural networks must be compared with other algorithms in the data assimilation task. Random forest and gradient boosting trees are strong candidates for this comparison. Indeed, these two methods are known for their performance and unlike neural networks, they also provide valuable information by computing the predictive power of each variable used as input. The predictive power or variable importance refers to how much a model relies on that variable to make accurate predictions. A variable with high predictive power means that its values have a significant impact on the prediction values. By contrast, a variable with low predictive power have a limited impact on the prediction values and it can be substracted from the model to make it simpler and faster.

To explore the performance of random forest and gradient boosting trees, we use as a test case the Basque coast (South west of France). Every winter, the basque coast faces numerous coastal flooding events. To prevent and mitigate the risk of flooding, wave forecast are used to compute the extreme run-up values either by using parametric models such as the formula of Stockdon et al. (2006) or process based models such as Xbeach (Vousdoukas et al., 2012; de Santiago et al., 2017). In both cases, the accuracy of this forecast is of utmost importance as the issuing of the early warning depends on it, especially during energetic wave climate where coastal flooding risk is the highest. In this study, we employ the error prediction method with the different machine learning algorithms and use local meteorological conditions and measured wave parameters from a local buoy to improve the wave forecast. Lastly, we investigate the potential benefits of using such corrections in the computation of extreme run-up values.

This study aims to present two alternatives (random forest and gradient boosting trees) to neural networks by comparing their performances when improving regional numerical models. A secondary objective is to show how to tune the hyperparameter values of machine learning algorithms with Bayesian Optimization. In machine learning, a hyperparameter is a parameter whose value is specified by the user before the learning process begins, it will affect how well a model trains and therefore it will have a non negligible impact on the final results. Bayesian optimization is an efficient hyperparameter optimization algo-

rithm and it is widely used to optimize the results of any given machine learning method.

Lastly, we investigate if the error prediction method makes a difference in a real application such as the computation of extreme run-up for the beach of Biarritz. Section 2 will introduce the study area, the data and all the statistical methods used. Results will be presented and discussed in Section 3. Finally, Section 4 will cover the conclusion.

2. Data and Methods

2.1. Study site and Data

The Basque coast is a 150 km long rocky coast facing the Bay of Biscay (Figure 1). Every winter, it is battered by numerous storm events. This results in frequent and sometimes intense coastal flooding which can severely damage seafront infrastructures. The city of Biarritz is particularly affected as the buildings and infrastructures are located right behind a sea wall that is located at the top of the beach. The damages associated with coastal flooding are costly for nearshore cities which try to prevent and mitigate the risks by developing early warning systems. Such systems rely on the knowledge of the sea state and its forecast.

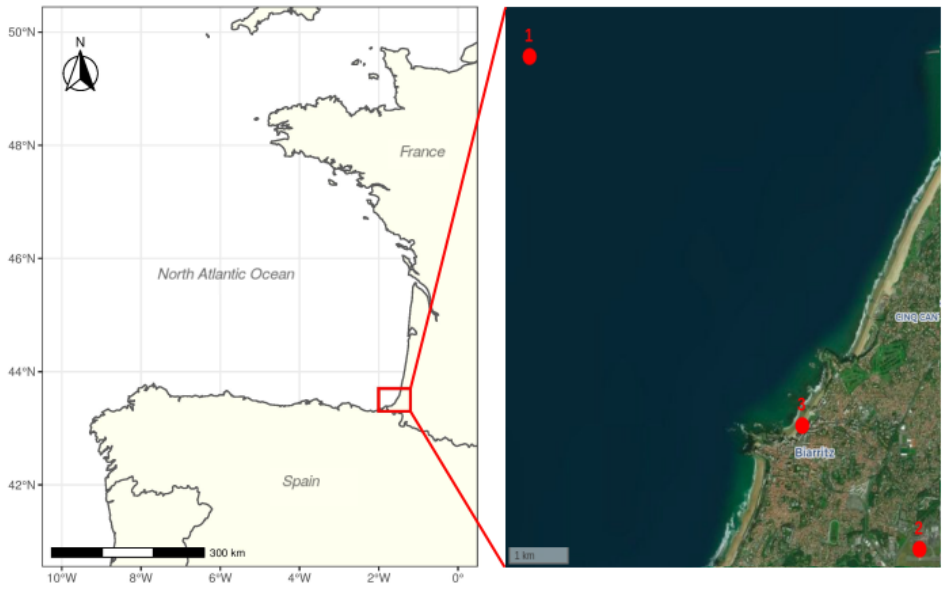


Figure 1: Map showing the location of the study site. The red dots show of the locations of the directional wave buoy (1), the meteorological station (2) and the beach called ”Grande Plage de Biarritz” (3).

This work focuses on the forecast improvement of three wave integrated parameters which describe the sea state in this area: the significant wave height (H_S), the peak period (T_p) and the peak wave direction (θ_p). Direct measurements of these parameters are obtained from the National Center for Archiving Swell Measurements (L'her et al., 1999). They were made by a directional wave rider buoy (DWR MKIII) operated by the Centre for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (CEREMA) and the University of Pau and Pays de l'Adour (UPPA). The buoy is located a few miles off the Basque Coast (Figure 1) at 50 meters water depth. Since its deployment in 2009, this buoy have been recording the parameters of interest every 30 minutes. The measuring range of this buoy is [-20m; 20m] for heave motion, [1.6s; 30s] for wave period and [0°; 360°] for wave direction. It has a resolution of 1 cm in heave motion and a directional resolution of 1.5°. To be consistent with the numerical wave data and meteorological data, a 1 hour time step was adopted for the buoy data.

The three parameters simulated at the buoy coordinates by the Meteo-France WAM model were provided by the Copernicus Marine Environment Monitoring Service. This reanalysis ("ibi_reanalysis_wav_005_006") covers the period 2007-2019 with a hourly time-step. The MFWAM model is derived from the third generation wave model WAM (Group, 1988). It is forced by wind fields obtained from a regional numerical weather prediction model (AROME). A more complete description of the MFWAM model can be found in Lefèvre and Aouf (2012).

Meteorological data, including average wind speed above 10 meters, wind direction and atmospheric pressure were furnished by the French national meteorological service MeteoFrance. The data were collected hourly by the meteorological station of the Biarritz airport, located only a few kilometers from the study site (Figure 1). It covers the period ranging from 2013-01-01 to 2018-12-31. By assembling the wave buoy data, the wind wave parameters and the meteorological data we obtain a dataset of 41439 hourly observations ranging from 2013-01-01 to 2018-12-31.

In this work, we are improving the wave forecast by correcting the systematic errors of the wind wave model. Therefore, we are not considering any temporal effects while improving H_S, T_p and θ_p. The dataset was randomly divided into 2 parts: the training part containing 70% of the observations (n = 28797) and the testing part containing the remaining 30% (n = 12342).

2.2. Error prediction method

The error prediction method consists in three steps:

- Step 1: Deviations between model predictions and measured values are computed:

$$E_{model} = X_{measured} - X_{modeled},$$

where E_{model} is the error of the model, $X_{measured}$ is the measured value of an output variable provided by the wave buoy and $X_{modeled}$ is the value of the same variable computed by the wave model.

- Step 2: E_{model} is predicted with an appropriate supervised machine learning algorithm.

- Step 3: The predicted error is added to the prediction of the wave model to obtain an updated numerical prediction:

$$X_{updated} = X_{modeled} + E_{predicted},$$

where $X_{updated}$ is the updated prediction of wave model and $E_{predicted}$ is the predicted error given by the supervised learning method.

This method is repeated separately for each output variable to improve (H_s, T_p, θ_p). The performance of this data assimilation method relies on two things: the quantity of data and the machine learning algorithm used. Since the machine learning algorithm are generally more suited to interpolate rather than extrapolate, the available data for learning process should cover as much as possible the range of all the probable events in the study area. Concerning the learning method, only neural networks have been used for the step (2) of the error prediction method to our knowledge(Makarynskyy et al., 2005; Moeini et al., 2012; Londhe et al., 2016). Because we want to compare the performance between different machine learning algorithms, we use random forest and gradient boosting trees. All the tested algorithms use the same input variables to improve the model accuracy: the three wave parameters (H_S, T_p, θ_p) given by the numerical model, the atmospheric pressure, the wind direction and speed.

2.3. Neural networks

Artificial neural networks have been extensively used in the domain of wave modelling (Deo et al., 2001; Makarynskyy et al., 2002; Makarynskyy, 2005; Mandal and Prabaharan, 2006) or wave parameters assimilation (Makarynskyy et al., 2005; Moeini et al., 2012; Londhe et al., 2016). It is why technical details will be avoided in this study and only the general concepts will be presented. The

readers can find more details and information on the working of neural networks in Liang and Bose (1996) or Friedman et al. (2001).

The most common class of neural networks is the multilayer perceptron. The neurons in this network are organized in three layers: the input layer that receive the input variables, the output layer that performs the final predictions and between these two layers there is the hidden layer. Neurons in the hidden layer transmit the signal to the output layer by transforming the weighted sum of the neurons present in the input layer with a non linear function called activation function. The weights between each neuron of the network are adjusted through the iterative process of backpropagation to minimize the error between the variable we want to predict and the variable predicted by the network (output layer).

As other machine learning methods, hyperparameters need to be specified before the training of neural networks. Some hyperparameters control the network architecture (number of neurons, layers, activation function used, etc...) while others control the training process (learning rate, bacth size, number of epochs, etc...). Hyperparameters must be tuned carefully in order to achieve optimal results with neural networks.

2.4. Tree based algorithms

Unlike neural networks, random forest and gradient boosting have never been used in the error prediction method. They are state-of-the-art ensemble learning techniques for classification and regression tasks. An ensemble learning technique commonly refers to a method that combines the predictions from multiple machine learning algorithms, called base learners, to produce more accurate predictions.

Random forest is an algorithm that builds many decision trees in parallel. These trees are the base learners for random forest and they have the following characteristics:

- Each tree is built using a different bootstrap sample of the data-set. This mechanism is called bagging.
- At each node, a given number (hereafter "mtry") of variables are randomly sampled as candidates at each split. The best split point is then selected within this random set of variables. This process is called feature sampling. The value "mtry" is fixed before growing the forest.
- Unlike the classification and regression trees of Breiman et al. (1984), the trees in random forest are fully grown (no pruning step).

Bagging and feature sampling are the core principles of random forest. They are two randomizing mechanisms which ensure that the trees are independent and are less correlated with each other. The final prediction of a random forest is obtained by averaging the results of all the independent trees in case of regression or using the majority rule in case of classification.

The most important hyperparameters in random forest are the number of trees and "mtry": the number of variables randomly sampled as candidates at each split when building the trees.

Gradient boosting is an algorithm that trains many weak learners sequentially to provide a more accurate estimate of the response variable. A weak learner is a machine learning model that perform slightly better than chance. In case of gradient boosting trees, the weak learners are shallow decision trees. Each new tree added to the ensemble model (combination of all the previous trees) minimizes the loss function associated with the ensemble model. The loss function depends on the type of the task performed and can be chosen by the user. For regression, the standard choice is the squared loss. By adding sequentially trees that minimize the loss function (i.e. follow the gradient of the overall loss function), the overall prediction error decreases. Technical details about gradient boosting trees can be found in (Friedman, 2001).

Many hyperparameters have to be tuned for gradient boosting trees, some of them control the gradient boosting process, such as the learning rate, the number of trees to be used whereas others regulate the construction process of the trees: minimal node size, sample of the dataset to be used, maximum depth.

2.5. Hyperparameter tuning

Hyperparameters influence significantly the training of the machine learning algorithms and therefore the quality of their predictions. The objective of hyperparameter tuning is to find the values of hyperparameters that yield the lowest error (RMSE in our case) for unseen data. Two types of methods exist to find the optimal values of hyperparameters: uninformed or informed.

In uninformed methods, many combinations of hyperparameter values are tested one after the other and the best combination is the one that yields the lowest error on unseen data. The values of hyperparameters are either sampled randomly (random search) or sampled along a grid (grid search). In both cases, each combination tested are independent from another. With grid and random search, it is not guaranteed to find the optimal set of hyperparameters and it usually requires a lot of iterations (combinations tested).

In informed methods, the results obtained by the past combinations are used to choose the next combination to evaluate. Bayesian optimization algorithm

is an informed method that aims to minimize an objective function, in our case the errors of the machine learning algorithms on unseen data. First, it builds a probability model (Gaussian process) of the objective function. Then it uses this surrogate model to select the most promising values of hyperparameters to evaluate. Once the promising combination of values have been evaluated, the probability model is updated and searched again for the most promising combination. This process is repeated several times. This method is employed in this article because it is very efficient for tuning hyperparameter values and it usually requires less iterations than uninformed methods (Bergstra et al., 2011). In-depth details of this method are given in the works of Snoek et al. (2012); Marchant and Ramos (2012) and Shahriari et al. (2015).

2.6. Training the algorithms

The machine learning algorithms described above are trained to predict the deviations of H_s, T_p or θ_p (one model for each variable), using 6 input variables: the three wave parameters (H_S, T_p, θ_p) given by the numerical model, the atmospheric pressure, the wind direction and speed.

The neural networks are built and trained with the R package **keras**. The input variables are centered and scaled to improve the result of neural networks and the weights are updated with the adam optimization algorithm (Kingma and Ba, 2014). Random forest and gradient boosting model are fitted in R using respectively the **ranger** package which provide fast implementation of Random Forests (suited for high dimensional data) and the **xgboost** package which is an efficient R implementation of the gradient boosting framework from Chen and Guestrin (2016). The input variables are not centered or scaled before the training of random forest and gradient forest because it does not influence the training of these algorithms.

The training is done twice: once with the default values of the hyperparameters in the R packages and once with the optimal values found with the Bayesian optimization method.

The best hyperparameter values are found by the means of Bayesian optimization method coupled with a 5-fold cross validation in the training dataset. That is, the training data are split into five equal-sized partitions and a machine learning model is recursively built on four partitions (80% of the training data) with a given hyperparameter combination. A performance metric, in our case the root mean square error is assessed on the remaining partition (20% of the training data). The resulting five performance metrics are averaged to provide an estimated out-of-sample performance of the respective hyperparameter combination. The objective function to minimize for the Bayesian optimization method is

Table 1: Statistical metrics for the three variables of interest before the hyperparameter tuning. "Ann" stands for artificial neural networks, "Rf" for random forest and "Gb" for gradient boosting tree.

	Hs				Tp				θ_p			
	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.
	Computed with all data											
Biais	-0.201	0.005	-0.002	-0.004	0.712	0.002	-0.026	0.005	-0.249	0.698	0.976	0.765
RMSE	0.399	0.306	0.248	0.267	1.839	1.603	1.282	1.388	13.803	12.391	9.749	10.645
SI	0.166	0.148	0.120	0.129	0.153	0.145	0.116	0.125	0.045	0.041	0.032	0.035
Cor	0.954	0.962	0.975	0.972	0.78	0.804	0.880	0.857	0.330	0.366	0.455	0.386
	Computed with data where $H_s > 3m$											
Biais	-0.536	-0.156	-0.124	-0.126	0.683	-0.026	-0.090	-0.041	1.925	-0.561	0.052	0.046
RMSE	0.766	0.515	0.420	0.433	1.348	1.083	0.956	1.022	7.351	5.769	4.449	4.931
SI	0.133	0.120	0.098	0.101	0.089	0.083	0.073	0.078	0.024	0.019	0.015	0.016
Cor	0.818	0.857	0.902	0.898	0.832	0.850	0.886	0.869	0.589	0.604	0.790	0.726

the average out-of-sample performance value. The Bayesian optimization for our data is performed using the R package **RBayesianOptimization**. First, random combinations of hyperparameter values are evaluated to serve as search base for the informed method (5 in this study), then an acquisition function (upper confidence bound) is used to find the next combination values to evaluate (this step is repeated 25 times).

3. Results and Discussion

3.1. Model comparison

To assess the accuracy of the numerical model and the proposed corrections, several metrics are computed including the root mean square error (RMSE), the correlation coefficient, the bias and the scatter index (SI). The bias represents the average error between the observed and modeled data and allows one to detect under or over estimation of the value of one parameter. The scatter index is a measure of the error normalized by the observation values. It is a standard metric for wave model inter-comparison (Londhe et al., 2016). More details about the computation of these two metrics can be found in the work of Mentaschi et al. (2013). The metrics are computed twice: once with the whole test data and once with a subset of the test data where $H_s > 3m$ because the underestimation of H_s is known to become larger above this height (Arnoux et al., 2018).

Table 1 presents the metrics obtained with no assimilation (numerical model) and after a preliminary data assimilation with the three machine learning algorithms. The term "preliminary" refers to the lack of hyperparameter tuning. The learning has been done using the default hyperparameter values given in table 2.

For significant wave height, the numerical model shows a negative bias. This indicates that the MFWAM model has a tendency to underestimate H_s such as other wind wave models (Moeini et al., 2012; Arnoux et al., 2018). The negative bias increases as the value of H_s becomes larger ($H_s > 3m$), meaning that H_s is more likely to be underestimated during energetic events. For the peak period, the numerical model shows a positive bias. When $H_s > 3m$, the bias and the RMSE for this parameter are smaller. The predictions of T_p are therefore better during energetic conditions. For wave direction, a small bias is observed in average and is greater when the waves are larger. The large difference in RMSE computed with data where $H_s > 3m$ and with all data is explained by the distribution of the wave direction according to the wave height. When the significant wave height is below 2 meters, the wave direction at the buoy is more variable (Figure S1, supplementary material) and the spectral wave model has more difficulties to predict correctly the direction. This can be confirmed by looking at the θ_p errors of the numerical model: we see that they are larger and occur more often when $H_s < 2m$ (Figure S2, supplementary material). A potential explanation of this phenomenon could be that below 2 meters, the sea state is more likely to be influenced by local wind conditions which are difficult to reproduce by the spectral wave model (Rascle and Ardhuin, 2013). When the significant wave height is above 2 meters, the wave directions are a lot less variable and the predictions of the spectral wave model are more accurate.

When we look at the metrics computed with all data, we see that the correction made by the three machine learning algorithms removes the bias and greatly reduces the RMSE and the scatter index for H_s and T_p. For θ_p, the mean bias is slightly larger after data assimilation for all algorithms. The correction of the machine learning algorithm could be less efficient for θ_p due to the high variability of the observed deviations they try to model (see the explanation in the paragraph above). However, lower value of RMSE and scatter index and larger correlation coefficients still indicate that the corrected data are closer to the observed values at the buoy.

For the metrics computed with data where $H_s > 3m$, the correction does not remove the bias for H_s and T_p but reduces it greatly. For the wave direction, the updated parameters are closer to the reality. Indeed, bias and RMSE obtained by the corrections are smaller than the numerical model and the correlation coefficients are larger for corrected data.

For this preliminary assimilation, random forest yields the best results for all the parameters. It reduces the RMSE values computed with all test data by 37.7%, 30% and 29% respectively for H_s, T_p and θ_p. Gradient boosting trees is

Table 2: Default values, ranges and selected value of hyperparameters for the machine learning algorithms

Machine learning algorithms	Hyperparameters	Default value	Range searched	Selected value for Hs	Selected value for Tp	Selected value for Dir
Neural networks	No. of units in hidden layer	13 ($2 \times h + 1$)	{1-40}	26	20	40
	Activation function	sigmoid	{relu,sigmoid,tanh}	sigmoid	sigmoid	relu
	Learning rate	0.001	{0.0001-0.1}	0.021	0.016	0.005
	Epochs	30	{10,30,50,100,150}	50	100	150
	Batch size	32	{16,32,64,128}	32	64	64
Gradient Boosting trees	Number of trees	100	{100-2000}	560	1150	1990
	Learning rate	0.3	{0.0001-0.3}	0.072	0.028	0.069
	Max depth	6	{1-20}	14	20	20
	Minimal node size	1	{1-15}	7	1	1
	Subsample	1	{0.5-1}	0.57	0.82	0.79
	Col sample	1	{0.5-1}	0.99	0.85	0.9
Random forest	Number of trees	500	{100,200,500,800,1000}	1000	1000	1000
	Mtry	2 ($\sqrt{h}$)	{2-6}	2	2	2

Note: h corresponds to the number of input variables (6 in our case).

close second and decreases the RMSE values by 33%, 24.5% and 22.8%. Finally, data assimilation with neural networks decreases the RMSE of H_s, T_p and θ_p by 23%, 12.8% and 10.2%.

As stated earlier, the performance of machine learning algorithms might depend on the choice of the hyperparameter values. The Bayesian optimization was therefore performed and optimal values were selected (Table 2). The selected hyperparameter values are quite different from the default values. Indeed, for neural networks, the best results were obtained with more epochs and more neurons in the hidden layer. For random forest, only the number of trees seems to have some effect on the results and models with a large number of trees performs better. Finally, for gradient boosting trees, models with a large number of trees and a small learning rate are preferred.

Metrics calculated with data corrected by the tuned machine learning algorithms are presented in Table 3. Overall, tuning the hyperparameter values has improved the results of all the algorithms. However, the degree of improvement differs depending on the algorithm. We observe the smallest improvements for random forest where the RMSE of every parameters seems to decrease by less than 1% in average. For neural networks, tuning hyperparameter values has a more significant effect by reducing the RMSE by 2 to 3% in average. The largest effect of tuning the hyperparameters are observed with gradient boosting trees. The RMSE is 8 to 11% lower for every parameter. The only exception is θ_p computed with all data where we have a small increase (2%) of RMSE. In general, the mean bias for H_s, T_p and θ_p remains the same before and after hyperparameter tuning expect for the bias of H_s computed when $H_s > 3m$ which is significantly

Table 3: Statistical metrics for the three variables of interest after the hyperparameter tuning."Ann" stands for artificial neural networks, "Rf" for random forest and "Gb" for gradient boosting tree.

	Hs				Tp				θ_p			
	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.	Numerical model	Ann Corr.	Rf Corr.	Gb Corr.
	Computed with all data											
Biais	-0.201	0.026	-0.002	-0.001	0.712	0.007	-0.022	0.003	-0.249	0.790	0.979	0.714
RMSE	0.399	0.300	0.246	0.240	1.839	1.553	1.269	1.231	13.803	12.07	9.646	9.501
SI	0.166	0.144	0.118	0.116	0.153	0.140	0.114	0.111	0.045	0.04	0.032	0.031
Cor	0.954	0.964	0.976	0.977	0.78	.817	0.882	0.889	0.330	0.36	0.461	0.421
	Computed with data where $H_s > 3m$											
Biais	-0.536	-0.117	-0.120	-0.099	0.683	0.032	-0.084	-0.051	1.925	-1.114	0.062	0.056
RMSE	0.766	0.495	0.417	0.404	1.348	1.064	0.950	0.943	7.351	5.820	4.412	4.365
SI	0.133	0.117	0.097	0.095	0.089	0.081	0.072	0.072	0.024	0.019	0.015	0.015
Cor	0.818	0.861	0.903	0.908	0.832	0.856	0.888	0.889	0.589	0.609	0.793	0.793

lower after the tuning.

For this dataset, gradient boosting algorithm shows the best performances for all parameters. Assimilation with this algorithm decreases the RMSE values computed with all test data by 39.8% for H_s, 33% for T_p and 31% for θ_p. For H_s and θ_p, the reduction are even lower for the RMSE values computed with $H_s > 3m$: 47% for the significant wave height and 40% for wave direction. The performances of random forest for H_s, T_p and θ_p are slightly better than the results obtained before tuning the hyperparameters: respectively 38.3%, 30.9%, 30.1%. The performances are also better for neural networks after hyperparameter tuning: it decreases the RMSE values by 24.8% for H_s, 15.5% for T_p and 12.5% for θ_p. The differences in efficiency between neural networks and ensemble learning techniques could be explained by the architecture chosen for the neural networks. Indeed, this work shows the results for multilayer perceptrons with only one hidden layer which is the typical choice in the literature (Londhe et al., 2016; Moeini et al., 2012). By choosing an architecture with more hidden layers, the networks might be able to model more complex phenomena and bring a better improvement for the three wave parameters.

The distribution of the errors after the different corrections are presented in the figure 2. For all wave parameters, the distributions of the errors after a correction have narrowed and are now more centered in zero. The differences in performance between algorithms are confirmed with these violin plots. Indeed, when the correction is made with random forest or gradient boosting trees, the distributions of the errors are more narrow than the distributions of the errors obtained with neural networks. The difference in efficiency between random forest

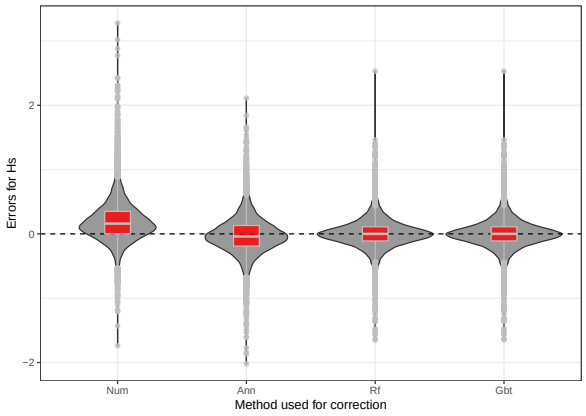
and gradient boosting trees is not distinguishable graphically. It is expected as the metrics of the two algorithms only differ by a few percents. For H_s and T_p, the corrections have also removed the bias observed for numerical model. The large errors of θ_p for the numerical model (Figure 2) are observed when $H_s < 2m$ and are not corrected by the machine learning algorithms. Figures showing the observed values versus the corrected values are available in the supplementary material for the three wave parameters.

3.2. Predictive power of the input variables

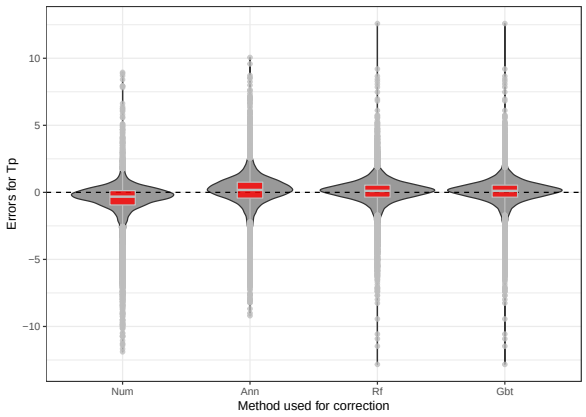
In addition to their performance, random forest and gradient boosting algorithms can provide a measure of importance for each variable used as input. This importance indicates the predictive power of the variable. It can be used to sort variable from most to least predictive, allowing one to have more insight on the problem and to perform feature selection when there are too many input variables. The figure 3 shows the importance measure of each variable computed by the random forest depending on the parameter to improve. For H_s and T_p, the most important variables are the value of H_s and T_p modeled by the wind wave model. It is different for the direction where the most important variables are the value of θ_p and H_s given by the model. The predictive power of local meteorological variables is quite low, suggesting that local and instantaneous meteorological variables does not bring valuable information in the assimilation process. The wind wave formation process is not instantaneous and occurs in large regional scale, therefore using meteorological variables from the past (several days before) and from different locations (located in the ocean) could lead to a better predictive power which means better updated wave predictions.

3.3. Example of application

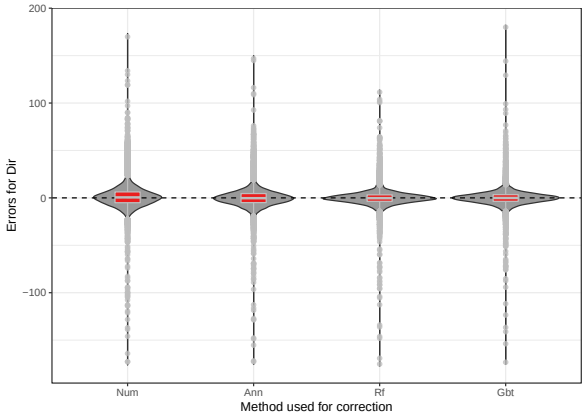
To investigate the potential effect of the different corrections in a real case scenario, the extreme wave run-up $R_{2\%}$ at the Grande Plage de Biarritz has been computed for the test period with the Stockdon formula (Stockdon et al., 2006) which uses H_s and T_p and the beach slope as parameters. The beach slope is fixed to 8% according to the work of Morichon et al. (2018). Using the extreme wave run-up calculated with the buoy data as reference, the metrics presented previously have been computed for the numerical model and the different corrections (Table 4). From this table, it is evident that the data corrected with machine learning algorithms provide wave run up values that are closer to the "real" values with lower RMSE, Scatter index and greater correlation coefficient. Although the bias remains, the correction made by the gradient boosting



(a) H_s correction



(b) T_p correction



(c) θ_p correction

Figure 2: Distribution of the errors computed between values observed at the buoy and values corrected or not with the different machine learning algorithms. ”Num” stands for numerical model (no correction), ”Ann” for artificial neural networks, ”Rf” for random forest and ”Gb” for gradient boosting trees. The horizontal lines in the red boxplots represent from top to bottom: the third quartile, the median and the first quartile.

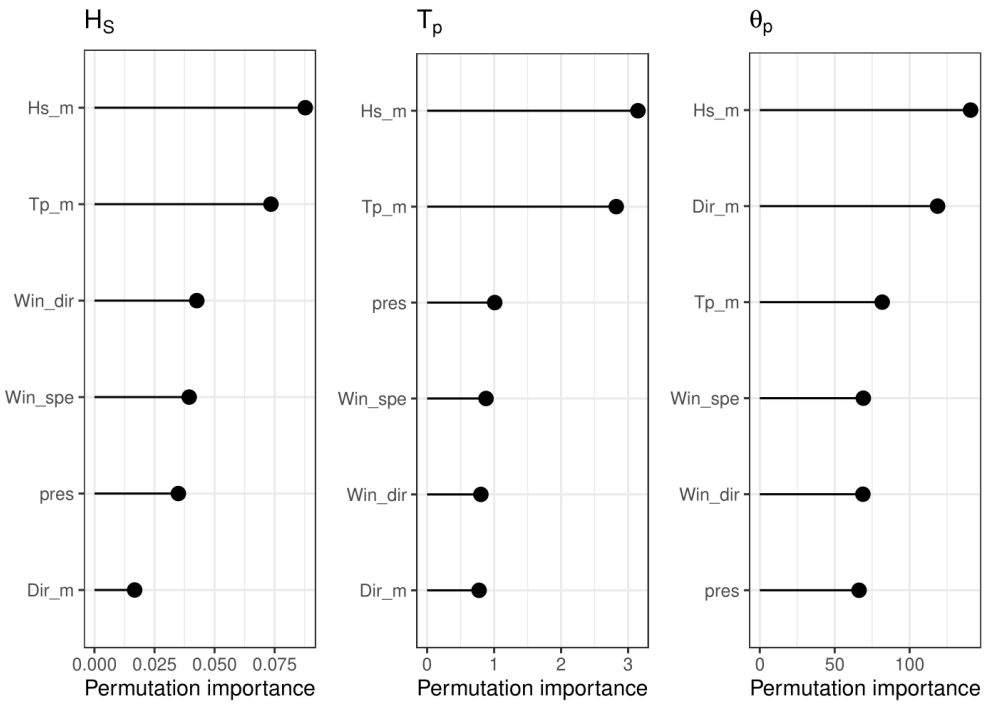


Figure 3: Variable importance for the correction of the three wave integrated parameters.

Table 4: Statistical metrics of the R2% calculated with Stockdon's formula. These results are obtained by taking the R2% computed with buoy data as reference. "Ann" stands for artificial neural networks, "Rf" for random forest and "Gb" for gradient boosting tree.

	Numerical model	Ann Corrected	Rf Corrected	Gb Corrected
	Computed with all data			
Biais	0.003	0.019	0.002	0.003
RMSE	0.223	0.209	0.175	0.172
SI	0.145	0.136	0.114	0.112
Cor	0.943	0.950	0.965	0.966
	Computed with data where Hs > 3m			
Biais	-0.042	-0.030	-0.054	-0.042
RMSE	0.313	0.282	0.248	0.242
SI	0.119	0.108	0.093	0.092
Cor	0.854	0.862	0.897	0.901

tree algorithm decreases the RMSE of the extreme wave run-up by 22% (for all data and data where $H_s > 3m$). Random forest shows almost the same reduction of RMSE values: 21.5% for all data and 20.7% for data where $H_s > 3m$. The correction obtained by neural networks is less efficient: it reduces the RMSE computed with all data and data where $H_s > 3m$ by 6.2 and 9.9% respectively.

4. Conclusion

In this work, random forest and gradient boosting trees were employed for the first time in the error prediction method. These ensemble learning techniques based on decision trees performed better than neural networks for improving the wave forecast of the Basque Coast. The correction made by gradient boosting trees yielded the best results for all the wave parameters: it reduced the RMSE values by nearly 40% for H_s, 33% for T_p and 31% for θ_p. The reduction of RMSE values for random forest was only a few percents lower than gradient boosting trees. The corrections made by neural networks were significant but yielded reductions in RMSE not as high as the two ensemble learning techniques: 24.8% for H_s, 15.5% for T_p and 12.5% for θ_p.

As expected, tuning the hyperparameters of the machine learning algorithms had a positive effect on the final results. However, the effect of the tuning differed

depending on the algorithms. Indeed, random forest was less affected as it only reduced the RMSE values by 1% in average. The tuning had more effect on neural networks reducing the RMSE values by 2 to 3%. Gradient boosting tree algorithm was the most affected by hyperparameter tuning as the results were improved by 8 to 11% in average. One of the main advantage of random forest over gradient boosting trees is that it doesn't need this tuning step in order to yield great results. This is not negligible as hyperparameter tuning step can be time consuming and computationally demanding depending on the complexity of the search (number of hyperparameters).

Contrary to neural networks, Random forest and Gradient boosting trees provided valuable insights by giving the predictive power of each input variable. The predictive power of variable brings interpretability to the model and can give a better understanding of the variable we try to predict. For example, we know that the significant wave height modelled by the numerical wave model was the most important variable in the correction of the three parameters. In cases where there are a lot of input variables, knowing their associated predictive power helps developing more parsimonious models by keeping the pertinent variables and subtracting the less informative ones from the model.

The error prediction method has proven to be useful in improving wave forecast. This had an impact in a real life application by improving the accuracy of the extreme run-up computed at the Grande Plage de Biarritz. Here again the corrections brought by random forest and gradient boosting tree were better than the correction made by neural networks. The decrease in RMSE values was around 22% for the two ensemble techniques and 6.2% for the neural networks. Even though the differences in performance might not appear significant, it can make a difference when using these corrections in an early warning system. It is especially true when dealing with storm events where H_s and T_p are large.

The differences between machine learning algorithms observed in this article are specific to Biarritz site. The results might differ for another study site. Therefore, we can only advise to test and compare several machine learning algorithms to find the optimal one associated with the site of interest.

Finally, the assimilation made in this study did not account for the temporal aspect in the errors of the numerical model, it only corrected systematic errors of the wave model. In the future, this work could be extended by adding input variables containing temporal aspect. This could be the values of a modeled parameter at previous time steps such as the work of Londhe et al. (2016). In this framework, neural networks could perform better as they are known to handle efficiently time series. Other input variables could be also used to improve

the wave forecast such as the meteorological data from the past or at different locations. Because the success of the error prediction method depends on the quantity of data, it would be also interesting to perform a sensitivity analysis on the quantity of data used in the training process. This could give us some insights on the minimal quantity of data required to obtain a desirable assimilation procedure.

Acknowledgments

Funding was provided by the Energy Environment Solutions (E2S-UPPA) consortium and the BIGCEES project from E2S-UPPA ("Big model and Big data in Computational Ecology and Environmental Sciences"). The authors would like to thank the French national meteorological service "MeteoFrance" and Copernicus Marine Environment Monitoring Service for providing data.

Reproducibility

Meteorological data used in this article are private and can not be provided by the authors. However, the R code to perform the analysis and an example of data assimilation on wave forecast data (used in operational) are provided in this Github repository.

References

Arnoux, F., Abadie, S., Bertin, X., Kojadinovic, I., 2018. A database to study storm impact statistics along the Basque Coast. Journal of Coastal Research 85, 806–810.

Babovic, V., Cañizares, R., Jensen, H.R., Klinting, A., 2001. Neural networks as routine for error updating of numerical models. Journal of Hydraulic Engineering 127, 181–193.

Babovic, V., Sannasiraj, S.A., Chan, E.S., 2005. Error correction of a predictive ocean wave model using local model approximation. Journal of Marine Systems 53, 1–17.

Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization , 2546–2554.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees Chapman & Hall. New York .

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, ACM. pp. 785–794.

de Santiago, I., Morichon, D., Abadie, S., Reniers, A.J., Liria, P., 2017. A comparative study of models to predict storm impact on beaches. Natural Hazards 87, 843–865.

Deo, M.C., Jha, A., Chaphekar, A.S., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean engineering 28, 889–898.

Deshmukh, A.N., Deo, M.C., Bhaskaran, P.K., Nair, T.B., Sandhya, K.G., 2016. Neural-network-based data assimilation to improve numerical ocean wave forecast. IEEE Journal of Oceanic Engineering 41, 944–953.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. volume 1. Springer series in statistics New York.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Annals of statistics , 1189–1232.

Group, T.W., 1988. The WAM model—A third generation ocean wave prediction model. Journal of Physical Oceanography 18, 1775–1810.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Lefèvre, J.M., Aouf, L., 2012. Latest developments in wave data assimilation, in: ECMWF Workshop on Ocean Waves, pp. 25–27.

L'her, J., Goasguen, G., Rogard, M., 1999. CANDHIS database of in situ sea states measurements on the French coastal zone, in: The Ninth International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers.

Liang, P., Bose, N.K., 1996. Neural network fundamentals with graphs, algorithms and applications. Mac Graw-Hill .

Londhe, S.N., Shah, S., Dixit, P.R., Nair, T.M.B., Sirisha, P., Jain, R., 2016. A Coupled Numerical and Artificial Neural Network Model for Improving Location Specific Wave Forecast. Applied Ocean Research 59, 483–491. doi:10.1016/j.apor.2016.07.004.

Makarynskyy, O., 2005. Neural pattern recognition and prediction for wind wave data assimilation. Pac Oceanogr 3, 76–85.

Makarynskyy, O., Pires-Silva, A.A., Makarynska, D., Ventura-Soares, C., 2002. Artificial neural networks in the forecasting of wave parameters, in: 7th International Workshop on Wave Hindcasting and Forecasting. Banff, Alberta, Canada, pp. 514–522.

Makarynskyy, O., Pires-Silva, A.A., Makarynska, D., Ventura-Soares, C., 2005. Artificial neural networks in wave predictions at the west coast of Portugal. Computers & Geosciences 31, 415–424.

Mandal, S., Prabaharan, N., 2006. Ocean wave forecasting using recurrent neural networks. Ocean engineering 33, 1401–1410.

Marchant, R., Ramos, F., 2012. Bayesian optimisation for intelligent environmental monitoring, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE. pp. 2242–2249.

Mentaschi, L., Besio, G., Cassola, F., Mazzino, A., 2013. Problems in RMSE-based wave model validations. Ocean Modelling 72, 53–58.

Moeini, M.H., Etemad-Shahidi, A., Chegini, V., Rahmani, I., 2012. Wave data assimilation using a hybrid approach in the Persian Gulf. Ocean Dynamics 62, 785–797.

Morichon, D., de Santiago, I., Delpey, M., Somdecoste, T., Callens, A., Liquet, B., Liria, P., Arnould, P., 2018. Assessment of flooding hazards at an engineered beach during extreme events: Biarritz, SW France. Journal of Coastal Research 85, 801–805.

Rakha, K.A., Al-Salem, K., Neelamani, S., 2007. Hydrodynamic atlas for Kuwaiti territorial waters. Kuwait Journal of Science and Engineering 34, 143.

Rascle, N., Ardhuin, F., 2013. A global wave parameter database for geophysical applications. part 2: Model validation with improved source term parameterization. Ocean Modelling 70, 174–188.

Refsgaard, J.C., 1997. Validation and intercomparison of different updating procedures for real-time forecasting. Hydrology Research 28, 65–84.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2015. Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE 104, 148–175.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, pp. 2951–2959.

Stockdon, H.F., Holman, R.A., Howd, P.A., Sallenger Jr, A.H., 2006. Empirical parameterization of setup, swash, and runup. Coastal engineering 53, 573–588.

Vousdoukas, M.I., Ferreira, Ó., Almeida, L.P., Pacheco, A., 2012. Toward reliable storm-hazard forecasts: XBeach calibration and its potential application in an operational early-warning system. Ocean Dynamics 62, 1001–1015.

Wolpert, D.H., 2002. The supervised learning no-free-lunch theorems, in: Soft Computing and Industry. Springer, pp. 25–42.

Supplementary material: Using Random forest and Gradient boosting trees to improve wave forecast at a specific location

Aurélien Callens[a,*], Denis Morichon[b], Stéphane Abadie[b], Matthias Delpey[c], Benoit Liquet[a,d]

[a] *Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Pau, France*
[b] *Université de Pau et des Pays de l'Adour, E2S UPPA, SIAME, Anglet, France*
[c] *Centre Rivages Pro Tech SUEZ EAU FRANCE, Bidart, France*
[d] *Department of Mathematics and Statistics, Macquarie University, Sydney, Australia*

Distribution of θ_p measured at the buoy.

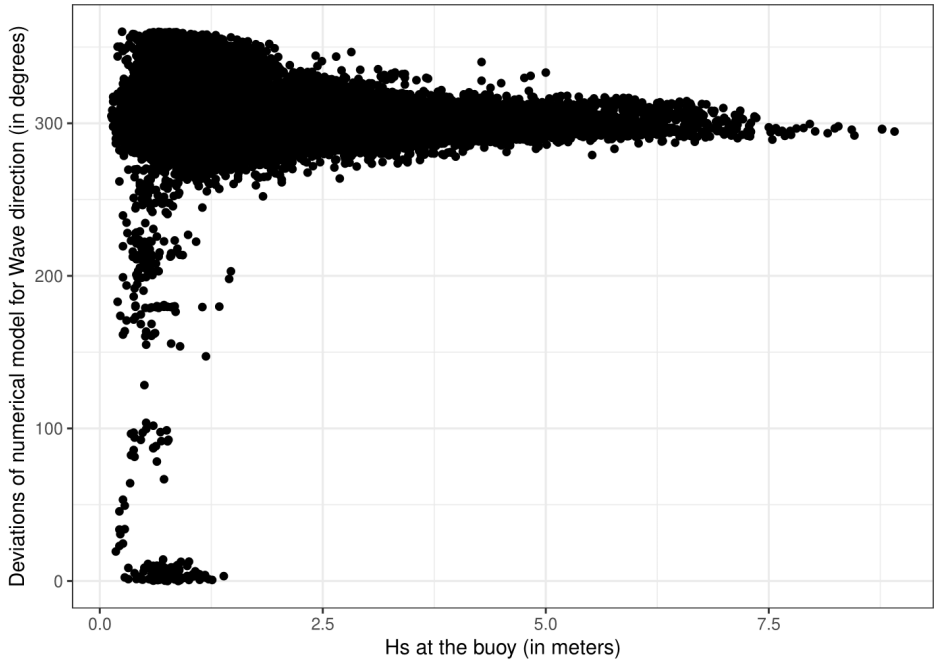


Figure 1: Wave direction vs significant wave height. Both parameters are measured at the buoy of Anglet.

*Corresponding author
Email address: aurelien.callens@univ-pau.fr (Aurélien Callens)

Errors of the numerical model for θ_p depending on significant wave height measured at the buoy.

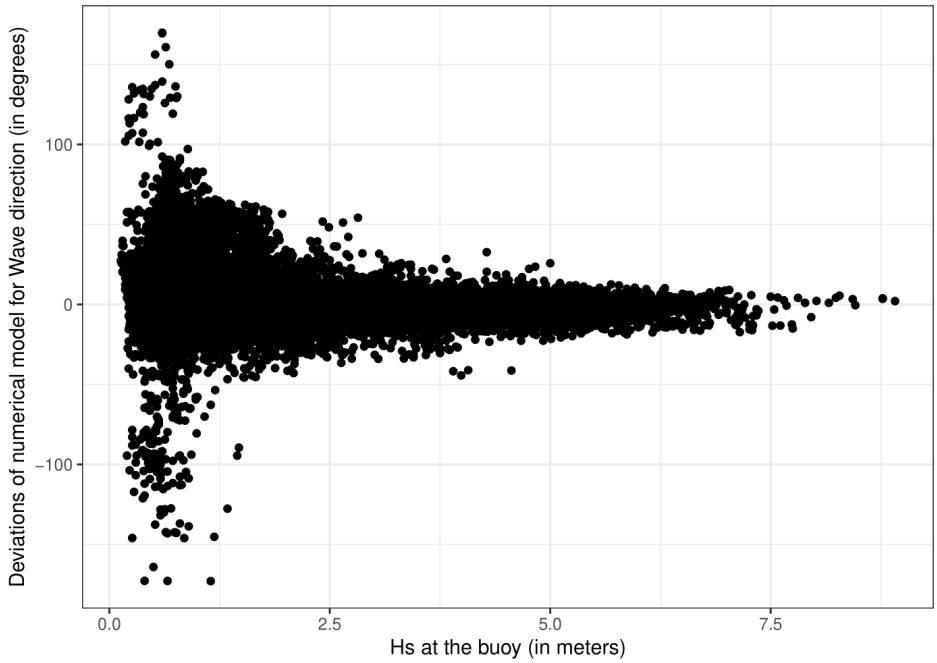


Figure 2: Errors of the numerical model for wave direction vs significant wave height at the buoy of Anglet.

Observed data vs data corrected with different machine learning algorithms

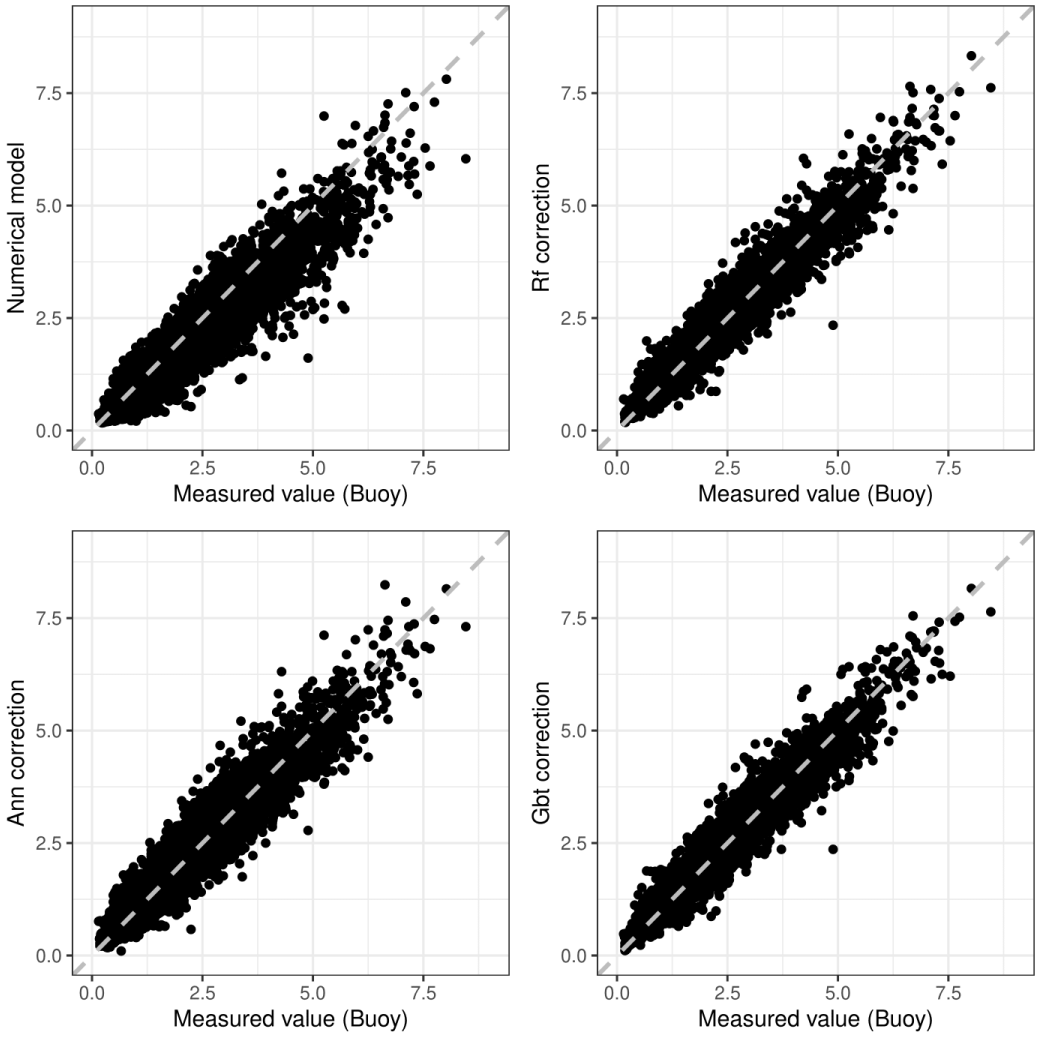


Figure 3: H_s values observed at the buoy vs H_s values corrected with different machine learning algorithms

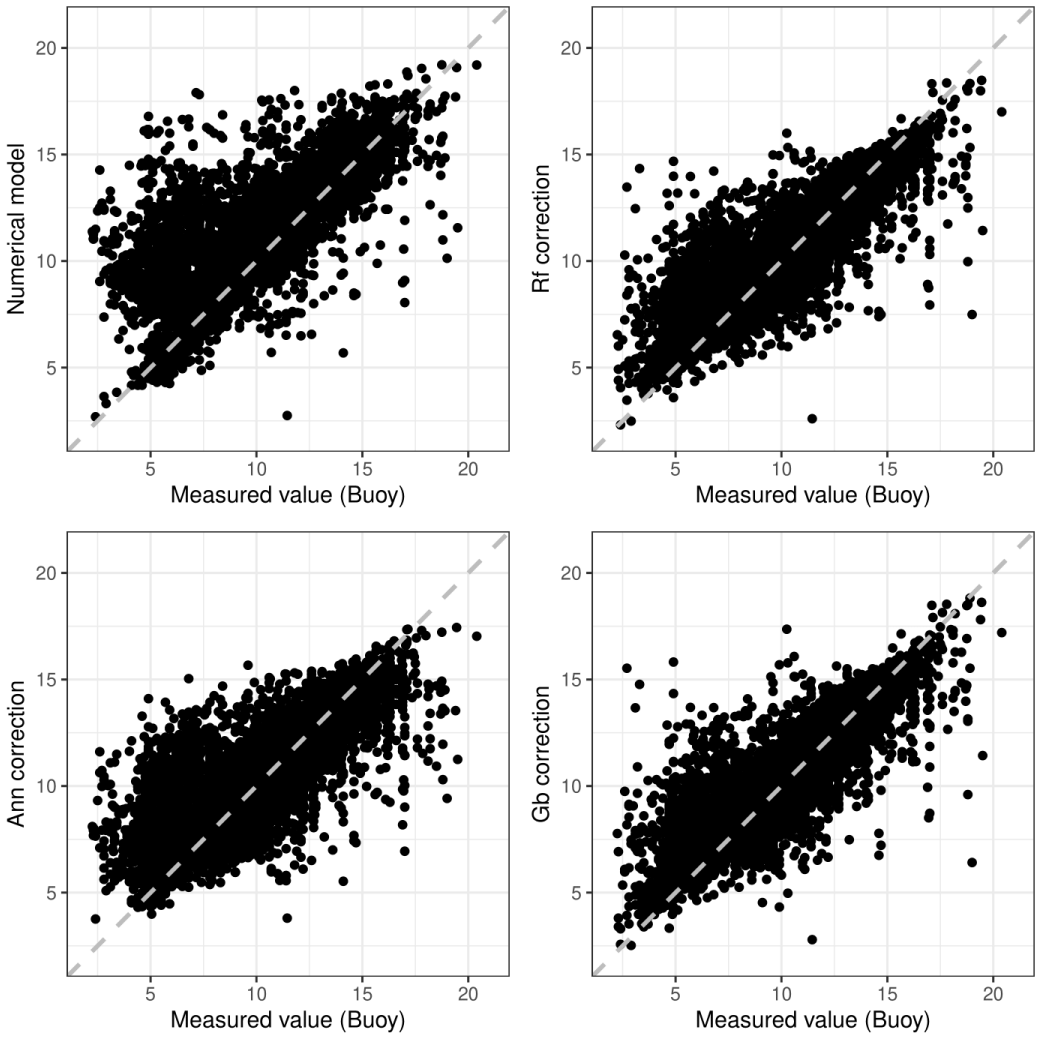


Figure 4: T_p values observed at the buoy vs T_p values corrected with different machine learning algorithms

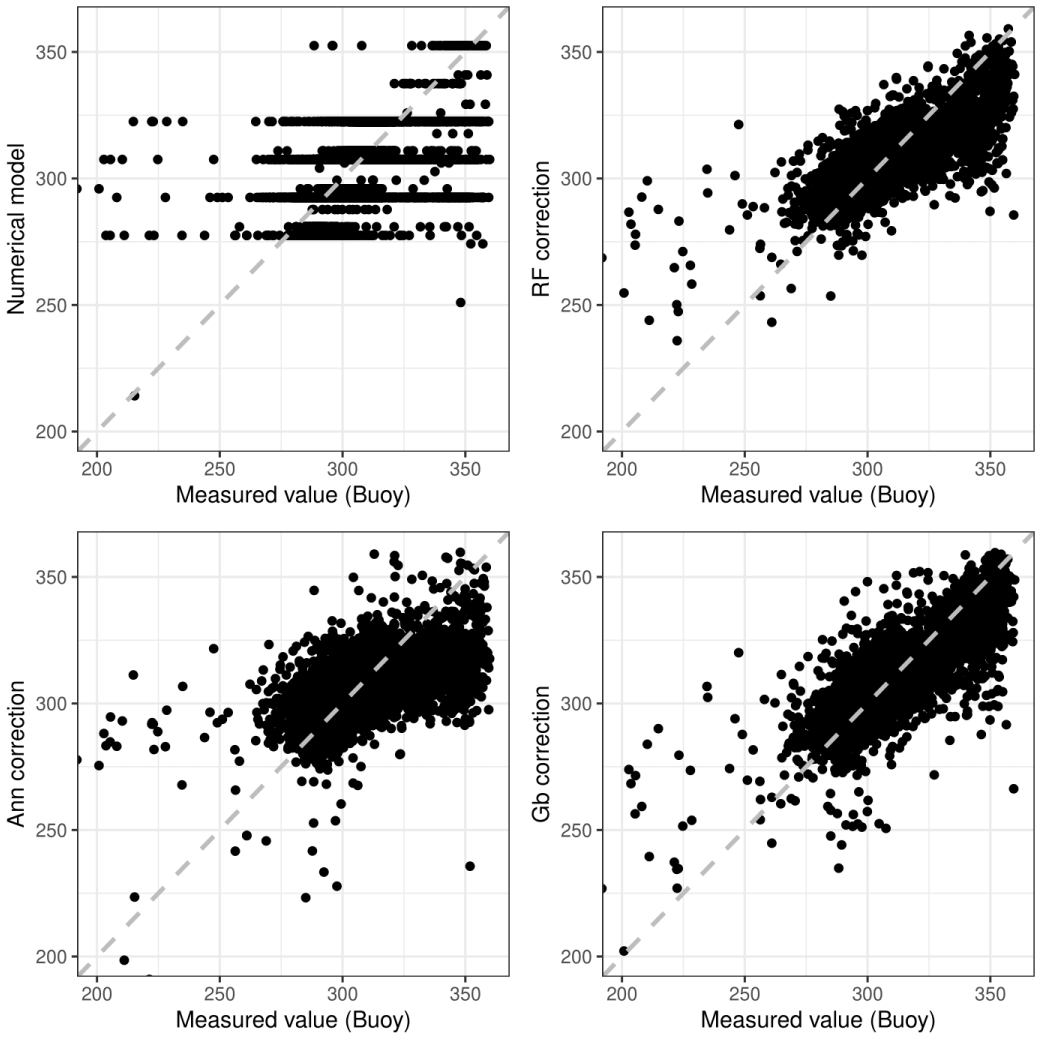


Figure 5: θ_p values observed at the buoy vs θ_p values corrected with different machine learning algorithms

3.3 Conclusion

In this chapter, we showed that SLM can be employed to improve significantly the predictions of numerical wave models at a specific location. We also proved the importance of comparing different SLM and choosing the optimal hyper-parameters in order to obtain the best results. With at least several years of wave buoy data, this method can be easily implemented to improve the predictions of numerical wave models in an EWS.

The methodology presented in this chapter only corrects the systematic errors of the numerical model at a specific location. This work could be therefore extended by including temporality in the error prediction method. To that end, errors of previous time steps could be included in the explanatory variable or methods specialized in time series prediction such as recurrent neural networks (RNN) could be employed in the same spirit as Zhang et al. (2021). In this work, they demonstrate that RNN can be used to improve the predictions of a numerical wind wave model by integrating both the local data and the temporality in the errors of the numerical wind wave model. Their proposed methodology with RNN led to a better accuracy than the error prediction method performed with usual machine learning algorithms. The work presented in this chapter could also be extended by correcting the whole wave field (multiple locations) predicted by the numerical model.

4. Automatic creation of storm impact database based on video monitoring and convolutional neural networks

4.1 Introduction

The objective of this chapter is to demonstrate that SLM can be employed to create automatically a storm impact database with images from video monitoring networks. Data about the impact of a hazard is usually the limiting factor when building a data-driven model translating tide, wave and meteorological conditions into impact on the shore. Unlike data about the tide, wave or meteorological conditions, data about storm impact are rare, sparse and mostly come from archives or insurances data. So far, no methods exist to collect routinely storm impact data. We propose to use Convolutional neural networks (CNN), which are deep learning methods, to classify the video monitoring images into three storm impact regimes which are categories of coastal flooding risk. Several CNN architectures and methods to deal with class imbalance are tested on two sites (Biarritz and Zarautz) to find the best practices for this classification task. Transfer learning is also investigated to facilitate the transferability of this method to new sites. Once trained, the CNN can predict the storm impact regimes of newly created timestacks, generating an incremental storm impact database.

Scientific output:

This work resulted in the publication of an article (presented below) and in a communication:

Publication:

- Callens, A., Morichon, D., Liria, P., Epelde, I., & Liquet, B. (2021). Automatic Creation of Storm Impact Database Based on Video Monitoring and Convolutional Neural Networks. Remote Sensing, 13(10), 1933.

Communication:

- "Automatic Creation of Storm Impact Database Based on Video Monitoring and Convolutional Neural Networks", Online seminar for the Probability and Statistics research team of the LMAP (July 2021).

4.2 Article: Automatic creation of storm impact database based on video monitoring and convolutional neural networks

remote sensing

MDPI

Article

Automatic Creation of Storm Impact Database Based on Video Monitoring and Convolutional Neural Networks

Aurelien Callens [1,*], Denis Morichon [2], Pedro Liria [3], Irati Epelde [3] and Benoit Liquet [1,4]

[1] LMAP, Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, 64000 Pau, France; benoit.liquet@univ-pau.fr
[2] SIAME, Université de Pau et des Pays de l'Adour, E2S UPPA, 64600 Anglet, France; denis.morichon@univ-pau.fr
[3] AZTI, Marine Research Division, KOSTARISK, 20110 Pasaia, Spain; pliria@azti.es (P.L.); iepelde@azti.es (I.E.)
[4] Department of Mathematics and Statistics, Macquarie University, Sydney, NSW 2006, Australia
* Correspondence: aurelien.callens@univ-pau.fr

Abstract: Data about storm impacts are essential for the disaster risk reduction process, but unlike data about storm characteristics, they are not routinely collected. In this paper, we demonstrate the high potential of convolutional neural networks to automatically constitute storm impact database using timestacks images provided by coastal video monitoring stations. Several convolutional neural network architectures and methods to deal with class imbalance were tested on two sites (Biarritz and Zarautz) to find the best practices for this classification task. This study shows that convolutional neural networks are well adapted for the classification of timestacks images into storm impact regimes. Overall, the most complex and deepest architectures yield better results. Indeed, the best performances are obtained with the VGG16 architecture for both sites with F-scores of 0.866 for Biarritz and 0.858 for Zarautz. For the class imbalance problem, the method of oversampling shows best classification accuracy with F-scores on average 30% higher than the ones obtained with cost sensitive learning. The transferability of the learning method between sites is also investigated and shows conclusive results. This study highlights the high potential of convolutional neural networks to enhance the value of coastal video monitoring data that are routinely recorded on many coastal sites. Furthermore, it shows that this type of deep neural network can significantly contribute to the setting up of risk databases necessary for the determination of storm risk indicators and, more broadly, for the optimization of risk-mitigation measures.

Keywords: convolutional neural networks; storm impact database; transfer learning; video monitoring

Citation: Callens, A.; Morichon, D.; Liria, P.; Epelde, I.; Liquet, B. Automatic Creation of Storm Impact Database Based on Video Monitoring and Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 1933. https://doi.org/10.3390/rs13101933

Academic Editors: Magaly Koch, Yukiharu Hisaki, Xiaofeng Li, Zhixiang Fang, Quanyi Huang and Jaroslaw Tęgowski

Received: 8 April 2021
Accepted: 11 May 2021
Published: 15 May 2021

1. Introduction

Databases containing information on past storm characteristics and their impacts on the coast are essential for the disaster-risk-reduction process. They enable scientists and coastal stakeholders to better understand the storm hazard in a specific area, to identify potential trends, and most importantly to assess coastal risks (present or future) through their use in the development and validation of early warning systems [1,2].

In these databases, storm impact is mostly represented as a qualitative variable with different categories. The different categories of storm impact are called "regimes" and are defined according to the Sallenger's scale [3]. This scale was originally derived to classify storm impact intensity based on the relation between wave-induced maximum water level and topographic elevations of the different sections of a natural beach. Recently, this approach has been extended to the estimation of storm impact intensity at an engineered beach backed by a seawall [4].

Due to the extreme and episodic nature of storms, databases covering a long period of time are necessary. Observed data about storm characteristics such as tide, wave, and

Due to the extreme and episodic nature of storms, databases covering a long period of time are necessary. Observed data about storm characteristics such as tide, wave, and wind are abundant and have been collected routinely for decades. In addition, numerous reanalyses and hindcasts are available for these variables. On the contrary, data on storm impacts are more sparse and mostly come from archives [5–7] or insurance data [8,9]. A few examples of storm impact databases are: the RISC-KIT database, which contains storm impact information for nine study sites in Europe [10]; the SurgeWatch database [5] for the UK; and a database for the Basque coast [7]. Even though archives and insurance provide information, there are some limitations including the heterogeneity, the incompleteness of the data sources, and the consequent amount of work needed [7]. A solution to routinely create a storm impact database could be to use images provided by coastal monitoring stations that are now widely used worldwide to survey and study coastal processes.

In recent decades, video monitoring systems have proven to be valuable assets in the study of the coastal zone due to their cost-efficiency and their ability to provide a continuous stream of data including intense storm conditions. Video monitoring systems are generally composed of one or several cameras operated by a monitoring software such as Argus [11], HORUS, Kosta (www.kostasystem.com, accessed on may 2021) or Sirena [12]. The reader is referred to the work of Nieto et al. [12] for a comparison of the cited monitoring systems. These systems generate different types of images that can be applied to study coastal processes such as beach morphology changes, wave runup, and coastal currents [13,14]. Among the different types of images generated by the video monitoring system, timestacks images represent the time-varying pixel intensities along a particular cross-shore transect in the camera's field of view. They are used to perform wave runup parametrization [15–17], wave breaking detection [18], or intertidal topography [19] and also to estimate wave characteristics [20,21], sea level [22], and bathymetry [23,24]. Timestacks images have also been employed in the study of storm impact. In the work of Thuan et al. [25], they quantify the impact of two typhoons on the longshore-averaged shoreline changes based on the analysis of a series of timestacks images. To our knowledge, timestacks have not been used to directly measure storm impact regimes as defined previously. Image processing techniques are usually employed to transform the information contained in the images into quantitative measurements (runup elevation, wave height, shoreline). In this article, we propose extracting storm impact regimes (qualitative data) directly from the timestacks.

The storm impact regimes can be extracted from timestacks using two methodologies. The first methodology can be qualified as deterministic; it relies on image processing techniques and consists of two steps. First, the water line position is found by segmenting the image. The storm impact regime is then deduced by comparing the position of the waterline with the position of the defense infrastructure in the timestack. Different methods can be used to extract the waterline from timestacks images [26,27]. For example, Otsu's method [26] divides the pixels into two groups depending on their intensity values. It is not always robust and depends on the quality and lighting of the images. Most of the time, it requires rigorous and tedious human verification and correction [16]. The second methodology, presented in this article, relies on deep learning with convolutional neural networks (CNNs). CNNs are a class of deep neural networks, specializing in imagery analysis, that perform well on specific problems such as image classification and segmentation. First, timestack images are classified into storm impact regimes by human operators. Then, the CNN is trained to classify timestacks into storm impact regimes using the annotated dataset. During the training process, the CNN learns to simultaneously classify the images and which features to detect in order to achieve the best classification accuracy. Once the neural network has learned on the training dataset, it can be used to routinely analyze the timestacks produced by the video monitoring system and therefore create incrementally a qualitative storm impact database. This second methodology, based on a self-learning algorithm (CNN), allows for more automation compared to the first methodology because it does not require site-specific calibration [17].

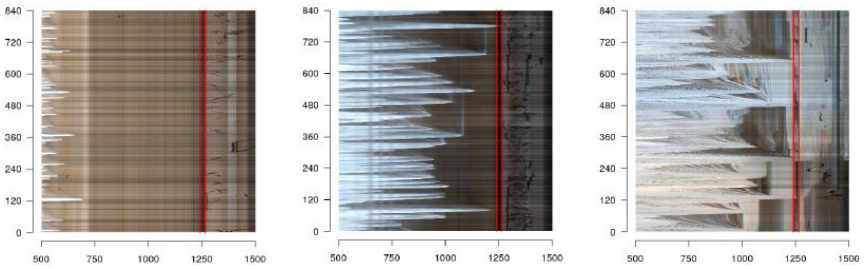
has already been employed in coastal engineering domain [14,29] and usually results in faster training and better accuracy. In the case of storm impact recognition, where images of extreme storm impact regimes are rare by nature, this method can significantly improve the performances of CNNs. Moreover, it is reasonable to think that knowledge acquired at one site can be used to improve the performances on another site. This could be a non-negligible asset for the application of the method to a new site.

This paper aims to demonstrate the high potential of CNN methods to constitute a storm impact database using timestacks images provided by coastal video monitoring stations. Different methods are tested using images collected at two study sites. The best practice and the transferability of knowledge gained at one site to another are studied. In the following sections, the study sites and the video dataset are first described. The main features of the CNN implementation procedure are then shown in Section 3. Results and transferability of the CNN between the study sites are presented in Section 4 and discussed in Section 5. The main results are finally presented in the conclusion, Section 6.

2. Study Sites and Data

In this study, the storm impact intensity is classified into three storm impact regimes (Figure 1) derived from the Sallenger's scale [3]. The following three categories have been adapted for the timestack images:

- Swash regime: all the waves in the timestack are confined to the beach;
- Collision regime: at least one wave in the timestack collides with the bottom of the seawall ;
- Overwash regime: at least one wave in the timestack completely overtops the seawall.



(**a**) (**b**) (**c**)

Figure 1. The three categories of storm impact regimes estimates from timestack images. *y*-axis represents the time in seconds, *x*-axis represent the pixel index (cropped images), and red lines represent the sea wall bottom and top positions. (**a**) Swash regime. (**b**) Collision regime. (**c**) Overwash regime.

The CNNs were trained on timestacks images collected by video monitoring stations operating on two sites along the basque coast (Figure 2), namely the Grande Plage of Biarritz (GPB), and the Zarautz beach (ZB). The use of two sets of data acquired from sites with different geological and morphological characteristics and distinct responses to oceanic forcing makes it possible to assess objectively the ability of CNN to detect storm impact regimes.

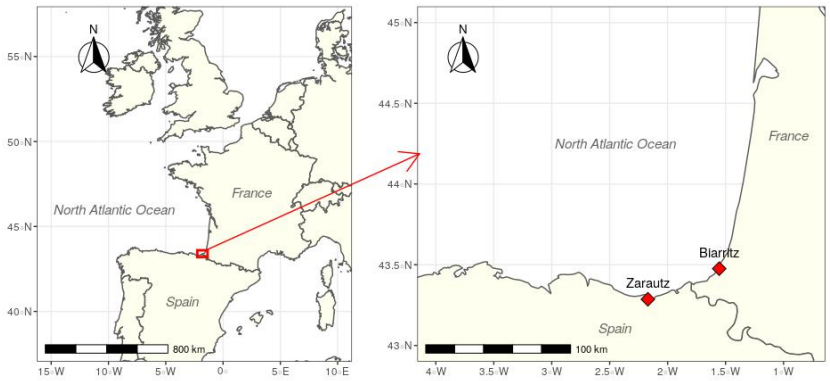


Figure 2. Map showing the locations of the two study sites.

2.1. Grande Plage de Biarritz

2.1.1. Site Characteristics

The Grande Plage of Biarritz (GPB) is an urban embayed beach that is 1.2 km long, located on the southern Aquitanian coast of France (Figure 2). It has a high socio-economic importance for the city of Biarritz due to its tourist appeal, its historical heritage, and its location near the city center. In terms of characteristics, the GPB is an intermediate-reflective beach with typically a steep foreshore slope of 8–9% and a gentle nearshore slope of 2–3% [30]. It is a mesotidal beach with 4.5 m spring tidal range around a mean water level of 2.64 m. This narrow beach is backed by a seawall with an alongshore elevation varying between 7 and 8 m. This seawall serves as defense infrastructure for back beach buildings.

The beach is predominantly exposed to waves coming from the WNW direction. The offshore wave climate is moderately to highly energetic. The annual average significant wave height and peak period are, respectively, H_s = 1.5 m and T_p = 10 s [30]. In this region, an event is qualified as a storm event when H_s and T_p are, respectively, greater than 3.5 m and 13.8 s. Such events correspond to 7.24% of the offshore wave climate [31] and are responsible for several overwash events each year.

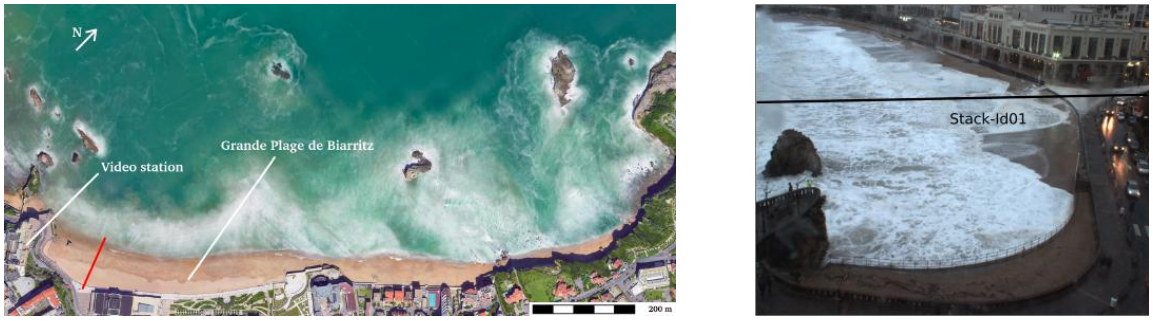
This site has been equipped with a coastal video monitoring station since 2017. The station includes 4 cameras with different lenses to ensure the coverage of the entire beach with a sufficient spatial resolution. The cameras are operated by the open source software SIRENA [12]. For this site, one transect is monitored by the camera pointing to the beach and seawall location (transect Stack-Id01 in Figure 3). The timestack images correspond to pixel intensities recorded along this transect over 14 min with a sampling frequency of 1Hz. Among the 70,000 images of this database, only 8172 images were kept to be part of the ground truth dataset. Indeed, the timestacks generated in summer months were excluded as the human activities negatively affect the quality of the images. The images where the tide level was below 2.8 m were excluded as they corresponded to timestacks images without visible swash.

2.1.2. Timestack Image Preprocessing

The ground truth dataset was built by labeling the 8172 images. There are two methods to annotate the images: by hand or in a semi-automatic way. The annotation by hand is the most straightforward but also the most time-consuming method. The semi-automatic method consists of two steps. First, the position of the waterline is extracted automatically by segmenting the image using Otsu's thresholding method [16]. Then, the storm impact regime is identified by comparing the position of the waterline with the one of the defense infrastructure. This method is faster than the annotation by hand; however, it still requires an operator because it is not always robust and highly depends on the lightning conditions

of the image. To employ this method, the position of the defense infrastructure in the image must be known. This is the case for the Grande Plage de Biarritz; therefore, semi-automatic annotation was performed.

After verification and correction by an operator, the result of the annotation was 7907/211/54 (Swash/Collision/Overwash). The classes are highly imbalanced, and this could have some effect on the classification accuracy of the CNN. Methods to deal with this problem are presented below. Before the training process, the images were resized to fit to the input dimensions of the CNNs tested in this study (224 × 224).



(**a**) (**b**)

Figure 3. (**a**) Satellite view of the site of Biarritz from Google Earth with red lines representing the transects on the site. (**b**) Transect on the Grande Plage de Biarritz.

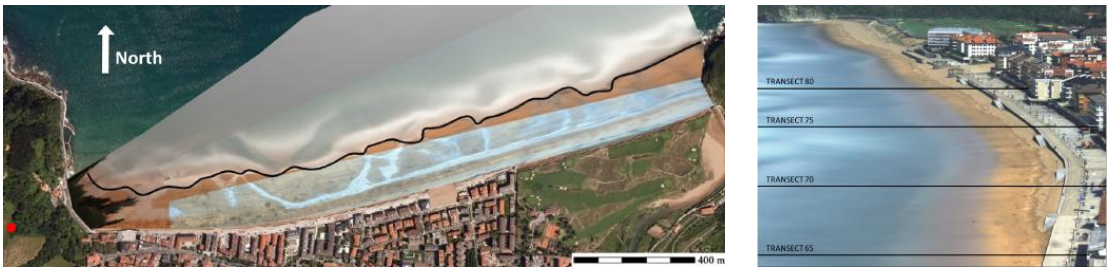
2.2. Zarautz

2.2.1. Site Characteristics

The beach of Zarautz is an embayed beach of 2.3 km long located on the Basque coast (northern Spain) in the SE Bay of Biscay, approximately 70 km southwest of GPB. The beach, facing north (345 degrees), can be divided into two parts (Figure 4): 30% of the beach in the eastern part presents a large and well-preserved dune system, with a maximum height of approximately 10 m above the minimum astronomic tide. The remaining 70% is an engineered urban beach, backed by a concrete seawall and the village of Zarautz.

In terms of characteristics, the beach of Zarautz is an intermediate-dissipative [32] and mesotidal beach with a 4 m spring tidal range. It is composed of fine–medium sand with a mean slope of around 2%. The annual average significant wave height and peak period are, respectively, 1 m and 9 s. Like the GPB, the beach of Zarautz is also exposed to highly energetic waves and storms coming from the WNW and NW directions. The seawall backing 70% of the beach has an along-shore elevation varying between 6.5 m in the western part and 8 m in the center of the beach. This seawall serves as a defense infrastructure for the buildings near the beach, and overtopping events are common at high tide during winter storms.

A video monitoring station, like the one used on the GPB site was installed in 2010. The station has 4 cameras of 1.4 Megapixels. Two of the cameras are equipped with 12 mm lens and have a panoramic view, and another 2 equipped with 25 mm lens cover with more resolution the mean high and low tide coastline positions. For the Zarautz dataset, 4 transects are monitored by the camera covering the supra-tidal beach with higher resolution (Figure 4). The transects are perpendicular to the seawall and are named corresponding to the elevation of the seawall in the point of intersection (i.e., transect 65 corresponds to the part of the seawall with 6.5 m elevation).



(**a**) (**b**)

Figure 4. (**a**) Satellite view of the site of Zarautz from Google Earth, with the red point representing the location of the video monitoring station. (**b**) Positions of the transects on the site of Zarautz.

2.2.2. Timestacks Images Preprocessing

Images from the site of Zarautz were annotated by hand. This method of annotation was preferred over semi-automatic annotation because: (i) the position of the seawall varied between timestack images, making the semi-automatic method more laborious, and (ii) the presence of strong winds and gust negatively impacted the quality of certain images, making the semi automatic method less robust. A simple web application was developed to facilitate the annotation for the operator and is accessible in a public GitHub repository (link in Data Availability Statement section). After classification by hand, the result of the annotation was 19,596/2776/162 (Swash/Collision/Overwash). Like the images of Biarritz, images of Zarautz present class imbalance, and they were resized to fit to the input dimensions of the CNNs before the training.

3. Convolutional Neural Networks

3.1. General Concept

CNNs are a type of neural networks widely used to perform tasks related to imagery analysis such as image segmentation, classification, or object detection. For classification problems, a CNN takes as input images with three channels (RGB), from which they output probabilities of belonging to specified categories, in our case storm impact regimes. Like a classical neural network, a CNN is a stacking of neurons that are organized in different layers. The structure of a CNN can be divided into two parts. The first part contains mostly convolutional and pooling layers and aims to learn specific features that help to classify the images correctly. The second part contains fully connected layers and the output layer. It uses the specific features extracted in the first part to output probabilities of belonging to specified categories.

In the feature extraction part, the convolutional layers detect features inside an image. They convolve their input with one or more filters, which results in one or more feature maps (one for each filter). The feature maps represent the activation of a specific filter at every spatial position of the input image. During the learning process, the network will learn filters that activate when they see specific visual features that help to correctly classify the training images. Usually, convolutional layers are stacked inside a CNN. The early layers detect simple features such as edges, whereas the deeper layers can detect more complex features.

Pooling layers are commonly found between convolutional layers. These layers also rely on convolutional operations and aim to reduce the dimensionality of the feature maps in order to increase the learning speed of the network and to control the overfitting of the CNN. If a CNN is overfitted, it would indicate that the network has learned exactly the characteristics of the training images and cannot generalize to new data. By stacking several convolutional and pooling layers inside a CNN, the complexity of the extracted information increases as we go deeper in the network with more feature maps with smaller dimensions.

The output of these specific layers serves as input to the second part of the network, which aims to classify the image into the correct category. In the classification part, neurons are organized in layers and are connected to the previous layers through weights (hence the name fully connected layers). To prevent overfitting, drop-out regularization can be applied on these layers. This method randomly ignores neurons during the training process, making the network learn more sparse and robust representation of the data. Finally, the output layer estimates the probabilities of belonging to the specified categories for the input image with a "softmax" activation function.

The CNNs are trained with backpropagation in the same manner as classical neural networks: the weights in the convolutional and fully connected layers are updated iteratively to minimize the errors between the prediction of the network and the ground truth. The ground truth dataset for such a network is made by annotating images. Details on the annotation of the timestacks can be found in the "Study Sites" section. Only the general ideas about CNN have been presented above; for a detailed description on CNN and their training, the reader is referred to the work of Bengio et al. [33].

There are many CNN architectures, each with different complexity and characteristics. In order to keep the computation time reasonable, it was decided to limit the comparison of performances between four architectures of increasing depth and complexity:

- A custom architecture inspired by the work of LeCun et al. [34] adapted for bigger images. The architecture is presented in the appendix (Table A1).
- AlexNet [35], which won the ImageNet challenge in 2012. Its architecture contains more convolutional layers and dense layers (Table A2). The number of filters is also larger than that of the custom architecture.
- VGG16 [36], which is a very deep CNN that uses 13 convolutional layers and three dense layers (Table A3).
- Inception v3, an improved version of the GoogleNet from Szegedy et al. [37] which won the ILSVRC in 2014. It relies on inception modules, which perform convolutions with filters of multiple size and concatenate their results (Table A4). In addition, the convolution operation with filters of large size inside an inception module are made by using $1 \times n$ filters to reduce computational cost. This results in deeper networks with significantly fewer parameters to learn.

3.2. Training the CNN

3.2.1. Data Processing

The datasets of both sites were divided into training, validation and testing sets containing, respectively, 65%, 15%, and 20% of the data (common proportions in the literature). Stratified random sampling was used to ensure that each part contains the same class proportions. The training set is used to fit the CNN. The validation set is used to stop the training for the CNN (early stopping). At last, the test part is used to evaluate the performance of the neural network on unseen data (not used in the training step).

During the training, each training image is seen multiple times by the CNN. This can be a problem as the network can learn exactly the characteristics of the training images and might not generalize to new data. To avoid this problem, called overfitting, data augmentation is employed during the training of the CNN. This method consists in making small changes to images in the training set before feeding them to the CNN. By generating modified images, this method artificially increases the number of images in the minority classes and makes the models more robust to overfitting.

The following changes were made:

- Random vertical flip: new timestack with inverted time;
- Random shift in the RGB image color to decrease the dependence on lighting conditions;
- Normalization of pixel values to 0–1 for faster training

3.2.2. Class Imbalance Problem

The datasets from both sites suffer from the class imbalance problem. Indeed, the distributions of storm impact regimes are highly imbalanced. For the Biarritz site, 96.8% of the images display swash regimes, 2.6% collision regimes, and 1% of overwash regimes. For Zarautz, 87% of the images are swash regimes, 12% collision regimes, and 1% overwash regimes. This class imbalance problem was expected as we are studying rare events.

It has been proven that class imbalance can negatively affect the performances of machine learning models in classification tasks [38]. Methods to deal with this problem are well known [38–40] and can be divided into two categories: data-level methods and classifier-level methods.

The data-level methods aim to modify the class distribution in order to reduce the imbalance ratio between classes. The most popular methods in this category are oversampling and undersampling. Oversampling consists in replicating random samples from minority classes until the imbalance problem is removed. In contrast, undersampling consists in removing random samples from the majority class until the balance between classes is reached.

The classifier-level methods aim to modify the training or the output of the machine learning algorithm. They include cost-sensitive learning, which is a method that gives more weights during learning to examples belonging to minority classes, and the thresholding method, which adjusts the output probabilities by taking into account the prior class probabilities [39].

3.2.3. Transfer Learning

For complex and deep CNN, it is common to use transfer learning to speed up the learning process and to improve performances. Transfer learning methods consist in using knowledge gained on a specific task to solve a different task. There are different methods of transfer learning for CNN; for an exhaustive listing, readers are referred to the work of Pan and Yang [28]. The method used in this article is "pre-training". It consists of using the weights of a CNN trained on a first task as initialization weights for a second CNN that will perform on a second task. The efficiency of pre-training was tested by using the pre-trained weights of VGG16 and Inceptionv3 on the ImageNet dataset, which is one of the largest labeled image dataset [41]. Then, transfer learning was performed between sites to see if the knowledge gained on one site is beneficial for the learning on the second site.

3.2.4. Application to the Datasets

The workflow for this study is presented in Figure 5. For each site, the four CNNs with different architectures were fitted without and with the two methods related to the class imbalance problem: oversampling and cost-sensitive learning (class weights). Transfer learning was used on the more complex architectures (VGG16, Inceptionv3) and only for the best performing method to cope with class imbalance. Data augmentation was used during the training of all the CNNs.

The networks were trained on a laptop equipped with a GPU (Quadro RTX 4000) using Keras (tensorflow GPU 1.12.0/Keras 2.3.1/Python 3.6.1), an open-source python library designed for building and training neural networks. The scripts used in this article are available on a public GitHub repository (link in Data Availability Statement section). The optimizer used is Mini-Batch gradient descent algorithm with batch size of 32 and a learning rate of 0.001 that decays by a factor of 2 every 10 epochs. The training is stopped at 100 epochs or earlier when the value of the validation loss does not decrease over 10 epochs (early stopping).

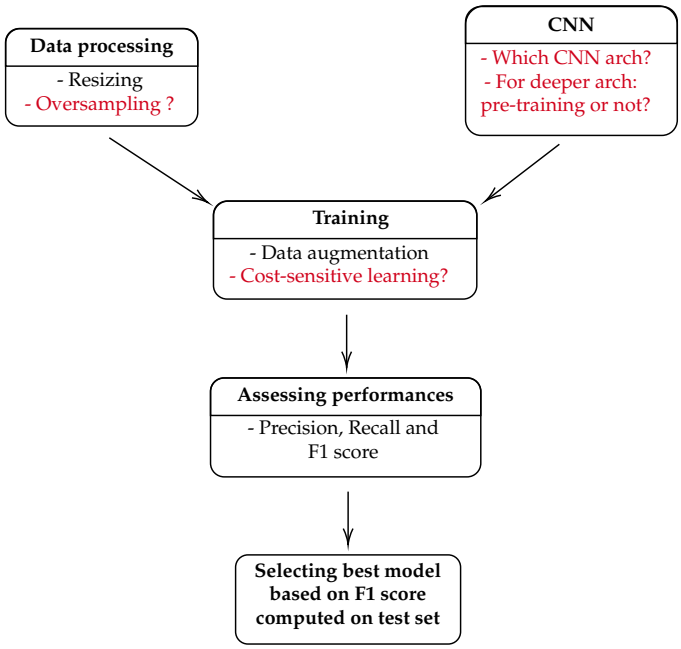


Figure 5. Workflow for this study. Items inside the boxes that are highlighted in red represent the choices tested in this study, whereas the items in black are the methods applied in every cases.

3.3. CNN Accuracy Assessment

To compare the performance of the different networks, the F_1-score is computed with the following formula:

$$F_1 = \frac{Precision \times Recall}{Precision + Recall}$$

where

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

and

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}.$$

The precision, recall, and F_1-score are computed for each storm impact regime and are averaged in order to have one global metric for each CNN. The F_1-score varies between 0 and 1, with 1 representing the best value. Unlike the global accuracy (number of correct predictions divided by the total number of predictions), the F_1 metric is not biased when data present a class imbalance.

4. Results

The results are organized into four subsections. Firstly, the performances of the different combinations of CNN architectures, methods to cope with class imbalance, and transfer learning are compared. Secondly, the prediction errors of the best CNN for each site are investigated. Thirdly, we present results related to transferability between sites. Finally, a sensitivity analysis is presented for the site of Zarautz.

4.1. CNN Performances

4.1.1. Architectures

Table 1 regroups the training time, number of epochs, and also the performance metrics (accuracy, recall, F_1-score) for different CNN architectures, methods to cope with

class imbalance, and with or without pre-training with ImageNet dataset for both sites. For both sites and for every methods used to cope with class imbalance (or not), CNNs with deeper and more complex architectures yielded better results (higher values of precision, recall, and F_1-score). Indeed, these kind of architectures tend to learn more complex features that lead to better performance in harder tasks. The downside of these networks is the training time, which is significantly higher than simpler and shallower models.

Table 1. CNN performances for both sites. Best models are in bold font.

(a) Biarritz						
CNN	**Training Time (min)**	**Epochs**	**Time per Epoch (s)**	**Precision**	**Recall**	**F_1-Score**
Baseline						
Custom CNN	16.1	100	9.7	/	0.333	0.328
AlexNet	17.2	100	10.3	/	0.333	0.328
VGG16	81.4	89	54.9	/	0.481	0.476
Inception v3	40.0	69	34.8	0.721	0.714	0.713
Class weights						
Custom CNN	4.6	28	9.9	/	0.603	0.474
AlexNet	3.8	21	10.9	0.568	0.777	0.609
VGG16	43.4	46	56.6	0.574	0.832	0.645
Inception v3	23.2	39	35.8	0.563	0.798	0.631
Oversampling						
Custom CNN	10.6	26	24.6	0.642	0.880	0.718
AlexNet	11.8	27	26.1	0.716	0.885	0.777
VGG16	69.6	28	149.1	0.783	0.851	0.813
VGG16 Transfer	**49.9**	**20**	**149.6**	**0.869**	**0.865**	**0.866**
Inception v3	59.6	38	94.1	0.679	0.767	0.717
Inception v3 Transfer	34.5	21	98.6	0.777	0.786	0.780
(b) Zarautz						
CNN	**Training Time (min)**	**Epochs**	**Time per Epoch (s)**	**Precision**	**Recall**	**F_1-Score**
Baseline						
Custom CNN	22.5	49	27.5	/	0.637	0.616
AlexNet	24.0	48	30.0	/	0.628	0.616
VGG16	202.1	72	168.4	/	0.635	0.617
Inception v3	108.7	64	101.9	/	0.630	0.614
Class weights						
Custom CNN	11.6	26	26.7	0.666	0.846	0.720
AlexNet	22.7	45	30.3	0.671	0.817	0.716
VGG16	81.9	30	163.7	0.680	0.844	0.732
Inception v3	89.3	53	101.1	0.654	0.838	0.710
Oversampling						
Custom CNN	38.7	36	64.5	0.769	0.804	0.783
AlexNet	22.6	19	71.3	0.756	0.797	0.775
VGG16	146.6	22	399.8	0.775	0.812	0.792
VGG16 Transfer	**86.5**	**13**	**399.1**	**0.897**	**0.834**	**0.858**
Inception v3	97.7	24	244.2	0.777	0.801	0.784
Inception v3 Transfer	65.3	16	245.0	0.869	0.835	0.849

4.1.2. Class Imbalance

Without coping with class imbalance problem, CNNs tend to predict all the images as the majority class, resulting in poor classification results. Between the two methods tested, oversampling seems to perform better, with F_1-scores on average 30% better than the ones obtained with cost-sensitive learning (class weights). The superior performance of oversampling method on this dataset might be due to the fact that the CNNs see more

images during training in the oversampling method than in the cost sensitive learning method, resulting in better classification accuracy.

4.1.3. Pre-Training

Finally, models using pre-trained weights (transfer learning) train faster (fewer epochs) and yield better classification results than models trained from scratch. Indeed, the F1-scores obtained with the pre-trained models are, respectively, 6 to 8% higher for the VGG16 and Inceptionv3 models. Even though the images from the ImageNet dataset have different characteristics than the timestack images that are being classified, the pre-trained weights might contain knowledge about general features that are helpful to better classify the timestacks.

4.1.4. Best Models

For GPB, the best model is the pre-trained VGG16 with an F1-score of 0.866. The pre-trained Inception model trains faster but shows a lower F1-score (0.780). For the Zarautz site, the best model is also the pre-trained VGG16 with an F1-score of 0.858, but this time the performance of the pre-trained Inception v3 model was very close with an F1-score of 0.849.

4.2. Investigating the Errors

The confusion matrices on the test sets are presented in Table 2. In general, the minority classes tended to have higher error rate. This is expected as minority classes contain fewer examples than majority classes.

Table 2. Confusion matrices obtained by the best models for both sites.

(a) Biarritz (best model: OV VGG16 Transfer)				
			Predicted	
		Swash	**Collision**	**Overwash**
Observed	Swash	1576	7	0
	Collision	4	34	2
	Overwash	1	2	9
(b) Zarautz (best model: OV VGG16 Transfer)				
			Predicted	
		Swash	**Collision**	**Overwash**
Observed	Swash	4265	40	0
	Collision	13	617	8
	Overwash	0	25	30

The prediction errors made by the CNN were manually inspected to gain an understanding of common error types. Prediction errors made on the GPB test set are presented in Figure 6 and in the appendix (Table A5). Among the 16 errors made on the test set, five errors came from human misclassification, five errors may have been caused by the presence of specific features such as vertical lines usually associated with collision regimes, two errors were made on images that were displayed in between category of storm impact regimes. The remaining errors correspond to images that were hard to classify, where lighting and meteorological conditions were poor. The misclassification errors were corrected for the test and validation sets, and the best network was trained once again, resulting in slightly lower results but this time without human misclassification error (Table 3).

Table 3. Classification results after correcting the misclassification in the test and validation sets.

CNN	Time (min)	Epochs	Time per Epoch (s)	Precision	Recall	F1 Score
Biarritz						
Best model before corr.	49.9	20	149.6	0.869	0.865	0.866
Best model after corr.	60	24	150	0.895	0.833	0.860
Zarautz						
Best model before corr.	86.5	13	399.1	0.897	0.834	0.858
Best model after corr.	88.5	13	408.5	0.917	0.859	0.883

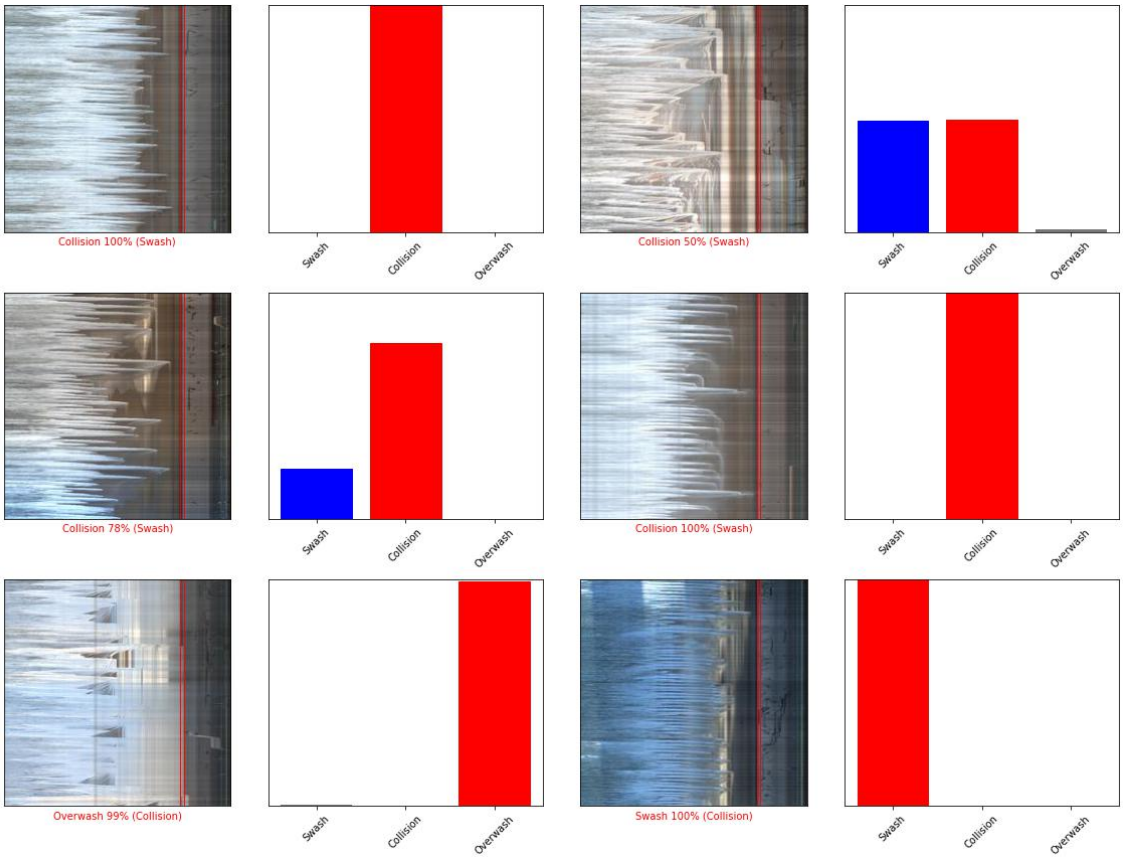


Figure 6. Prediction errors on the test set of Biarritz by the best CNN. The storm impact regime predicted is written under each timestack, and the ground truth is written in parentheses. The probabilities of belonging predicted by the CNN are represented on the right side of each timestack (red = prediction made by CNN, blue = ground truth). The red line in the timestack represents the position of the seawall.

The errors made on the Zarautz dataset were also analyzed (Table A6). A large number of errors were made on images that were in between the categories of storm impact regimes: either the images displayed a swash regime, which was very close to the collision regime, or the images displayed a regime impact, with one small overtopping of the wall. Some misclassification errors were made. The rest of the errors may have come from lighting conditions (large horizontal band, lighter in the images). The misclassification errors were corrected for the test and validation, and the best network was trained once again, resulting in a slightly better result for this site (Table 3).

4.3. Transferability between Sites

The interest in transfer learning for CNN training has been highlighted in Section 4.1.3. The models using pre-trained weights from ImageNet trained faster (fewer epochs) and yielded better classification results. In this section, we investigate if the knowledge acquired on one site (CNN weights) could be transferred to another site with pre-training. Pre-training is the most common way to transfer knowledge between tasks. It consists in using the weights of a neural network trained on one site as initialization weights for the training on the second site. The weights of the best CNN for each site are used for the pre-training of a CNN on the other site. The performances of these CNN are presented in Table 4.

Table 4. Performances of CNN learning from scratch, pre-trained with ImageNet, or pre-trained with the other site for Biarritz and Zarautz. "OV" stands for oversampling.

CNN	Time (min)	Epochs	Time per Epoch (s)	Precision	Recall	F_1-Score
Biarritz						
VGG16 (OV)	69.6	28	149.1	0.783	0.851	0.813
VGG16 (OV) Pretraining with ImageNet	49.9	20	149.6	0.869	0.865	0.866
VGG16 (OV) Pretraining with Zarautz data	47	19	148.4	0.826	0.832	0.823
Zarautz						
VGG16 (OV)	81.9	30	163.7	0.680	0.844	0.732
VGG16 (OV) Pretraining with ImageNet	86.5	13	399.1	0.897	0.834	0.858
VGG16 (OV) Pretraining with Biarritz data	92	14	394.2	0.909	0.867	0.885

The weights of the best model on Zarautz data (VGG16 transfer) were used as initialization weights for the training on Biarritz data. This resulted in classification results better than the learning from scratch with a higher precision and F_1-score (Table 4). However, the values of precision, recall, and F_1-score obtained with pre-training on Zarautz data remained slightly lower than the ones obtained with pre-training on ImageNet data.

Pre-training method was also applied on Zarautz data, where the weights of the best model on the Biarritz site were used as initialization weights. The classification results were better than learning from scratch and learning with pre-trained weights from ImageNet with higher F_1-score (Table 4).

4.4. Sensitivity Analysis

A sensitivity analysis was performed on the dataset of Zarautz to highlight the effect of the size of the training images dataset on the classification accuracy. The dataset of Zarautz was divided into three smaller datasets. Each of these datasets was divided into the training/validation/test sets with the proportions described in Section 3.2.1. Finally, a CNN model was trained on each smaller datasets (VGG16 with transfer learning ImageNet and oversampling).

The averaged F1-metric for these three models was 0.805. This value is slightly lower than the one obtained with the full dataset, which is 0.858 (Table 1). These results confirm what was already known in the literature: CNN performances tend to increase with the training set size [42].

5. Discussion

Even though we showed the strong potential of CNN to automatically generate storm impact regime database, the proposed methodology can be improved in several ways. More attention could be paid to the choice of CNN architectures and hyperparameters. Other CNN architectures need to be tested, especially recent architectures such as ResNet, MobileNet, or Xception. They could perform better than the architectures presented in

this work. For instance, the ResNet architecture contains skip connections between layers, which allows the training of much deeper and performant networks. Hyperparameters are parameters whose values are specified by the user before the training process begins; they affect the structure of a CNN and how well it trains. They have a non negligible impact on the final results. Several optimization algorithms such as Bayesian optimization could be employed to select the optimal hyperparameters [43], which has not been done in this study.

In addition to hyperparameter tuning, other methods of data augmentation could be used to improve the performances of the CNN. The analysis of prediction errors can help in the choice of other data augmentation methods. In our case, many errors were related to lighting conditions; it would be wise to test various data augmentation methods affecting the lighting or brightness of the images. This could make the CNN more robust to lighting conditions and therefore improve its performances.

It is worth noting that the performances of a CNN model implemented at a given site are expected to increase with time as more timestacks are collected by the video monitoring system. With more training images, the minority classes will contain more images, and this will lead to less classification errors for these classes. Moreover, if enough timestacks are collected, intermediate storm impact regime classes could be created. These classes could reduce the errors on the images displaying impact regimes not corresponding to the three regimes presented in this work.

One very interesting feature of CNNs models is their transferability. We showed that using the knowledge acquired from another site can lead to improved classification results when using pre-training (especially for Zarautz site). The weights of the best CNNs for both sites are available in a GitHub repository (link in Data Availability Statement section) and could be used as initialization weights for a CNN applied to a new site. The only requirement is to annotate timestacks from the new site, which will serve as training data.

Despite the promising performance, this methodology has some limitations, mainly related to the image annotation, an obligatory step for CNN training. The first limitation of this method is the lack of knowledge about its sensitivity. We showed for the site of Zarautz that CNNs yield lower performances when trained on a smaller training set. However, we do not know the minimum number of timestacks to annotate for a new site in order to have satisfactory accuracy. A sensitivity analysis should be performed to find this minimum threshold and to make some recommendations on the use of this method in the case of new sites with a small number of timestacks.

The second limitation of this method is the annotation process itself, which is tedious and time-consuming. An alternative solution could be to use the domain-adaptation approach presented in the work of Ganin and Lempitsky [44]. They propose a specific CNN architecture that can be trained simultaneously on a large number of labeled data from a source domain (one site) and unlabeled data from a target domain (new site). At the end of the training, the CNN is able to classify correctly images from both sites even though only images from one site have been labeled.

Finally, the performances of the proposed method must be compared objectively with human-level performance and other methodologies. Assessing the human-level performance on this task is essential and would give precious insights into how to improve further the CNN performances [45]. For example, a CNN performance lower than the human-level performance could indicate the presence of a bias, which can be avoided by using deeper models or by training more slowly and for longer. It would be of great interest to compare this methodology based on CNNs with methodologies based only on traditional imagery analysis. As stated in the introduction, a possible methodology could be to first extract the waterline position using Otsu's segmentation [16,26] or using the radon transform [27] and then compare its position with the position of defense infrastructure to define the storm impact. Another methodology could be based on the analysis of pixel intensity such as the works of Simarro et al. [46] and Andriolo et al. [47]. The methodologies based on simple image processing algorithms could have some advantages over CNNs.

Indeed, they would not require the building and training of a CNN structure, which is time-consuming, and the whole decisional process is known to be contrary to CNN, which can be considered as a "black box". However, these methodologies would need to be adapted for each site by indicating the position of the defense infrastructure, which is not needed with CNN. In addition, the simple image processing algorithms could be more affected by the erratic brightness of the timestacks than CNNs, which are trained with data augmentation.

This work is a first step in the analysis of storm impact with video monitoring. Numerous extensions can be envisaged, particularly on the type of information extracted and the type of image analyzed. Indeed, the CNN could be used to count the number of collision or overwash events in one timestack. This technique could be also extended to analyze other types of images produced by video monitoring systems such as oblique and/or rectified images. Finally, it can be employed to analyze images from already existing cameras such as surfcam [48,49]. This could constitute a low-cost monitoring method with a large spatial coverage for the qualitative study of storm impact. Many questions arose with this work, especially about the minimum number of images to annotate to have satisfactory accuracy or the lack of comparison with the current method or human level performance. More questions will arise during the operational implementation and use of the CNNs concerning the verification of predictions, the prediction error handling, or how often we need to re-train the neural networks with the newly classified images.

6. Conclusions

In this paper, we presented an innovative methodology based on convolutional neural networks and coastal imagery that could be used to collect storm impact data routinely. We described the methodologies associated with CNNs, including the annotation of the dataset, the training of the networks, or transfer learning. We also introduced the problem of class imbalance, which is due to the extreme nature of the storm impact regimes, and we proposed and compared different solutions such as oversampling or cost-sensitive learning.

The proposed methodology was tested on two sites: Biarritz and Zarautz. We showed that convolutional neural networks are well adapted for the classification of timestacks into storm impact regimes. Overall, we found that more complex and deeper architectures yielded better results. Best performances were achieved with the VGG16 architecture for both sites with F-scores of 0.866 for the site of Biarritz and 0.858 for the site Zarautz. For the class imbalance problem, the method of oversampling showed better classification accuracy than the cost-sensitive learning method, with F-scores on average 30% higher. Finally, we showed that the method can be easily applied to a new site with optimal efficiency using transfer learning. Indeed, training a CNN using pre-trained weights (ImageNet or weights of another site) resulted in better accuracy than training a CNN from scratch (F-scores on average 6 to 8% higher).

With convolutional neural networks, we can take full advantage of the large number of data produced by video monitoring systems. We showed that they are able to transform images into usable qualitative data about storm impact. Even if the data are not continuous (only day time and winter months), this method could be, without a doubt, a real asset in the future for coastal researchers and stakeholders by routinely collecting storm impact data, which are rare at present. These data are essential in the disaster risk reduction chain, and they have many uses. They can serve as validation data for impact models or early warning systems based on numerical modeling. They can also be used to train early warning system based on Bayesian networks [50,51]. Finally, statistical analysis can be performed to find relationships between observed storm impact regimes and local conditions such as wave characteristics, tide, or meteorological conditions.

Author Contributions: Conceptualization, A.C., D.M. and B.L.; methodology, A.C. and B.L.; software, A.C.; validation, A.C.; formal analysis, A.C.; investigation, A.C.; resources, D.M., P.L. and I.E.; data curation, A.C.; writing—original draft preparation, A.C.; writing—review and editing, A.C., D.M., P.L. and B.L.; visualization, A.C.; supervision, D.M. and B.L.; project administration, D.M. and B.L.;

funding acquisition, D.M. and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was provided by the Energy Environment Solutions (E2S UPPA) consortium and the BIGCEES project from E2S-UPPA ("Big model and Big data in Computational Ecology and Environmental Sciences").

Data Availability Statement: The images used to train the CNN are not publicly available due to the large size of the files. SIAME laboratory owns the Biarritz images data; they can be provided upon request. Images of Zarautz's site used in this work are from Azti and the Zarautz Town Council (shared). They can also be provided upon request. The python scripts and weights of the best CNN are available here: https://github.com/AurelienCallens/CNN_Timestacks, accessed on 1 May 2021. The web application used to label the images of Zarautz site (R programming language) is available here: https://github.com/AurelienCallens/Shiny_Classifier, accessed on 1 May 2021.

Acknowledgments: The authors gratefully acknowledge European POCTEFA Program funding under the research project MARLIT EFA344/19 and the complementary program of OCA. The authors would like to thank Zarautz Town Council and the Directorate of Emergencies and Meteorology of the Basque Government for their continuous support of Zarautz's video station.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural networks
GBP	Grande Plage of Biarritz

Appendix A. CNN Architectures

Appendix A.1. Custom CNN

Table A1. Architecture of the "custom" CNN.

Layer (Type)	Output Shape	Param
Block1 Conv (Conv2D)	(None, 111, 111, 32)	896
Block1 Pool (MaxPooling2D)	(None, 55, 55, 32)	0
Block2 Conv2d (Conv2D)	(None, 53, 53, 64)	18,496
Block2 Pool (MaxPooling2D)	(None, 26, 26, 64)	0
Block3 Conv2d (Conv2D)	(None, 24, 24, 128)	73,856
Block4 Pool (MaxPooling2D)	(None, 12, 12, 128)	0
Block5 Conv2d (Conv2D)	(None, 10, 10, 256)	295,168
Block5 Pool (MaxPooling2D)	(None, 5, 5, 256)	0
Flatten (Flatten)	(None, 6400)	0
Dense1 (Dense)	(None, 512)	3,277,312
Dropout1 (Dropout)	(None, 512)	0
Dense2 (Dense)	(None, 256)	131,328
Dropout2 (Dropout)	(None, 256)	0
Dense3 (Dense)	(None, 128)	32,896
Dropout3 (Dropout)	(None, 128)	0
Output (Dense)	(None, 3)	387

- Total params: 3,830,339
- Trainable params: 3,830,339
- Non-trainable params: 0

Appendix A.2. AlexNet

Table A2. Architecture of the AlexNet.

Layer (Type)	Output Shape	Param
Block1 Conv (Conv2D)	(None, 54, 54, 96)	34,944
Block1 Pool (MaxPooling2D)	(None, 27, 27, 96)	0
Block2 Conv (Conv2D)	(None, 17, 17, 256)	2,973,952
Block2 Pool (MaxPooling2D)	(None, 8, 8, 256)	0
Block3 Conv (Conv2D)	(None, 6, 6, 384)	885,120
Block3 Conv (Conv2D)	(None, 4, 4, 384)	1,327,488
Block4 Conv (Conv2D)	(None, 2, 2, 256)	884,992
Block4 Pool (MaxPooling2D)	(None, 1, 1, 256)	0
Flatten (Flatten)	(None, 256)	0
Dense1 (Dense)	(None, 4096)	1,052,672
Dropout1 (Dropout)	(None, 4096)	0
Dense2 (Dense)	(None, 4096)	16,781,312
Dropout2 (Dropout)	(None, 4096)	0
Output (Dense)	(None, 3)	12,291

- Total params: 23,952,771
- Trainable params: 23,952,771
- Non-trainable params: 0

Appendix A.3. VGG16

Table A3. Architecture of the CNN used based on VGG16.

Layer (Type)	Output Shape	Param
Input (Input Layer)	(None, 224, 224, 3)	0
Block1 Conv1 (Conv2D)	(None, 224, 224, 64)	1792
Block1 Conv2 (Conv2D)	(None, 224, 224, 64)	36,928
Block1 Pool (MaxPooling2D)	(None, 112, 112, 64)	0
Block2 conv1 (Conv2D)	(None, 112, 112, 128)	73,856
Block2 Conv2 (Conv2D)	(None, 112, 112, 128)	147,584
Block2 Pool (MaxPooling2D)	(None, 56, 56, 128)	0
Block3 Conv1 (Conv2D)	(None, 56, 56, 256)	295,168
Block3 Conv2 (Conv2D)	(None, 56, 56, 256)	590,080
Block3 Conv3 (Conv2D)	(None, 56, 56, 256)	590,080
Block3 Pool (MaxPooling2D)	(None, 28, 28, 256)	0
Block4 Conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
Block4 Conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
Block4 Conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
Block4 Pool (MaxPooling2D)	(None, 14, 14, 512)	0
Block5 Conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
Block5 Conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
Block5 Conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
Block5 Pool (MaxPooling2D)	(None, 7, 7, 512)	0
Flatten (Flatten)	(None, 2048)	0
Dense1 (Dense)	(None, 512)	262,656
Dropout1 (Dropout)	(None, 512)	0
Output (Dense)	(None, 3)	1539

- Total params: 14,978,883
- Trainable params: 14,978,883
- Non-trainable params: 0

Appendix A.4. Inception v3

Table A4. Architecture of the CNN used based on Inception v3.

Layer (Type)	Output Shape	Param
Inceptionv3 (Model)	(None, 2048)	21,802,784
Flatten (Flatten)	(None, 2048)	0
Dense1 (Dense)	(None, 512)	1,049,088
Dropout1 (Dropout)	(None, 512)	0
Output (Dense)	(None, 3)	1539

The inception model was imported with Keras with the following function: `keras.applications.InceptionV3()`.

The architecture is not displayed due to readability; the reader is referred to the original work of Szegedy et al. [37] for more details.

- Total params: 22,853,411
- Trainable params: 22,818,979
- Non-trainable params: 34,432

Appendix B. Investigating the Errors

Table A5. Errors explanation for Biarritz data. "Misclass." stands for misclassification during annotation, "Splash" corresponds to an intermediate storm regime between impact and overwash, "Vertical" corresponds to the presences of vertical features of runup.

Test					
Splash	Lighting	Misclass.	Hard to classify	Vertical	
2	1	5	3	5	
Validation					
Splash	Misclass.	Sand bags ?	Splash	Hard to classify	Vertical
1	2	1	2	2	4

Table A6. Errors explanation for Zarautz data. "Misclass." stands for misclassification during annotation, "Splash" corresponds to an intermediate storm regime between impact and overwash, "Vertical" corresponds to the presences of vertical features of runup. Finally, "SI" corresponds to an intermediate storm regime between swash and impact that was very close to the sea wall but did not impact.

Test						
Hard to classify	Lighting	Misclass.	SI	SI + Light	Splash	Vertical
10	30	7	22	2	14	1
Validation						
Hard to classify	Lighting	Misclass.	SI	SI + Light	Sandbag ?	Splash
5	27	3	26	2	4	4

References

1. Valchev, N.; Andreeva, N.; Eftimova, P.; Trifonova, E. Prototype of Early Warning System for Coastal Storm Hazard (Bulgarian Black Sea Coast). *CR Acad. Bulg. Sci.* **2014**, *67*, 977.
2. Van Dongeren, A.; Ciavola, P.; Martinez, G.; Viavattene, C.; Bogaard, T.; Ferreira, O.; Higgins, R.; McCall, R. Introduction to RISC-KIT: Resilience-increasing strategies for coasts. *Coast. Eng.* **2018**, *134*, 2–9. [CrossRef]
3. Sallenger, A.H.J. Storm Impact Scale for Barrier Islands. *J. Coast. Res.* **2000**, *16*, 890–895.
4. de Santiago, I.; Morichon, D.; Abadie, S.; Reniers, A.J.; Liria, P. A Comparative Study of Models to Predict Storm Impact on Beaches. *Nat. Hazards* **2017**, *87*, 843–865. [CrossRef]
5. Haigh, I.D.; Wadey, M.P.; Gallop, S.L.; Loehr, H.; Nicholls, R.J.; Horsburgh, K.; Brown, J.M.; Bradshaw, E. A user-friendly database of coastal flooding in the United Kingdom from1915 to 2014. *Sci. Data* **2015**, *2*, 1–13. [CrossRef]
6. Garnier, E.; Ciavola, P.; Spencer, T.; Ferreira, O.; Armaroli, C.; McIvor, A. Historical analysis of storm events: Case studies in France, England, Portugal and Italy. *Coast. Eng.* **2018**, *134*, 10–23. [CrossRef]
7. Abadie, S.; Beauvivre, M.; Egurrola, E.; Bouisset, C.; Degremont, I.; Arnoux, F. A Database of Recent Historical Storm Impact on the French Basque Coast. *J. Coast. Res.* **2018**, *85*, 721–725. [CrossRef]
8. André, C.; Monfort, D.; Bouzit, M.; Vinchon, C. Contribution of insurance data to cost assessment of coastal flood damage to residential buildings: Insights gained from Johanna (2008) and Xynthia (2010) storm events. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2003. [CrossRef]
9. Naulin, J.P.; Moncoulon, D.; Le Roy, S.; Pedreros, R.; Idier, D.; Oliveros, C. Estimation of insurance-related losses resulting from coastal flooding in France. *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 195–207. [CrossRef]
10. Ciavola, P.; Harley, M.; den Heijer, C. The RISC-KIT storm impact database: A new tool in support of DRR. *Coast. Eng.* **2018**, *134*, 24–32. [CrossRef]
11. Holman, R.A.; Stanley, J. The history and technical capabilities of Argus. *Coast. Eng.* **2007**, *54*, 477–491. [CrossRef]
12. Nieto, M.A.; Garau, B.; Balle, S.; Simarro, G.; Zarruk, G.A.; Ortiz, A.; Tintoré, J.; Álvarez-Ellacuría, A.; Gómez-Pujol, L.; Orfila, A. An open source, low cost video-based coastal monitoring system. *Earth Surf. Process. Landforms* **2010**, *35*, 1712–1719. [CrossRef]
13. Splinter, K.D.; Harley, M.D.; Turner, I.L. Remote sensing is changing our view of the coast: Insights from 40 years of monitoring at Narrabeen-Collaroy, Australia. *Remote Sens.* **2018**, *10*, 1744. [CrossRef]
14. Buscombe, D.; Carini, R.J. A data-driven approach to classifying wave breaking in infrared imagery. *Remote Sens.* **2019**, *11*, 859. [CrossRef]
15. Senechal, N.; Coco, G.; Bryan, K.R.; Holman, R.A. Wave runup during extreme storm conditions. *J. Geophys. Res. Ocean.* **2011**, *116*. [CrossRef]
16. Vousdoukas, M.I.; Wziatek, D.; Almeida, L.P. Coastal Vulnerability Assessment Based on Video Wave Run-up Observations at a Mesotidal, Steep-Sloped Beach. *Ocean. Dyn.* **2012**, *62*, 123–137. [CrossRef]
17. den Bieman, J.P.; de Ridder, M.P.; van Gent, M.R. Deep learning video analysis as measurement technique in physical models. *Coast. Eng.* **2020**, *158*, 103689. [CrossRef]
18. Stringari, C.E.; Harris, D.L.; Power, H.E. A Novel Machine Learning Algorithm for Tracking Remotely Sensed Waves in the Surf Zone. *Coast. Eng.* **2019**. [CrossRef]
19. Valentini, N.; Saponieri, A.; Molfetta, M.G.; Damiani, L. New algorithms for shoreline monitoring from coastal video systems. *Earth Sci. Inform.* **2017**, *10*, 495–506. [CrossRef]
20. Almar, R.; Cienfuegos, R.; Catalán, P.A.; Michallet, H.; Castelle, B.; Bonneton, P.; Marieu, V. A new breaking wave height direct estimator from video imagery. *Coast. Eng.* **2012**, *61*, 42–48. [CrossRef]
21. Andriolo, U.; Mendes, D.; Taborda, R. Breaking wave height estimation from Timex images: Two methods for coastal video monitoring systems. *Remote Sens.* **2020**, *12*, 204. [CrossRef]
22. Ondoa, G.A.; Almar, R.; Castelle, B.; Testut, L.; Leger, F.; Sohou, Z.; Bonou, F.; Bergsma, E.; Meyssignac, B.; Larson, M. Sea level at the coast from video-sensed waves: Comparison to tidal gauges and satellite altimetry. *J. Atmos. Ocean. Technol.* **2019**, *36*, 1591–1603. [CrossRef]
23. Holman, R.; Plant, N.; Holland, T. cBathy: A robust algorithm for estimating nearshore bathymetry. *J. Geophys. Res. Ocean.* **2013**, *118*, 2595–2609. [CrossRef]
24. Simarro, G.; Calvete, D.; Luque, P.; Orfila, A.; Ribas, F. UBathy: A new approach for bathymetric inversion from video imagery. *Remote Sens.* **2019**, *11*, 2722. [CrossRef]
25. Thuan, D.H.; Binh, L.T.; Viet, N.T.; Hanh, D.K.; Almar, R.; Marchesiello, P. Typhoon impact and recovery from continuous video monitoring: A case study from Nha Trang Beach, Vietnam. *J. Coast. Res.* **2016**, *75*, 263–267. [CrossRef]
26. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
27. Almar, R.; Blenkinsopp, C.; Almeida, L.P.; Cienfuegos, R.; Catalan, P.A. Wave runup video motion detection using the Radon Transform. *Coast. Eng.* **2017**, *130*, 46–51. [CrossRef]
28. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
29. Valentini, N.; Balouin, Y. Assessment of a smartphone-based camera system for coastal image segmentation and sargassum monitoring. *J. Mar. Sci. Eng.* **2020**, *8*, 23. [CrossRef]
30. Morichon, D.; de Santiago, I.; Delpey, M.; Somdecoste, T.; Callens, A.; Liquet, B.; Liria, P.; Arnould, P. Assessment of Flooding Hazards at an Engineered Beach during Extreme Events: Biarritz, SW France. *J. Coast. Res.* **2018**, *85*, 801–805. [CrossRef]

31. Abadie, S.; Butel, R.; Mauriet, S.; Morichon, D.; Dupuis, H. Wave Climate and Longshore Drift on the South Aquitaine Coast. *Cont. Shelf Res.* **2006**, *26*, 1924–1939. [CrossRef]
32. De Santiago, I.; Morichon, D.; Abadie, S.; Castelle, B.; Liria, P.; Epelde, I. Video monitoring nearshore sandbar morphodynamics on a partially engineered embayed beach. *J. Coast. Res.* **2013**, *65*, 458–463. [CrossRef]
33. Bengio, Y.; Goodfellow, I.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2017; Volume 1.
34. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
35. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
38. Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study. *Intell. Data Anal.* **2002**, *6*, 429–449. [CrossRef]
39. Buda, M.; Maki, A.; Mazurowski, M.A. A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef]
40. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 27. [CrossRef]
41. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009; pp. 248–255.
42. Uchida, S.; Ide, S.; Iwana, B.K.; Zhu, A. A further step to perfect accuracy by training CNN with larger data. In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 405–410.
43. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv* **2012**, arXiv:1206.2944.
44. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 1180–1189.
45. Dodge, S.; Karam, L. A study and comparison of human and deep learning recognition performance under visual distortions. In Proceedings of the 2017 26th International Conference on Computer Communication and Networks (ICCCN), Vancouver, BC, Canada, 31 July–3 August 2017; pp. 1–7.
46. Simarro, G.; Bryan, K.R.; Guedes, R.M.; Sancho, A.; Guillen, J.; Coco, G. On the use of variance images for runup and shoreline detection. *Coast. Eng.* **2015**, *99*, 136–147. [CrossRef]
47. Andriolo, U.; Sánchez-García, E.; Taborda, R. Operational use of surfcam online streaming images for coastal morphodynamic studies. *Remote Sens.* **2019**, *11*, 78. [CrossRef]
48. Mole, M.A.; Mortlock, T.R.; Turner, I.L.; Goodwin, I.D.; Splinter, K.D.; Short, A.D. Capitalizing on the surfcam phenomenon: A pilot study in regional-scale shoreline and inshore wave monitoring utilizing existing camera infrastructure. *J. Coast. Res.* **2013**, 1433–1438. [CrossRef]
49. Andriolo, U. Nearshore wave transformation domains from video imagery. *J. Mar. Sci. Eng.* **2019**, *7*, 186. [CrossRef]
50. Poelhekke, L.; Jäger, W.; van Dongeren, A.; Plomaritis, T.; McCall, R.; Ferreira, Ó. Predicting Coastal Hazards for Sandy Coasts with a Bayesian Network *Coast. Eng.* **2016**, *118*, 21–34. [CrossRef]
51. Plomaritis, T.A.; Costas, S.; Ferreira, Ó. Use of a Bayesian Network for Coastal Hazards, Impact and Disaster Risk Reduction Assessment at a Coastal Barrier (Ria Formosa, Portugal). *Coast. Eng.* **2018**, *134*, 134–147. [CrossRef]

4.3 Conclusion

This chapter showed the ability of deep learning methods, in particular Convolutional neural networks (CNN) to create automatically a storm impact database with images from video monitoring stations. We demonstrated that deeper CNN architectures associated with oversampling method yielded best classification results for two study sites which are Biarritz and Zarautz. This methodology is easily transferable on new sites by using the pre-trained weights of other sites. Even though the database constituted by this methodology is only qualitative, it can already be used to train or validate statistical models aiming to predict storm impact. In the future, CNN could be essential assets as the monitoring networks are expected to be more dense with new technological advances and lower cost of the components (Xu et al., 2019; Marcelli et al., 2021).

Many extensions of this work can be envisaged. As of now, this methodology only extracts maximum storm impact regime for a timestack image resuming 15 minutes of video. A lot of information about coastal flooding are therefore not taken into account such as the number of collision or overwash events, or the time when they happen. One possible extension of this work could be to detect and count each collision/overwash in each timestack, leading to better risk characterization.

5. Bayesian networks to model storm impact using data from both monitoring networks and statistical learning methods

5.1 Introduction

In this chapter, we develop a storm impact model for the Grande Plage de Biarritz based on Bayesian networks (BN). In the literature, most of the applications of BN in storm impact modeling are relying on process-based modeling to constitute the training database for the BN. Unlike previous works on this subject, we only use observational data collected by diverse monitoring networks (tide gauge, wave buoy, weather station) as training database. Because observations of storm impact and atmospheric surge are limited, we also present and test a methodology based on SLM to extend the dataset. This methodology is based on cross validation and aims to select the SLM with the best generalizing ability. In total, two BNs are trained in this chapter, one exclusively on the observational data and one with both observational and predicted data. Their performances are compared by predicting on the same events.

Scientific output:

This work resulted in a communication during an international conference:

- "Developing a Bayesian network to predict coastal flooding on the Grande

Plage de Biarritz based on observational data and statistical models.", XVII International Symposium on Oceanography of the Bay of Biscay (ISOBAY 17), Online conference (June 2021).

This work is in preparation for submission to *Natural Hazards* journal.

5.2 Article: Bayesian networks to model storm impact using data from both monitoring networks and statistical learning methods

Bayesian networks to model storm impact using data from both monitoring networks and statistical learning methods

Aurélien Callens[a], Denis Morichon[b], Benoit Liquet[a]

[a]*Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Pau, France*
[b]*Université de Pau et des Pays de l'Adour, E2S UPPA, SIAME, Anglet, France*

Abstract

Bayesian networks (BNs) are probabilistic graphical models that are increasingly used to translate hydraulic boundary conditions during storm events into onshore hazard. However, comprehensive databases representative of the extreme and episodic nature of storms are needed to train the BNs. Such databases do not exist for many sites and many BNs are trained on data generated by process-based model. To our knowledge, BNs have not been trained exclusively on observational data in coastal engineering domain.

This study aims to explore the performance of a BN exclusively based on observational data in coastal flooding prediction. To this end, we use as a test case the Grande Plage of Biarritz (South west of France). The BN is trained using data from several monitoring networks located near the study site. Because observational data about storm impact regime and atmospheric surge are limited, a second aim of this work is to propose a methodology based on statistical learning methods to extend the data about these variables. This methodology aims to select the statistical learning method with the best generalizing ability with a cross validation. Two BNs are trained, one exclusively on the observational data and one with both observational and predicted data. To compare the two networks, their performances are evaluated on the same events.

We demonstrated that it was possible to predict coastal flooding risk in a qualitative manner with a BN based only on observational data with a F_1-score of 0.628. However, the predictive skill of this network is questionable for the most intense storm impact regimes which are impact and overwash regime. Storm impact and atmospheric surge data were both extended by random forest method which is the method that showed the best generalizing ability in the two cross validation. This extension of the database led to a better BN in terms of predictive skill, with precision, recall and F_1-score in average 20% higher than the BN

trained only on observational data.

Keywords: Bayesian networks, Coastal flooding, Observational data, Statistical learning methods.

---

1. Introduction

The frequency and intensity of coastal flooding is expected to increase in the future due to sea level rise and the continuous development of coastal areas (Vousdoukas et al., 2018; Taherkhani et al., 2020). The development of early warning systems (EWS) is therefore mandatory to mitigate the risk related to coastal flooding events. A crucial step in an operational EWS is the translation of the hydraulic boundary conditions into onshore hazard. This translation is performed most of the time by phase-resolving wave models. This type of wave models is employed to describe the free surface elevation at the scale of single waves. They have the ability to represent complex processes such as wave set-up, re-circulation, and infra-gravity waves as a result of the free surface evolution (Roeber et al., 2019). However, these models require a fine resolution to perform well and are therefore computationally expensive. This complicates their use in EWS which requires fast predictions. Recently, more attention have been given to bayesian networks (BNs) for the development of the storm impact model.

BNs are a class of probabilistic graphical models which allow for an intuitive representation of a set of random variables and their conditional dependencies (Pearl, 1988; Scutari and Denis, 2014). They can learn causal effect from observational data, represent complex systems with intuitive graphical structure and include uncertainty. For these reasons, applications of BNs have multiplied recently in various domains spanning from ecology (McCann et al., 2006; Landuyt et al., 2013) to coastal engineering (Poelhekke et al., 2016; Jäger et al., 2018; Beuzen et al., 2018).

In coastal engineering, BNs are appreciated methods due to their low computational cost and their ability to represent complex systems by integrating different sources of data. Their intuitive representation is also a non negligible advantage compared to other modeling approaches (e.g. process-based models) as it facilitates the communication between scientists and stakeholders in management applications (Henriksen et al., 2007). Most of the applications of BNs in coastal engineering domain concern the translation of forcing variables (e.g. wave, weather and tide conditions) into impact and damages on the shore during storm events. Indeed, they have been employed to predict coastal cliff erosion

(Hapke and Plant, 2010), shoreline retreat (Beuzen et al., 2018), dune retreat and erosion (Palmsten et al., 2014; den Heijer et al., 2012) and barrier island response (Plant and Stockdon, 2012; Wilson et al., 2015; Poelhekke et al., 2016) resulting from coastal storms.

A challenge in the development of BNs for storm impact modeling is related to the training database. The training database must be comprehensive and cover a long time period in order to be representative of the extreme and episodic nature of storms. However, such databases do not exist for many coastal sites, therefore, previous applications of BNs rely on process-based modeling to either simulate the forcing parameters or to translate the forcing conditions into impact on the shore using a numerical model to simulate nearshore waves transformation and their impact at the coast. For instance, Poelhekke et al. (2016) and Plomaritis et al. (2018) create synthetic storm events by using copula statistical method fitted on the limited observations of offshore hydrodynamic parameters. The impact of these synthetic events are then simulated using the nearshore wave propagation model Xbeach (Roelvink et al., 2010). The BNs are then trained to link the characteristics of the synthetic storm events to their simulated impact. Lately, with the increasing amount of data collected by diverse monitoring networks, observational data are being incorporated in the training data of the BNs. In Beuzen et al. (2018), the BN is trained on 10 years of storm wave data simulated by process-based models (spectral wind wave model) and also on shoreline retreat data derived from a coastal imaging station.

To our knowledge, BNs have not been trained exclusively on observational data in coastal engineering domain. This study aims to explore the performance of a BN exclusively based on observational data in coastal flooding prediction. To this end, we use as a test case the Grande Plage of Biarritz (South west of France). Every winter, this embayed beach faces numerous storm events which often result in coastal flooding. To prevent and mitigate the risk of flooding, a BN is developed using data from several monitoring networks located near the study site. Forcing parameters are collected by a directional wave buoy, a tide gauge and a weather station. Data about storm impact on the beach are determined from the video monitoring station installed on the site, following the methodology presented in Callens et al. (2021).

Due to the recent installation of the video monitoring system (in 2017), the storm impact data is limited. It is why we propose a methodology based on machine learning methods to extend the database. The potential gain of performance with the extension of the database is investigated by building two different BNs: one only using observational data and the other based on both observational

data and data predicted by the ML algorithms. Their performances are compared by testing their accuracy on the same events. Section 2 will introduce the study area, section 3 will introduce the theory behind bayesian networks and the data used to train them. Results will be presented and discussed in Section 4. Finally, Section 5 will cover the conclusion.

2. Study site

The Grande Plage of Biarritz (GPB) is an urban embayed beach of 1,2 km long located on the southern Aquitanian coast of France (Figure 1). It has a high socio-economic importance for the city of Biarritz due to its tourist appeal, its historical heritage and its location near the city center. In terms of characteristics, the GPB is an intermediate-reflective beach with typically a steep foreshore slope of $8-9\%$ and a gentle nearshore slope of $2-3\%$. It is a mesotidal beach with 4.5 m spring tidal range around a mean water level of 2.64 m. This narrow beach is backed by a seawall with an alongshore elevation varying between 7 and 8m. This seawall serves as defense infrastructure for back beach buildings.

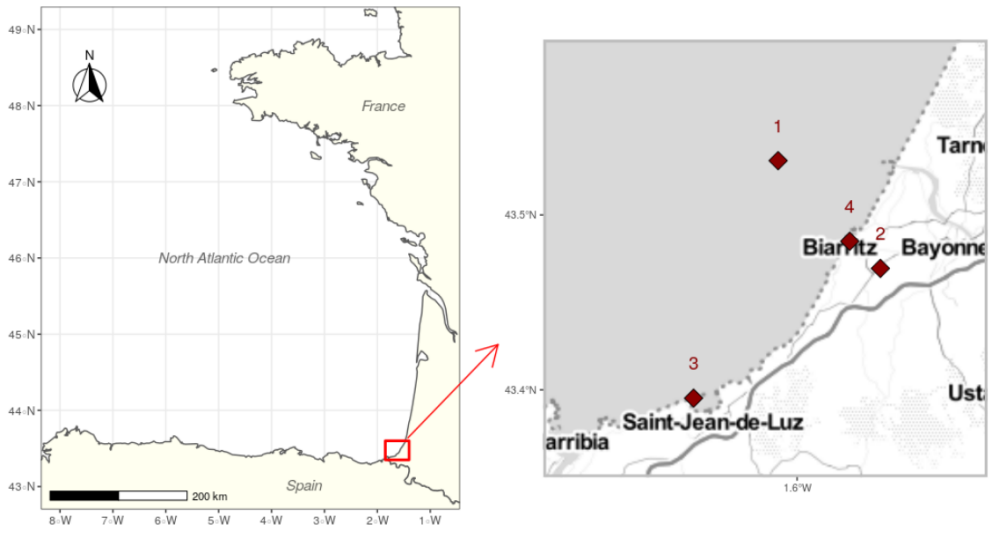


Figure 1: Map showing the location of the study site. The red dots show of the locations of the directional wave buoy (1), the weather station (2), the tide gauge (3) and the beach called "Grande Plage de Biarritz" where a video monitoring station is installed (4).

The beach is predominantly exposed to wave coming from the WNW direction. The annual average significant wave height and peak period are respectively

$H_s = 1.5m$ and $T_p = 10s$. Storms in the region are considered as those in Hs and Tp are respectively greater than 3.5 m and 13.8 s. Such events correspond to 7.24% of the offshore wave climate (Abadie et al., 2006) and are responsible for several coastal flooding events each year.

This site is equipped with a coastal video monitoring station since 2017. The station includes 4 cameras with different lens to ensure the coverage of the entire beach with a sufficient spatial resolution. The cameras are operated by the open source software SIRENA (Nieto et al., 2010).

3. Methodology

3.1. Bayesian networks

In this section, we give a general overview of the theory behind BNs. The reader is referred to Pearl (1988); Scutari and Denis (2014) for in-depth details and examples of applications of BNs .

A bayesian network represents a set of random variables $X = \{X_1, ..., X_n\}$ and their dependencies via a directed acyclic graph (Pearl, 1988). In this directed acyclic graph (DAG), each variable is represented as a *node*. The *nodes* can be connected together by *arcs* that represent potential dependence between the variables. This arc is directed depending the direction of the influence from *parent* to *child* node. In a BN, the arcs must not form a cycle, hence the name of directed acyclic graph.

A fundamental property of the BN is that the global distribution of the set of random variables can be economically factorized through the chain rule (Jäger et al., 2018):

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i)), \tag{1}$$

with $pa(X_i)$ denoting the parent nodes of X_i and $P(X_i|pa(X_i))$ specified by the *conditional probability tables* (CPTs), or the *probability tables* (PTs) when the nodes do not have parents. In discrete bayesian networks, the variables are discretized into states, in this case, the CPTs and PTs quantify the probability of a node being in its particular states.

Given an evidence, the BN can update the states of all the nodes connected to this evidence. This update relies on Bayes' rule:

$$p(R_i|O_j) = \frac{p(O_j|R_i)p(R_i)}{P(O_j)} \tag{2}$$

where $p(R_i|O_j)$ is the updated or conditional probability of the response R_i given O_j a set of observations/evidences, $p(R_i)$ is the probability of response R_i happening, $p(O_j)$ represents the probability of the observations O_j.

The structure, the CPTs and PTs of a BN can be learned from a dataset. In this work, the structure of the BNs and their parameters are learned with the academic version of GeNIe Modeler software from BayesFusion (`http://www.bayesfusion.com/`). In this software, the structure of the BNs is learned using Bayesian Search which is a score-based algorithm (BayesFusion, 2017). This learning algorithm is based on hill-climbing and uses the out-of-sample classification accuracy (computed with 5-fold cross validation) as the scoring function to find the optimal graph. For parameters estimation (learning CPTs and PTs from the data), GeNIe Modeler uses the EM algorithm (Dempster et al., 1977).

3.2. Data sources

Table 1 summarizes the characteristics of the data available for the study site. The majority of the variables has an hourly frequency. However this is not the case for the buoy measurements and the storm impact regime variable. Each data source is described more precisely below.

Table 1: Table summarizing all the data available for the site of Biarritz.

Data source	Variables	Period covered	Frequency
Wave buoy	Measured wave characteristics: H_s, T_p, θ_p	11/2009-04/2020	30 minutes
Reanalysis and forecast wave models improved	Simulated wave characteristics: H_s, T_p, θ_p	01/1993-05/2021	Hourly
Weather station	Meteorological conditions: Atm. Pressure, Wind speed and direction	01/1993-11/2020	Hourly
Tide gauge	Astronomical tide predicted with harmonical analysis.	01/1993-05/2021	Hourly
Tide gauge	Atm. Surge extracted from measurements	05/2011-06/2020	Hourly
Video monitoring station	Storm impact regime extracted from images	03/2017-03/2021	15 minutes

3.2.1. Wave data

We are interested only in the three wave integrated parameters that play a major role in coastal flooding: the significant wave height (H_s), the peak period (T_p) and the peak wave direction (θ_p).

Measured data

Direct measurements of these parameters are obtained from the National Center for Archiving Swell Measurements (L'her et al., 1999). They were made

by a directional wave rider buoy (DWR MKIII) operated by the Centre for Studies and Expertise on Risks, Environment, Mobility, and Urban and Country Planning (CEREMA). The buoy is located a few miles off the Basque Coast (Figure 1) at 50 meters water depth. Since its deployment in 2009, this buoy have been recording the parameters of interest every 30 minutes.

Reanalysis data improved with error prediction method

The wave parameters are also simulated at the buoy coordinates using two hindcast sea-states database computed with the spectral wave model MFWAM developed by MeteoFrance (Lefèvre and Aouf, 2012), each covering two distinct period between 1993 and 2021:

- "IBI_MULTIYEAR_WAV_005_006". This reanalysis covers the period 1993-2019 with a hourly time-step. It is forced by the ERA 5 reanalysis wind data from ECMWF.
- "IBI_ANALYSIS_FORECAST_WAV_005_005" covers the period 2020-2021. It is forced with the ECMWF hourly wind data (forecast).

The spectral wave models are known to underestimate wave characteristics during energetic conditions (Rakha et al., 2007; Moeini et al., 2012; Arnoux et al., 2018). When measured data are available, it is common to perform data assimilation to improve the values of the wave parameters. The error prediction method presented in Callens et al. (2020) is therefore applied to both models in order to improve the predicted wave parameters. The performances before and after the data assimilation on both reanalysis are shown in appendix (Tables A1 and A2).

3.2.2. Tide data

The water level data come from the tide gauge located a few kilometers south of the study site (Figure 1) and were provided by the french Naval Hydrographic and Oceanographic Service (SHOM). The data range from 2011 to nowadays with a hourly time step. In addition to provide the data, the SHOM also documented each dysfunction of the tide gauge. This allowed for the proper removal of aberrant data. From tide data measurements, the astronomical tide was estimated with harmonic analysis performed by the R package **TideHarmonics**. The atmospheric surge was calculated by subtracting the astronomical tide to the measured water level.

3.2.3. Meteorological data

Meteorological data, including average wind speed above 10 meters, wind direction and atmospheric pressure were furnished by the French national meteorological service MeteoFrance. The data were collected hourly by the meteorological station of the Biarritz airport, located only a few kilometers from the study site (Figure 1). It covers a period ranging from 1993-01-01 to 2020-10-31.

3.2.4. Storm impact data

The storm impact regimes for the study site of Biarritz have been extracted from timestack images created by the video monitoring system of the Grande Plage of Biarritz (Figure 1). They are extracted with convolutional neural networks (CNN) following the methodology presented in Callens et al. (2021). In this method, each timestack image is classified into 3 storm impact regimes representing increasing categories of coastal flooding risk:

- Swash regime: all the waves are confined to the beach
- Impact regime: at least one wave collides with the bottom of the seawall
- Overwash regime: at least one wave completely overtops the seawall

The storm impact data ranges from 2017-03-23 to 2021-03-28. It contains 9550 swash regimes, 220 impact regimes and 54 overwash regimes.

3.3. Statistical models for database extension

Data about storm impact and atmospheric surge cover a shorter period compared to the other variables. In order to extend the data of these two variables, statistical learning methods (SLMs) are trained on the available data and employed to predict the occurrences where observations are missing.

For each variable, the performances of several SLMs are compared through a 5-fold cross validation on the available data. The aim of this comparison is to find the method with the best generalization performance. The hyperparameters of these methods are calibrated with bayesian optimization with the aim to maximize the generalization performance. For each variable, the model (and associated hyperparameters) showing the best generalization performance is used to predict the data when observations are not available. The cross validation and hyperparameter search is performed with the R package **Tidymodels**. Once the best predictive model is found, it is trained on all the available data and can be employed to predict data on the period of interest.

Concerning the atmospheric surge model, the explanatory variables are: the astronomical tide, the meteorological conditions (atmospheric pressure, wind speed and direction) and wave characteristics improved by the data assimilation method (H_s, T_p, θ_p). Three statistical learning methods are compared: gradient boosting trees, random forest and shallow neural networks.

For storm impact model, the explanatory variables are the astronomical tide, the wave characteristics improved by the data assimilation method (H_s, T_p, θ_p) and the atmospheric surge. Once again, three SLMs are compared through cross validation: gradient boosting trees, random forest and multinomial models.

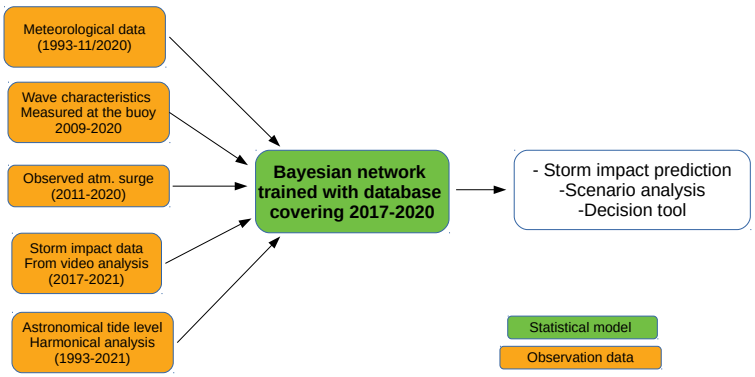
3.4. Training the BNs

In order to assess the gain of performance with the extension of the database with statistical learning methods, we have to build two bayesian networks: one only based on observation data (Figure 2a) and the other based on observation and predicted data (Figure 2b).

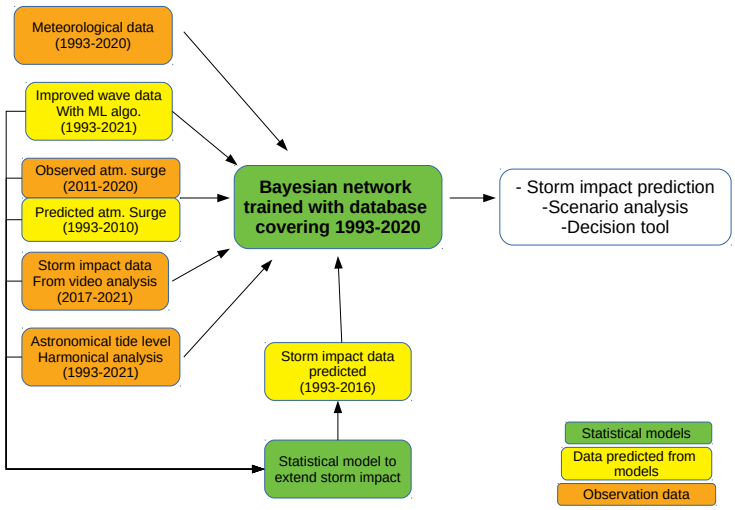
The observational database is built by gathering the wave characteristics at the buoy, the observed atmospheric surge and astronomical tide extracted from the tide gauge, the weather conditions and the storm impact extracted from video monitoring network (Figure 2a). In total, we have 6358 observations with 6166 swash events, 151 impact and 41 overwash events. The test set is built with random stratified sampling and represents 20% of the observations (1234 swash, 31 impact and 9 overwash). The training set contains 4932 swash events, 120 impact and 32 overwash events.

The database extended with statistical learning methods is built by gathering the wave characteristics from the improved wave predictions, the astronomical tide extracted from the tide gauge, the weather conditions, the atmospheric surge and storm impact data (Figure 2b). For the the atmospheric surge and storm impact data, predictions of the statistical learning methods are included when observations are not available. The events of the test set are removed from the extended database to ensure an objective performance evaluation. The training set contains 237479 swash events, 860 impact and 292 overwash events.

Before training the bayesian networks, all the variables are discretized. Because the methods available to automatically discretize variables often find meaningless or questionable thresholds (Chen et al., 2012), the discretization of the variables in this study is based on expert judgment and visible thresholds observed in the data. The discretization thresholds for each parameter can be found in appendix (Table A5).



(a)



(b)

Figure 2: Diagrams of the development of the bayesian networks based on observation data (a) and on observation and predicted data (b).

In addition, partial undersampling is performed on both training set to minimize the imbalance ratio between class. Partial undersampling consists in randomly sampling a percentage of the observations belonging to the majority class.

It is an effective method to deal with the negative effect of class imbalance on the training of statistical learning methods (Japkowicz and Stephen, 2002). For the observational training set, only 20% of the swash events are kept resulting in 86.7% swash events, 10.5% of impact events and 2.8% overwash events. For the training set containing both observational and predicted data, only 2.5% of the swash events are kept to match the class distribution of the first training set (83.8% swash, 12.1% of impact and 4.1% overwash).

3.5. Evaluating the BNs

The descriptive and predictive skill of both BNs are evaluated. The descriptive skill corresponds to the predictive performance of the BN on the training set while the predictive skill corresponds to the predictive performance on unseen data (test set). The performances are evaluated with F_1-score, precision and recall:

$$F_1 = \frac{Precision \times Recall}{Precision + Recall}$$

where

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

and

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}.$$

These metrics are computed for each storm impact regime and are averaged in order to have one global metric for each BN. The F_1-score varies between 0 and 1, with 1 representing the best value. Unlike the global accuracy (number of correct predictions divided by the total number of predictions), the F_1 metric is not biased when data present a class imbalance.

4. Results

4.1. Statistical models to extend the database

Atmospheric surge model

The performances of random forest, gradient boosting trees and neural networks are presented in Table 2. The optimal hyperparameters found for these algorithms are shown in appendix (Table A3).Random forest shows the best performance for atmospheric surge modeling with a mean RMSE of 0.0565. This method is followed by gradient boosting trees and neural networks.

Table 2: Out of sample performances and optimal hyperparameters for the different models for the prediction of atmospheric surge. Results from 5-fold cross validation.

	RMSE	SE RMSE
Random forest	0.0565	0.0004
Gradient boosting trees	0.0605	0.0004
Shallow neural networks	0.1094	0.0007

Random forest with the optimal hyperparameters is trained on the totality of the atmospheric surge observations (2011-2020) and then used to predict the period 1993-2011 that does not have observations for atmospheric surge.

Storm impact model

The performances of the three SLMs for storm impact prediction are presented in Table 3. The optimal hyperparameters found for these algorithms are shown in appendix (Table A4) Kappa metric was used to compare the performances of the SLMs. This metric measures the degree of agreement between the true values and the predicted values by a classifier and has the advantage to not being affected by class imbalance. A value of 1 represents a perfect agreement between predictions and true values whereas 0 represents chance agreement. Here again, random forest showed the best generalization performance with a kappa metric of 0.68.

Table 3: Out of sample performances and optimal hyperparameters for the different models for the storm impact model. Results from 5-fold cross validation

	Kappa	SE Kappa
Random forest	0.6788	0.0265
Gradient boosting trees	0.5648	0.0226
Multinomial model	0.4734	0.0153

Random forest with its optimal hyperparameters is trained on all the storm impact data available (2017-2021). To have an idea about the descriptive performance of this model, it has been used to predict on the training set. The confusion matrix between the predictions of the model and the real observations of storm impact is presented in Table 4. The predictions of this model are not perfect and some events are under- or overestimated. However, a majority of the events, including events from the minority classes (impact and overwash) are well classified.

Table 4: Performance of the best model (Random forest) on the training set

	Swash	Impact	Overwash
Swash	8143	19	1
Impact	69	131	3
Overwash	4	11	39

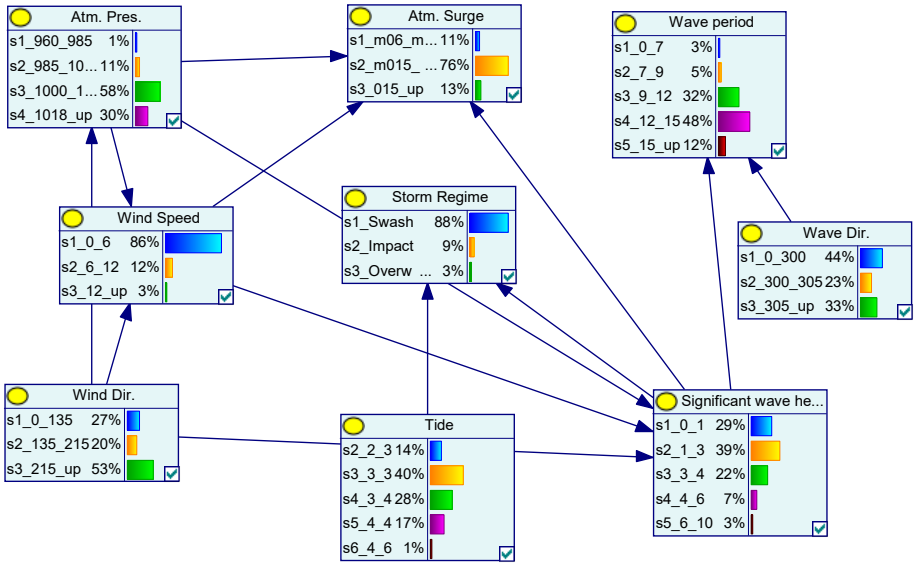
The model is then used to predict storm impact hourly on the period without observations (1993-2016). For the period 1993-2016, this model has predicted 232547 Swash events, 741 Impact events and 260 Overwash events.

4.2. Bayesian network

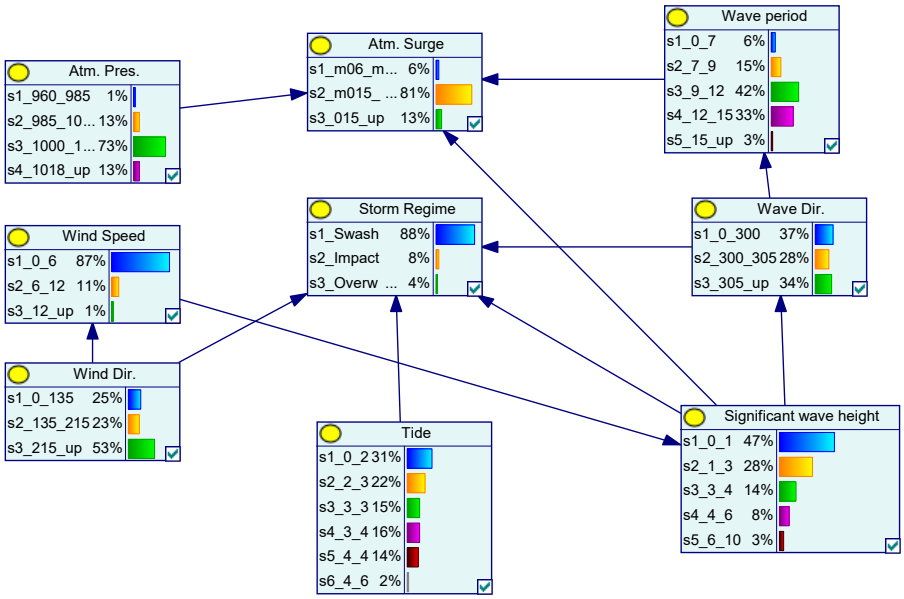
4.2.1. Structure learning

Figure 3 shows the BN structures for the two datasets found by the Bayesian search implemented by GeNIe Modeler. This score-based method aim to find the structure that maximizes the out-of-sample classification accuracy while taking into account a list of forbidden arcs. For this study, we forbid all the arcs that are not logical in a physical point of view: storm impact regime can not influence any of the variables, atmospheric surge can not influence any of the variables expect storm impact regime, the waves parameters can not influence the weather conditions and finally the astronomical tide can not be influenced by any of the variables.

Both BNs are similar in terms of structure. Indeed, the storm impact variable is influenced by the tide and the significant wave height in both networks. This is coherent with the literature as the tide and waves are known to play a significant role in coastal flooding (Yang and Liu, 2020). In the network (b), two other variables seem to influence the storm impact regime: wind and wave direction. About the contribution of these two variables, no references have been found, the dependencies between these variable and the storm impact must have been kept for a question of prediction accuracy. Concerning atmospheric surge, both BN found atmospheric pressure and significant wave height as influencing variables. This is also coherent with the literature as the storm surge is the results of the interaction between atmospheric pressure (Harris, 1963), wave characteristics (Bertin et al., 2015) and wind conditions (Arnaud and Bertin, 2014).



(a)



(b)

Figure 3: Bayesian network structures found for (a) the dataset based on observational data and (b) the dataset based on both observational and predicted data.

4.2.2. Evaluation of the performances

Once the structure and the parameters of the BNs are learned, we can perform the predictions. To that end, the networks are conditioned with the observations of explanatory variables and they output the probabilities of belonging to the different storm impact regimes. For a set of observations, the predicted storm impact regime is the one with the highest probability of belonging. Table 5 presents the descriptive and predictive performances of the BNs. In both cases, the second BN, built on observational and predicted data, shows better performances than the first BN built only on observational data. In the descriptive case, the second BN has better precision and recall resulting in a F_1-score 21.5% higher than the first BN. The confusion matrices of both networks are presented in appendix (Table A6).

Table 5: Descriptive and predictive performances of the two BNs.

Descriptive performance (training set)			
	Precision	Recall	F1-score
BN based on obs. data	0.743	0.709	0.72
BN based on obs. and pred data	0.871	0.88	0.875

Predictive performance (test set)			
	Precision	Recall	F1-score
BN based on obs. data	0.594	0.673	0.628
BN based on obs. and pred data	0.63	0.789	0.691

Concerning the predictive performance, the precision of both BN are nearly similar, however the recall is higher for the second BN which results in a F_1-score 10% higher for the second BN. Confusion matrices computed on the test set are presented in Table 6. In general, the minority classes (Impact and Overwash) tend to have higher error rate for both BNs. This is expected as minority classes are more difficult to represent as they contain fewer examples than majority classes. It is worth noting that even if the second BN have better predictive performance it still has false positives and false negatives (underestimation or overestimation) for the Impact and Overwash class. This can be problematic for a potential use in an operational EWS.

Table 6: Confusion matrices on the test set (predictive accuracy) obtained by both Bayesian networks.

(a) BN based on observational data				
		Predicted		
		Swash	Impact	Overwash
Observed	Swash	1209	24	1
	Impact	10	15	6
	Overwash	3	1	5

(b) BN based on observational and predicted data				
		Predicted		
		Swash	Impact	Overwash
Observed	Swash	1205	23	6
	Impact	10	19	2
	Overwash	0	2	7

5. Discussion

In this section we reflect on a number of aspects related to the advantages and limitations of (i) the prediction of coastal flooding risk with BN based on observational data and (ii) the extension of the database with SLMs.

The main advantage of training BNs exclusively on observational data is that we avoid potential biases related to data generated by process-based models. Indeed, process-based models are not perfect and bias might come from different sources: simplifying assumptions, errors in input data, discretization of the domain (Babovic et al., 2001, 2005). This is especially true for spectral wave models which are known to underestimate wave parameters during storm events (Rakha et al., 2007; Moeini et al., 2012; Arnoux et al., 2018). However by using exclusively observational data, we depend on the precision and data availability of the diverse monitoring systems.

The BN based only on observational data is able to predict coastal flooding risk in a qualitative manner. However, the predictive accuracy of this BN is questionable especially for the minority regimes (impact and overwash) which are the most interesting regimes in terms of coastal flooding prediction. This result was expected as we try to predict extreme events with a database covering only 4 years. For an implementation in a EWS, the predictive accuracy of

the BN should be further investigated. More attention could be given to probabilities of belonging for each regime instead of studying only the most likely regime. In addition, sensitivity analysis on decision thresholds could be performed to minimize the number of false positives and negatives for impact and overwash regimes. This step is essential for operational implementation as false positives could lead to the unnecessary implementation of costly mitigation measures such as the installation of sandbag protection while false negatives could lead to significant damages with the absence of mitigation measures.

Concerning the extension of database with SLMs, it led to better results in terms of metrics, nevertheless the BN trained with the extended database still presents false positives and negatives for overwash regime which can be problematic for the use of this BN in an operational EWS. Here again, sensitivity analysis on decision thresholds must be done to minimize the number false positives and negatives. The only pitfall of the extension of the database by SLMs concerns the storm impact model. Even though this model has been chosen due to its generalization performance, it has not been validated yet on historical events. Therefore, we do not know the performance of this model in the historical reconstitution of the storm impact variable. The validation of this model could be performed with databases made from press or insurance archives such as the database proposed by Abadie et al. (2018).

The main limitation of the proposed methodology is related to the storm impact data which are too limited. The proposed extension of the database by SLMs only allows a better representation of the historical events. However, a good predictive model should be able to predict unseen storm events. The observations should be completed by synthetic events and their modeled impact by process-based in the same spirit as Poelhekke et al. (2016). Databases from archives and insurance could be also integrated in the bayesian networks (Abadie et al., 2018) to complete the storm impact data. Another limitation is the inherent assumption of time invariance of the coastal flooding process. Indeed, by not taking into account the change in the site morphology, we consider that the conditions leading to coastal flooding do not differ over time. A last limitation lies in the fact that this model do not predict the coastal flooding risk on the entirety of the Grande Plage of Biarritz but only on the transect used to create the timestacks from which the qualitative storm impact data is extracted.

6. Conclusion

The aim of this study was to assess the performance in coastal flooding prediction of a BN exclusively based on data measured by diverse monitoring net-

works. Because observational data about storm impact regime and atmospheric surge were limited, we proposed a methodology based on statistical learning methods to extend the data about these variables. This methodology was based on cross validation and aimed to select the statistical learning method and associated hyperparameters with the best generalizing ability. Two BNs were trained, one exclusively on the observational data and one with both observational and predicted data. To compare the two networks, their performances were evaluated on the same events.

We demonstrated that it is possible to predict coastal flooding risk in a qualitative manner with a BN based only on observational data with a F_1-score of 0.628. However, the predictive skill of this network is questionable for the most intense storm impact regimes which are impact and overwash regime. Storm impact and atmospheric surge data were both extended by random forest method which is the method that showed the best generalizing ability in the two cross validation. The extension of the database led to a better BN in term of predictive skill, with precision, recall and F_1-score in average 20% higher than the BN trained only on observational data.

Even though the predictive skill of the two BNs on the most likely class is not perfect, they can be without a doubt great tools in the prediction of coastal risk. For an operational implementation, it would be more interesting to look at the probabilities of belonging instead of looking at the most likely regime. In addition, a sensitivity analysis could be performed on these probabilities to select tailored decision thresholds for the study site. In any case, the observational data should be completed with synthetic events and their impact simulated with process-based model in order to predict potential unseen events.

Acknowledgments

Funding was provided by the Energy Environment Solutions (E2S-UPPA) consortium and the BIGCEES project from E2S-UPPA ("Big model and Big data in Computational Ecology and Environmental Sciences"). The authors would like to thank the French national meteorological service "MeteoFrance" and Copernicus Marine Environment Monitoring Service for providing data.

References

Abadie, S., Beauvivre, M., Egurrola, E., Bouisset, C., Degremont, I., Arnoux, F., 2018. A database of recent historical storm impact on the french basque coast. Journal of Coastal Research , 721–725.

Abadie, S., Butel, R., Mauriet, S., Morichon, D., Dupuis, H., 2006. Wave climate and longshore drift on the South Aquitaine coast. Continental Shelf Research 26, 1924–1939.

Arnaud, G., Bertin, X., 2014. Contribution du setup induit par les vagues dans la surcote associée à la tempête Klaus. XIII emes Journées Nationales Génie Côtier Génie Civil, Paralia Ed., Dunkerque, France , 2–4.

Arnoux, F., Abadie, S., Bertin, X., Kojadinovic, I., 2018. A database to study storm impact statistics along the Basque Coast. Journal of Coastal Research 85, 806–810.

Babovic, V., Caňizares, R., Jensen, H.R., Klinting, A., 2001. Neural networks as routine for error updating of numerical models. Journal of Hydraulic Engineering 127, 181–193.

Babovic, V., Sannasiraj, S.A., Chan, E.S., 2005. Error correction of a predictive ocean wave model using local model approximation. Journal of Marine Systems 53, 1–17.

BayesFusion, L., 2017. Genie modeler. User Manual. Available online: https://support. bayesfusion. com/docs/(accessed on 21 October 2019) .

Bertin, X., Li, K., Roland, A., Bidlot, J.R., 2015. The contribution of short-waves in storm surges: Two case studies in the Bay of Biscay. Continental Shelf Research 96, 1–15.

Beuzen, T., Splinter, K.D., Marshall, L.A., Turner, I.L., Harley, M.D., Palmsten, M.L., 2018. Bayesian Networks in coastal engineering: Distinguishing descriptive and predictive applications. Coastal Engineering 135, 16–30.

Callens, A., Morichon, D., Abadie, S., Delpey, M., Liquet, B., 2020. Using random forest and gradient boosting trees to improve wave forecast at a specific location. Applied Ocean Research 104, 102339.

Callens, A., Morichon, D., Liria, P., Epelde, I., Liquet, B., 2021. Automatic creation of storm impact database based on video monitoring and convolutional neural networks. Remote Sensing 13, 1933.

Chen, W.B., Liu, W.C., Hsu, M.H., 2012. Predicting typhoon-induced storm surge tide with a two-dimensional hydrodynamic model and artificial neural network model. Natural Hazards and Earth System Sciences 12, 3799–3809. doi:10.5194/nhess-12-3799-2012.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological) 39, 1–22.

Hapke, C., Plant, N., 2010. Predicting coastal cliff erosion using a bayesian probabilistic model. Marine geology 278, 140–149.

Harris, D.L., 1963. Characteristics of the Hurricane Storm Surge. Department of Commerce, Weather Bureau.

den Heijer, C.K., Knipping, D.T., Plant, N.G., de Vries, J.S.v.T., Baart, F., van Gelder, P.H., 2012. Impact assessment of extreme storm events using a bayesian network. Coastal Engineering Proceedings , 4–4.

Henriksen, H.J., Rasmussen, P., Brandt, G., Von Buelow, D., Jensen, F.V., 2007. Public participation modelling using bayesian networks in management of groundwater contamination. Environmental Modelling & Software 22, 1101–1113.

Jäger, W.S., Christie, E.K., Hanea, A.M., den Heijer, C., Spencer, T., 2018. A Bayesian network approach for coastal risk analysis and decision making. Coastal Engineering 134, 48–61.

Japkowicz, N., Stephen, S., 2002. The class imbalance problem: A systematic study. Intelligent data analysis 6, 429–449.

Landuyt, D., Broekx, S., D'hondt, R., Engelen, G., Aertsens, J., Goethals, P.L., 2013. A review of bayesian belief networks in ecosystem service modelling. Environmental Modelling &

Software 46, 1–11.
Lefèvre, J.M., Aouf, L., 2012. Latest developments in wave data assimilation, in: ECMWF Workshop on Ocean Waves, pp. 25–27.
L'her, J., Goasguen, G., Rogard, M., 1999. CANDHIS database of in situ sea states measurements on the French coastal zone, in: The Ninth International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers.
McCann, R.K., Marcot, B.G., Ellis, R., 2006. Bayesian belief networks: applications in ecology and natural resource management. Canadian Journal of Forest Research 36, 3053–3062.
Moeini, M.H., Etemad-Shahidi, A., Chegini, V., Rahmani, I., 2012. Wave data assimilation using a hybrid approach in the Persian Gulf. Ocean Dynamics 62, 785–797.
Nieto, M.A., Garau, B., Balle, S., Simarro, G., Zarruk, G.A., Ortiz, A., Tintoré, J., Álvarez-Ellacuría, A., Gómez-Pujol, L., Orfila, A., 2010. An open source, low cost video-based coastal monitoring system. Earth Surface Processes and Landforms 35, 1712–1719.
Palmsten, M.L., Splinter, K.D., Plant, N.G., Stockdon, H.F., 2014. Probabilistic estimation of dune retreat on the gold coast, australia. Shore & Beach 82, 35–43.
Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.
Plant, N.G., Stockdon, H.F., 2012. Probabilistic prediction of barrier-island response to hurricanes. Journal of Geophysical Research: Earth Surface 117.
Plomaritis, T.A., Costas, S., Ferreira, Ó., 2018. Use of a Bayesian Network for coastal hazards, impact and disaster risk reduction assessment at a coastal barrier (Ria Formosa, Portugal). Coastal Engineering 134, 134–147.
Poelhekke, L., Jäger, W.S., van Dongeren, A., Plomaritis, T.A., McCall, R., Ferreira, Ó., 2016. Predicting coastal hazards for sandy coasts with a Bayesian network. Coastal Engineering 118, 21–34.
Rakha, K.A., Al-Salem, K., Neelamani, S., 2007. Hydrodynamic atlas for Kuwaiti territorial waters. Kuwait Journal of Science and Engineering 34, 143.
Roeber, V., Pinault, J., Morichon, D., Abadie, S., Azouri, A., Guiles, M., Luther, D., Delpey, M., Danglade, N., 2019. Improving wave run-up forecasts–benefits from phase-resolving models. Coastal Structures 2019 , 752–761.
Roelvink, D., Reniers, A., Van Dongeren, A., Van Thiel de Vries, J., Lescinski, J., McCall, R., 2010. Xbeach model description and manual. Unesco-IHE Institute for Water Education, Deltares and Delft University of Tecnhology. Report June 21, 2010.
Scutari, M., Denis, J.B., 2014. Bayesian networks: With examples in r .
Taherkhani, M., Vitousek, S., Barnard, P.L., Frazer, N., Anderson, T.R., Fletcher, C.H., 2020. Sea-level rise exponentially increases coastal flood frequency. Scientific reports 10, 1–17.
Vousdoukas, M.I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L.P., Feyen, L., 2018. Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. Nature communications 9, 1–12.
Wilson, K.E., Adams, P.N., Hapke, C.J., Lentz, E.E., Brenner, O., 2015. Application of bayesian networks to hindcast barrier island morphodynamics. Coastal Engineering 102, 30–43.
Yang, T.H., Liu, W.C., 2020. A general overview of the risk-reduction strategies for floods and droughts. Sustainability 12, 2687.

Appendix A. Improvement of the numerical wave model predictions

Such as the methodology of Callens et al. (2020), we tested random forest and gradient boosting tree to improve the numerical wave model predictions. With 5-fold cross validation on the data, we perform hyperparameter tuning and we choose the best generalizing model for each parameter.

Appendix A.1. Reanalysis 1993-2020

For this data assimilation task, random forest was the best generalizing algorithm for each of the 3 wave parameters. The metrics with and without the error prediction method are given in the table below. They are computed on all data and on data where there are stormy conditions (H_s above 3 meters) because the wave model is known to greatly underestimate wave parameters in these conditions.

Table A1: Metrics computed for MFWAM model (a) and after the data assimilation method with random forest (b) on the reanalysis covering the period 1993-2019. Bias, RMSE and correlation coefficient are computed using buoy data as reference.

(a)

Metrics	Hs	Tp	Dir
Bias	0.007	0.893	-0.225
RMSE	0.272	2.262	14.784
Cor.	0.969	0.721	0.313
	$H_s > 3\ meters$		
Bias	-0.173	0.746	1.567
RMSE	0.474	1.407	7.559
Cor.	0.888	0.859	0.676

(b)

Metrics	Hs	Tp	Dir
Bias	0.001	-0.002	0
RMSE	0.103	0.539	3.597
Cor.	0.997	0.975	0.954
	$H_s > 3\ meters$		
Bias	-0.02	-0.01	-0.01
RMSE	0.159	0.321	1.836
Cor.	0.978	0.981	0.941

Appendix A.2. Reanalysis 2020-2021

For this data assimilation task, random forest was the best generalizing algorithm for each of the 3 wave parameters. The metrics with and without the error prediction method are given in the table below. They are computed on all data and on data where there are stormy conditions (H_s above 3 meters) because the wave model is known to greatly underestimate wave parameters in these conditions.

Table A2: Metrics computed for MFWAM model (a) and after the data assimilation method with random forest (b) on the reanalysis covering the period 2020-2021. Bias, RMSE and correlation coefficient are computed using buoy data as reference.

(a)

Metrics	Hs	Tp	Dir
Bias	0.047	0.705	0.588
RMSE	0.261	2.172	10.576
Cor.	0.981	0.659	0.594
	$H_s > 3\ meters$		
Bias	-0.129	0.503	1.643
RMSE	0.365	0.906	6.263
Cor.	0.902	0.911	0.553

(b)

Metrics	Hs	Tp	Dir
Bias	0.001	-0.002	0.017
RMSE	0.103	0.535	3.544
Cor.	0.997	0.975	0.956
	$H_s > 3\ meters$		
Bias	-0.02	-0.012	-0.015
RMSE	0.16	0.32	1.847
Cor.	0.978	0.981	0.939

Appendix B. Extending the database

Appendix B.1. Atmospheric model

Table A3: Optimal hyperparameters found with the 5-fold cross validation.

	Optimal hyperparameters
Random forest	Ntrees = 500, mtry = 5
Gradient boosting trees	Ntrees = 446, mtry = 6, lr = 0.011
Shallow neural networks	Hidden units = 19, dropout = 0.232

Ntrees represents the number of trees, *mtry* represents the number of feature selected when forming each split in a single tree, *lr* designates the learning rate, *hidden units* corresponds to the number of neurons in the hidden layers and *dropout* represents the drop out rate of the neurons in the hidden layer.

Appendix B.2. Storm impact model

Table A4: Optimal hyperparameters found with the 5-fold cross validation.

	Optimal hyperparameters
Random forest	Ntrees = 263, mtry = 3
Gradient boosting trees	Ntrees = 428, mtry = 3, lr = 0.038
Multinomial model	Penalty= 2.05e-10

Ntrees represents the number of trees, *mtry* represents the number of feature selected when forming each split in a single tree, *lr* designates the learning rate, *Penalty* corresponds to the penalty of the multinomial model fitted by the R package **glmnet**.

Appendix C. Training the bayesian networks

Appendix C.1. Discretization of the variables

Table A5: Discretization thresholds chosen for the explanatory variables.

Variables	Number of class	Ranges for class	Metric
Significant wave height	5	[0,1.5]; [1.5,3]; [3,4.5]; [4.5,6]; [6,10]	m
Wave period at peak	5	[0,7]; [7,9]; [9,12]; [12,15]; [15,22]	s
Wave direction	3	[0,300]; [300,305]; [305,360]	degree
Tide	6	[0,2]; [2,3]; [3,3.5]; [3.5,4]; [4,4.5]; [4.5,6]	m
Atm. pressure	4	[960,985]; [985,1000]; [1000, 1018]; [1018, 1035]	hPa
Wind speed	3	[0,6]; [6,12]; [12,24]	m/s
Wind direction	3	[0,135]; [135,215]; [215,360]	degree
Atm. surge	3	[-0.6 , -0.15]; [-0.15, 0.15]; [0.15, 0.7]	m

Appendix C.2. Confusion matrix on the training set

Table A6: Confusion matrices on the training set (descriptive accuracy) obtained by both Bayesian networks.

(a) BN based on observational data				
		Predicted		
		Swash	Impact	Overwash
Observed	Swash	963	22	1
	Impact	42	63	15
	Overwash	10	2	20
(b) BN based on observational and predicted data				
		Predicted		
		Swash	Impact	Overwash
Observed	Swash	5883	47	6
	Impact	59	728	73
	Overwash	2	56	234

5.3 Conclusion

In this chapter, we showed that BN can predict coastal flooding in a qualitative manner using exclusively observational data collected by diverse monitoring networks. We also demonstrated that the extension of the database by SLM leads to better predictive performance at the expense of more false positives for "overwash" regime.

Before implementing the presented BN in an operational EWS, further investigations must be made. A sensitivity analysis on the probabilities of belonging must be performed to select tailored decision thresholds for the study site. Concerning the extension of the database with SLM, the performance of the storm impact model must be investigated. The presented storm impact models could be extended to incorporate the temporal dynamic of coastal flooding by using Dynamic bayesian networks (DBN) which are BN including the concept of time. Finally, an interesting work could be to compare the performances of a storm impact model based on BN or DBN with the performances of a storm impact model based on process-based models.

6. Conclusion

This thesis proposed innovative methodologies based on SLM which contribute to the improvement of coastal risk assessment tools and to the development of an early warning system which aims to reduce coastal flooding risk.

Firstly, we employed SLM to model two coastal processes involved in coastal flooding namely: atmospheric surge and wave runup. We showed that supervised models can be used to predict accurately the coastal processes (in the training data variability) and that knowledge can be acquired by performing variable importance analysis on these models. We also highlighted the ability of unsupervised learning methods to detect groups and patterns from big data, which helped in the characterization of the local sea state and its seasonality. Predictive models and knowledge about coastal processes at a local scale are essential in the development of EWS and consequently in the DRR process.

Secondly, we proposed a methodology based on SLM to improve the forecast of process-based models that are commonly used in operational settings. We focused more particularly on the improvement of the wave forecast made by spectral wave model which are known to underestimate wave parameters during storm events. With local data and a data assimilation method based on SLM, we were able to improve significantly the forecast of wave parameters during stormy conditions. By improving the wave forecast, this methodology contributes to the improvement of the EWS predictions. This improvement is all the more important because the risk of coastal flooding is at its highest during storm events.

Thirdly, we proposed a new methodology based on deep learning methods to collect storm impact data routinely. We employed Convolutional neural networks (CNN) to classify images from the video monitoring station into

storm impact regimes which are categories of coastal flooding risk. Once the CNN are trained, they can be employed to classify the newly created images and therefore generate incrementally a storm impact database. Storm impact databases are crucial in the DRR process as they are used to develop and validate storm impact model or EWS.

Finally, we developed a predictive model translating offshore hydraulic boundary conditions into onshore hazards with SLM. This storm impact model, based on BN, was trained using only observational data collected by different monitoring networks and was able to predict qualitatively the storm impact. In addition, we showed that SLM can be employed to extend the database which can results in a BN with better predictive skill. This storm impact model can be included in an EWS to complement current storm impact model. It has the advantages to avoid the computation time and the potential bias related to the process-based models commonly used.

Scientific contributions

The objective of this research work was to propose innovative solutions to the different scientific issues discussed in each chapter.

The first chapter acted as an "introduction" about the use of SLM in the study of coastal processes. The aim of this chapter was to highlight the ability of SLM to provide better knowledge and understanding about coastal processes at a local scale. Our contribution is related to the fact that we employed methodologies that have never been applied to our study site. In the previous applications of SLM for the modeling of coastal processes, the prediction accuracy was the main focus and little attention was given on the explicability of the SLM. This is why, we also decided to deal with the explicability of the models by performing variable importance analysis.

The second chapter focused on the underestimation of wave parameters by spectral wave model during storm events. This problem is well known in

literature and many solutions have been already proposed, including the data assimilation method we employed. However, shallow neural networks were the off-the-shelves SLM in all the previous works using the error prediction method. In addition, little attention was given to hyperparameter tuning in these works. Consequently, we proposed to use ensemble methods (random forest and gradient boosting trees) as alternatives to the neural networks and we performed hyperparameter tuning with Bayesian optimization. With this work, we proved the importance of comparing different statistical learning methods and choosing the optimal hyperparameters in order to obtain the best results.

The third chapter concerned the generation of storm impact databases. These databases are essential in the DRR process because they are used to train storm impact models or to validate EWS. However, these databases are rare, sparse and mostly come from archives or insurances data. Therefore, we proposed a methodology to collect routinely data about storm impact. This innovative methodology is based on convolutional neural networks (CNN) and images created by video monitoring stations. It aims to classify the images into different storm impact regimes which are categories of coastal flooding risk. To find the best practices for this classification task, we compared several CNN architectures and methods to cope with class imbalance which is a problem related to the extreme nature of storms. We also investigated on the transferability of this method by looking at "pretraining". We showed that a CNN pretrained on a study site can lead to better classification results in fewer epochs for a new study site. Therefore, we shared our pretrained CNN and code with the community to facilitate the application of this methodology to new sites.

The last chapter was about the development of a storm impact model with bayesian network based exclusively on data acquired with diverse monitoring networks. Even though BNs have been employed extensively in the development of storm impact models, they all rely on process-based models or empirical formula to generate the training data. This can be problematic as

process-based models and empirical formula are not perfect and can be biased. This chapter aimed to address this problem by building a storm impact model for coastal flooding prediction with a BN exclusively on observational data. In addition, we proposed and tested a methodology based on SLM to extend the storm impact data as they were limited for our study site.

Finally, a great emphasis was placed on the dissemination and reproducibility of our scientific work. We developed functions in the R package **rlmDataDriven** related to an article published at the beginning of this thesis (Appendix B). In this article, we present a robust estimation procedure for auto-regressive models with heterogeneity (Callens et al., 2020). Concerning the improvement of the wave forecast, the code and a short tutorial were made available on a GitHub repository (Link to the tutorial). For the generation of the storm impact database, the code, the weights of the CNN and even the code of the shiny application developed to annotate the timestacks were made available on Github repositories (Link for code and weights of CNN, Link for the application).

Perspectives

This research work opens up a number of perspectives that we have mentioned on different occasions in this manuscript. We extend on some of these perspectives below.

Extreme nature of storm

The main limitation encountered in this thesis is related to the extreme nature of storms. Extreme events are rare by definition and even if the monitoring networks have been recording for years or decades, we always have a limited number of observations of these events. This is a problem in the training of SLM as the storm events, the events of interest, are under-represented in the training dataset. Methods to deal with class imbalance were employed

in this thesis (chapter 3 and 5) to obtain satisfactory results. However, we saw the limitations of these methods in chapter 5, where we have only a few observations of storm events. It would be interesting to extend the work of this chapter by combining the observational database with a synthetic database created in the same spirit as proposed by Poelhekke et al. (2016) which we have exploited in Morichon et al. (2018) for the Grande Plage Biarritz. In the study of Poelhekke et al. (2016), they employed a statistical method (copula), which capture the dependencies between the storm characteristic variables, to simulate synthetic storm events. The impact of these storm events were then simulated by a process-based model. This methodology combining a statistical method and a process-based model could be the solution to improve the storm impact model presented in last chapter. In any case, this limitation highlights the need to maintain existing monitoring networks and to install new ones in order to train better statistical learning models for the prediction of coastal risks.

Temporal aspect

In this research work, the temporal aspect has not been treated. It could be interesting to integrate this temporal aspect particularly in the improvement of wave forecast made by spectral wave model. Indeed, it is reasonable to think that the errors of the spectral wave model are temporally correlated. To include this temporal aspect, we could use errors of the previous time step as explanatory variables and we could also employ SLM that are adapted for time series modeling such as recurrent neural networks (Zhang et al., 2021) or random forest adapted to time series (Goehry et al., 2021). The temporal aspect could also be included in the storm impact model by employing dynamic bayesian networks which are bayesian networks including the concept of time (Murphy and Russell, 2002).

Operational implementation of the methods

So far, none of the proposed methods in this thesis has yet been implemented in an operational context. Therefore, the operational implementation of these methods is the logical continuation of this thesis. During this step, numerous interrogations will emerge. The first interrogation is about the frequency of "retraining" of the SLM. Because the monitoring networks keep collecting new data, it is legitimate to know the frequency at which the SLM need to be "retrained" to include the information of the newly collected data. This frequency will surely be different for each method proposed in this thesis. Concerning the generation of the storm impact database with CNN, the new timestacks will need to be labeled in order to "retrain" the CNN. To facilitate and even avoid the tedious labeling step of the new timestacks, a special attention could be given to semi-supervised learning which aim to jointly learn from labeled and unlabeled samples (Baur et al., 2017). Concerning the storm impact model, a sensitivity analysis on the probabilities of belonging must be performed to select optimal decision thresholds. This sensitivity analysis is crucial for operational implementation as it aims to minimize the false positives for the storm impact regimes representing the highest risk.

Bibliography

Alqushaibi, A., Abdulkadir, S. J., Rais, H. M., Al-Tashi, Q., Ragab, M. G., and Alhussian, H. (2021). Enhanced weight-optimized recurrent neural networks based on sine cosine algorithm for wave height prediction. *Journal of Marine Science and Engineering*, 9(5):524.

Arnaud, G. and Bertin, X. (2014). Contribution du setup induit par les vagues dans la surcote associée à la tempête Klaus. *XIII emes Journées Nationales Génie Côtier Génie Civil, Paralia Ed., Dunkerque, France*, pages 2–4.

Babovic, V., Caňizares, R., Jensen, H. R., and Klinting, A. (2001). Neural networks as routine for error updating of numerical models. *Journal of Hydraulic Engineering*, 127(3):181–193.

Barbier, E. B. (2013). Valuing ecosystem services for coastal wetland protection and restoration: Progress and challenges. *Resources*, 2(3):213–230.

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C., and Silliman, B. R. (2011). The value of estuarine and coastal ecosystem services. *Ecological monographs*, 81(2):169–193.

Baur, C., Albarqouni, S., and Navab, N. (2017). Semi-supervised deep learning for fully convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 311–319. Springer.

Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep learning*, volume 1. MIT press Massachusetts, USA:.

Bengtsson, L., Hodges, K. I., and Keenlyside, N. (2009). Will extratropical storms intensify in a warmer climate? *Journal of Climate*, 22(9):2276–2301.

Benshila, R., Thoumyre, G., Najar, M. A., Abessolo, G., Almar, R., Bergsma, E., Hugonnard, G., Labracherie, L., Lavie, B., Ragonneau, T., et al. (2020). A deep learning approach for estimation of the nearshore bathymetry. *Journal of Coastal Research*, 95(SI):1011–1015.

Berbić, J., Ocvirk, E., Carević, D., and Lončar, G. (2017). Application of neural networks and support vector machine for significant wave height prediction. *Oceanologia*, 59(3):331–349.

Bergstra, J. S., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. pages 2546–2554.

Bertin, X., Bruneau, N., Breilh, J.-F., Fortunato, A. B., and Karpytchev, M. (2012). Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay. *Ocean Modelling*, 42:16–30.

Bertin, X., Li, K., Roland, A., and Bidlot, J.-R. (2015). The contribution of short-waves in storm surges: Two case studies in the Bay of Biscay. *Continental Shelf Research*, 96:1–15.

Beuzen, T., Splinter, K. D., Marshall, L. A., Turner, I. L., Harley, M. D., and Palmsten, M. L. (2018). Bayesian Networks in coastal engineering: Distinguishing descriptive and predictive applications. *Coastal Engineering*, 135:16–30.

Bezuglov, A., Blanton, B., and Santiago, R. (2016). Multi-Output Artificial Neural Network for Storm Surge Prediction in North Carolina. *arXiv preprint arXiv:1609.07378*.

Breilh, J.-F., Bertin, X., Chaumillon, É., Giloy, N., and Sauzeau, T. (2014). How frequent is storm-induced flooding in the central part of the Bay of Biscay? *Global and Planetary change*, 122:161–175.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Browne, M., Castelle, B., Strauss, D., Tomlinson, R., Blumenstein, M., and Lane, C. (2007). Near-shore swell estimation from a global wind-wave model: Spectral process, linear, and artificial neural network models. *Coastal Engineering*, 54(5):445–460.

Burkart, N. and Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317.

Buscombe, D. and Carini, R. J. (2019). A data-driven approach to classifying wave breaking in infrared imagery. *Remote Sensing*, 11(7):859.

Caires, S., Swail, V. R., and Wang, X. L. (2006). Projection and analysis of extreme wave climate. *Journal of Climate*, 19(21):5581–5605.

Callens, A. (2017). *Analyse des forçages physiques et climatiques influant sur le transport des micropolluants dans l'estuaire de l'Adour*. Master Thesis, Université de Pau et des Pays de l'Adour.

Callens, A., Fu, L., Liquet, B., et al. (2020). Robust estimation procedure for autoregressive models with heterogeneity. *Environmental Modeling and Assessment*.

Camus, P., Cofiño, A. S., Mendez, F. J., and Medina, R. (2011). Multivariate Wave Climate Using Self-Organizing Maps. *Journal of Atmospheric and Oceanic Technology*, 28(11):1554–1568.

Camus, P., Menéndez, M., Méndez, F. J., Izaguirre, C., Espejo, A., Cánovas, V., Pérez, J., Rueda, A., Losada, I. J., and Medina, R. (2014). A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research: Oceans*, 119(11):7389–7405.

Camus, P., Rueda, A., Méndez, F. J., and Losada, I. J. (2016). An atmospheric-to-marine synoptic classification for statistical downscaling marine climate. *Ocean Dynamics*, 66(12):1589–1601.

Chaumillon, E., Bertin, X., Fortunato, A. B., Bajo, M., Schneider, J.-L., Dezileau, L., Walsh, J. P., Michelot, A., Chauveau, E., and Créach, A. (2017). Storm-induced marine flooding: Lessons from a multidisciplinary approach. *Earth-Science Reviews*, 165:151–184.

Chen, X., Zhang, X., Church, J. A., Watson, C. S., King, M. A., Monselesan, D., Legresy, B., and Harig, C. (2017). The increasing rate of global mean sea-level rise during 1993–2014. *Nature Climate Change*, 7(7):492–495.

da Silva, P. G., Coco, G., Garnier, R., and Klein, A. H. (2020). On the prediction of runup, setup and swash on beaches. *Earth-Science Reviews*, 204:103148.

den Heijer, C. K., Knipping, D. T., Plant, N. G., de Vries, J. S. v. T., Baart, F., and van Gelder, P. H. (2012). Impact assessment of extreme storm events using a bayesian network. *Coastal Engineering Proceedings*, (33):4–4.

Deo, M. C., Jha, A., Chaphekar, A. S., and Ravikant, K. (2001). Neural networks for wave forecasting. *Ocean engineering*, 28(7):889–898.

Deshmukh, A. N., Deo, M. C., Bhaskaran, P. K., Nair, T. B., and Sandhya, K. G. (2016). Neural-network-based data assimilation to improve numerical ocean wave forecast. *IEEE Journal of Oceanic Engineering*, 41(4):944–953.

Di Nunno, F., Granata, F., Gargano, R., and de Marinis, G. (2021). Forecasting of extreme storm tide events using narx neural network-based models. *Atmosphere*, 12(4):512.

Diwedar, A. (2016). Investigating the effect of wave parameters on wave runup. *Alexandria Engineering Journal*, 55(1):627–633.

Elgohary, T., Mubasher, A., and Salah, H. (2017). Significant deep wave height prediction by using support vector machine approach (alexandria as case of study). *Int. J. Curr. Eng. Tech*, 7:135–143.

Espejo, A., Camus, P., Losada, I. J., and Méndez, F. J. (2014). Spectral ocean wave climate variability based on atmospheric circulation patterns. *Journal of Physical Oceanography*, 44(8):2139–2152.

Etemad-Shahidi, A. and Mahjoobi, J. (2009). Comparison between m5 model tree and neural networks for prediction of significant wave height in lake superior. *Ocean Engineering*, 36(15-16):1175–1181.

European Commission (2012). The challenge of climate change to the european coastal areas: State of coasts in the context of global climate change.

Feser, F., Barcikowska, M., Krueger, O., Schenk, F., Weisse, R., and Xia, L. (2015). Storminess over the North Atlantic and northwestern Europe—A review. *Quarterly Journal of the Royal Meteorological Society*, 141(687):350–382.

Fiedler, J. W., Smit, P. B., Brodie, K. L., McNinch, J., and Guza, R. (2018). Numerical modeling of wave runup on steep and mildly sloping natural beaches. *Coastal Engineering*, 131:106–113.

Fisher, A., Rudin, C., and Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177):1–81.

Flanders Marine Institute (2021). *Wave Runup - Coastal wiki*. `http://www.coastalwiki.org/wiki/Wave_run-up`.

Flather, R. A. (2000). Existing operational oceanography. *Coastal Engineering*, 41(1-3):13–40.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Goehry, B., Yan, H., Goude, Y., Massart, P., and Poggi, J.-M. (2021). Random forests for time series.

Granata, F. and Di Nunno, F. (2021). Artificial intelligence models for prediction of the tide level in venice. *Stochastic Environmental Research and Risk Assessment*, pages 1–12.

Hapke, C. and Plant, N. (2010). Predicting coastal cliff erosion using a bayesian probabilistic model. *Marine geology*, 278(1-4):140–149.

Harris, D. L. (1963). *Characteristics of the Hurricane Storm Surge*. Department of Commerce, Weather Bureau.

Hashemi, M. R., Spaulding, M. L., Shaw, A., Farhadi, H., and Lewis, M. (2016). An efficient artificial intelligence model for prediction of tropical storm surge. *Natural Hazards*, 82(1):471–491.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.

He, Y., Shen, H., and Perrie, W. (2006). Remote sensing of ocean waves by polarimetric sar. *Journal of Atmospheric and oceanic technology*, 23(12):1768–1773.

Hegermiller, C. A., Antolinez, J. A., Rueda, A., Camus, P., Perez, J., Erikson, L. H., Barnard, P. L., and Mendez, F. J. (2017). A multimodal wave spectrum–based approach for statistical downscaling of local wave climate. *Journal of Physical Oceanography*, 47(2):375–386.

Hill et al. (2020). *Oceanography*. Geosciences, LibreTexts.

Huang, W., Murray, C., Kraus, N., and Rosati, J. (2003). Development of a regional neural network for coastal water level predictions. *Ocean Engineering*, 30(17):2275–2295.

Idier, D., Paris, F., Le Cozannet, G., Boulahya, F., and Dumas, F. (2017). Sea-level rise impacts on the tides of the European Shelf. *Continental Shelf Research*, 137:56–71.

Kennedy, A. B., Westerink, J. J., Smith, J. M., Hope, M. E., Hartman, M., Taflanidis, A. A., Tanaka, S., Westerink, H., Cheung, K. F., and Smith, T. (2012). Tropical cyclone inundation potential on the Hawaiian Islands of Oahu and Kauai. *Ocean Modelling*, 52:54–68.

Kidd, C., Levizzani, V., and Bauer, P. (2009). A review of satellite meteorology and climatology at the start of the twenty-first century. *Progress in Physical Geography*, 33(4):474–489.

Kim, J., Kim, J., Kim, T., Huh, D., and Caires, S. (2020). Wave-tracking in the surf zone using coastal video imagery with deep neural networks. *Atmosphere*, 11(3):304.

Kim, S.-W., Lee, A., and Mun, J. (2018). A Surrogate Modeling for Storm Surge Prediction Using an Artificial Neural Network. *Journal of Coastal Research*, 85:866–870.

Klemas, V. (2012). Remote sensing of coastal and ocean currents: An overview. *Journal of Coastal Research*, 28(3):576–586.

Klemas, V. V. (2009). The role of remote sensing in predicting and determining coastal storm impacts. *Journal of Coastal Research*, 25(6 (256)):1264–1275.

Kopp, R. E., Horton, R. M., Little, C. M., Mitrovica, J. X., Oppenheimer, M., Rasmussen, D., Strauss, B. H., and Tebaldi, C. (2014). Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites. *Earth's Future*, 2(8):383–406.

Larson, M. (2005). *Numerical Modeling*. Springer Netherlands, Dordrecht.

Lavidas, G. and Venugopal, V. (2018). Application of numerical wave models at european coastlines: A review. *Renewable and Sustainable Energy Reviews*, 92:489–500.

Lee, H., Kim, S., and Jun, K. (2018). The Study for Storm Surge Prediction Using Generalized Regression Neural Networks. *Journal of Coastal Research*, 85(sp1):781–785.

Lee, T.-L. (2004). Back-propagation neural network for long-term tidal predictions. *Ocean Engineering*, 31(2):225–238.

Lee, T.-L. (2006). Neural network prediction of a storm surge. *Ocean Engineering*, 33(3-4):483–494.

Lee, T.-L. (2008). Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan. *Engineering Applications of Artificial Intelligence*, 21(1):63–72.

L'her, J., Goasguen, G., and Rogard, M. (1999). CANDHIS database of in situ sea states measurements on the French coastal zone. In *The Ninth International Offshore and Polar Engineering Conference*. International Society of Offshore and Polar Engineers.

Lionello, P., Sanna, A., Elvini, E., and Mufato, R. (2006). A data assimilation procedure for operational prediction of storm surge in the northern Adriatic Sea. *Continental shelf research*, 26(4):539–553.

Little, C. M., Horton, R. M., Kopp, R. E., Oppenheimer, M., Vecchi, G. A., and Villarini, G. (2015). Joint projections of us east coast sea level and storm surge. *Nature Climate Change*, 5(12):1114–1120.

Londhe, S. N., Shah, S., Dixit, P. R., Nair, T. M. B., Sirisha, P., and Jain, R. (2016). A Coupled Numerical and Artificial Neural Network Model for Improving Location Specific Wave Forecast. *Applied Ocean Research*, 59:483–491.

Mafi, S. and Amirinia, G. (2017). Forecasting hurricane wave height in gulf of mexico using soft computing methods. *Ocean Engineering*, 146:352–362.

Makarynska, D. and Makarynskyy, O. (2008). Predicting sea-level variations at the Cocos (Keeling) Islands with artificial neural networks. *Computers & Geosciences*, 34(12):1910–1917.

Makarynskyy, O. (2005). Neural pattern recognition and prediction for wind wave data assimilation. *Pac Oceanogr*, 3(2):76–85.

Makarynskyy, O., Makarynska, D., Kuhn, M., and Featherstone, W. (2004). Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia. *Estuarine, Coastal and Shelf Science*, 61(2):351–360.

Makarynskyy, O., Pires-Silva, A. A., Makarynska, D., and Ventura-Soares, C. (2005). Artificial neural networks in wave predictions at the west coast of Portugal. *Computers & Geosciences*, 31(4):415–424.

Mandal, S. and Prabaharan, N. (2006). Ocean wave forecasting using recurrent neural networks. *Ocean engineering*, 33(10):1401–1410.

Marcelli, M., Piermattei, V., Gerin, R., Brunetti, F., Pietrosemoli, E., Addo, S., Boudaya, L., Coleman, R., Nubi, Q., Rick, J., et al. (2021). Toward the widespread application of low-cost technologies in coastal ocean observing (internet of things for the ocean). *Mediterranean Marine Science*, 22(2):255–269.

Marchant, R. and Ramos, F. (2012). Bayesian optimisation for intelligent environmental monitoring. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2242–2249. IEEE.

Marcos, M., Jordà, G., Gomis, D., and Pérez, B. (2011). Changes in storm surges in southern europe from a regional model under climate change scenarios. *Global and Planetary Change*, 77(3-4):116–128.

Masselink, G., Hughes, M., and Knight, J. (2014). *Introduction to coastal processes and geomorphology*. Routledge.

MEa, M. E. A. (2005). Ecosystems and human well-being: wetlands and water synthesis.

Mehvar, S., Filatova, T., Dastgheib, A., De Ruyter van Steveninck, E., and Ranasinghe, R. (2018). Quantifying economic value of coastal ecosystem services: a review. *Journal of marine science and engineering*, 6(1):5.

Mellado-Cano, J., Barriopedro, D., García-Herrera, R., Trigo, R. M., and Hernández, A. (2019). Examining the north atlantic oscillation, east atlantic pattern, and jet variability since 1685. *Journal of Climate*, 32(19):6285–6298.

Mengel, M., Levermann, A., Frieler, K., Robinson, A., Marzeion, B., and Winkelmann, R. (2016). Future sea level rise constrained by observations and long-term commitment. *Proceedings of the National Academy of Sciences*, 113(10):2597–2602.

Meyer, V., Priest, S., and Kuhlicke, C. (2012). Economic evaluation of structural and non-structural flood risk management measures: examples from the mulde river. *Natural Hazards*, 62(2):301–324.

Moeini, M. H., Etemad-Shahidi, A., Chegini, V., and Rahmani, I. (2012). Wave data assimilation using a hybrid approach in the Persian Gulf. *Ocean Dynamics*, 62(5):785–797.

Morichon, D., de Santiago, I., Delpey, M., Somdecoste, T., Callens, A., Liquet, B., Liria, P., and Arnould, P. (2018). Assessment of flooding hazards at an engineered beach during extreme events: Biarritz, SW France. *Journal of Coastal Research*, 85(sp1):801–805.

Munk, W. H. and Traylor, M. A. (1947). Refraction of ocean waves: a process linking underwater topography to beach erosion. *The Journal of Geology*, 55(1):1–26.

Murphy, K. P. and Russell, S. (2002). Dynamic bayesian networks: Representation, inference and learning.

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding-a global assessment. *PloS one*, 10(3):e0118571.

Nicolle, A. (2006). *Modélisation Des Marées et Des Surcotes Dans Les Pertuis Charentais*. PhD Thesis, Université de La Rochelle.

Nicolle, A., Karpytchev, M., and Benoit, M. (2009). Amplification of the storm surges in shallow waters of the pertuis charentais (bay of biscay, france). *Ocean Dynamics*, 59(6):921.

Nieto, M. A., Garau, B., Balle, S., Simarro, G., Zarruk, G. A., Ortiz, A., Tintoré, J., Álvarez-Ellacuría, A., Gómez-Pujol, L., and Orfila, A. (2010). An open source, low cost video-based coastal monitoring system. *Earth Surface Processes and Landforms*, 35(14):1712–1719.

Palmsten, M. L., Splinter, K. D., Plant, N. G., and Stockdon, H. F. (2014). Probabilistic estimation of dune retreat on the gold coast, australia. *Shore & Beach*, 82(4):35–43.

Pérez, J., Méndez, F. J., Menéndez, M., and Losada, I. J. (2014). EStelA: A method for evaluating the source and travel time of the wave energy reaching a local area. *Ocean Dynamics*, 64(8):1181–1191.

Perez, J., Menendez, M., Camus, P., Mendez, F. J., and Losada, I. J. (2015). Statistical multi-model climate projections of surface ocean waves in Europe. *Ocean Modelling*, 96:161–170.

Pickering, M., Horsburgh, K., Blundell, J., Hirschi, J.-M., Nicholls, R. J., Verlaan, M., and Wells, N. (2017). The impact of future sea-level rise on the global tides. *Continental Shelf Research*, 142:50–68.

Pirazzoli, P. A. (2000). Surges, atmospheric pressure and wind change and flooding probability on the atlantic coast of france. *Oceanologica Acta*, 23(6):643–661.

Plant, N. G. and Stockdon, H. F. (2012). Probabilistic prediction of barrier-island response to hurricanes. *Journal of Geophysical Research: Earth Surface*, 117(F3).

Plant, N. G. and Stockdon, H. F. (2015). How well can wave runup be predicted? comment on laudier et al.(2011) and stockdon et al.(2006). *Coastal Engineering*, 102:44–48.

Plomaritis, T. A., Costas, S., and Ferreira, Ó. (2018). Use of a Bayesian Network for coastal hazards, impact and disaster risk reduction assessment at a coastal barrier (Ria Formosa, Portugal). *Coastal Engineering*, 134:134–147.

Poelhekke, L., Jäger, W. S., van Dongeren, A., Plomaritis, T. A., McCall, R., and Ferreira, Ó. (2016). Predicting coastal hazards for sandy coasts with a Bayesian network. *Coastal Engineering*, 118:21–34.

Quintana, G. I., Tandeo, P., Drumetz, L., Leballeur, L., and Pavec, M. (2021). Statistical forecast of the marine surge. *Natural Hazards*, pages 1–13.

Rogers, D. and Tsirkunov, V. (2011). Costs and benefits of early warning systems. *Global assessment rep*.

Roland, A., Zhang, Y. J., Wang, H. V., Meng, Y., Teng, Y.-C., Maderich, V., Brovchenko, I., Dutour-Sikiric, M., and Zanke, U. (2012). A fully coupled 3D wave-current interaction model on unstructured grids. *Journal of Geophysical Research: Oceans*, 117(C11).

Rueda, A., Cagigal, L., Antolínez, J. A., Albuquerque, J. C., Castanedo, S., Coco, G., and Méndez, F. J. (2019). Marine climate variability based on weather patterns for a complicated island setting: The New Zealand case. *International Journal of Climatology*, 39(3):1777–1786.

Rueda, A., Camus, P., Tomás, A., Vitousek, S., and Méndez, F. J. (2016). A multivariate extreme wave and storm surge climate emulator based on weather patterns. *Ocean Modelling*, 104:242–251.

Sadeghifar, T., Nouri Motlagh, M., Torabi Azad, M., and Mohammad Mahdizadeh, M. (2017). Coastal wave height prediction using recurrent neural networks (rnns) in the south caspian sea. *Marine Geodesy*, 40(6):454–465.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). Ncep climate forecast system reanalysis (cfsr) 6-hourly products, january 1979 to december 2010.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., ya Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E. (2011). Ncep climate forecast system version 2 (cfsv2) 6-hourly products.

Savitha, R., Al Mamun, A., et al. (2017). Regional ocean wave height prediction using sequential learning neural networks. *Ocean Engineering*, 129:605–612.

Seto, K. C., Fragkias, M., Güneralp, B., and Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PloS one*, 6(8):e23777.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

Shao, W., Zhang, Z., Li, X., and Li, H. (2016). Ocean wave parameters retrieval from sentinel-1 sar imagery. *Remote Sensing*, 8(9):707.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959.

Solomatine, D. P. and Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of hydroinformatics*, 10(1):3–22.

Souza, A. J., Brown, J. M., Williams, J. J., and Lymbery, G. (2013). Application of an operational storm coastal impact forecasting system. *Journal of Operational Oceanography*, 6(1):23–26.

Splinter, K. D., Harley, M. D., and Turner, I. L. (2018). Remote sensing is changing our view of the coast: Insights from 40 years of monitoring at narrabeen-collaroy, australia. *Remote Sensing*, 10(11):1744.

Stockdon, H. F., Holman, R. A., Howd, P. A., and Sallenger Jr, A. H. (2006). Empirical parameterization of setup, swash, and runup. *Coastal engineering*, 53(7):573–588.

Taherkhani, M., Vitousek, S., Barnard, P. L., Frazer, N., Anderson, T. R., and Fletcher, C. H. (2020). Sea-level rise exponentially increases coastal flood frequency. *Scientific reports*, 10(1):1–17.

Tsai, C.-P. and Lee, T.-L. (1999). Back-propagation neural network in tidal-level forecasting. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 125(4):195–202.

Tsai, C.-P., Lin, C., and Shen, J.-N. (2002). Neural network for wave forecasting among multi-stations. *Ocean engineering*, 29(13):1683–1695.

Tseng, C.-M., Jan, C.-D., Wang, J.-S., and Wang, C. (2007). Application of artificial neural networks in typhoon surge forecasting. *Ocean Engineering*, 34(11-12):1757–1768.

UNGA, U. (2016). Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction.

Valchev, N., Andreeva, N., Eftimova, P., and Trifonova, E. (2014). Prototype of early warning system for coastal storm hazard (Bulgarian Black Sea coast). *CR Acad Bulg Sci*, 67(7):977.

Van Dongeren, A., Ciavola, P., Martinez, G., Viavattene, C., Bogaard, T., Ferreira, O., Higgins, R., and McCall, R. (2018). Introduction to risc-kit: Resilience-increasing strategies for coasts. *Coastal Engineering*, 134:2–9.

Vitousek, S., Barnard, P. L., Fletcher, C. H., Frazer, N., Erikson, L., and Storlazzi, C. D. (2017). Doubling of coastal flooding frequency within decades due to sea-level rise. *Scientific reports*, 7(1):1–9.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Bianchi, A., Dottori, F., and Feyen, L. (2018a). Climatic and socioeconomic controls of future coastal flood risk in europe. *Nature Climate Change*, 8(9):776–780.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., Jevrejeva, S., Jackson, L. P., and Feyen, L. (2018b). Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard. *Nature communications*, 9(1):1–12.

Vousdoukas, M. I., Wziatek, D., and Almeida, L. P. (2012). Coastal vulnerability assessment based on video wave run-up observations at a mesotidal, steep-sloped beach. *Ocean Dynamics*, 62(1):123–137.

Wang, X. L. and Swail, V. R. (2006). Climate change signal and uncertainty in projections of ocean wave heights. *Climate Dynamics*, 26(2-3):109–126.

Wei, P., Lu, Z., and Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety*, 142:399–432.

Wilson, K. E., Adams, P. N., Hapke, C. J., Lentz, E. E., and Brenner, O. (2015). Application of bayesian networks to hindcast barrier island morphodynamics. *Coastal Engineering*, 102:30–43.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

Xu, G., Shi, Y., Sun, X., and Shen, W. (2019). Internet of things in marine environment monitoring: A review. *Sensors*, 19(7):1711.

Yang, T.-H. and Liu, W.-C. (2020). A general overview of the risk-reduction strategies for floods and droughts. *Sustainability*, 12(7):2687.

You, S. H. and Seo, J.-W. (2009). Storm surge prediction using an artificial neural network model and cluster analysis. *Natural Hazards*, 51(1):97–114.

Zhang, X., Li, Y., Gao, S., and Ren, P. (2021). Ocean wave height series prediction with numerical long short-term memory. *Journal of Marine Science and Engineering*, 9(5):514.

A. Extracting storm surge data with harmonic analysis

The observed storm surge is obtained by subtracting the astronomical tide from the measured water level of a tide gauge. The astronomical tide has been studied for a long time and many methods allow for the prediction of the tide level. Harmonical analysis is the most commonly used method and the one presented below. This method aims to represent the tidal water level as a sum of basic harmonic constituents.

Data

The water level data from the tide gauge located in Socoa (city in the south west of France) were provided by the french Naval Hydrographic and Oceanographic Service (SHOM). The data range from 2011 to nowadays with a hourly time step. In addition to provide the data, the SHOM also documented each dysfunction of the tide gauge. This allowed for the proper removal of aberrant data. The final data are presented in the Figure A.1.

To assess the accuracy of the harmonic model on unseen data, the dataset has been divided in two part : the training set corresponding to 80% of the data and ranging from 2011-04-26 to 2016-10-03 and the test set corresponding to the remaining 20% and ranging from 2016-10-04 to 2020-06-12. It is worth noting that the period 2016-2020 contains a lot of missing data due to a dysfunction of the tide gauge (Figure A.1).

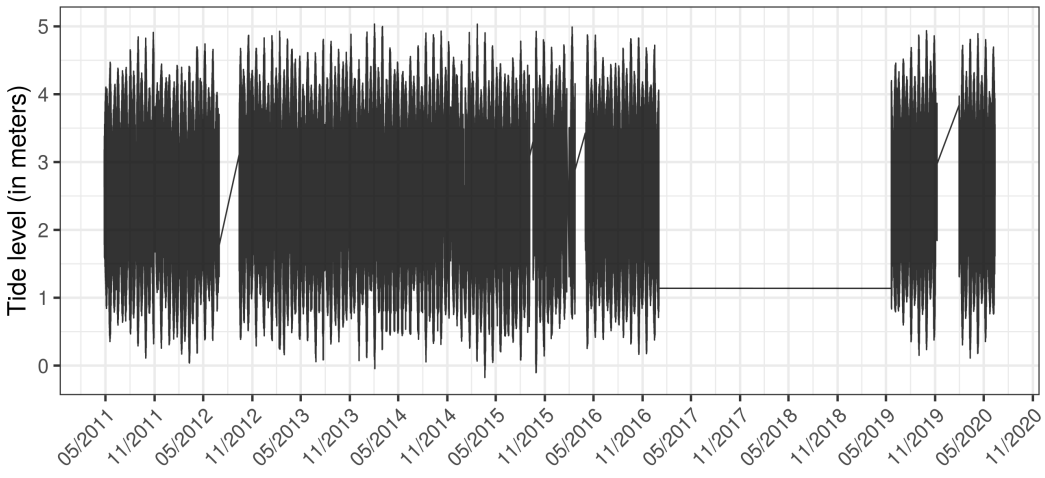


Figure A.1: Tide data from the Socoa tide gauge.

Method

Harmonic analysis is the most common method to model astronomical tide with water level from tide gauge. In this method, the tidal level is modeled as a sum of harmonical constituents, the majority of which being related to the gravitational forces of the Moon and Sun (Nicolle, 2006).

At time t, the tide level can be written as :

$$h(t) = Z_0 + \sum_{i=1}^{n} [A_i cos(v_i t + \phi_i)], \tag{A.1}$$

where Z_0 corresponds to the mean water level, A_i, v_i, ϕ_i to the amplitude, the frequency and the phase of the wave respectively. Finally, n corresponds to the number of constituents chosen.

The unknown parameters A_i and ϕ_i can be estimated with the least square method for each constituent. An implementation of this method in R code is available in the package **TideHarmonics**. For the harmonic analysis method, the number of harmonic constituents must be chosen: a higher number means a better accuracy. Because we want the most accurate model, we chose to estimate the coefficient of the first 114 harmonic constituents which is the

maximum number implemented in the R package **TideHarmonics**.

Table A.1: Tide level prediction with artificial neural networks in the literature

Authors	Year	Type of neural networks	Input	Output
Tsai and Lee	1999	Multilayer Perceptron	-Past sea level observations -Past error of prediction Both from t-1h to t-2h	Sea level at time (t)
Huang et al.	2003	Multilayer Perceptron	Past sea level observations from t-1h to t-4h	Sea level at time (t)
Lee	2004	Multilayer Perceptron	Main Tidal constistuants (from 4 to 7) at time (t)	Tidal level at time (t)
Makarynskyy et al.	2004	Multilayer Perceptron	Past sea level observations from t-1h to -12/24/36/48/60/70h	Sea level forecast from t+1h to t+24h
Makarynska and Makarynskyy	2008	Multilayer Perceptron	Past sea level observations from t-1h to t-120h	Sea level forecast from t+1h to t+120h

Neural networks could have also been used to model the tide level. Generally, they use as input the past observations of sea level and they forecast the sea level for the next time steps. Some of them forecast the sea level related to the tide up to 5 days in advance. Few examples are presented in table A.1.

Tide harmonic method was chosen over neural networks for the modelling of the tide level because once the model is fitted, past observations are not needed to make a prediction unlike neural networks. This is convenient as the data from Socoa tide gauge contains a lot of missing data.

Results and discussion

The estimated amplitudes and phases for the 10 first harmonic components of Socoa are presented in the table A.2. The 2 harmonic components with the highest amplitudes are M2 and S2 which are semi-diurnal components. This indicates that the type of the tide in the area of Socoa is semi-diurnal (Nicolle, 2006).

Once all the harmonic amplitudes and phases are estimated, they can be

Table A.2: Amplitude and phase of the first 10 harmonics estimated with the **TideHarmonics** package

Harmonic components	Amplitude	Phase
M2	1.337	291.5
S2	0.468	306.1
N2	0.283	281.9
K2	0.132	301.6
O1	0.071	69.7
K1	0.063	157.4
nu2	0.053	281.9
mu2	0.050	268.8
2N2	0.041	270.6
Sa	0.040	208.9

used to predict the astronomical tide (without the meteorological effect). The root mean square error (RMSE) computed on the test data is 10.4 centimetres. This result is consistent with the literature, for example Makarynskyy et al. (2004) obtained RMSE values between 8 and 15 centimetres for different neural networks architectures. An example of one month forecast with the harmonic analysis method is showed in the Figure A.2.

For the modelling of the tidal water level, there will always be some discrepancies between the predicted and the measured values at the tide gauge. Indeed, the tide gauge measures the total water level which is composed by the addition of the astronomical tidal level and the water level which is induced by the meteorological conditions. The low pressure and high winds during a storm lead to an increase in water level which is called "storm surge".

Observed storm surge can be extracted by subtracting the modelled astronomical tide from the measured water level of the tide gauge. The storm surge obtained for Socoa tide gauge is presented Figure A.3. Such as the tide data,

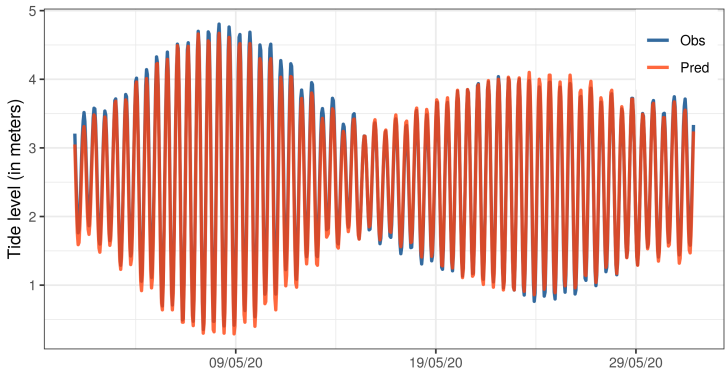


Figure A.2: Water level measured and modelled for the month of May 2020.

the storm surge data range from 2011 to nowadays with a hourly time step and contains missing data for the period 2016-2020.

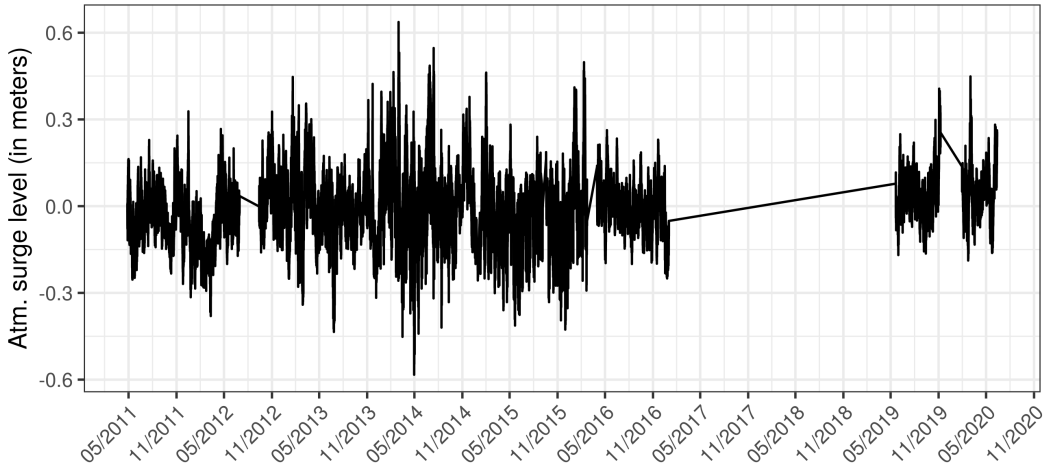


Figure A.3: Atmospheric surge level extracted from the water level of Socoa tide gauge.

Conclusion

Harmonical analysis was used to model the tide with the water level data from the tide gauge data of Socoa. It was found that the type of the tide in

this area was semi-diurnal. In addition, this method allowed us to compute the observed storm surge for the site of Socoa by subtracting the modelled astronomical tide from the water level observed at the tide gauge.

B. Robust estimation procedure for autoregressive models with heterogeneity

This appendix section presents the accepted version of the article. The final publication is available at https://link.springer.com/article/10.1007/s10666-020-09730-w

Noname manuscript No.
(will be inserted by the editor)

Robust estimation procedure for autoregressive models with heterogeneity

A. Callens · Y-G. Wang · L. Fu · B. Liquet

Received: 7 January 2020 / Accepted: 24 August 2020 / Published online: 1 October 2020

Abstract In environmental studies, regression analysis is frequently performed. The classical approach is the ordinary least squares method which consists in minimizing the sum of the square of the residuals. However, this method relies on strong assumptions that are not always satisfied. In environmental data, the response variable often contains outliers and errors can be heteroscedastic. This can have significant effects on parameter estimation. To solve this problem, the weighted M-estimation was developed. It assumes a parametric function for the variance, and, estimates alternately and robustly, mean and variance parameters. However, this method is limited to the independent errors case, and does not apply to time series data. Therefore, we in-

A. Callens
Laboratoire de Mathématiques et de leurs Applications de Pau, Université de Pau et des Pays de l'Adour, UMR CNRS 5142, E2S-UPPA, France
ARC Centre of Excellence for Mathematical and Statistical Frontiers and School of Mathematical Science Queensland University of Technology, Brisbane, Australia
E-mail: aurelien.callens@univ-pau.fr

Y-G. Wang
ARC Centre of Excellence for Mathematical and Statistical Frontiers and School of Mathematical Science Queensland University of Technology, Brisbane, Australia
School of Mathematical Sciences, Queensland University of Technology, Australia

L. Fu
School of Mathematics and Statistics, Xi'an Jiaotong University, China.

B. Liquet
Laboratoire de Mathématiques et de leurs Applications de Pau, Université de Pau et des Pays de l'Adour, UMR CNRS 5142, E2S-UPPA, France
ARC Centre of Excellence for Mathematical and Statistical Frontiers and School of Mathematical Science Queensland University of Technology, Brisbane, Australia

troduce a new estimation procedure which adapts the weighted M-estimation to environmental time series data, while selecting optimal value for the tuning parameter present in the M-estimation. We compare the efficiency of our procedure on simulated data to other usual regression methods. Our estimation procedure outperforms the other methods providing estimates with lower variance and mean squared errors. Finally we illustrate the proposed method using an air quality dataset from Beijing. This method has been implemented in the R package RlmDataDriven.

Keywords Heteroscedasticy · Model selection · Robust estimation · Temporal correlations

1 Introduction

In environmental modelling, pure homoscedasticity is uncommon. For example, residuals of hydrological [8], or air pollution models [26] are usually heteroscedastic. Ignoring this problem and performing an ordinary least squares regression would result in regression parameters with biased covariance matrix and hence would lead to erroneous inference.

One method to deal with this biased covariance matrix is to use the White's estimator [25] which provides a heteroscedasticity consistent covariance matrix. In case of a time series, the Newey-West estimator [17] can provide heteroscedasticity and an autocorrelation-consistent covariance matrix. Both of these methods account for heteroscedasticity but do not give information on the variability of the data generation process.

Another approach to cope with heteroscedasticity is to perform a weighted analysis, where the assumed underlying model is:

$$y_i = x_i^T \beta + \sigma_i \epsilon_i,$$

with y_i the observed response, x_i the associated covariates, β the vector containing the parameters to be estimated, ϵ_i the independently and identically distributed error terms with mean 0 and unknown symmetric distribution function and σ_i the term accounting for heteroscedasticity. Usually, a parametric function is assumed for this term [1, 4, 7, 9]. It can be a power function of the mean as proposed by Box et al. [3]: $\sigma_i = \phi |x_i^T \beta|^\gamma$ or some functions of the covariates. Parameters in the variance function are not known and have to be estimated by maximum likelihood method.

Estimation of parameters in models exhibiting heterogeneous variance is performed by an iterative procedure. A preliminary estimate of mean parameters is obtained by the least squares method. Residuals of this model are then used to estimate variance parameters. Finally, a weighted least squares

method is performed with the estimated variance as weight. Unlike White and Newey-West methods, modeling heterogeneous variance enables one to get better estimates for the mean parameters and also to gain information on the variability of the data generation process [7, 9].

Such as least squares method, the estimation method for heteroscedastic models has a low breakdown point of $1/n$, meaning that only one outlier in the observed response can have a large effect on the estimation of the mean parameters [20, 23, 27]. The estimation of the variance parameter is also affected as maximum likelihood methods are very sensitive to outliers [9, 21].

In practice, outliers are prevalent in environmental dataset. They can be found in both response variable and covariates but, in this work, our interest lies only in outliers that are present in the response variable. In the presence of outliers, robust methods must be applied. They aim to produce reliable estimates that are not seriously affected by outliers, extreme values or small deviations from model assumptions [13].

When both heteroscedasticity and outliers are present in regression analysis, one can use the method described by Carroll et al. [4]. This iterative method allows one to robustly estimate mean and variance parameters. For the mean parameters, they perform a weighted M-estimation with variance as weight. M-estimation is a robust method which consists in minimizing a loss function that is slowly varying for abnormal residuals instead of squared residuals [24]. This loss function is controlled by a tuning parameter c which "regulates the amount of robustness" [11]. For the variance, they assume a parametric function and robustly estimate its parameters with high-breakdown point estimators. A limitation of this method is that errors must be independently and identically distributed, which is not the case when regression is performed on temporal data.

In this article, we propose an estimation procedure which adapts the weighted M-estimation method of Carroll et al. [4] to time series by taking into account temporal correlations, and, chooses the value of the tuning constant by minimizing the variance of the estimators such as the work of Wang et al. [23, 24]. We perform numerical studies to compare the proposed method with other usual regression methods. Finally, we illustrate our methodology with a dataset on fine particle matter pollution ($PM_{2.5}$) in the city of Beijing. This dataset is particularly interesting for two reasons. First, variability of $PM_{2.5}$ concentration is not likely to be constant. Second, largest concentrations, which are related to increased coal consumption during cold days, may have a significant impact on the estimation of regression parameters.

2 Method

2.1 The regression model

Let $y = (y_1,\ y_2,\ \dots, y_n)^T$ be the observed response measured over n equivalent time periods and $x_i\ = (x_{i1},\ \dots,\ x_{ik})^T$ be the set of k associated predictors (x_{i1} is equal to 1 if intercept is considered as a predictor). We assume data are generated from the following heteroscedastic linear model :

$$y_i = x_i^T\beta + \sigma_i\epsilon_i, \tag{1}$$

in which β is the vector collecting the parameters to be estimated, ϵ_i are the error terms following an autoregressive process of order p (AR(p)) which takes into account temporal correlations, and

$$\sigma_i = \phi g(x_i^T\beta, \gamma),$$

where $g(\cdot)$ is a known function of the mean ($x^T\beta$) with unknown parameter vector γ, and unknown dispersion parameter ϕ. Several choices for σ_i are possible, few examples are :

* $\sigma_i = \phi(1 + |x_i^T\beta|)^\gamma$ or $\sigma_i = \phi|x_i^T\beta|^\gamma$ proposed by Box et al. [3],
* $\sigma_i = \phi e^{\gamma x_i^T\beta}$ proposed by Bickel et al. [2].

2.2 Estimation of the parameters

For given or estimated value $\hat{\sigma}_i$, the robust M-estimation for β minimizes :

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T\beta}{\hat{\sigma}_i}\right),$$

where ρ is a loss function that is slowly varying for abnormal residuals (outliers). The most known is Huber's loss function :

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } |u| \le c \\ c|u| - \frac{1}{2}c^2 & \text{if } |u| > c \end{cases}. \tag{2}$$

Here, c is a tuning parameter chosen between 0 and 3, which controls the degree of robustness. Default value for Huber's function is 1.345 to ensure 95% asymptotic relative efficiency when data are normally distributed. More examples of loss functions can be found in Wang et al. [23].

Taking derivatives of (2) leads to the following estimating equation of β:

$$U(\beta) = \sum_{i=1}^{n} \left(\frac{x_i}{\hat{\sigma}_i}\right) \psi\left(\frac{y_i - x_i^T\beta}{\hat{\sigma}_i}\right) = 0,$$

where $\psi(x) = min\{c, max\{x, -c\}\}$ is the derivative of Huber's loss function.

To solve this estimating function, one can rewrite $U(\beta)$ as :

$$U(\beta) = \sum_{i=1}^{n} x_i W_i r_i = 0,$$

where $W_i = \psi(r_i)/r_i$ are weighting terms, and $r_i = (y_i - x_i^T\beta)/\hat{\sigma}_i$ are the Pearson residuals. Now, for a given weight W_i, the robust estimator of β can be obtained by the following formula :

$$\hat{\beta} = \{\sum_{i=1}^{n} x_i W_i x_i^T\}^{-1}\{\sum_{i=1}^{n} x_i W_i y_i\}. \tag{3}$$

An iterative approach is needed as W_i is a function of β and σ. This approach is derived from the pseudolikelihood approach and consists in fixing alternatively parameters of the variance (γ and ϕ) and the regression parameters (β).

The variance parameters are also robustly estimated and are given by :

– A high breakdown estimator for γ :

$$\sum_{i=1}^{n} \chi\Big(\frac{y_i - x_i^T\hat{\beta}}{\hat{\phi} g(x_i^T\hat{\beta}, \gamma)}\Big) \frac{g'(x_i^T\hat{\beta}, \gamma)}{g(x_i^T\hat{\beta}, \gamma)} = 0, \tag{4}$$

where $\chi(\cdot)$ is a bounded function. Croux [6] and Bianco et al. [1] used $\chi(x) = min(x^2/c_1^2, 1) - 0.5$ with $c_1 = 1.041$ to obtain a 50% breakdown estimator of γ under the normality assumption.

– The MAD estimator for the dispersion parameter :

$$\hat{\phi} = \text{Median}\Big\{\frac{|y_i - x_i^{\mathrm{T}}\hat{\beta}|}{g(x_i^T\hat{\beta}, \hat{\gamma})}\Big\}/0.6745. \tag{5}$$

Wang et al. [24] showed that under some regularity conditions, the robust estimator $\hat{\beta}$ obtained by the iterative procedure is Fisher consistent. Moreover, when $n \to \infty$ the covariance matrix is given by [4, 10, 11] :

$$var(\hat{\beta}) = K^2 \frac{[1/(n-k)] \sum_{i=1}^{n} \psi(r_i)^2}{[(1/n) \sum_{i=1}^{n} \psi'(r_i)]^2} \times \sum_{i=1}^{n} \{\hat{\sigma}_i^2 (x_i^{\mathrm{T}} x_i)^{-1}\},$$

where $r_i = (y_i - x_i^T \hat{\beta})/\hat{\sigma}_i$, and

$$K = 1 + \frac{p}{n} \frac{var(\psi'(r_i))}{(E\psi'(r_i))^2}.$$

with p the number of unknown parameters.

2.3 A data-dependent tuning constant

The tuning parameter c associated with the loss function regulates the amount of robustness in the estimation. When observations are normally distributed and without outliers, optimal value of this parameter should be $c = +\infty$. On the other hand, for heavy tailed distributions, optimal value should be around or smaller than 1. The value of the tuning parameter should be chosen carefully since robustness comes at the price of efficiency. Indeed, a smaller value than needed, would result in considering more usual observations as outliers and would lead to a loss of efficiency in the estimation of the regression parameters.

As in Wang et al. [23], we define the best tuning constant as the one which minimizes the variance of the regression parameters. Therefore we propose to repeat the estimating procedure with different values of c between 0 and 3 for the Huber's function [23] and choose the one which minimizes the sum of the estimated variances of the regression parameters.

2.4 Accounting for temporal correlations

So far we have only considered the independent model, we now need to incorporate the autoregressive process of order p present in the error terms. We write ϵ_i as $\sum_{j=1}^{p} (\alpha_j \epsilon_{i-j}) + \xi_i$ where ξ_i are independent errors and we rewrite the model (1) as :

$$y_i = x_i^T \beta + \sum_{j=1}^{p} \alpha_j \sigma_i \epsilon_{i-j} + \sigma_i \xi_i.$$

Because the ϵ_i are unobserved, we propose to use the Pearson residuals from the initial model (1), say, $\hat{\varepsilon}_i$, and we now have the following linear model with roughly independent errors,

$$y_i = x_i^T \beta + \sum_{j=1}^{p} \alpha_j \hat{\sigma}_i \hat{\varepsilon}_{i-j} + \eta_i,$$

where $(\hat{\sigma}_i\hat{\varepsilon}_{i-1}, \hat{\sigma}_i\hat{\varepsilon}_{i-2}, ..., \hat{\sigma}_i\hat{\varepsilon}_{i-j},)$ are the augmented additional covariates including p lagged residuals, and $(\beta, \alpha_1, ..., \alpha_p)$ are the new parameters to be estimated including p lag parameters, $\hat{\sigma}_i$ is an estimate of σ_i. In the iterative procedure to be described below, this $\hat{\sigma}_i$ will be estimated from the variance function using the previous parameter estimates for (ϕ, γ, β). Here, η_i represents the resulting error which should be close to $\sigma_i\xi_i$. We fit this augmented model with the optimal value of the tuning parameter to obtain the final estimate of β.

In the application section, we will demonstrate how we choose the order p of the autoregressive process.

2.5 The estimation procedure

The complete estimation procedure is summarized in the following algorithm :

1. Obtain an initial robust estimate $\hat{\beta}_0$ assuming a constant variance $g(x^T\beta, \gamma) = 1$ and using M-estimation with the default value of c (rlm function).
2. By fixing $\hat{\beta} = \hat{\beta}_0$, the robust variance parameters $(\hat{\phi}, \hat{\gamma})$ are estimated with (4) and (5) respectively.
3. By fixing the variance parameters equal to their robust estimates, we update $\hat{\beta}$ with (3).
4. To find the best tuning parameter, the steps 2-3 are repeated for a range of c values between 0 and 3. The best tuning constant is the one which minimizes the sum of the estimated variance of the regression parameters.
5. The model is fitted using the best value of the tuning constant $(\hat{c})$ from the steps 2-3. Then, temporal correlations are added by following the procedure described previously.

3 Numerical studies

In this section, we investigate the performance of our procedure. We compare mean bias and mean square errors of the estimates obtained by different methods such as least squares (`lm` function in R), generalized least

squares method (`gls` function from the `nlme` package),M-estimation with $c = 1.345$ (`rlm` function from the `MASS` package), the weighted M-estimation (`whm` function from `rlmDataDriven` package) with $c = 1.345$ and our data-driven method (`rlmDD_het` function from `rlmDataDriven` package).

For one simulation, we generate a multivariate normal dataset ($n = 500$) using the model (1). In our case, $x_i^T\beta = \beta_0 + \beta_1 x_{i1}$ where we fix the value of β_0 and β_1 to 10 and x_{i1} comes from a uniform distribution on $(0, 1)$. For σ_i, we test two functions: the power function $\sigma_i = |x_i^T\beta|^\gamma$ with $\gamma = 0.2$ and the exponential function $\sigma_i = e^{\gamma|x_i^T\beta|}$ with $\gamma = 0.01$. For the term ϵ_i, we consider two cases: (i) an autoregressive process of order 1 with $\alpha = 0.5$ and (ii) an autoregressive process of order 2 with $\alpha_1 = 0.5$, $\alpha_2 = 0.2$. These processes can be written as follows: $\epsilon_i = \sum_{j=1}^{p} \alpha_j \epsilon_{i-j} + \xi_i$ where ξ_i are independent and normally distributed errors following a standard normal distribution $N(0, 1)$. In order to add outliers, the data is randomly contaminated by adding a value sampled from a uniform distribution on $(0, 10)$. Several contamination rates are considered: $\lambda = 0\%, 5\%, 10\%$.

Table 1 shows the means and the standard errors of the estimate bias computed accross all the simulations for the exponential function. The results obtained with the power function are included in the online supplementary material (Table S-1).

For both variance functions, average value of the data-dependent tuning constant decreases as contamination rate increases. This tuning constant has the expected behaviour: as the proportion of outliers becomes larger, more values should be considered as outliers therefore the tuning constant is smaller.

The bias and associated standard errors both increase with the proportion of outliers. However it is less important for robust methods, especially for the proposed method. Indeed, when data are contaminated, the method yields estimators with lower mean bias and associated standard errors.

This method also provides estimates with lowest MSE in all the cases when contamination is present. Differences in efficiency with the other methods augment with the contamination rate, reaching MSE values fives fold lower than the ones obtained by non robust methods (Figure 1). The same figure for the power variance function is available in the supplementary material (Figure S-1).

4 Application to Air Quality Data

4.1 Context of the application

We apply the proposed robust procedure to analyse fine particle matter ($PM_{2.5}$) concentrations from the US embassy situated in Beijing, China. Such as other China's mega cities, Beijing has been suffering from chronic air pollution [5]. The main constituents of this pollution are suspended particle matters under 2.5 μm of diameter, widely known as $PM_{2.5}$ [15]. Concentration of $PM_{2.5}$ is highly variable and depends on sources of emission, secondary chemical generation processes and meteorological conditions. According to numerous studies [15, 18, 28], these suspended particles affect climate, visibility and human health in many ways.

Given the severity of the pollution and the potential hazardous effects, China's State Council aimed to reduce the $PM_{2.5}$ pollution by at least 25% for the period 2012-2017. Our objectives are (i) to show the efficiency of our procedure on a real dataset compared to a conventional estimation method (least squares method), (ii) to analyze the potential effect of the decision of China's State Council on the concentration of $PM_{2.5}$. Our method is highly desirable in this case study since the variance of $PM_{2.5}$ concentrations is not likely to be homogeneous [26] and outliers may be present.

4.2 Regression analysis

The hourly data come from a previous study lead by Liang et al . [15] (Figure 2). $PM_{2.5}$ concentrations were taken at the US Embassy of Beijing and meteorological measurements at the Beijing Capital Airport. Both time series covers the period from 1 January 2010 to 31 December 2014.

To evaluate changes in $PM_{2.5}$ concentration after the decision of China's State Council, we created two new variables : Policy and Time Policy. Both variables take the value 0 before 2012, however Policy take the value 1 after 2012 to detect any shifts in the intercept and Policy Time take the value of the time lapsed in days after 2012 to test and quantify the trend after that decision.

We modelled $PM_{2.5}$ concentration as a linear combination of the following covariates : dew point (°C), temperature (°C), atmospheric pression (hPa), combined wind direction (3 factors), cumulated wind speed (m/s), cumulated hour of snow (mm), cumulated hour of rain (mm), Time policy, Policy and seasonal patterns with *sin* and *cos* functions for the 3,2 years and 6,4,3 months cycles. To perform the regression analysis, the following numerical covariates have been standardized: dew point, temperature, atmospheric

pression, cumulated wind speed, cumulated hour of snow, cumulated hour of rain.

The response variable $PM_{2.5}$ is a concentration; therefore a transformation was needed in order to avoid predictions lower than zero with the regression analysis. Two transformations were tested on the dataset: square-root and logarithm transformation (Figure 2). For both transformations, mean squared and mean absolute errors were computed after re-transformating fitted values in original scale. Althougth logarithm transformation is the classical transformation for air quality data, we chose the square root transformation as it yielded slightly lower errors (Supporting information, Table S-3).

The model is fitted with least squares method by the lm function from R statistical software [19]. The residuals vs. fitted value plot (Figure 3) indicates the presence of heteroscedasticity with larger residuals as fitted values increase. This heteroscedasticity does not lead to biased estimators but to estimators with biased covariance matrix. This could result in underestimation of standard errors, erroneous Z-values and therefore erroneous hypothesis tests. The normal probability plot in Figure 3 reveals that distribution of residuals is skewed, indicating the presence of outliers which may have influenced the estimation of regression parameters. Moreover, temporal correlations in the residuals have been found using the ACF and pACF plots.

In this present case, our method is highly desirable as we have the presence of heteroscedasticity, temporal correlations and outliers.

4.3 Robust regression analysis with the proposed method

We fitted the same regression model to the data with the proposed method. This method has been implemented in a R function: `rlmDD_het`, available in the package `RlmDataDriven`. The R code for the presented analysis can be found in the supporting information file.

In the literature, we did not find previous papers or indications on how to model variance of $PM_{2.5}$ concentration, consequently, we used common variance functions for the analysis such as power or exponential functions. Hereafter, we present the result obtained by considering $\sigma_i = \phi e^{\gamma |x^T \beta|}$.

First, our method chooses best tuning constant by testing a range of values for the tuning parameter between 0 and 3. As stated earlier, the best value of c will be the value that minimizes the sum of the variance of the regression parameters. For the $PM_{2.5}$ data, optimal value of the tuning parameter is found around 1.5.

Then, we use ACF and pACF plots of the robustified residuals to determine the order of the autoregressive process. The robustified residuals are

defined as $\psi((y_i - x_i^T\hat{\beta})/\hat{\sigma}_i)$. Two significant lags are found in the pACF plot, therefore, we consider that $p = 2$ and we add two lagged terms to the regression model. The lagged terms are:

$$\alpha_1\hat{\sigma}_i\hat{\varepsilon}_{i-1} + \alpha_2\hat{\sigma}_i\hat{\varepsilon}_{i-2} + \eta,$$

where $\hat{\sigma}_i\hat{\varepsilon}_{i-1}, \hat{\sigma}_i\hat{\varepsilon}_{i-2}$ are lagged terms built from the initial model estimated with the best tuning constant ($\hat{c} = 1.5$) and η is the assumed independent error term. The term $\hat{\varepsilon}$ corresponds to Pearson residuals of the initial model and $\hat{\sigma}$ is the estimated variance function.

The normal probability plot (Figure 3) of the residuals clearly illustrates that our robust procedure has taken care of outliers successfully. The residuals vs. fitted values plot (Figure 3) seems to indicate that heteroscedasticity has been lowered.

The results of both methods are listed in Table 2. In this table, the covariance matrix of the regression parameters obtained by least squares method has been estimated with the `NeweyWest` function (`sandwich` package) which gives a heteroscedasticity and autocorrelation consistent estimation of the covariance. This estimation was necessary to obtain corrected standard errors as the residuals exhibited heteroscedasticity and temporal correlations.

Our estimation method drastically reduced the variance of the parameters. It is worth noting that the covariates Policy and Time policy are not significantly different from 0 in the least squares method contrary to our robust method. Years after 2011 are characterized by an positive shift in the intercept. However, coefficient of the variable Time policy is slightly negative. After one year, this coefficient outweights the positive shift in the intercept, meaning that the $PM_{2.5}$ concentration decreases slightly compared to years before 2012.

Finally, we can see in the Table 2 that lagged terms are significant. This indicates that the two previous terms contribute significantly to the output and were, therefore, necessary to consider.

5 Discussion

This method is data-dependent by the optimal choice of the tuning constant and it incorporates temporal correlations by adding lagged terms in the covariates. The numerical study showed that this procedure outperforms the other usual regression methods when data are contaminated by providing more precise estimates for the mean parameters. In the application with the $PM_{2.5}$ concentration dataset, we proved that our method results in estimates with significantly lower variance compared to the ones obtained by

least squares estimation, leading to better hypothesis testing. This method is well suited for environmental dataset due to the frequent presence of heterogeneity and outliers. Hyslop et al. [12] utilized Thiel-Sen robust regression to evaluate long-term trends in aerosol concentrations via the historical PM2.5 element measurements. Van Donkelaar et al. [22] used a geographically weighted regression (GWR) statistical model to represent bias of fine PM2.5 concentrations over North America . Knibbs et al. [14] utilized the land-use regression (LUR) to estimate PM2.5 at continental scale and explained the most spatial variability in PM2.5 in Australia . These three methods did not consider the heterogeneity and autoregressive errors. In the future, the proposed method could be generalized to these three models. Furthermore, the proposed method could be extended to model time-series of counts by using a link function such as the generalized linear model [16]. Concerning the variance, it would be helpful to consider more flexible approaches for the estimation of the parameters and to provide guidance for choosing the most appropriate variance function for a dataset.

Supplementary Material

Online supplementary material includes :

- Table S-1 : Mean bias from true parameters and standard errors of the estimators obtained by several regression methods for the case of the power variance function.
- Table S-2 : Number of non convergence removed for the calculation of the mean bias and standard deviation for the simulation study.
- Table S-3 : Mean Absolute Error and Root mean squared error for different transformations of $PM_{2.5}$ and different models.
- Figure S-1 : Mean square errors of the estimated parameters obtained by several regression methods for the power variance function.
- Page 5 : Information on the dataset with downloading link and indications on the different steps to tidy the dataset before analyzing it.
- The R script used to obtain the results in the application section.

Acknowledgements

This research was partially funded by the Australian Research Council project (DP160104292).

References

1. Bianco, A., Boente, G., Di Rienzo, J.: Some results for robust GM-based estimators in heteroscedastic regression models. Journal of Statistical Planning and Inference **89**(1-2), 215–242 (2000)
2. Bickel, P.J.: Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. The Annals of Statistics **6**(2), 266–291 (1978)
3. Box, G.E., Hill, W.J.: Correcting inhomogeneity of variance with power transformation weighting. Technometrics **16**(3), 385–389 (1974)
4. Carroll, R.J., Ruppert, D.: Robust estimation in heteroscedastic linear models. The annals of statistics pp. 429–441 (1982)
5. Chan, C.K., Yao, X.: Air pollution in mega cities in China. Atmospheric environment **42**(1), 1–42 (2008)
6. Croux, C.: Efficient high-breakdown M-estimators of scale. Statistics & Probability Letters **19**(5), 371–379 (1994)
7. Davidian, M., Carroll, R.J.: Variance function estimation. Journal of the American Statistical Association **82**(400), 1079–1091 (1987)
8. Evin, G., Kavetski, D., Thyer, M., Kuczera, G.: Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. Water Resources Research **49**(7), 4518–4524 (2013)
9. Giltinan, D.M.: Bounded influence estimation in heteroscedastic linear modelS. Ph.D. thesis, Citeseer (1983)
10. Huber, P.J.: Robust regression: Asymptotics, conjectures and Monte Carlo. The Annals of Statistics **1**(5), 799–821 (1973)
11. Huber, P.J.: Robust Statistics. New York, Chichester, Brisbane. Toronto, Singapore: John Wiley & Sons, Ltd (1981)
12. Hyslop, N.P., Trzepla, K., White, W.H.: Assessing the suitability of historical pm2. 5 element measurements for trend analysis. Environmental science & technology **49**(15), 9247–9255 (2015)
13. Jiang, Y., Wang, Y.G., Fu, L., Wang, X.: Robust Estimation Using Modified Huber's Functions With New Tails. Technometrics (just-accepted), 1–32 (2018)
14. Knibbs, L.D., van Donkelaar, A., Martin, R.V., Bechle, M.J., Brauer, M., Cohen, D.D., Cowie, C.T., Dirgawati, M., Guo, Y., Hanigan, I.C., et al.: Satellite-based land-use regression for continental-scale long-term ambient pm2. 5 exposure assessment in australia. Environmental science & technology **52**(21), 12445–12455 (2018)
15. Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., Chen, S.X.: Assessing Beijing's PM2. 5 pollution: Severity, weather impact, APEC and winter heating. Proc. R. Soc. A **471**(2182), 20150257

(2015)
16. McCullagh, P., Nelder, J.A.: Generalized Linear Models, vol. 37. CRC press (1989)
17. Newey, W.K., West, K.D.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica **55**(3), 703–708 (1987). URL http://www.jstor.org/stable/1913610
18. Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D.: Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama **287**(9), 1132–1141 (2002)
19. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2018). URL https://www.R-project.org/
20. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection, vol. 589. John wiley & sons (2005)
21. Stefanski, L.A., Carroll, R.J., Ruppert, D.: Optimally hounded score functions for generalized linear models with applications to logistic regression. Biometrika **73**(2), 413–424 (1986)
22. Van Donkelaar, A., Martin, R.V., Spurr, R.J., Burnett, R.T.: High-resolution satellite-derived pm2. 5 from optimal estimation and geographically weighted regression over north america. Environmental science & technology **49**(17), 10482–10491 (2015)
23. Wang, N., Wang, Y.G., Hu, S., Hu, Z.H., Xu, J., Tang, H., Jin, G.: Robust Regression with Data-Dependent Regularization Parameters and Autoregressive Temporal Correlations. Environmental Modeling & Assessment pp. 1–8 (2018)
24. Wang, Y.G., Lin, X., Zhu, M., Bai, Z.: Robust estimation using the Huber function with a data-dependent tuning constant. Journal of Computational and Graphical Statistics **16**(2), 468–481 (2007)
25. White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica **48**(4), 817–838 (1980). URL http://www.jstor.org/stable/1912934
26. Wu, E.M.Y., Kuo, S.L.: Air quality time series based garch model analyses of air quality information for a total quantity control district. Aerosol and Air Quality Research **12**(3), 331–343 (2012)
27. Zhao, J., Wang, J.: Robust testing procedures in heteroscedastic linear models. Communications in Statistics—Simulation and Computation®**38**(2), 244–256 (2009)
28. Zheng, M., Salmon, L.G., Schauer, J.J., Zeng, L., Kiang, C.S., Zhang, Y., Cass, G.R.: Seasonal trends in PM2. 5 source contributions in Beijing, China. Atmospheric Environment **39**(22), 3967–3976 (2005)

Table 1 Mean bias and associated standard errors of the estimated parameters obtained by several regression methods for the exponential variance function. Results based on 500 replications.

$\sigma_i = e^{\gamma\lvert x_i^T\beta\rvert}, \gamma = 0.02$, AR(1), $\alpha = 0.5$						
	$\lambda = 0\%$		$\lambda = 5\%$		$\lambda = 10\%$	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Least square	0.02(0.01)	-0.02(0.01)	-0.72(0.01)	2.42(0.01)	-1.32(0.01)	4.62(0.02)
M-estimation	0.02(0.01)	-0.02(0.01)	-0.04(0.01)	0.12(0.01)	-0.22(0.01)	0.74(0.01)
Generalized least square	0.02(0.01)	-0.02(0.01)	-0.21(0.01)	1.09(0.01)	-0.32(0.01)	1.9(0.02)
Weighted M-estimation	0.02(0.01)	-0.02(0.01)	0.1(0.01)	-0.38(0.01)	0.13(0.01)	-0.51(0.01)
Proposed method	0.02(0.01)	-0.02(0.01)	-0.01(0.01)	-0.18(0.01)	-0.08(0.01)	-0.07(0.02)
$\bar{c}$	2.11		0.93		0.74	
$\sigma_i = e^{\gamma\lvert x_i^T\beta\rvert}, \gamma = 0.02$, AR(2), $\alpha_1 = 0.5, \alpha_2 = 0.2$						
	$\lambda = 0\%$		$\lambda = 5\%$		$\lambda = 10\%$	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Least square	-0.01(0.01)	-0.02(0.01)	-0.69(0.01)	2.37(0.01)	-1.26(0.01)	4.53(0.02)
M-estimation	-0.01(0.01)	-0.02(0.01)	-0.04(0.01)	0.12(0.01)	-0.23(0.01)	0.76(0.02)
Generalized least square	-0.01(0.01)	-0.02(0.01)	-0.22(0.01)	1.18(0.02)	-0.36(0.01)	2.11(0.02)
Weighted M-estimation	-0.01(0.01)	-0.02(0.01)	0.1(0.01)	-0.4(0.01)	0.12(0.01)	-0.51(0.02)
Proposed method	-0.01(0.01)	-0.02(0.01)	-0.06(0.01)	-0.17(0.01)	-0.15(0.01)	-0.07(0.02)
$\bar{c}$	2.11		0.91		0.72	

Fig. 1 Mean square errors of the estimated parameters obtained by several regression methods for the exponential variance function (log-scale). Results based on 500 replications. Gls corresponds to generalized least squares, Lm to least squares, RlmDDh to the proposed method, Rlm to M-estimation and Whm to weighted M-estimation.

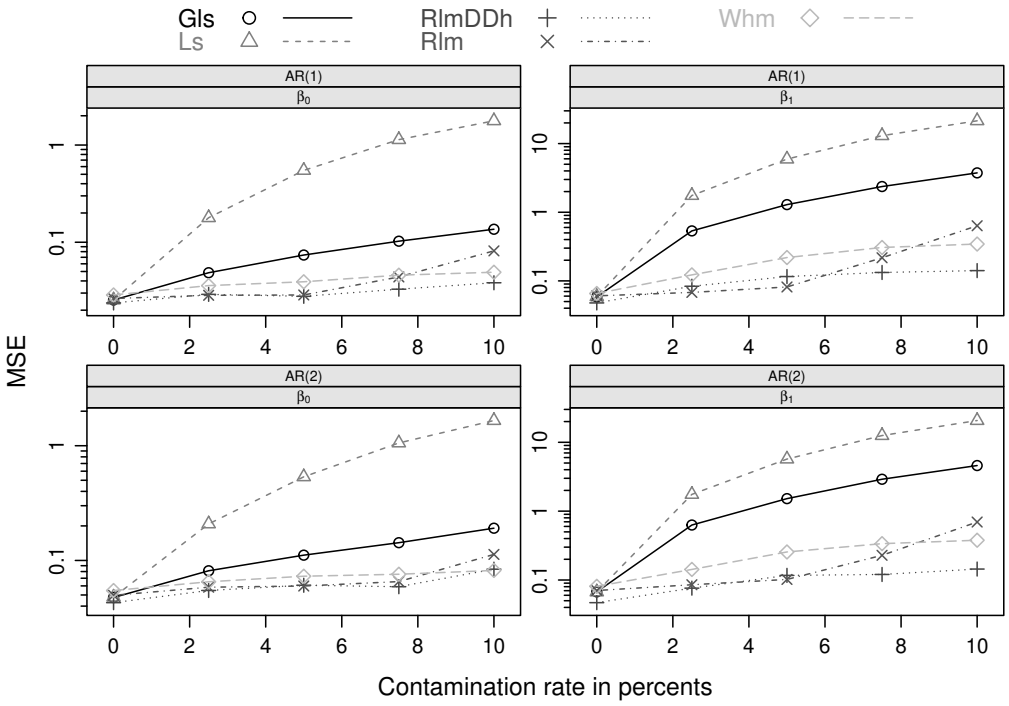


Fig. 2 Time series of the $PM_{2.5}$ concentration with different transformation. The first panel shows the raw data, the second and last ones show the square-root and logarithm transformation respectively.

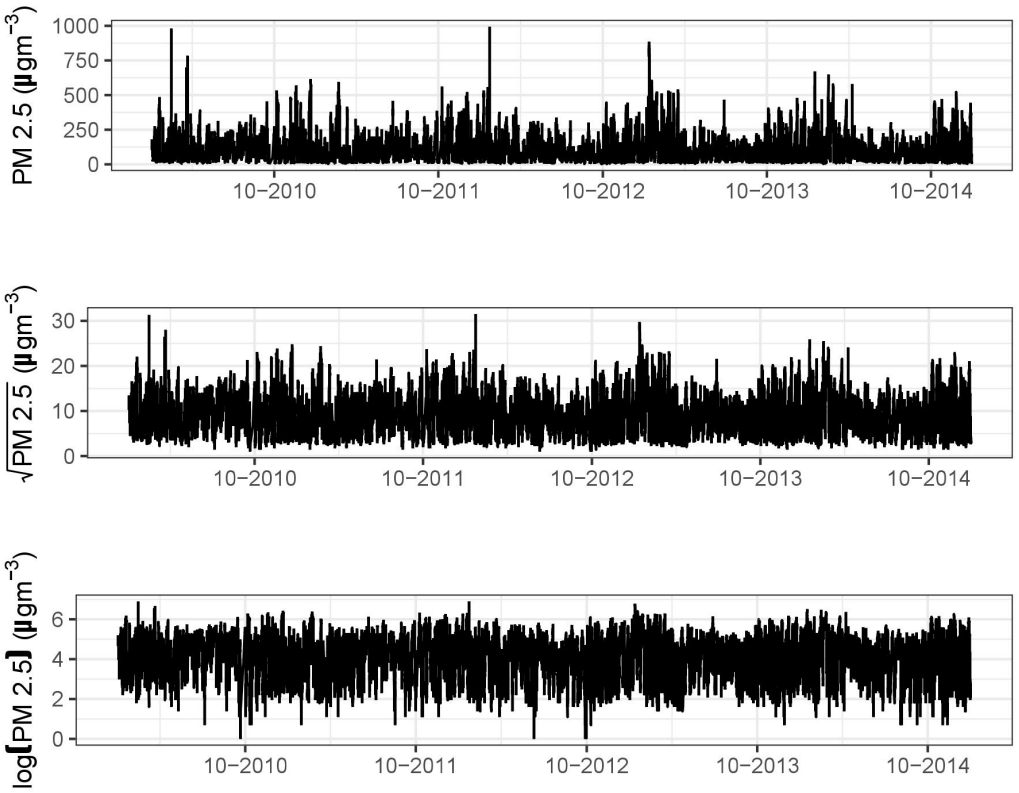


Fig. 3 Regression diagnostic plots for the least squares model (upper panel) and the proposed method (lower panel). In the residuals vs fitted values plot for the method, 6 residuals lie outside the plot area.

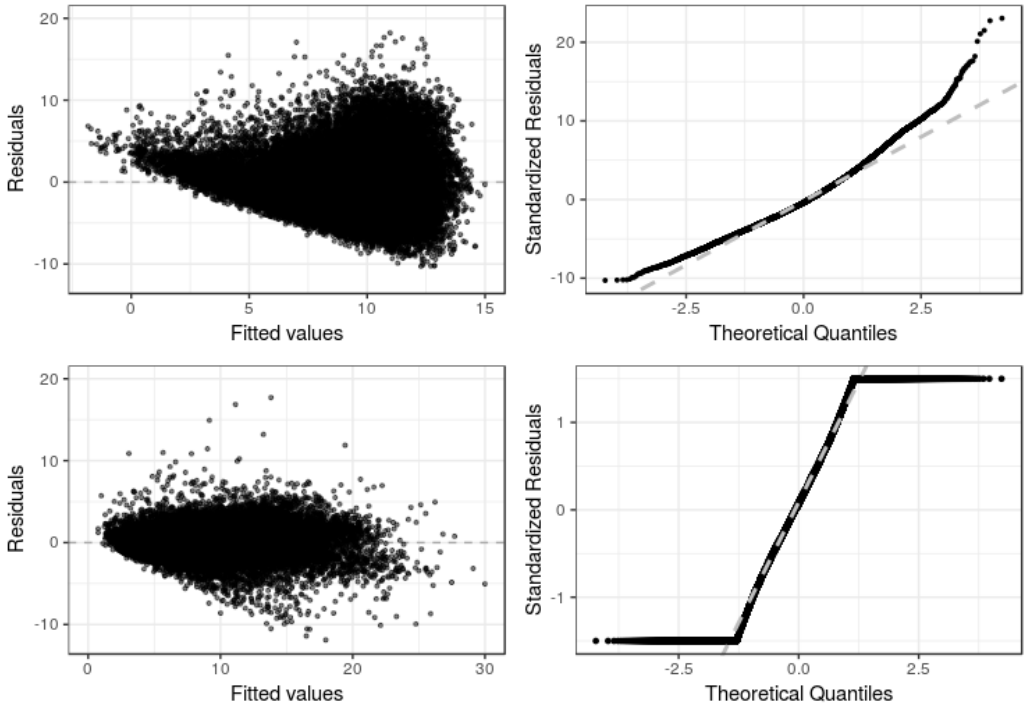


Table 2 Parameter estimates (β), their standard errors and z-values for the least squares and proposed method. The critical region of the significance test is of the full model $|z| > 1.96$.

	Least squares method			Proposed method with $\hat{c} = 1.5$ and $\hat{\sigma}_i = 1.03e^{0.131\|x_i\hat{\beta}\|}$		
	Estimate	Std.Error	Z-value	Estimate	Std.Error	Z-value
Intercept	9.378	0.157	59.765	8.784	0.018	493.763
Dew point	3.214	0.122	26.446	2.953	0.011	268.731
Temperature	-3.978	0.148	-26.845	-3.416	0.013	-260.880
Pressure	-0.882	0.149	-5.909	-0.877	0.011	-77.211
Cbwd NE	-1.455	0.104	-13.952	-1.117	0.021	-52.428
Cbwd NW	-1.716	0.099	-17.284	-1.400	0.018	-77.689
Cbwd SE	0.512	0.081	6.306	0.666	0.019	35.685
Iws	-0.577	0.058	-9.888	-0.410	0.005	-89.071
Is	-0.086	0.054	-1.575	-0.027	0.008	-3.475
Ir	-0.453	0.039	-11.677	-0.380	0.004	-88.028
Policy	0.362	0.343	1.058	0.222	0.023	9.464
Time Policy	-7×10^{-4}	5×10^{-4}	-1.373	-6.10^{-4}	3×10^{-5}	-18.052
Cos 3 years cycle	-0.340	0.157	-2.164	-0.238	0.010	-22.878
Sin 3 years cycle	0.547	0.174	3.155	0.509	0.010	48.993
Cos 2 years cycle	0.206	0.133	1.549	0.345	0.008	43.472
Sin 2 years cycle	0.084	0.123	0.683	0.104	0.009	11.361
Cos 6 months cycle	-0.342	0.123	-2.776	-0.355	0.008	-42.542
Sin 6 months cycle	-0.365	0.119	-3.069	-0.425	0.008	-51.631
Cos 4 months cycle	-0.184	0.120	-1.539	-0.038	0.008	-4.722
Sin 4 months cycle	-0.035	0.121	-0.287	-0.084	0.008	-10.509
Cos 3 months cycle	0.032	0.120	0.263	0.047	0.008	5.947
Sin 3 months cycle	0.025	0.119	0.212	-0.005	0.008	-0.651
lag1	/	/	/	0.746	0.005	162.733
lag2	/	/	/	0.102	0.004	23.260