Facultad de Ciencias

VARIABILIDAD ESPACIAL Y TEMPORAL DEL RECURSO SURF EN CANTABRIA
(Spatial and temporal variability of surfing in Cantabria)

Trabajo de Fin de Máster
para acceder al

MÁSTER EN CIENCIA DE DATOS

Autor: Javier Tausía Hoyal

Director\es: Fernando Javier Méndez Incera
Juan José González Trueba

Septiembre – 2020

Resumen:

El surf es un deporte que ha presentado un crecimiento enorme en los últimos años, aumentando exponencialmente el número de personas que lo practican y disfrutan. En este ámbito, Cantabria es una de las regiones punteras de España, y es por esta razón por la que en este trabajo se busca dar una respuesta cuantitativa al valor y calidad del recurso surf y su aplicación en la gestión y conservación de sus espacios de interés en la región.

Se ha desarrollado un método de reducción de escala híbrido que permite evaluar con precisión las características del oleaje incidente en las principales rompientes del litoral cántabro, siendo este escalable a toda España y también mundialmente. Haciendo uso de herramientas comúnmente usadas en oceanografía, en ingeniería de costas y en ciencia de datos, el registro histórico de variables oceanográficas (altura de ola, periodo...) de los últimos 40 años que se obtiene permite analizar estadística y climatológicamente las playas con mejores condiciones de surf, lo que resulta de utilidad en la gestión y ordenación de los espacios de interés para el mismo, sirviendo este trabajo de base para un posible proyecto de protección legal de las rompientes o surfbreaks.

Palabras clave: protección del surf, hindcast, minería de datos, base de datos de oleaje

Abstract:

Surfing is a sport that has presented an enormous growth in the last years, increasing exponentially the number of people who practice and enjoy it. In this area, Cantabria is one of the leading regions in Spain, and it is for this reason that this work seeks to provide a quantitative response to the value and quality of the surfing resource and its application in the management and conservation of its areas of interest in the region.

A hybrid downscaling method has been developed that allows the precise evaluation of the characteristics of the incident waves in the main surfbreaks of the cantabric coast, being this scalable to all Spain and also worldwide. Making use of tools commonly used in oceanography, coastal engineering and data science, the historical record of oceanographic variables (wave height, period ...) of the last 40 years that is obtained allows statistical and climatic analysis of the beaches with better surfing conditions, which is useful in the management and planning of areas of interest to it, serving this work as a basis for a possible project of legal protection of the surfbreaks.

Keywords: surfing protection, hindcast, data mining, waves database

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

It has been 20 years since the beginning of the 21st century, and if a list is made that includes the fastest growing sports in this short but intense period of time, surfing would be one of them without a doubt. So much so that it will be an Olympic sport at the next Olympics in Tokyo. It should be noted that surfing encompasses a large number of similar water sports, but which actually require different skills and abilities. In this work we will focus on traditional surfing where the surfer catches the waves standing up, and without the help of kites, foils or other instruments... Even so, most of the study carried out can be extrapolated to other water sports that require the existence of waves and wind.

Surfbreaks and surfing in general are becoming more and more recognized in our society. In addition to their natural or intrinsic values, as a heritage that can be preserved due to its exceptional nature, we must add a high added cultural, recreational and sportive value, becoming a socioeconomic resource in Cantabria that has been constantly growing in the last few years, and which has already shown itself capable of providing great benefits to the coastal communities where they are located. Waves are the raw material for surfing, their quality is directly proportional to their demand and capacity to attract specific tourism, with the particularity of being a model that contributes to break the seasonality of the "sun and beach" model, and which existence does not generate any impact on the environment where it is developed. However, it is necessary to consider its dual value as a heritage and a resource, in order to shape policies that tend to guarantee its conservation. Its loss or degradation would also translate into the loss and degradation of the natural heritage of Cantabria and of the quality and capacity of an attractive resource of the cantabric coast in general.

Waves are an essential element of the sea. Their presence is a significant component, icon of the ocean and coastal landscape. They are part of its value and condition. Despite having an ephemeral existence, they are a basic element in the marine and planetary energy balance. The energy of the sun transmitted to the atmosphere in the form of wind and storms that generate the waves, propagates in the form of waves that after a long and hazardous journey through the high seas reach the shallow waters of the coast, and in combination with other agents, discharge their energy by shaping the coasts. They do not exist only for our benefit. They are a fundamental agent in the natural system.

Their singular, ephemeral, dynamic, fragile and changing character, demands an integral treatment and the consideration not only of the waves themselves, but of all those conditioning factors that participate in the breaking of the waves, which can be integrated under a more global concept such as the surf break or "Surfbreak". A surf break is defined as an area in which factors such as the swell from the open sea, currents, sea level and variable depth associated with the tides, the seabed and wind, interact to give rise to the formation of a surfable wave. When considering the operation of a surf break, it is necessary to include the swell corridor, located offshore. Thus, a surf break is that strip of the coastal environment in which the combination of marine hydrodynamics, meteorology and coastal morphology generate waves with a form of break suitable for surfing (González Trueba, 2012). By surfable waves we mean those waves that present a way of breaking such that from an initial point or "peak", the wave offers a stretch of unbroken crest that does so in a progressive and continuous way from the peak, allowing the surfer to take it and slide through the wall of the wave in a lateral way (generating right and left waves). Bearing in mind that most coasts do not produce good surfable waves, their existence makes them an exceptional and unique element.

FIGURE 1.1: This lefty wave was legally protected in 2016, being the first legally protected wave in the history and settling the beginning to this surfing legal revolution. Source: https://www.surferrule.com/la-ola-chicama-peru/. Paper: https://olasperu.com/blog/resolucion-chicama-la-ola-mas-larga-del-mundo-legalmente-protegida/8003.

All this and more, makes up the intrinsic or natural value of the surf breaks. But in addition, the human being has given them wide and deep senses and meanings, a whole cultural deposit.

It is necessary to carry out a process of patrimonialization and conservation: To turn an element into a heritage is to show and justify its value as such. Together with this, it is a question of showing the particularities of an element such as the highly complex, dynamic and fragile surfing breaks, in which a large number of components and factors interact. It is not about protecting the waves in an isolated way, with their particular specific character, but the set of elements that cause the unique break of those waves. Waves are an ephemeral phenomena of the hydrosphere, they are not present in a continuous way in space, hence the conceptual complexity and the need to resort to a more integral term and phenomenon such as the breaker, in which

the parameters of the sea bed and the underwater topography, physical oceanography and marine hydrodynamics, meteorology or the entrance corridor of the wave towards the breaker, among others, interact. Not all waves are surfable, only some have a unique configuration that makes them highly valuable and rare elements. The world's coastline is occupied only by its 20 %, making them highly valued spaces on a scale that ranges from local to international. The exceptional characteristics of the waves are quantifiable in parameters such as the length of the breaking wall, mode and angle of opening, associated to the type of refraction of the wave line, the size of the waves, from small to huge, the shape or profile of the wall of the wave and their energy and breaking speed are all qualities that condition the type and degree of surfing and the exceptionality of the natural phenomenon. Furthermore, beyond their intrinsic or natural values, they are endowed with an extrinsic or added value, especially appreciated and valued for recreational, sports practice or their aesthetic contemplation. Just as not all the mountains have the same scenic value or attraction for mountaineering or climbing, not all the waves and breaks have the same values and meanings. There are some exceptional ones that need to be preserved.

FIGURE 1.2: This right wave is considered one of the best waves of the region by the local surfers (it is located in Cantabria). Santa Marina, as this is the name the surfbreak receives for the island where it breaks, offers very highly rated waves that attract the interest of a lot of surfers along the year, achieving the realization of professional competitions as it is the case of the photo, where Aritz Aranburu is catching a 10 point wave to win the championship. Source: https://www.federacioncantabradesurf.com/aritz-aranburu-renueva-su-trono-en-la-isla/.

The protection of the breaks of interest for surfing with its own legal entity, through different figures of legal protection according to the existing scopes and legislations, has been a reality for a decade in countries like Australia, USA, New Zealand or Peru. In Australia, the first "national surfing reserve" with its own legal entity was declared four decades ago, and at the beginning of the 21st century the list has increased to around twenty enclaves. In New Zealand, in the last revision of the National Coastal Law, the surf breaks or "surfbreaks" have been included as a natural space representative of the national coast, and as such, susceptible to be managed

and conserved in the face of possible threats and impacts. In Peru, a South American country with a long tradition of surfing, and given the high number of breaks and their interest as a heritage and endogenous resource for coastal communities, a national law has been declared to guarantee their conservation and management. In the USA, for the last few years, a line of declaration of some exceptional surfing breaks as UNESCO heritage has been established. In Europe, progress is beginning to be made in countries such as Portugal, the United Kingdom and France, which share with Spain the most valuable and unique coastal strips.

In our case, in Spain, and specifically in Cantabria, during the last years a growing social demand has been formed by the main agents and entities of the Cantabrian and Spanish surfing: national and regional federations, clubs, associations, schools, shops, factories, brands, groups, NGOs..., which has favoured the publication in 2012 of the *Manifest for the Protection of Waves*, and the creation of the first Surfing Reserves in Spain and Europe declared by several Spanish municipalities. The first three "surf reserves", understood as a name that the local administrations approve in their municipal plenary sessions to highlight the high heritage and strategic value of these natural areas, were declared in Cantabria: Municipalities of Ribamontán al Mar (2013), Suances (2013) and Noja (2015). Later in the Basque Country, the Mundaka reserve was declared (2015) and again in Cantabria, the San Vicente de la Barquera Town Hall, at the proposal of the Cantabrian Surfing Federation and the Saja - Nansa Local Action Group, declared its breaks as a "surfing reserve" in 2016. Such initiatives are in line with a future horizon that has already begun in other autonomous communities and that opens the horizon to initiatives such as the "Surfing Europe Plan of the European Commission", a line of action of the "Blue Economy" that promotes the development of sustainable activities with the environmental quality of the European Atlantic coasts. Given the unique nature of some interesting surfing breaks in the Cantabrian coast, which constitute a natural and cultural heritage, as well as their high value as a recreational, sports and socioeconomic strategic resource for our coastal municipalities, the need for detailed knowledge, management and conservation of these natural spaces in the face of possible threats and damage is evident.

In this context, demand and interest, this work is framed to advance in the knowledge of the surf resource in Cantabria, analyzing the spatial-temporal variability of the waves and the surfability of its breaks. In this way, the regional government has entrusted the University of Cantabria with the project: *HYDRODIVERSITY, BLUE HERITAGE AND SUSTAINABLE DEVELOPMENT: DIAGNOSIS AND PROPOSALS FOR ACTION ON SURF IN CANTABRIA*. This work is focused on Cantabria, although it has a potential projection for the whole country, given the existing information gap. In the current context, the first drafts and working documents of the *Plan of Ordering of the Spanish Maritime Area* are being prepared by the sub-directorate of Sustainability of the Coast of the *Ministry for Ecological Transition and the Demographic Challenge*.

In this way, a method capable of carrying out this study of the surfing conditions with break precision on the coast of Cantabria is presented in this work, and this is where the second existing motivation for carrying out this work comes in, because as will be explained below, the existing techniques for carrying out this process require enormous data processing, where more or less complex techniques of both data mining and machine learning have to be applied.

The field of artificial intelligence (AI) is developing by leaps and bounds since the computer

FIGURE 1.3: Ribamontán al Mar (Somo - Langre - El Puntal, among others) zone. This zone was declared the first spanish surf zone in 2013. Source: https://www.cantabriarural.com/que-ver/ajo-isla-noja-somo/ribamontan-al-mar/.
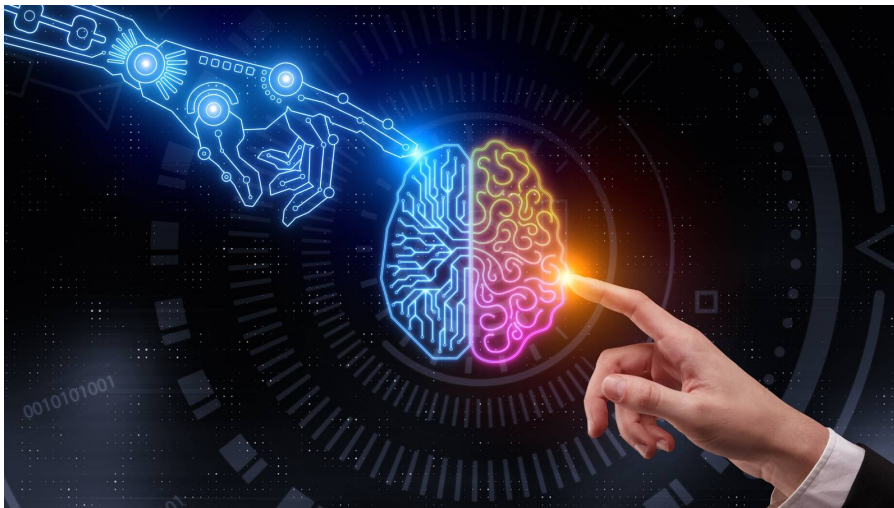

FIGURE 1.4: The importance of the artificial intelligence revolution and its relationship with humans is underlined. Source: https://www.martechadvisor.com/articles/machine-learning-ai/we-are-not-robots-5-ways-leading-companies-combine-ai-and-human-intelligence-2/.

revolution in the recent decades. The enormous developments in the semiconductor materials industry have generated a giant boom that has allowed technology companies to create increasingly powerful and, more importantly, cheaper computers, which is what allows more and more people to have access to large machines with which they can carry out their revolutionary ideas. This change in the way of seeing technology has made the already existing neural networks (invented around the middle of the last century with the creation of the perceptron by Frnak Rosenblatt) take on a gigantic importance, thus passing from the winter of artificial intelligence to the golden age of the last few decades. Advances have been enormous, in some cases even recognizing or creating totally human faces, or winning the world champions of Go or chess (Alpha Go and Deep blue AIs respectively). This is the new world and companies as Google, Microsoft or Netflix know that, so the principal job new offers of these important companies are always related with this new branch of technology, the Artificial Inteligence.

But artificial intelligence, machine learning or even deep learning must not be seen as black boxes that work or as things that only those super intelligent programmers can understand. Nowadays, the development of this new kind of science offers an overestimated vision of the reality. Machine learning and data mining techniques are just additions and multiplications, sometimes some nonlinear equations, that try to imitate the functioning of the human brain. The functioning of the brain is not explained yet, but we know that neurons receive information and pass information and in these processes, they learn. This is what neural networks do, they receive information, pass information through additions and multiplications, and in this process, they save that information that is useful, or if talking of a computer, that is the solution of those additions and multiplications. The purpose of this short paragraph is to show how these "intelligent" algorithms are very important and powerful, but also very understandable and friendly.

It is for these two reasons (protection of surfing and machine learning) that this study has been decided to be carried out, and although not yet finished, that presents very good ways of becoming part of this new quantitative revolution when it comes to demonstrating the real importance of beaches in Cantabria, Spain and the whole world.

1.2 State of Art

The existing literature presents different possible ways to perform this study in a practical manner, although the dynamics used can be more or less summarized in (Camus, Mendez, and Medina, 2011; Camus et al., 2013). In the world of surfing, the normal surfer consults daily the different web pages of forecasts to see if good waves can exist or not in the beach (technically speaking surfbreak, as the beach is just the morphodinamic element, but it does not generate waves by its own) to which he wants to go, but the problem with these web pages is that the forecasts, being global, present a very bad precision in many areas of the world as it is the case of Cantabria. The problem with this global computation is that it is not possible to make a study with a very good precision and at the same time in all the beaches of the world, and this is why many areas that do have good surfing qualities, do not have good predictions. The main problem that this generates is that in these areas of bad predictions, or very global predictions

(regional scale), the sportsman has to learn to read this part or global prediction, and to use it in an intelligent way, knowing that in a certain surfbreak, the part is going to be similar to the one that shows the magazine or web page that is consulted, but that in some other spot, the conditions are going to be very different. In this work, it has been performed a detailed historical study of the main surfbreaks of the cantabric coast, and of course this can be extended to each and every one of them. This study will be extended to predict future surfing conditions in one or two weeks too, although this task is still pending. In this way, the surfer with a desire to catch good waves, whether amateur or professional, will be able to consult whether the waves on the beach where he/she is or where he/she intends to go will be of good quality, without depending on prior knowledge of the area, as he/she may not be from that town or region, or may simply be starting out and not know how to read the current surfing reports well. This historical study will be the perfect key to establish the required quantitative level to the protection of the surfbreaks natural heritage.

In order to carry out this project in an orderly manner, different stages have been proposed, which are commented on below.

Firstly, the variables that are taken into account when thinking about a swim or surfing session with good quality are selected, as this will be the main objective, to obtain an index or numerical value representative of each historical moment according to the values of these variables that will allow us to see on which beaches (spatial variability), and in which months or seasons (temporal variability), the waves have been good or bad. To do this, it will be necessary to have a rigorous oceanographic database ("CAWCR Wave Hindcast 1979-2010"), as this will be the basis on which the whole study will be built. This main database will be entirely formed by the historical data of the last 40 years, will be provided by the CSIRO wave hindcast, and will be corrected by a satellite calibration method that will be explained later.

Once the database has been calibrated (Mínguez et al., 2011; Albuquerque et al., 2018), it must then be validated with data from existing deep water buoys, and then a hybrid (statistical and dynamical) downscaling technique will be developed to obtain this time record for each of the desired beaches on the coast, this method will be also explained later. Finally, but more importantly, and once the historical surfing conditions in the main points of interest (surfbreaks with importance according to the Cantabrian surfing community) are known, a surfing quality index will be developed, both spatially and temporally, which will give an idea of when and where the waves are good throughout Cantabria. Once this index has been obtained, the pending task for future research will be to be able to make wave forecasts at these same points, providing these predictions with this same assessment, so that as has already been mentioned, the surfer will be able without the need for prior knowledge, to know if on a certain day at a certain time, he or she will be able to take a bath on a certain beach.

1.3 Objectives

Both for Cantabria and for the whole of the Spanish coast, the different public administrations find a gap in knowledge regarding the surfing resource, the surfing breaks are not inventoried in official sources and nor is their different degree of interest or level of surfing been studied. There

are surfing breaks with different levels of interest, from local scale, to national and international. Quantifying such quality and surfability, and characterizing the spatial-temporal variability of the surfing resource has been one of the objectives of the present work.

Thus, the main objective of this study is to quantitatively demonstrate the importance of surfing in this region, but the other important objective is to provide the surfers with a tool with which they can know as accurately as possible whether or not waves will be encountered at their nearest beaches. Nowadays, most websites or surfing magazines, being the first ones by far the most used by the highest percentage of the surfing community, base their predictions or forecasts on numerical and statistical models that compute globally all over the world, making special emphasis on certain areas of "good" conditions for surfing such as those so worldwide known for their "surfer" culture Hawaii Islands or many areas of the Pacific, because it is in these places where we find the best waves, and where most of the world circuit championships take place of course. This study that is thought to be able to be performed all over the world, wants to eliminate this discrimination.

Another important consequence of this study and from which the user can benefit is the discovery of new spots and I explain myself. Today, many professional surfers spend hours in front of the computer trying to find new waves along the entire world coastline, this process is usually carried out using web applications and services such as Google Earth, because it allows you to see if on a beach, estuary or cliff, break or not waves, and also allows you to see how they break these waves, because you see the angle that the wave forms with the coast, but this method has many gaps as expected. Now, if that surfer could get a historical record of 40 years with high spatial resolution that would allow him/her to assess the surfing conditions at that point where he has seen that there may be a possibility of finding a new spot, then things change drastically.

Chapter 2

Methodology

The following chapter defines the methodology, preliminary concepts, variables and data resources used as well as the general solution to the main objectives of the project.

As previously mentioned, the main illusion of this work is to give the surfing a quantitative protection and also to provide the regional surfer with a tool capable of analyzing, and in the future predicting, good or bad surfing conditions anywhere they break waves, and all this in an optimal way (Wright and Short, 1984). To do this, first we must define what it means to find good conditions for surfing, and this means explaining each and every one of the different variables that have been used and that are very important in a surfing session. In this way, these variables will be explained below, so the reader can understand how it is known today when a session is going to have good or bad conditions.

FIGURE 2.1: Bad (left) and good (right) surfing conditions. Source: (Espejo Hermosa, 2011).

But first, some preliminary concepts must be explained so the other parts can be correctly understood, which can be found in Appendix A.

2.1 Parameters that define surfing at breaker scale

The term surfability, applied on a breaker scale, refers to the degree of suitability for the practice of the sport of surfing. A detailed review of the parameters that govern wave surfing in the breaker zone was conducted by (Black et al., 2003). The three most important parameters for wave characterization are: wave height at break, break or peel angle and break intensity. In addition to the above-mentioned parameters, the following factors have been incorporated in this section as variables that may affect the above, these are the wind and the cleanliness of the

swell, whose effect on the quality of the wave breakage can be decisive. The peak period is also very important in the quality of the breaking waves, so it has been already described in the Preliminary Concepts section (Appendix A), and in the surfbreak, it is not more than the propagation of this variable to shallower waters, which will be explained in detail in the next chapter.

In this section, the transfer from deep water to the surfbreak (notice that this is not the same as the beach) and its effect on the waves will be briefly explained, as our main goal is to obtain the quality of the waves in this point.

2.1.1 Wave height at break, H_B

Returning to the height of breaking waves, of interest to surfers, this will not be more than the transfer of this significant wave height in deeper water to coast, plus the effect of a possible slope on the beach at that point, the reef if we find a point break or the shape of the estuary when this is the case. This is due to the fact that waves, which more or less propagate gently from deep water to shallower areas with phenomenons such as refraction and diffraction, undergoes a very abrupt final change when they really approach the coast (less than 10 m depth approximately), suffering brutal refractions and a drastic change in the height of the wave, since the bottom of the beach at that moment makes it break, generating the surfeable wave (H_B, see Figures A.1 and B.1). This final step is described by Equation 2.1.

$$\gamma \, h_B = H_S \, \left(\frac{h_1}{h_b} \right)^{1/4} \sqrt{ \frac{\cos \theta_1}{\cos \arcsin \sqrt{\frac{h_b}{h_1}} \sin \theta_1} } \tag{2.1}$$

where must be noticed that this equation will be solved using a numerical method and its solution is the wave height at break, H_B, as our unknown quantity is h_b, the depth of the profile that the wave sees while approaching shore. For the rest, H_S is the significant and propagated wave height, already known, h_1 is the depth at the propagated point, θ_1 is the mean direction of the waves before the final refraction occurs (θ_m, explained below) and γ is a variable parameter that modifies the solution of the equation, it depends on the characteristics of the surfbreak and the waves. The propagation to this approximated 10 m depth point is essential, as the main variables already described could suffer significant changes, the wave height, the period and the direction are affected by refraction, which is shown in Fig 3.10. Thus, propagating the wave to a point near shore and then using this equation, the wave height at break is obtained.

2.1.2 Break type and intensity, I_R

Once the main variables involved in a surfing normal session have been understood, fact that can be checked in any surfing page or magazine as windguru, since they are the standard characteristics or predicted variables that everybody looks at before going to the beach. It will be explained below how an equation that takes into account several parameters is able to give a value of the type and intensity of the breaking wave or what is the same, to the shape of the wave when it breaks. This "magic" equation is shown and explained below:

TABLE 2.1: Different types of breaking according to its *Iribarren number*.

Break	Range
Spilling	$I_R < 0.4$
Plunging	$0.4 > I_R < 2$
Collapsing	$2 < I_R$

$$I_R = \frac{\tan \beta}{\sqrt{\frac{H_B}{L_0}}} \tag{2.2}$$

where β is the slope of the beach, H_B is the already mentioned breaking wave height and L_0 is the wavelength. This magic number receives the name Iribarren number for its creator. The slope of the beach, although commented on previously, has not been explained in detail yet, and it is time.

The morphology of the beaches is constantly changing due to different factors, but one of them is the height and period of the waves. In this way, when in the winter season the beaches suffer, or at least on the cantabrian coast as it is our place of study, strong and constant attacks of energetic incident train of waves, the morphology of the beach responds in an expected way. This is presenting a very flat and constant beach profile, which in coastal engineering terminology is called a dissipative profile, being not capable of generating good breaks in the incident waves and dissipating their energy. At the other extreme, when the summer season arrives, the beaches are less exposed to these strong waves, and this makes the sand pile up more in the areas near the shore, causing steps in the profile that will make good waves in many cases when conditions allow it.

This is the explanation why in months like September there are mostly every year good conditions, as more energetic waves and an orderly and more steep beach profile, or scientifically speaking reflective profile, coincide. The slope of the beach, which in beaches of this region is around 0.02, or what is the same, a height of 2 meters of elevation when advancing 100, will be a determining factor in the type of the surfable wave (Moragues, Clavero, and Losada, 2020).

As for the denominator of Equation 2.2, already explained as H_B is the wave height at break and the wavelength is just $L_0 = \frac{g}{2\pi} T_P^2 \approx 1.5 \cdot T_P^2$, this gives an idea of the waves at the moment of breakage. All this can be summarized in three groups or types and intensities of waves according to their breakage clearly differentiated, see Table 2.1.

- **Spilling**: This type of break is characterized by the formation of waves with a little and difficult wall to surf as it does not allow too many maneuvers. This type of wave usually occurs on beaches with a very flat and dissipative profile, which is not able to perk up the waves enough and therefore makes them end up breaking by inertia without form or foundation. This type of waves are not that bad in sports such as longboarding, kayaking, windsurfing and kitesurfing, although they can also surf any other types.

- **Plunging**: The *wave*. What every more or less experienced surfer is looking for in his/her daily life, a wave that allows him/her to do both maneuvers and catch barrels, which is the

maneuver par excellence in surfing. This is the most difficult type of wave to surf, as it requires certain ideal conditions that do not just involve presenting an adequate value in this magic parameter, and are easier to find in point breaks or reef breaks than in sandy beaches (beach breaks), but everything is possible. Notice that the majority of the beaches in Cantabria are sandy beaches, nevertheless, we also have point breaks with very good and powerful plunging waves such us Santa Marina o La Vaca Gigante.

- **Collapsing**: When the profile of the beach is very steep, or the change occurs suddenly, then the wave can collapse and break quickly and abruptly, without giving the surfer enough time to perform the relevant maneuvers. This regime is common on the reef bottom, when the tide is too low and the bottom is not sufficiently submerged. This type of breakage is not considered surfable, although those waves with intermediate breakage between volute (plunging) and collapse are preferred in bodyboarding mode. In our case, the beach with the most similar conditions could be the Sardinero beach at low tide, where we usually observe waves that break very fast and with difficult tubes, although these could be very good conditions for those experienced surfers that can manage to finish their maneuvers.

- **Surging**: The last type of break could even be classified as no break, since it includes those waves that are not able to see the bottom, and reach the shore without breaking, oscillating until they crash without the capacity to transfer their energy to a clean and surfable wave. Figure 2.2 summarizes what has been explained in this subsection.

FIGURE 2.2: Spilling and plunging waves, from up to down. The other is a croquis where the different types of breaking are simply explained. Source: https://en.wikipedia.org/wiki/Breakingwave.

2.1.3 Wind intensity and direction, W and θ_W

Another very important aspect to consider when evaluating the quality of a surfing session is the wind, both in magnitude and direction. The wind is the only one responsible for the creation of the waves, as they are generated when strong winds blow for long periods of time and manage to generate these waves that, after long distances, are capable of reaching the coasts

of the entire world for surfers to enjoy them. In Figure 2.3, it can be seen a wave generation area where the wind is blowing intensely, generating waves that will be surfable in the newt week/s.

It should be noted that this process of generating waves is not at all fast, the large swells that are surfed in islands such as Hawaii in the months of December-January have periods of creation of even days, and once generated, these swells can take another few days to reach the coast, which would allow the surfer even to travel to the islands simply because he has seen that the waves will be good the next week, which would have been impossible, or at least more difficult a few years ago.

FIGURE 2.3: In the left, a waves generation area can be found while in the right, a good and mature swell is reaching shore. Source: (Espejo Hermosa, 2011).

But the wind is not just the main factor in generating waves, it is also one of the main factors in determining how waves die on shore. The wind is capable of making the waves break very cleanly, even generating waves with tubes when the conditions described in the previous subsection were not totally ideal, but it can also spoil a good session if its magnitude and direction are adverse. In this way, "good" wind is defined as that which presents a direction totally opposite to the orientation of the beach, or *offshore*, and a magnitude not too great, since very strong winds can make the waves take a very long time to break, even making them collapsing. On the other hand, it is said that there is "bad" wind when the direction of the wind is towards the beach, or *onshore*, regardless of its magnitude, although as is evident, the greater the magnitude of the wind, the greater the negative influence on the swell, see Figure 2.4.

FIGURE 2.4: Bells Beach beach with bad (left) and good (right) wind conditions and similar wave height and period conditions. It can be seen how the effect of the wind is so important in the daily training of a surfer. Source: (Espejo Hermosa, 2011).

2.1.4 Train of waves cleanliness or dispersion, σ_θ

It has been previously mentioned above how frequency dispersion affects the arrival of waves in groups of frequency or period similar to the coasts of the whole world, but this dispersion phenomenon, which can also be seen as a grouping, also affects the direction of the waves. When waves are generated in open sea, they travel lengths that can be hundreds or even thousands and tens of thousands of kilometers until they reach a beach where they can break, and in this course, the waves are grouped together in a directional manner, suffering both radial and angular dispersion, even presenting very well defined directions when the distance traveled is large enough.

In this way, when the direction is well defined, then it can be inferred that the swell has travelled several kilometres, being a ground swell generated far from the coast, and that it is not affected by sudden creation phenomenon, but that it is already a mature passage. It will be valued as positive, because the greater the order, the greater the cleanliness, and this is what is sought in a good surfing session. This directional dispersion can be observed in Figure 2.3, as the swell propagates very defined, but it can also be seen in Figure 2.5.

2.1.5 Wave peel angle, $\delta\theta$

Finally, the angle of break (can be also seen as the relative direction of the waves to the coast) will be defined. The relative direction of the wave to the beach is estimated in this study by subtracting the mean direction of the wave, θ_m, minus the direction of the beach (i.e. NNW for Liencres or E for El Sardinero), thus giving an idea of how the wave will break, see Figure 2.5. This is maybe not the best approximation as we are supposing the bathymetry as straight and parallel to the shore and surfbreaks are sometimes differently oriented than their respective beaches, but it works as a first try although more precise work will be done. Therefore, waves that present very high relative directions will have very stretched walls and will be easy to surf, while smaller directions will present faster walls and will require a higher level for the rider, even more if we add to this a type of break in volute.

In Figure B.1, the refraction phenomenon can be observed, thus giving an idea about this peel angle when the wave breaks. This parameter is very useful, as it can give the surfer an idea about the side of the wave, as they can be rights or lefts, but also about the difficulty of this wave.

Ending with this brief introduction to the variables used in the study, simply mention that this common quality index that we are trying to find will also depend on the surfer to which it is destined, and therefore a single global index will not be ideal, but rather an index that depends both on the quality of the surfer and his or her abrasion to risk, as well as the type of wave. If we put these two factors together and think of an index for each of the possible scenarios, a total of about 15 different indices comes out, which although necessary, would present common characteristics. For this reason, and due to the lack of time and resources, this work presents a single global index that will be tested only on sandy beaches and some more complex bottoms, leaving as a future task the implementation of all the existing casuistry.

FIGURE 2.5: Both images give an idea of how waves behave near shore, forming an angle of break known as peel angle in their path to die. Here $\delta\theta$ is defined as β, the other variables used are trivial. Source: https://www.surf30.net/.

2.2 Parameters estimation along the coast

Understanding the variables involved is fundamental, but even more essential is the use of a robust and comprehensive database that allows all data mining techniques, as well as the numerical models applied to obtain consistent and contrasted results. To this end, three main databases have been used in this work, with the overall objective of creating a user-friendly flow of information. This ensures that the data, which in many cases is very dispersed, as the same variable can be found in a million different ways in different databases, has a common structure and is easy to handle. The three main repositories are a wave reanalysis information, satellite data and buoys. These real and virtual measures are located as it is shown in Figure 2.6 and their properties and usage is described below. Have an eye at this picture, as it is essential in the correct understanding of the entire study.

2.2.1 CSIRO wave reanalysis

This is the most useful tool when a coastal engineering task is carried out. Reanalysis, commonly named as hindcasts, provide historical well temporal and space spaced data of different oceanongraphic variables as the significant wave height, the peak period, the wind magnitude and direction... It must be said that these hindcasts are obtained using numerical models, forced by some initial conditions and that solve equations numerically, so they are not exact.

For this case, the reanalysis used is CSIRO, ("CAWCR Wave Hindcast 1979-2010"), which has a resolution of 0.4 × 0.4 degrees worldwide and has been obtained using the WaveWatch III v4.08 wave model forced with NCEP CFSR hourly winds.

This numerical approximation has a temporal coverage of 40 years, and just one virtual node that acts as a virtual buoy has been used. This virtual node is located in -3.6ºW, 44.4ºN, in the Cantabric Sea, see Figure 2.6.

This wave reanalysis will contain information of the significant wave height, H_S, the wave peak period, T_P, the wave mean direction, θ_m and the directional dispersion, σ_θ, for four different partitions at each time (hourly calculated data) and also the bulk or total values of these variables. This means that at each time, the hindcast gives us information of four different train of waves reaching shore, although NaNs (Not a number) can sometimes be found. The first

FIGURE 2.6: Explanatory map for the different databases and resources used. Detailed methods will be explained in the next chapter, but it is shown as an introduction to the methodology followed.

partition is always related to the windsea (waves with the same direction as the wind) and the other ones could be different swells. This reanalysis give us also information about the wind, the pressure... and a lot of other interesting variables that will be used or not depending on its applicability to the study.

2.2.2 Satellite data

Another key part in the development of this analysis is the use of satellite data. The only source of instrumental information that allows global and equal train of waves data is the one coming from the satellites, which remotely (with orbits of the order of 1000 km above the earth's surface) are able to determine parameters as well as the hindcast or the buoys do (Caires and Sterl, 2003). The wave reanalysis should be also equal all over the world, but it is not always the case.

The satellites can carry different instruments, but it is the radar altimeters those that work best and have been used for the longest time in measurements of waves. At present, reliable data is available from various satellite missions that have been launched and replaced over the years. With these satellite data, insisting that they are homogeneous throughout the globe, certain imperfections in the wave hindcast used as a basis for the study can be calibrated using a statistical model, since this is a numerical model that does not take into account the observed reality and is not perfect. A lot of other interesting and related investigations can be done using satellite imagery (Vos et al., 2019).

2.2.3 Buoys

This data will be our ground truth, and it will be what we will compare all the data obtained in the course of the investigation with. The buoys represent the best possible way to make an in situ measurement of sea level, and in this way, they are able to obtain historical records of wave height, peak period, mean wave direction, wind speed and direction... This is why the buoys are such a fundamental piece, because if after calibrating the previously obtained data with a super complex numerical model, the buoys show measurements far from what that sophisticated models say, then all the work becomes useless.

So if buoys are perfect and hindcasts are not that good, Why do we use them as the basis for the project? The answer to this question is simple, buoys are very accurate, but they are very disperse. This study takes into account data from around 40 years, and it wants to be made so it can be repeated all over the world. Buoys are very disperse, even inexistent in some regions, commonly in the south hemisphere, and they usually stop working, have to be repaired or maybe just left the place where they were located and get lost due to strong energetic waves, big currents... These are the reasons why wave reanalysis is used, it proportionates a continuous temporal and spatial representation of waves worldwide, and while computers can suffer an electric issue while computing the numerical models, it can be more or less easily solved (Reguero et al., 2012).

2.2.4 Bathymetry

Bathymetry is another key piece of the puzzle, as it is essential in the propagation of the waves from deep to shallower waters all along the cantabrian coast. This bathymetry has been proportioned by the SMC (Sistema de Modelado Costero in Spanish) and the way this was obtained is described in (González et al., 2007). It is long enough valid to made this propagation possible as the resolution of this bathymetry is perfectly accurate for the purpose. The numerical model used for the propagation is valid until a certain region where the waves stop behaving normally. At this step, the model stops being accurate, and it is not until this moment when the resolution of the existent bathymetry starts being not good enough. Another tools will be used for this last propagation step, see Equation 2.1 and 3. All the examples of the bathymetrys used for the propagations are shown in Figures 2.8 and 2.7, where the entire cantabrian coast is covered.

2.2.5 Sediment grain size

Related with this last propagation task appears the sediment grain size. As it was described in the previous pages, the morphology of a beach depends roughly on the incident train of waves, more precisely on the incident waves of the last month (Bernabeu, Medina, and Vidal, 2003). These waves change the shape of the beach, as it was also mentioned, and they also modify the size of the sediment grain size of the beach and the dynamics of the surfbreak. Knowing these two facts and using techniques developed and explained in the previously mentioned paper, including Equation 2.3, a method to calculate the slope of the beach using just the sediment grain size and the incident train of waves in the last month has been developed. This gives the study the ability to solve the problem of the last propagation process, as it simulates the
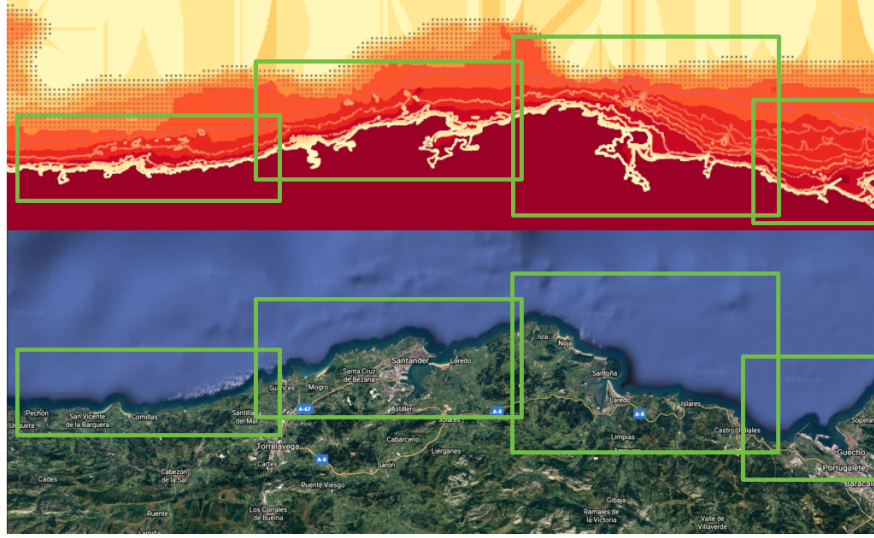
FIGURE 2.7: The total region is shown with smaller areas highlighted as they represent the subareas where the propagations have been done.

waves breakage correctly, see Equation 2.1. In this range, the next equation can be presented, Equation 2.3, as it will be the key point in the calculation of the beaches slope:

$$\Omega = \frac{H_B}{w \cdot T_P} \tag{2.3}$$

where w is the sediment velocity fall, calculated using the relationships seen in Figure 2.9, been D the sediment grain size, commonly known in a lot of scientific studies as D_{50} (Narra, Coelho, and Fonseca, 2015).

2.2.6 Beach profile estimation along the cantabric coast

Once the results propagated to coast have been validated and the previously mentioned spectra have been reconstructed, the last step to be taken to obtain the time series of the last 40 years of breaking wave heights, H_B, Iribarren number, I_R, relative wave direction $\delta\theta$, relative wind direction, $\delta\theta_W$ and directional dispersion, σ_θ (variables used in the estimation of the surfing index), is to estimate the profile of the beaches, obviously making use of what was explained in the sediment size part of the previous chapter and whose detailed explanation can be found in (Bernabeu, Medina, and Vidal, 2003).

As already explained in this section, the morphology of the beaches varies continuously due to the effect of the waves, see Equation 2.3. Using this Ω parameter and making use of the equations described in the previously mentioned paper and explained in detail below, it is possible to calculate a biparabolic estimated profile in a numerical way and which more or less represents for each moment of the year, taking into account that the dynamics are highly influenced by the previous month, the bottom existent in all the beaches of Cantabria. In this way, and using Equation 2.1, it is possible to calculate the slope of the profile at the point where the wave breaks, being then able to infer the Iribarren number, which will contribute the cherry
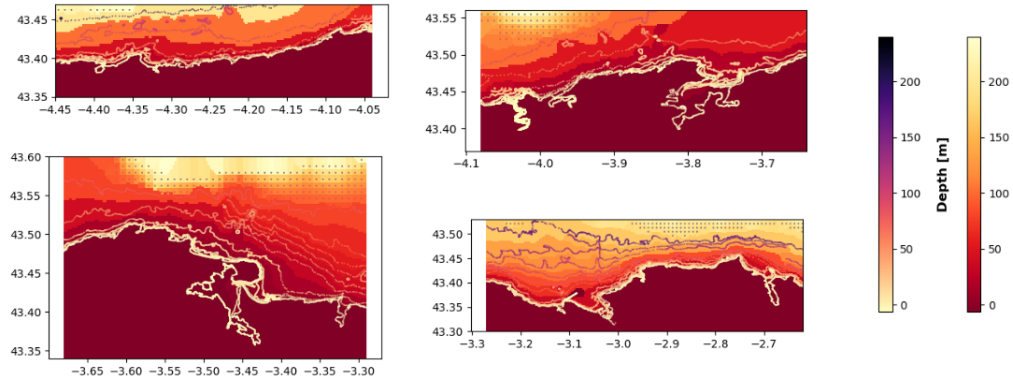
FIGURE 2.8: These are the four bathymetrys used in the propagations. The lines represent the raw data proportioned by the SMC while the colormap represents the interpolation that must had to be done for the use of the numerical model as it propagates the waves in constant and rectangular meshes. As it will be explained later, the numerical propagations from deep to shallow waters are made in four groups as a matter of computational capacity. From left-up to right-down, these plots cover the zones of San vicente de la Barquera - Oyambre, Suances - Liencres - Santander - Somo, Noja - Santoña - Laredo and Castro Urdiales - Bilbao - Getxo.

$$w \text{ (m/s)} = 1,1 * 10^6 \, D^2 \text{(m)} \qquad D < 0,1 \text{ mm}$$

$$w \text{ (m/s)} = 273 \, D^{1,1} \text{(m)} \qquad 0,1 < D < 1 \text{ mm}$$

$$w \text{ (m/s)} = 4,36 \, D^{0,5} \text{(m)} \qquad D > 1 \text{ mm}$$

FIGURE 2.9: Relantionship between the sediment diameter grain and its velocity fall for different sand sizes. Source: (Narra, Coelho, and Fonseca, 2015).

on top of the set of variables needed to later develop the surfability index. This approximate way of estimating the slope of the beach for each breaking wave is not unique, and in this same study other different methods have been carried out, although due to lack of time they have not been able to be perfected too much and leading to incoherent results, this method is described in (Vos et al., 2019).

The calculation of the profiles has been carried out following the methodology proposed in (Bernabeu, Medina, and Vidal, 2003). For this purpose, and once the omega variable has been calculated according to Equation 2.3 where $\Omega = \Omega_{surfing}$, the shape of the biparabolic profile is estimated numerically using the following equations:

$$\text{Surf profile: } x = \left(\frac{h}{A}\right)^{3/2} + \frac{B}{A^{3/2}} h^3 \quad 0 \leq x \leq x_r \tag{2.4}$$

$$\text{Shoaling profile: } x = \left(\frac{h}{C}\right)^{3/2} + \frac{D}{C^{3/2}} h^3 \quad x_r \leq x \leq x_a \tag{2.5}$$

where the surfing profile refers to the part closest to the coast and the shoaling profile refers to the deeper part, knowing that $\Omega_{shoaling} = 0.53 + 1.3 \cdot \Omega_{surfing}$ and calculating A, B, C and D using the following equations where $surfing = sf$.

$$A = 0.21 - 0.02 \cdot \Omega_{sf} \tag{2.6}$$

$$B = 0.89 \exp\left(-1.24 \cdot \Omega_{sf}\right) \tag{2.7}$$

$$C = 0.06 + 0.04 \cdot \Omega_{sf} \tag{2.8}$$

$$D = 0.22 \exp\left(-0.83 \cdot \Omega_{sf}\right) \tag{2.9}$$

It should be noted that this technique is not perfect, but it is used to estimate the surfing profiles on the different sandy beaches throughout the year. In Figure B.3 we can see how the profiles vary for a specific Ω and TR, which is the tidal range of the beach. The cantabrian coast has an average tidal range of about 3.6 m, but the Ω can vary enormously, as well as the size of the sediment, which I reiterate has not been calculated precisely, leaving this calculation as a future task.

Finally, and in order to understand well how the dynamics of this biparabolic profile can be, simply commenting that the results obtained agree with reality. What this means is that beaches like Liencres, see Figure 2.10, where exposure to strong waves is enormous if not the strongest of the entire cantabric sea, the profile of the beaches varies greatly from season to season. In winter, when the beaches are hit by strong swells, the beach responds with a flattened and dissipative profile but in summer, the swells impact with less energy, accumulating more sand at the end of the profile, and resulting in a much steeper profile that generates better surfing conditions. Moreover, in beaches like El Sardinero, which are less exposed, the profile is steeper in the shoaling part, and there is a very abrupt change from low to high tide, also observed in reality in this beach.

It should be noted that this part of the study is the most fragile in terms of precision, as the parameter that most affects the calculation of these profiles is the sediment grain size, which directly affects the Ω of the beaches, and in some of them, although in Liencres it has been measured, it has had to be estimated due to the lack of resources caused by the covid19 pandemic. However, having some knowledge of surfing, it can seen how the profiles calculated for different times of the year and different beaches suggest things that can be clearly observed in the reality of the cantabrian surfbreaks.

2.3 Development and validation of a surfing index

Once the different databases and the variables involved in them have been presented, a series of calibrations, propagations, validations... will be carried out, all with the aim of obtaining this continuous time series of data distributed temporarily along the entire coast of Cantabria during the last 40 years. Important aspects that must be taken into account when evaluating a wave will be also mentioned, even more if a rating wants to be given to it, and not forgetting that not everyone wants or is able to surf the same type of waves. For this purpose, basic aspects and characteristics of an easy wave and a difficult wave will be briefly commented.

First of all, an easy wave is characterized, according to the variables previously defined, by having a wave height at break not very big, below 1.5 m, independently of period, an Iribarren
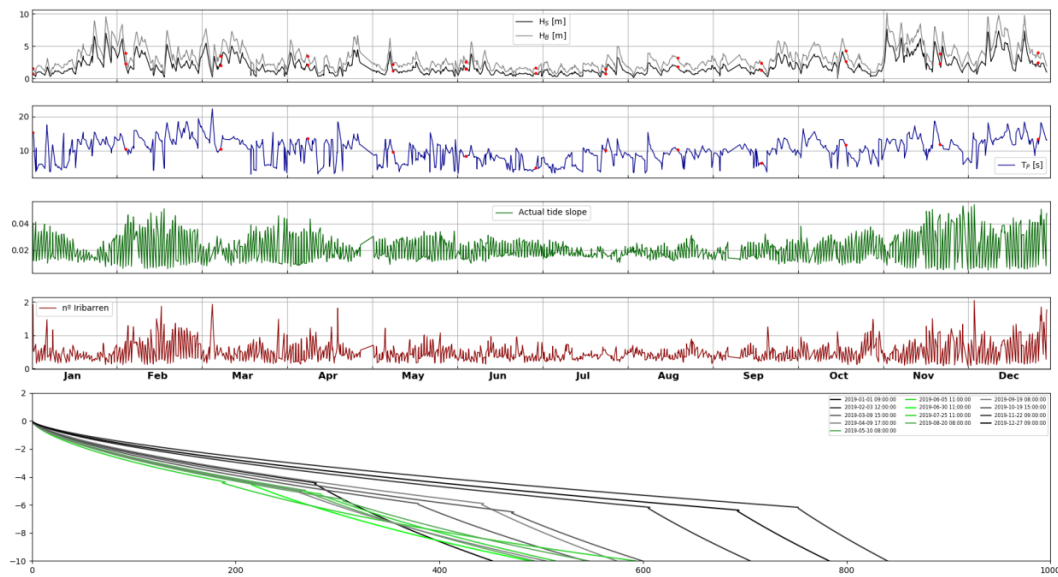
FIGURE 2.10: Associated beach profile to the different moments of the year 2019 in Liencres. As it can be seen in the figure, strong waves in winter provokes flattened profiles while relaxed summery waves end up forming steeper profiles that can generate good surfing conditions in months like September or October.

number (type of break) between 0.4 and 1 and a relative direction to the coast between 30 and 50 degrees. In the other side, a difficult wave will present heights that can be greater, although this is not the most important thing but the break, which will take higher values coming to generate tubes and fast waves, and directions relative to the coast smaller, placing its optimum at about 20 - 30 degrees, and resulting in waves again faster and more interesting to professional surfers. It will now be explained why waves with smaller coast-related directions are more difficult to surf and what implications this has on the quality of the training season, (Espejo Hermosa, 2011; Black et al., 2003).

When this series is obtained, the icing on the cake will be to quantify the quality of all these variables, both individually and jointly, according to their own values. In this way, it will be possible to evaluate objectively how the surfing conditions have been all this time, and it will be possible to observe in which beaches the surfing conditions are ideal most of the time, and in which beaches there are excellent conditions but only a few days a year. In addition, mediocre surfbreaks will be observed, but the set of all the data obtained will allow to offer a global vision of how the qualities and conditions of surfing are distributed along the coast. The development of this index is described in (Espejo Hermosa, 2011), although some changes have been made. This will be explained below, but if the importance of this regional surfing index has not been understood yet, just take into account how the most important websites on the world always try to quantify these surfing conditions, see Figure B.2.

The variables included in these index could be different, as it is explained in the caption of the figure, but in our case, the variables that will be used are: The wave height at break, H_B, the
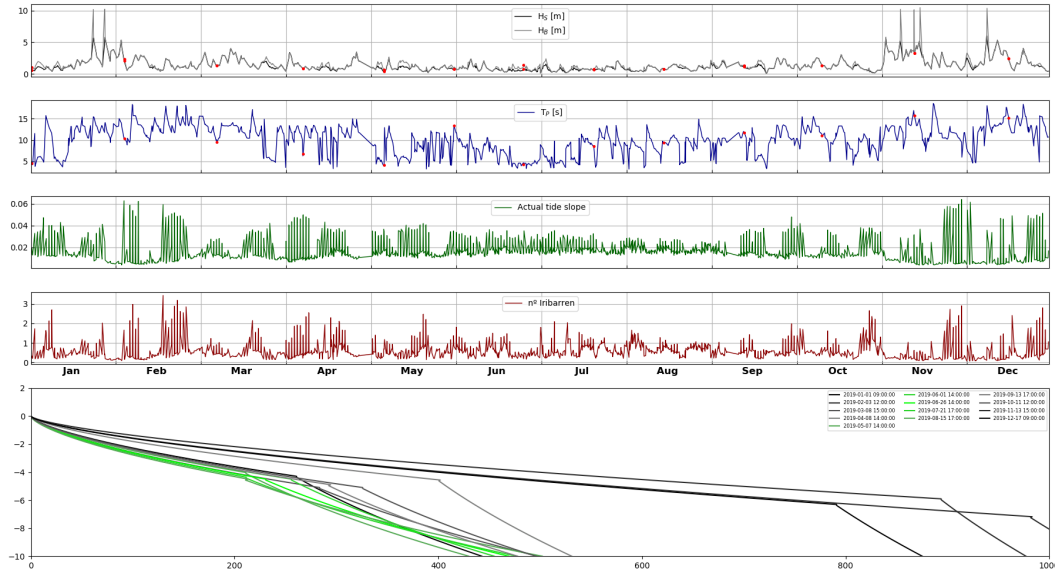
FIGURE 2.11: Associated beach profile to the different moments of the year 2019 in El Sardinero. As it can be seen in the figure, stronger waves in winter provokes flattened profiles and very abrupt changes from low to high tide while relaxed summery waves end up forming steeper profiles that can generate good surfing conditions in months like September or October.

peak period, T_P, the Iribarren number, I_R, the relative direction of the waves, $\delta\theta$, the directional dispersion of the waves, σ_θ, and the relative direction of the wind, $\delta\theta_W$. When a dataset of these variables is achieved for all the different surfbreaks of interest, which are for this first part of the project San Vicente de la Barquera, Oyambre, Los Locos, Liencres, El Sardinero, Somo, El Brusco y La Fortaleza (see Figure 3.13), a SOM (Self-Organizing Map), which will be described on detail later, will be performed, grouping all these conditions into different groups, giving a grade to each of them. This SOM is done to the entire dataset with the data of all the surfbreaks inside, but it has been also carried out in individual surfbreaks, so the temporal variability in each of them can be easily observed. Notice that with the grouped SOM, spatial variability can be observed in the other hand. This offers an enormous playground for the final study of the surfing conditions in all the region, even bigger if a probability of occurrence is set to each of these clusters or groups of similar conditions, which is also the case, adjusted to the beach or month if it is requested. With this tool, the quantitative evaluation useful for the protection of the different surfbreaks will be possible.

2.3.1 Regional Surfing Index (RSI)

The method used for the clustering will be explained later but it is not essential for the development of a historical time series of grades. Now, the index used must be described. As it has been said in the motivation section, our goal here is to give the regional surfer a tool that can study and in a future predict the conditions for surfing in the different beaches of Cantabria. With this tool, the professional and amateur sportsmen will be able to know the conditions of the beach he or she wants to go, and he or she will not have to has a previous knowledge about that beach. With this purpose, a Regional Surfing Index (RSI hereafter) has been developed so

TABLE 2.2: Different ranges used in the calculation of the surfing index for the different variables H_B, T_P, I_R and σ_θ.

H_B [m]		T_P [s]		I_R		σ_θ [º]	
Range	I_{H_B}	Range	I_{T_P}	Range	I_{I_R}	Range	I_{σ_θ}
< 0.4	0	< 6	2	< 0.3	0	< 10	10
0.4 - 0.8	4	6 - 8	5	0.3 - 0.6	4	10 - 14	9
0.8 - 1.0	6	8 - 11	7	0.6 - 0.9	7	14 - 18	8
1.0 - 1.5	8	11 - 14	8	0.9 - 1.2	10	18 - 22	6
1.5 - 2.5	10	14 - 16	9	1.2 - 1.6	7	> 22	4
2.5 - 3.0	9	> 16	10	1.6 - 1.9	5		
3.0 - 3.5	7			1.9 - 2.2	3		
3.5 - 4.5	4			> 2.2	0		
4.5 - 5.5	2						
> 5.5	0						

TABLE 2.3: Different ranges used in the calculation of the surfing index for θ_W.

θ_W [º]	W [m/s]	I_{θ_W}
> 120	< 15	1.5
	15 - 25	1.2
	> 25	0.9
90 - 120	< 15	1.2
	15 - 25	1.0
	> 25	0.8
< 90	< 15	0.8
	15 - 25	0.6
	> 25	0.4

depending on the values of the variables, a grade is assigned. Notice the fact that with this development of an index, now the previously commented SOM gains a lot of importance, as a grade can be assigned to each cluster, having this way a two dimensional representation of all the different surfing conditions in Cantabria, with an evaluation of all of them.

Is is also important to underline the ranges of the variables used for the index, which are described in Tables 2.2 and 2.3. These variables are not exactly the same as the ones used for the training of the neural network, the peak period is included in the index. In this case, the calculation of the discrete index is done using the variables H_B, T_P, I_R, $\sigma_t heta$ and θ_W. The peak period is included because it is a very common and important variable in the development of the surfing session, but it appears in the calculation of the Iribarren number, so it is not contained in the training. For the peel angle, it should be included in the index, but its calculation has not been perfected yet, so this will be included in future works.

The calculation of $I_{\delta\theta}$ has not been showed, as more work has to be done before it is correctly

calculated. As mentioned before in this study, the relative direction to the coast has been calculated subtracting the direction of the incident waves minus the direction of the beach, but more research has to be done to estimate this peel angle correctly. When this is done, waves with a peel angle of about 30 º will be ideal for the practice of the sport while waves with very low or high angles will be not that good. In the situation we are now, the peeling angles are always around 5-10 º, and this is not real, as beaches like Liencres or Somo, where good waves usually appear, seam to be very very bad in this fact.

For the rest, the calculation of the RSI is performed using the equation below (Equation 2.10):

$$I_{RSI} = \frac{I_{H_B} + I_{T_P} + I_{I_R} + I_{\sigma_\theta}}{40} \cdot I_{\theta_W} \tag{2.10}$$

where the variables described in Table 2.2 form the base of the grade while the wind improves or deteriorates the situation (see Figure 2.4).

Finally, a historical record of notes or grades is obtained for the different surfbreaks along the cantabric coast, having this way a waves database capable of describing all the temporal and spatial variability of surfing in the region.

Chapter 3

Development of a high resolution waves database

As mentioned in the previous chapter, the cornerstone on which the entire study will pivot is the numerical model database provided by CSIRO. It should be noted that this database is updated every month, so a technique has been developed that downloads the data making it easily usable by the software developed for the different steps that will be explained later. And I stop here to make a brief comment about the importance of the pre-processing, the cleanliness and the organization of the data (Wickham, 2014), because it is this task, and even more so when the amounts of data are large and come from different places, one of the most important parts if not the key part in the whole process. Good pre-processing leads to a correct understanding of the problem, which will be the best possible start for the development of any possible solution. On the other hand, a fuzzy beginning, with mixed and repeated data, of which 100% is not understood at all, has no other end than that of future chaos when new actions are piled up on this erroneous initial set. In most of the projects, and this is the case too, the most part of the time is employed in developing a good database so that it can be used easily and making a good task on the presentation and the explanation of the results obtained. In this work, both tasks took a lot of time, a lot more time than the creation of the models used for the other jobs, so these should not be forgotten although in this chapter, we will focus on these models and their mathematical formulation.

Then, once the methodology has been explained in the previous chapter and having in mind all these variables, databases and resources used, it is time to focus on the models and algorithms used. The workflow followed is described in Figure 3.1, where almost all the different actions carried out in the project are included, see also Fig. 2.6. Notice the importance of this whole set of actions, as it globally forms an entire and complete process that allows some reanalysis raw data in deep water virtual buoys, as the nodes in shallower waters usually show imperfections due to facts not took into account in the numerical approximation, to characterize the coastal wave climate all over the world where this reanalysis exists (Perez, Menendez, and Losada, 2017; Camus, Mendez, and Medina, 2011). Is is true that there are steps that require the existence of real buoys, but these steps, even if they are very very important for the validation of the method, could have been omitted if buoys did not work for example, and the invalidated results would have been exactly the same.
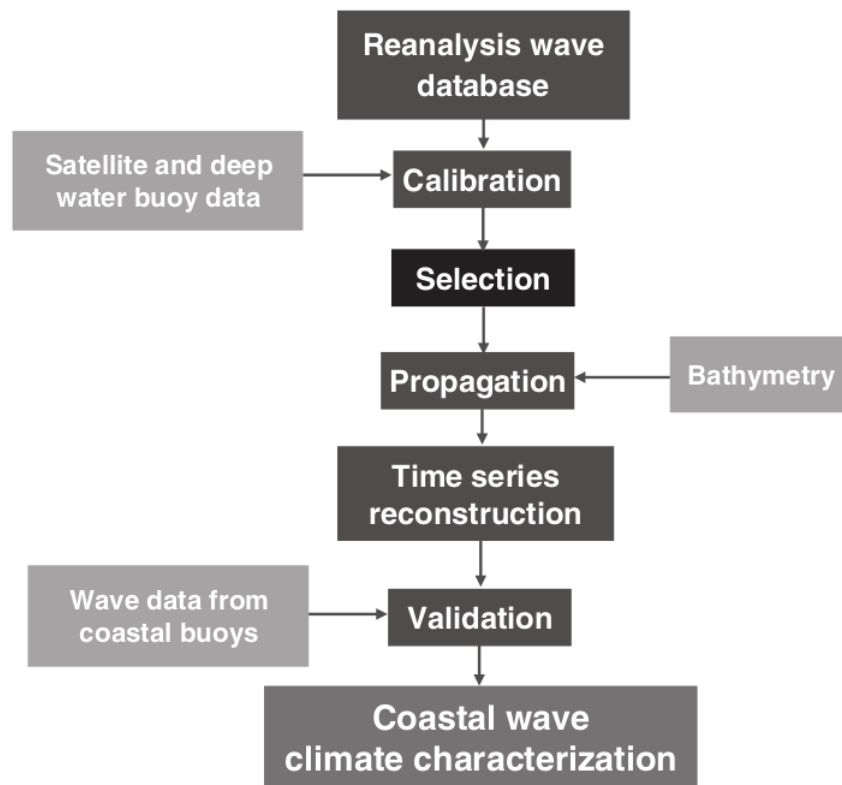
FIGURE 3.1: Graphical description of the workflow followed, this whole process
is also very well explained in (Camus et al., 2013).

3.1 Calibration with satellite data

A python software has been developed that is able to calibrate the data coming from the nu-
merical reanalysis and that is freely accessible in github. Waves can come from everywhere, and
although this allows us surfers to surf, worth the redundancy, waves of all kinds and in very
different places is not an easy fact to interpret by a numerical model. For this reason, a tool has
been developed that is capable of separating waves by direction of origin, and comparing them
through a linear regression model with the bulk (data is not separated in partitions as it is in
the hindcast) data from the satellite wave height database, a significant wave height calibration
coefficient is obtained for each direction of existing waves in the reanalysis, see Figure 3.2 for
a detailed graphical explanation. It must be said, and this is also mentioned in the sketch, that
the satellite data used to calibrate the reanalysis coincide in time and space with the hindcast
in intervals of one hour and 0.2 degrees maximum. This entire calibration process is implicitly
explained in (Albuquerque et al., 2018; Mínguez et al., 2011).

As it is the cantabric coast, the fact that the waves come from all sides is not entirely true,
and this generated imperfections in the model that have been corrected, forcing it to take into
account only those coefficients that are statistically representative, which does not happen in
cases such as when there is not a representative amount of data. This translates into using just
those coefficients that have a p-value less than 0.05 once the linear regression has been applied,
but also eliminating those coefficients that have been calculated with few data, even if their
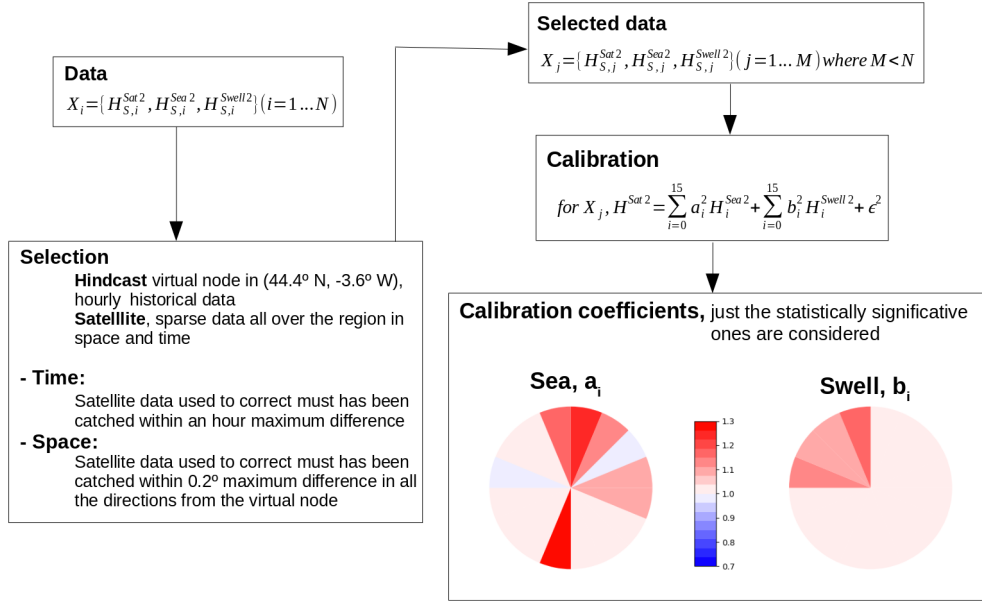
FIGURE 3.2: Explanatory sketch of the methodology used for the calibration of the reanalysis information. The circles in color represent the values for the directional and corrected calibration coefficients.

p-value is small enough, although it should not happen. In addition, wave hindcasts such as CSIRO provide information separated by partitions, as at one time there may be several waves coming from different locations arriving to the coast at the same time. Thus, if a swell dies on the coast and there is a wind that has the same direction, then this swell can be considered a windsea, and the other train of waves coming from different directions at that moment will be considered swells. This is not always the case, but it is the general criterion used. Other more sophisticated methods relating spectrums shape could be used (Qian et al., 2019). Following this reasoning, the model will be calibrated taking also into account this difference in the way the hindcast predicts these two different types of waves, thus obtaining a total of 32 different coefficients (16 direction · 2 types of waves), many of which will not be representative and whose value will be set to one, but which in many cases will provide very valid information for the calibration, see Fig. 3.3. The equation that explains the linear regression performed is Equation 3.1.

$$H^{Sat2} = \sum_{i=0}^{15} a_i^2 H_i^{Sea2} + \sum_{i=0}^{15} b_i^2 H_i^{Swell2} + \epsilon^2 \tag{3.1}$$

where the H^{Sat2} is the bulk squared significant wave height measured at each time and H^{Sea2} and H^{Swell2} represent the squared significant wave heights associated with each direction and partition. The objective is, as it is in all the linear regressions, to minimize ϵ^2. The quadratic sum is due to the fact that the energy depends proportionally on the square of the significant wave height, and to add the heights for the different partitions, the addition must be done this way. This method is capable of calibrating not only hindcast predictions with satellite data, but any resource that provides wave height data partitioned by different train of waves (seas and swells) at the same time, and that can be corrected with historical bulk wave height data. In fact,

this data has also been calibrated with the deep-water buoys, but this is not the aim of the work, as the idea is to offer a method that from start to end is applicable to everyone and everywhere.
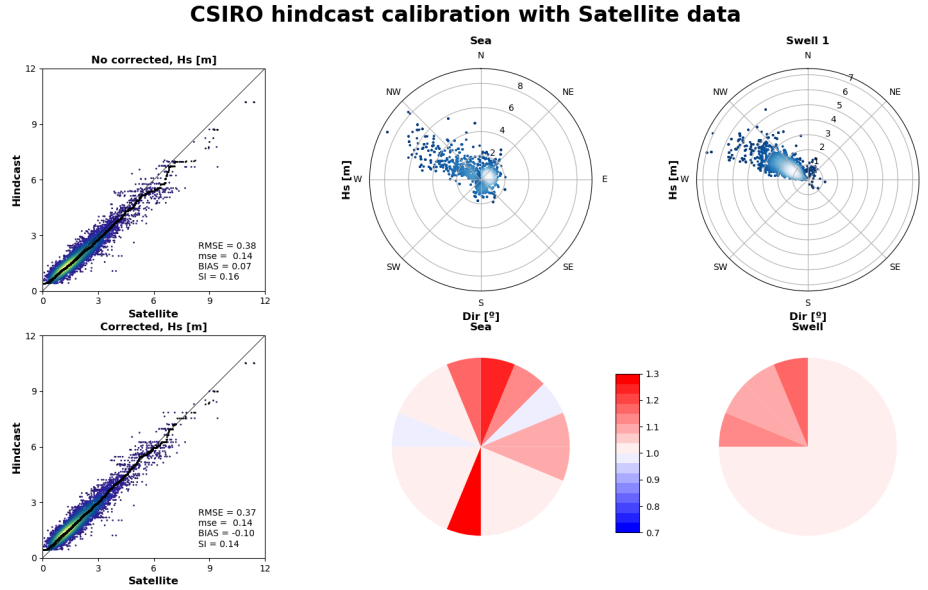

FIGURE 3.3: Hindcast calibration using satellite data. Different subplots are shown and the explanation will be done from left to right. The first two subplots in the left compare the relationship between the total significant wave heights of the numerical model and the satellite data, before (up) and after (down) the calibration has been done. Secondly, the top two figures on the right show how the significant wave height and the direction of these train of waves relate, both in windseas and swells, giving fact to the appearance of mostly W-NW swells in the cantabrian coast, but more disperse incident wave directions for the windseas. Lastly, the correction coefficients are shown for the seas and for the swells, the values set to one are due to the lack of statistical significance on their calculation. As it can be seen, the calibration shows and underestimation of the wave height by the reanalysis, H_S, as the majority of the directional coefficients have values over one.

3.2 Validation with deep water buoys

Once the data has been calibrated, it is time to validate the results obtained because as it was said in the previous chapter, the buoys are the absolute truth, and as long as they exist, results must be compared with them. Thus, a series of figures are presented, see Figure 3.4 and 3.5, in which the data of the satellite-calibrated reanalysis and the buoy are compared, being both very similar and thus demonstrating that the calibration has been timely.

These validation steps are the only ones that can not be extended worldwide as buoys, as mentioned in the Chapter 2, are not equally distributed all along the world. There is an enormous lack of this technology in regions principally located in the south hemisphere and buoys are not always working neither, what makes the validation step an important but not essential task in our study.

FIGURE 3.4: Calibrated hindcast validation with a deep water buoy located in the zone of Bilbao (43.40ºN, -3.14ºW). The two figures in the left show two scatter plots of the significant wave height H_S and the peak period T_P, and the four figures in the right show different comparisons between H_S, T_P and θ_m (Dir) for both the buoy and the calibrated model. All the figures show concordance thus showing the calibration has been performed correctly. Two comments will be done regarding these four figures on the right. First, the relationship between windseas coming from the E-NE-N and smaller peak periods and also the relationship of the previously commented swells that usually come from the W-NW with higher periods, which is totally observed in Cantabria and second, the existence of two principal groups of H_S and T_P, which is also observed in the region. In summer it is very usual to see wave heights of around 1 meter with short periods, but more energetic waves with bigger wave heights and better periods are also observed in other months.

3.3 Hybrid downscaling to shallow waters

The data is calibrated and validated, but we still have a single virtual node to carry out the analysis along the entire coast of Cantabria, so a hybrid statistical downscaling model is used to propagate all this data to the coast (Perez, Menendez, and Losada, 2017; Camus, Mendez, and Medina, 2011). The ideal scenario in these cases is the propagation of all the existing measures in the 40 years of hourly record to the coast, but this is impracticable at a computational level, so the process previously mentioned in the diagram showed in Figure 3.1 has been carried out and will be explained step by step below, making special emphasis on those parts that are more complex and more interesting from a data mining view. This process is basically a propagation of just a certain number of different cases to reconstruct afterwards the total dataset in all the required locations, which is faster than the other dynamical method and very similar in terms of accuracy.

All the techniques related with the generation of this waves database in the coast that will be explained below are performed in the cited papers except the creation of the SOM (Self-Organizing Map) for the grouping of the waves, so although these techniques will be explained carefully, more detailed information can be found in these attached resources.
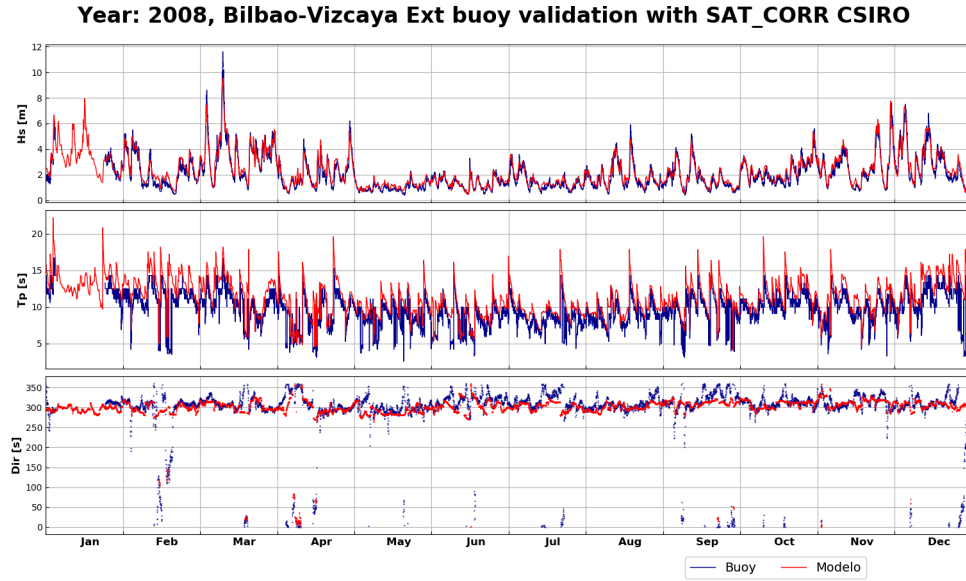
FIGURE 3.5: Bilbao deep water buoy validation in 2008 with the calibrated hind-cast. This image shows the temporal comparison of the reconstructed data and the buoy data, notice that the only thing that is calibrated here with the previously explained model is the significant wave height, but the other plots still work as a confirmation of the reliability of the wave reanalysis used. This time series show very correlated results that validate the work done.

3.3.1 Selection of the cases to propagate with MDA

As already mentioned, once we have the data calibrated and validated, we have to propagate a number of determined cases to coast to later reconstruct the time series in the different requested beaches using these propagated and individual cases. In this way, we apply a selection algorithm of maximum dissimilarity known as MDA (Maximum Dissimilarity Algorithm), which is explained in Figure 3.6 and chooses the 300 most disparate cases of the total set of 40 years. The fact that the number of data is exactly 300 does not have a strictly scientific reasoning beyond the fact that the number of data cannot be very small, because as will be explained later the reconstruction would not be accurate, but neither should it be too large, because we would return to the problem of computer overload, without improving the results too much neither (Camus et al., 2011a). In this way, the number of cases selected to propagate in the four different areas showed in Figure 2.8 is 300, as other numbers like 600 were tried, but 300 shows a very similar accuracy, improving the wasted time by more than the half. Notice also that this selection has to be done twice, one to obtain a subset of windseas and another time to obtain a subset of swells, as the propagation will be different for each case, obtaining a total of four different partitions in our final database that could be joined for different purposes as the development of the regional surfing index.

The aim of MDA is to select a representative subset of size M from a database of size N. Therefore, given a data sample $X = x_1, x_2, ..., x_N$ consisting of N n-dimensional vectors, a subset of M vectors $v_1, ..., v_M$ representing the diversity of the data is obtained by applying this algorithm. The selection starts initializing the subset by transferring one vector from the data sample v_1. The rest of the $M - 1$ elements are selected iteratively, calculating the dissimilarity
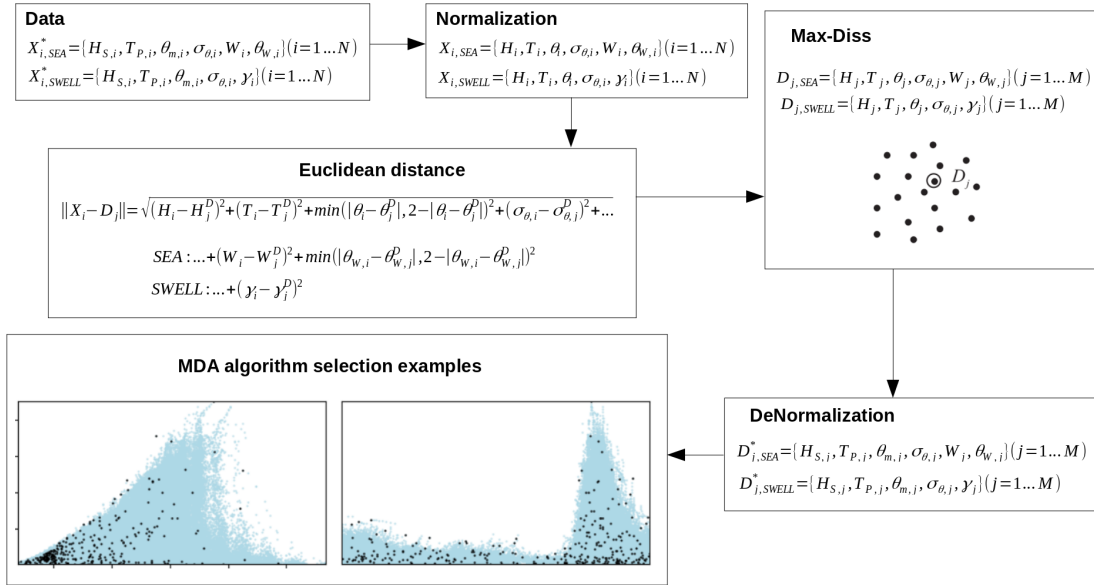
FIGURE 3.6: Explanatory sketch of the methodology used for the selection of disperse cases with MDA. All the steps done are covered in this figure for the selection of both the windseas and the swells cases, as they will need different variables in their propagation to coast. Two examples are shown of the performance of this algorithm where the black points represent the selected cases while the blue ones represent the total amount of data available in the historical record.

between each remaining data in the database and the elements of the subset and transferring the most dissimilar one to the subset. The process finishes when the algorithm reaches M iterations. This algorithm was first described by (Kennard and Stone, 2012).

For example, if the subset is formed by R ($R \leq M$) vectors, the dissimilarity between the vector i of the data sample $N - R$ and the j vectors belonging to the R subset is calculated:

$$d_{ij} = \left\| x_i - v_j \right\|; \quad i = 1, ..., N - R; \quad j = 1, ..., R \tag{3.2}$$

Subsequently, the dissimilarity $d_{i,subset}$ between the vector i and the subset R, is calculated as:

$$d_{i,subset} = \min \left\{ \left\| x_i - v_j \right\| \right\}; \quad i = 1, ..., N - R; \quad j = 1, ..., R \tag{3.3}$$

Once the $N - R$ dissimilarities are calculated, the next selected data is the one with the largest value of $d_{i,subset}$.

MDA has an expected time complexity of $O(M^2N)$ for M-member subsets from an N-member database. The more efficient algorithm $O(MN)$ developed by (Polinsky et al., 1996) has been considered. In this case, the definition of the distance $d_{i,subset}$ does not imply the calculation of the distance between the different vectors d_{ij}. For example, in the selection of the $(R + 1)$ vector, the distance $d_{i,subset}$ is defined as the minimum distance between the vector i of the data sample (consisting of $N - R$ vectors at this cycle) and the last vector transferred to the subset R, and the minimum distance between the vector i and the $R - 1$ vectors of the subset determined in the previous cycle:

$$d_{i,subset}^{min} = \max \left[d_{i,R}, d_{i,subset\,(R-1)}^{min} \right] \tag{3.4}$$

The subset of size $M = 300$ obtained by the maximum dissimilarity algorithm is shown in the figures below for both the seas and the swell cases. Notice that the selection of the cases is very good, as it properly represents all the possible scenarios for the different variables. There are regions where the density of the black points can be higher and lower, but it is just the effect of the density probability in the existent data. For example, for the H_S vs T_P plot, more black points are encountered in the bottom left region, but having a look at the axis, it is very rare to see those bigger heights and periods in the reality of Cantabria.

FIGURE 3.7: Sea cases selected using MDA. All the subplots show a good and homogeneous selection of the cases, black dots are the selected ones. The 6 variables used for the later propagations of the seas are H_S, T_P, θ_m, σ_θ, W_{speed} and θ_W.

To consider an n-dimensional space with the same axes, first the variables are normalized, differentiating between scalar and directional variables. The dissimilarities between the points are calculated by the Euclidean distance (see Figure 3.6) and we finally de-normalize the variables to return to the initial range. Scalar variables are normalized to values between [0, 1] with a simple linear transformation from the maximum and minimum of each variable. For the directional variables in radians, the maximum difference is π and the minimum 0, so the values are divided by π to have the normalized distance between [0, 2].

As a key point in this selection step, notice that this process will be applied to two different sets of data, as the way the cases are propagated depend on the characteristics of the waves, they can be windseas or swells. Then, the data used to obtain the subset of windseas to propagate
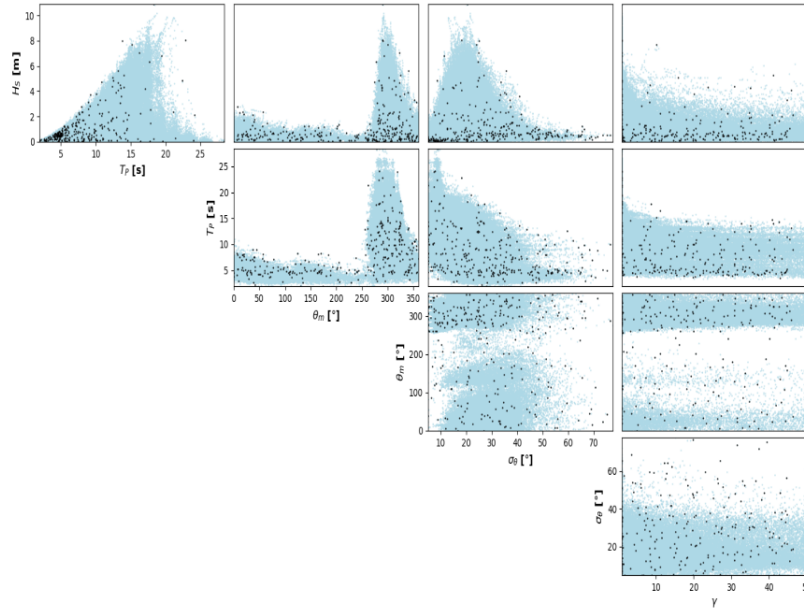
FIGURE 3.8: Swell cases selected using MDA. All the subplots show a good and homogeneous selection of the cases, black dots are the selected ones. The 5 variables used for the later propagations of the swells are H_S, T_P, θ_m, σ_θ and γ, being this a spectral parameter that gives an idea about the narrowness of the wave, it should be around 5 for very pronounced swells and bigger for more disperse waves.

will consist just on the partition of these type of waves while the data used for the swells will have all the three other partitions. Variables used in both cases are also shown in Fig. 3.6.

3.3.2 Propagation to shallow waters using SWAN

Once the selection has been made, the propagations are accomplished using the SWAN (Simulating WAves Nearshore) computer software, developed by the Technical University of Delf and used in this case to propagate the sea and swell cases to each corner of the cantabric coast with an accuracy of about 200 m. A resolution more than sufficient for the purpose of the study as it allows more than one point to be obtained on each beach of interest if necessary, and there will always be at least one point on any beach where the time series wants to be reconstructed. Bear in mind that the reconstructions will be carried out in areas where there is a minimum depth of about 10 metres, as this is when the computer software is reliable. Above this depth or maybe even sooner depending on the characteristics of the wave, the wave starts feeling the bottom more intensely, and other phenomenons come to play.

In linear wave theory, when the wind hits a certain region over a period of time, a sea of wind is generated and propagates with the group velocity $C_g = g/4\pi f$, equal to the distance travelled between the time. The wave spectrum is scattered by $f^2 = gk/(2\pi)^2$, where f is the frequency, g is the gravity acceleration and k is the wave number. When the wave approaches
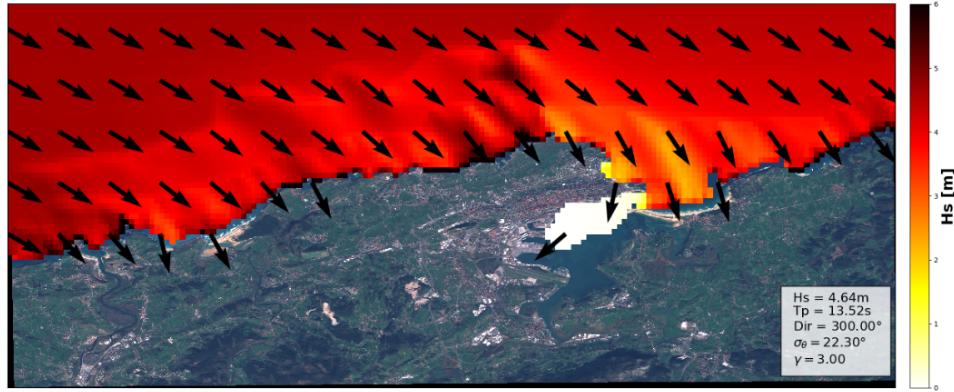
FIGURE 3.9: Example of a numerical propagation of the sea state explained in the bottom-right part of the figure in the zone of Santander, Cantabria. Several things can be observed, first, the expected reduction of the significant wave height in areas such us El Sardinero o Somo, where the beach is less exposed to W-NW waves, second, the increase of the significant wave height in the area of La Vaca Gigante, which is also seen in real life, and last, the phenomenon of diffraction, which is totally observed in this picture as waves reach the coast perpendicular to it.

the coast and begins to feel the bottom, about 200 m of depth, the processes of wave energy transformation are no longer linear and this is where this type of computational softwares as SWAN come into play.

It is now important to highlight the way in which the different types of waves propagate, which again are the wind seas and the ground swells. In this way, the wind seas will propagate taking into account the wind at that moment and the ground swells will only take into account the variables mentioned in the previous subsection. It should be noted that the propagations have no temporal value, since this is indifferent when it comes to reconstructing, the only thing that matters to us is to know the cases that has been propagated, and how this propagation has reached every corner of the coast. But another important fact must be commented again, and this is the propagation of all these 300 seas and 300 swells cases to the coast in four different regions, as it is shown in Figure 2.8. The propagation is done in four different steps because the computational resources are not good enough to put up with this high charge. Some examples of propagations to coast are shown in Figure 3.9 and more zoomed in Figure 3.10.

3.3.3 Coastal reconstruction of the cases with RBF

Cases are already propagated and we now have a lot of different waves with different characteristics propagated all along the coast of Cantabria covering 100% of the beaches, then the reconstruction of the wave climate all over the cantabrian coast must be done.

The reconstruction of the wave climate in shallow waters is carried out by an interpolation from the selected case series that have been propagated from undefined depths. The interpolation technique used is based on radial basis functions, very suitable for data with high dimensionality and not evenly distributed (Oqielat, 2017). There is a series of values of the real function $f(x_i)$ $i = 1, ..., N$ in the points $x_1, ..., x_N$. The RBF (Radial Basis Functions) interpolation technique considers that the RBF function approximation consists of a linear combination of
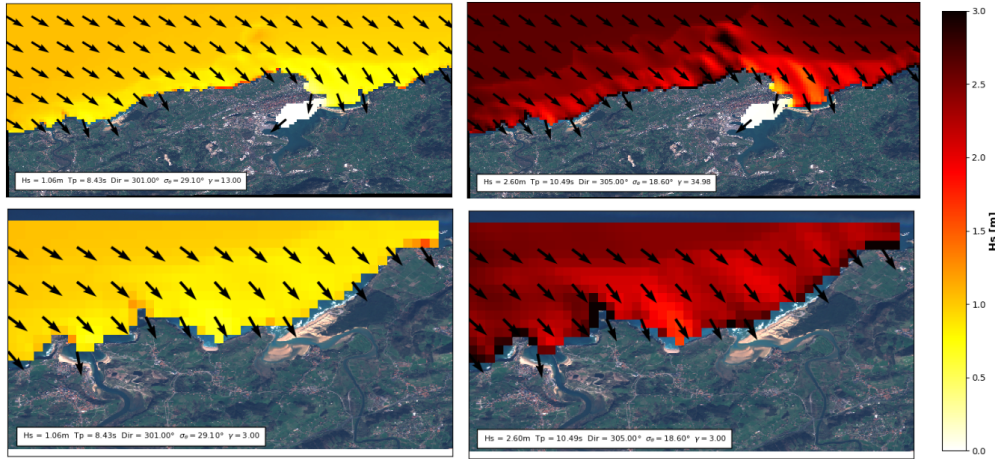
FIGURE 3.10: Examples of two different numerical propagations of two different sea states in the zone of Santander and zoomed in Playa de Los Locos and Playa de Liencres, both in Cantabria too. Similar aspects as the ones commented in the previous picture are observed, especially the refraction of the waves when seeing the bottom, as they reach the coast almost perpendicular to it.

symmetrical radial functions centered on the given points (Camus et al., 2013). The objective function has the expression showed in Equation 3.5.

$$RBF(x) = p(x) + \sum_{i=1}^{N} a_i \Phi\left(\|x - x_i\|\right) \tag{3.5}$$

interpolating the given values as follows:

$$RBF(x_i) = f_i \quad i = 1, ..., N \tag{3.6}$$

where RBF is the interpolation function, $p(x)$ is the linear polynomial in all the variables involved in the problem, a_i are the RBF adjustment coefficients, Φ is the basic radial function, $\|\cdot\|$ is the Euclidean norm and x_i are the centres of the RBF interpolation.

The polynomial $p(x)$ in the expression of the RBF interpolation function is defined as a monomial base $p_0, p_1, ..., p_d$. The first is a monomial, consisting of a number of grade one monomials equal to the dimensionality of the data, where $b = b_0, b_1, ..., b_d$ are the coefficients of these monomials.

The radially based functions can have different expressions. Some of these radial functions contain a shape parameter that plays a very important role in the precision of the technique. In the methodology of propagation of the maritime climate, it has been considered the Gaussian radial function that depend on a shape parameter. The optimal value of this parameter is obtained from the algorithm proposed by (Rippa, 1999).

Notice that the RBF reconstruction is done for each CSIRO reanalysis partition, but all these partitions must be grouped together for the buoy validation and the final estimations of both the profile of the beach and the surfing regional index, this is explained below.

3.4 Shallow water buoy validation and spectrum reconstruction

Finally, and once all the relevant reconstructions have been carried out on the different surf-breaks of interest (these surfbreaks have been mentioned in the "index" explanation section and are shown in Fig. 3.13), a grouping of the propagations is made for the different propagated seas and swells, as the parameters to be used in the index and the ones measured by the coastal buoys are bulk and not segregated in different partitions. In this way, bulk parameters are calculated using two different methods. Both methods demonstrate propagations have been done correctly, as it is shown in Figures 3.12 and 3.11, but the one that will be used later on this study is the spectral reconstruction, as this also gives us information about the directional dispersion σ_θ, which is crucial in the development of the index, and this takes into account the total reconstructed spectrum too.
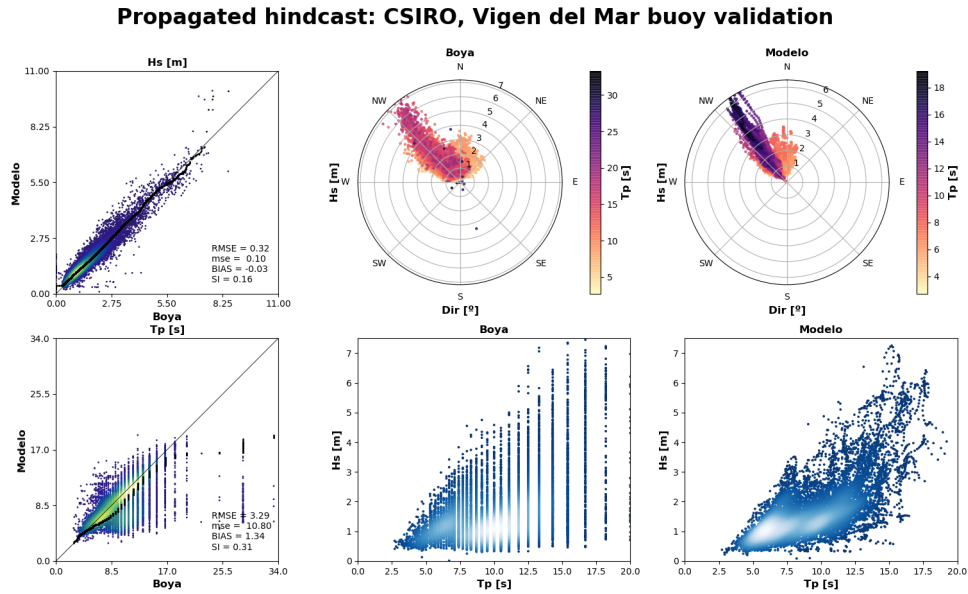

FIGURE 3.11: Virgen del Mar buoy validation with propagated, reconstructed and spectrally grouped data. The two figures in the left show two scatter plots of the significant wave height H_S and the peak period T_P and the four figures in the right show different comparisons between H_S, T_P and θ_m (*Dir*) for both the buoy and the propagated model. All the figures show concordance thus showing the propagation has been performed correctly.

3.4.1 Bulk parameters

The first method consists of energetically adding up each of the propagated partitions and this is done using the following equations: Eqs. 3.7, 3.8 and 3.9.

$$H_S = \sqrt{\sum_{i=1}^{N} H_{Si}^2} \tag{3.7}$$

$$T_m = \sqrt{\frac{\sum_{i=1}^{N} H_{Si}}{\sum_{i=1}^{N} \frac{H_{Si}}{T_{mi}^2}}} \tag{3.8}$$
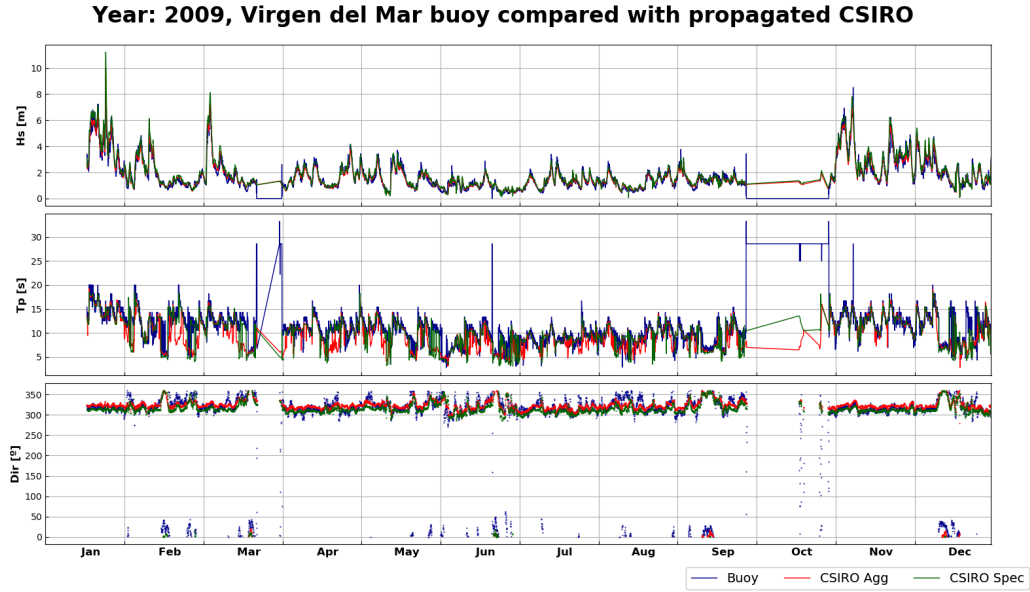
FIGURE 3.12: Virgen del Mar buoy validation in 2009 with propagated and reconstructed data. This image shows the temporal comparison of the reconstructed data using the both previously commented methods. Notice that for the bulk parameters calculated with the formulas (also known as aggregated or "Agg" in the figure legend), the variable compared is T_m and not T_P, which is the case in the two others, fact that makes sense once the plot is observed, as the mean period is always less than the peak period. This time series show very correlated results that validate the work done.

$$\theta_m = \arctan \frac{\sum_{i=1}^{N} H_{Si}^2 \, T_{mi} \, \sin \theta_i}{\sum_{i=1}^{N} H_{Si}^2 \, T_{mi} \, \cos \theta_i} \tag{3.9}$$

where H_S, T_m and θ_m can be now compared with the measured same variables in the coastal buoys of Cantabria. Notice that the variable that is used for the index is T_P, being this one of the reasons why the spectral reconstruction of the parameters is preferable.

3.4.2 Spectral parameters

This other method consists in reconstructing the complete spectrum with the propagated data. Spectra are ideal and are widely used in coastal engineering studies as they contain all the necessary information about the different train of waves that live together in a precise moment in time, as it can be seen in Figure A.5, and this allows the calculation off all the requested variables about the waves in that moment.

It could be messy as we are reconstructing the total spectrum for then recalculating the previously existed variables, but remember we propagate the sea and swell train of waves independently, so we do not have all those bulk parameters that will be used in the SOM neural network yet. Once the spectrum is obtained and saved, the different variables that have an important role in the posterior validation and creation of the surfing index are calculated using some formulas described in (Espejo Hermosa, 2011; Kumar et al., 2017), and that involve the

FIGURE 3.13: In this map, all the surfbreaks studied are shown.

calculation of the area below the spectrum curve or what is the same, the different integrals of order zero, one, two and so on.

Spectral variables are already calculated, but Figure A.5 shows another important job that can be done with these spectra. The free surface movement of the sea is also a very good way to see how waves reach the shore, as it gives us an idea about the cleanliness of the train of waves. This free surface elevation is calculated using Equation A.6 for each of the frequencies and directions existent in the spectrum to be analyzed (see Fig. A.3 for a graphical explanation). In this equation, $a_{i,j}$ represents the amplitude of each spectrum frequency and direction component or what is the same, the discrete amount of energy associated to this values, and $\alpha_{i,j}$ represents a random phase uniformly distributed between $[0, 2\pi]$. Nevertheless, the generation of these free surface plots is also very expensive, leaving this method as an example to show its importance on the visualization of the spectra, but not having them included in the methodology followed.

As a summary for this chapter and as an introduction for the next, have in mind the different aspects covered. First, the waves database has been obtained in all the different surfbreaks/beaches in Figure 3.13, having in all these spots an historical record of the surfing conditions available. With this historical record, the evaluation of this conditions is performed using the profile of the beach, thus calculating the type of break, and using the ranking formulas previously described in the previous chapter. Now that we have all the grades, an statistical analysis can be done, but also a neural network has been trained that group all these different surfing conditions to make them easily interpretable both spatially and temporally. With these machine learning technique, the regional and international surfer, without matter of his/her level of expertise, will observe all the situations that can occur in Cantabria, being able to choose the one he/she prefers when and where he/she wants.

Chapter 4

Grouping of the different surfing conditions

4.1 SOM (Self-Organizing Maps) explanation

Visual Techniques can be mainly divided into two different groups: univariate and multivariate. The univariate representations produce representations of one variable while the multivariate representations try to show the relationship between several variables. Some examples of univariate representations are histograms, box and whiskers plots. Within multivariate representations, Parallel Coordinates and Self-Organizing Maps must be highlighted as the most representative elements of this kind of techniques. In this way, it will be explained how Self-Organizing Maps (SOM hereafter) work and will be also demonstrated how powerful they are.
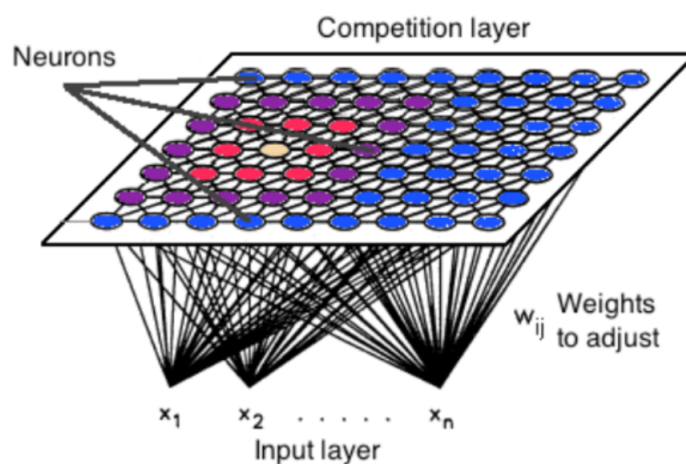

FIGURE 4.1: SOM layers graphical explanation where both the input and the competition layers are showed. It can be seen how all the input neurons connect with the neurons in the competition layer as it is explained. Source: https://ivape3.blogs.uv.es/2015/03/15/self-organizing-maps-the-kohonens-algorithm-explained/.

The SOM was proposed in 1984 by Teuvo Kohonen, a Finnish academician. It is based in the

process of task clustering that occurs in our brain a and it is a kind of neural network used for the visualization of high-dimensional data (Kohonen, Schroeder, and Huang, 2001). It compresses the information of high-dimensional data into geometric relationships onto a low-dimensional representation. In a SOM algorithm, the neurons are ordered in two layers: the input layer and the competition layer (see Fig. 4.1). The input layer is composed of N neurons, one for each input variable. The competition layer is composed of a topological low-dimensional grid of neurons (usually 2-dimensional and also in our case) geometrically ordered. The number of neurons in the output layer depends on the problem. It may fluctuate from a few to several thousands depending on the complexity of the problem, again in our case, the number of neurons in the competition layer is 400.
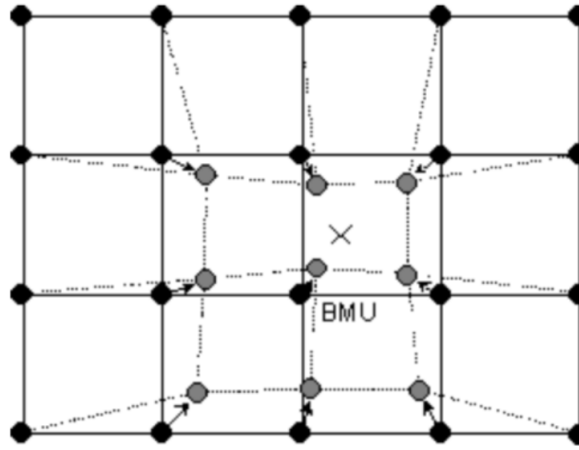

FIGURE 4.2: Weight update process for the different existent clusters. This figure shows how the best matching unit (BMU) and its neighbors move towards the input sample marked with *x*. The solid and dashed lines correspond to the situation before and after updating, respectively. Source: https://ivape3.blogs.uv.es/2015/03/15/self-organizing-maps-the-kohonens-algorithm-explained/.

Each input layer unit is connected with every neuron of the competition layer and next, an N-dimensional weight vector is assigned to each competition layer unit. This is the most important part for the understanding of the method, because N is the number of variables, which in our case is 5 (H_B, I_R, σ_θ, $\delta\theta$ and θ_W). As a preamble, just notice that the algorithm will be done for each instance of the data, in each neuron, so each neuron will receive a value from each variable (each input neuron) in every step, and will perform an operation with its 5 (N) weights (randomly initialized).

Thereby, a model (group of surfing conditions, for example: $H_B = 1.5$ m, $T_P = 8$ s...) of some observations is associated with each second layer unit. The algorithm calculates the models that best describe the domain of observations. The models are arranged in the two-dimensional grid so that similar models are closer each other than the different ones. The SOM keeps a neighborhood relation between the original distribution of N-dimensional data and the topological

low-dimensional grid.

Concerning to the nuts and bolts of the algorithm, a couple of choices have to be made: the map type (hexagonal or rectangular grid, which indicates the neighborhood relation or topology, hexagonal in our case), and the number of neurons (which defines the size of the low-dimensional grid, 20 by 20 in our case). These choices depend on the size of the input data and their dispersion. In this study, the decision of making a large two dimensional final layer has been based in the idea of exploring the totality of the data, but future researches may include less neurons in the competition layer. For the input layer, the number of variables included are 5, and it is searching the avoid of redundancy.

Afterwards, a learning algorithm is used in order to calculate the associated weights for each competition layer neuron. The first step in this algorithm is the weight initialization (at random, usually). The following step is to approach them to the optimum values using an iterative procedure. In each training step, a sampled input observation is randomly chosen, and the distance within its N-dimensional space position and the weights vectors associated with each competition layer unit is calculated using some distance measurement (usually the Euclidean distance). The neuron whose weight vector is nearest to the input observation is called winning neuron or Best-Matching Unit (BMU).

$$\left\| v_{w(i)} - x_i \right\| = \min_j \left\{ d(v_j - x_i) \ \ j = 1, ..., M \right\} \tag{4.1}$$

where $d()$ is the Euclidean distance defined in the sketch (Fig. 4.3), x_i stands for the data vectors, v_j corresponds to the different centroids and $1 \leq w(i) \leq M$ is the index of the winning reference vector. Right after, a weight update is performed so that the winning neuron is approached to the input observation. The BMU's topological neighbors are treated likewise. This adaptation method pulls the BMU and its topological neighbors towards the sampled observation (see Fig. 4.2). So for the Euclidean distance, the weight update is performed as follows:

$$v_j = v_j + \alpha \cdot h(w(i), j) \cdot (x_i - v_j) \ \ j = 1, ..., M \tag{4.2}$$

where $0 \leq \alpha \leq 1$ is the learning rate and controls the velocity of the adaptation process. The function $h(w(i), j)$ is a neighborhood kernel on the SOM lattice, which determines the rate of change around the winning centroid and which projects the topological relationships in the data space onto the lattice (Camus et al., 2011b). The training phase is usually performed in two stages: the rough training (in that a relatively large initial learning rate and neighborhood radius are used) and the fine-tune (in that, small learning rate and neighborhood radius are applied from the begining).

The self-organizing maps (bidimensional projections with spatial organization) can be rectangular or hexagonal, the number of neighbors being 4 or 6 respectively. Each cluster of a SOM is defined by two vectors: one in the data space v_j (prototype) and the other one (m_j, n_j) describing the position on the lattice (Fig. 4.3). For a given SOM of size $M = A \cdot B$, the jth index of a cluster is related with the lattice dimensions and its position in the lattice by the expression: $j = B \cdot (M - 1) + N$, where M usually equals N (notice that N is not now the number of variables included but the number of clusters or neurons in one direction of the two dimensional
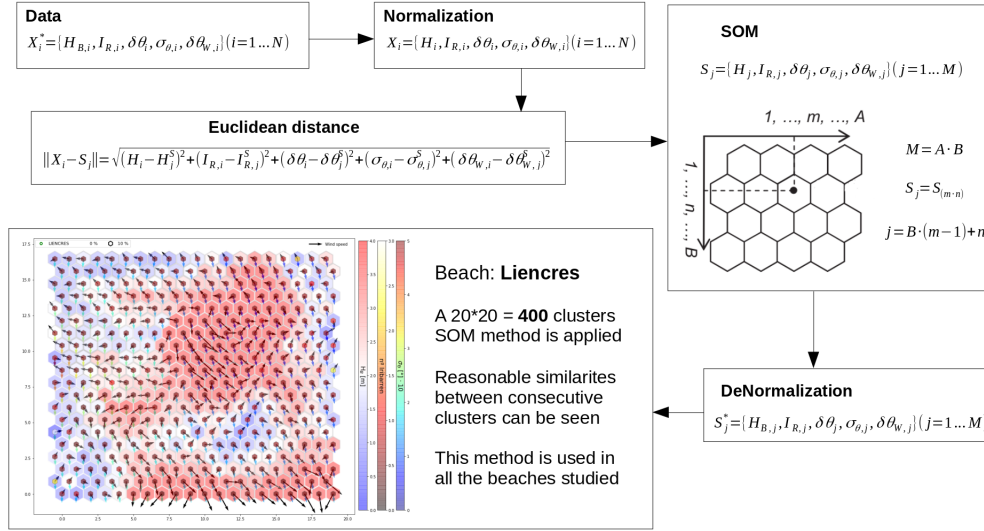
FIGURE 4.3: This sketch shows some of the steps took into account to develop the clustering of the data in the different beaches. The variables used and previously described are normalized, then used to perform the clusterization (using the Euclidean distance needed in Equation 4.2) and then de-normalized to be plotted. The subplot shows how this SOM algorithm finishes grouping the data into spatially related clusters that show similarities between them.

projection).

But this explained process must be performed over normalized data, as the used variables previously described in the study have very different intervals and in this way, the weights associated to each variable will depend on these ranges, which is not desirable. It is a very common task in a lot of data mining and machine learning techniques (Sola and Sevilla, 1997).

Once the map training is finished, the visualization of the two-dimensional map can be performed. It is also called the *components plane* and provides a qualitative information about how the input variables are related to each other for the given dataset (Fig. 4.4). Furthermore, a *hits map* can be performed. The topological structure of this map is the same as a component of the components plane. It represents the number of times each map unit, or neuron, was the BMU for each input register, so that the distribution of the BMU for a given data is represented. This gives an idea of the number of input observations that gather in each neuron. This makes it possible to compare the importance of each unit in the components plane and will be used for the calculation of the probabilities of each cluster (Figs. B.5 and B.6).

4.2 SOM and RSI results

Once the functioning of the neuronal network and the calculation of the surfing index have been both explained, it is time to make sense of the results obtained, since in Figure 4.4 it is explained how the values associated with each variable have been presented, but the correlation with reality that these results present must be also explained. In this way, the neural network trained with the set of data from all the beaches that represent the spatial and temporal variability of each and every one of them, highlights 400 clusters that encompass all the possible surfing
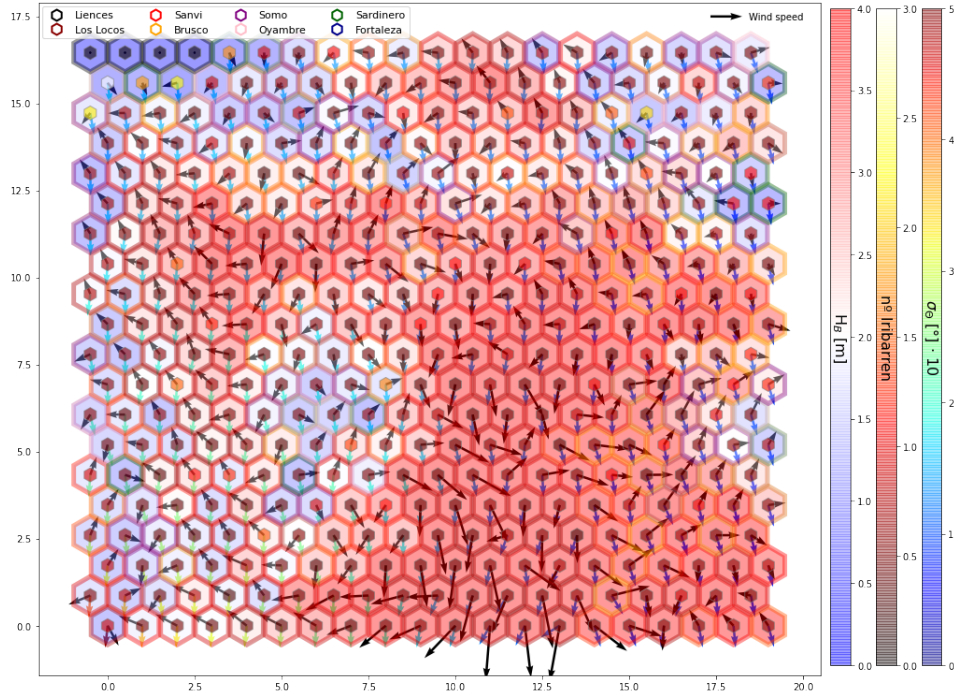
FIGURE 4.4: Results obtained when the SOM is applied over the different beaches mentioned. In this figure, the color of the big hexagons refers to the wave height at break H_B, the direction and color of the colored arrows refer to the relative wave direction and its directional dispersion respectively, denoted by θ_m and σ_θ, the smaller hexagon color gives an idea about the type of break and intensity, as it represents the Iribarren number, and the black arrows show the wind directions, being the magnitude represented by the length of these arrows. Notice that both the waves and the wind directions are relative to the coast, being the non-relative usual waves direction W-NW and the wind direction very disperse. A lot of different clusters with very different surfing conditions are showed, representing this way the whole casuistry available in the cantabric sea, or at least in the beaches selected, which more or less represent all the known one. This method is easily applicable to every single beach inside the range described in the previously described methodology (Chapter 2).

conditions. But not all the clusters present good conditions for the practice of the sport, in the index figure (Fig. B.4) we can observe the valuation that each of the clusters receive for the independent variables as well as at a global level (bottom-right of the figure). For this reason, clusters such as the ones located in the top, which present a wave height of about 2 meters, with not very pronounced wind and an Iribarren number in the range of 0.8-1.2, which presents a very good qualification, are presented as ideal for surfing, while clusters such as the ones in the bottom, where the wave heights are very big, the onshore wind is also high and the break type is bad, are presented as not ideal.

Moreover, the two probability figures below show how often the surfer can find those good or bad clusters and where and when. This is related with the previously defined hit maps, as it is just the counting of the number of times that a cluster appears in the data. With these three plots, the historical temporal and spatial variability of the surfing conditions in Cantabria are defined.

The development of this neural network is not essential, as the statistical evaluation can

be done with the historical record of discrete data, but it helps to study the variability of the different surfing conditions. In Figure 4.4 a lot of clusters are shown, but just some of them must be watched. The border of the hexagons give an idea about the beaches where those clusters occur the most, which can be also seen in the probability figure, Fig. B.5, and using this information together with the temporal probability in Figure B.6, examples of good waves can be obtained. For example, clusters as the ones situated in the top-right part of the two-dimensioanl representation, that have good ratings, appear in months like Septemer in beaches such us Liencres and San Vicente, while very good conditions in La Fortaleza can be seen in months like December, although they are rarely available. The detailed study of these pictures is essential, and more important, it represents very good the reality observed in the region. With this tool, we are capable of both giving the sportsman a tool that can tell when the surfing session is good, but we can also tell the Goverment of Cantabria in this case when waves are good and where, which are the two main objetives of the project. Finally, a future and highly expected protection of the surfbreaks will be achieved.

Chapter 5

Conclusions and Future Tasks

The objective of the project has been achieved, as a surfability index has been obtained which perfectly summarises the temporal and spatial variability on the coast of Cantabria. This score or assessment of the surfing conditions for the last 40 years on the different beaches has been calculated, and this gives the study an immense power when carrying out the statistical study, as the number of data is very large. In this way, the already commented hit maps of probability and the valuation have been constructed, where it is possible to see in which beaches and in which moments of the year, the conditions of surfing are suitable. But in addition to this study in Cantabria, a method has been developed that is applicable worldwide and that is capable of, given a virtual wave reanalysis node, studying these surfing conditions precisely at break level, which is a great advance in those areas where it is not known if the surfing conditions have been good. In this way, and having the bathymetry of that area, which can be obtained worldwide from the GEBCO website, the software developed makes it possible to obtain these spatial and temporal probability maps that, for a given area, will give an idea of the areas to be exploited.

It should be noted that the results obtained are both accurate and real or applicable. By accurate, and this was already commented on in the validation part of the development part (Chapter 3), I mean that the data observed with the buoys and with the human eye itself on the beaches, agree both globally and on a day-to-day level with the data propagated to the coast with the hybrid (dynamical-statistical) downscaling method. On the other hand, by real or applicable, I mean that the scores calculated with the proposed index also match the observed data, since days with good conditions on outstanding surfbreaks such as Liencres, El Sardinero or La Fortaleza have been chosen, and the data have been extracted both from surfing conditions and from quantitative assessment, both of which are in agreement with what is expected, although it is the index that makes this study real and applicable.

Despite the good results achieved, there are still certain aspects to be improved that will be tried to correct in the future, as well as new lines of research related to the project. Firstly, there are imperfections in the model such as the rude estimation of the grain size or the still pending task of calculating the correct peeling angle. Once measured, the sediment size is an essential parameter in the subsequent calculation of the beach profiles, since it is directly related to the omega (Ω) and as we have already seen in the text, this is the main cause of the dynamics of the profiles. In this way, a series of field campaigns will be carried out in which sand samples will be taken from the different beaches on the coast and will be analyzed, and after screening

these samples, the study can be carried out without any need to estimate this parameter. In addition, as far as the calculation of the peeling angle is concerned, satellite images will be taken using certain Google Earth Engines where waves are seen breaking on the coast in order to be able to estimate this angle of breakage accurately, as the method used in the study is based on calculating relative directions, but this is a very rude method as it does not even take into account the beach bars that make the waves break at a certain angle and generate surfable waves. Furthermore, using the new software developed and explained in (Vos et al., 2019), the average slope of the beaches will also be estimated and compared with the results obtained with the profiles, simply as a new validation that can also be extrapolated to the whole globe.

Moreover, not only sandy beaches but point breaks and estuaries must be also included in the study. A lot of very important surfbreaks rest over not sandy bottoms and it the protection, ideally legal, of the surfbreaks wats to be achieved, the study must be extended to all these types of cases. Once again, surfing is evolving immensely, thus proportioning the region enormous benefits both economically and at an entertainment level, and this can not be underestimated. Surfing is growing, and so has to be the protection and exploitation of it, always with the principal look over the interests of the people who practice it, the surfers.

Finally, another aspect if not the most important one to investigate in the future is the implementation of a prediction method capable of giving these good results with breaking precision, because websites like windguru or surf-forecast, already mentioned a couple of times in the work, give global predictions in offshore points at several kilometers of coast, but there is no study capable of giving with this precision information to one or two weeks of the waves that are going to be. As it can be imagined, this task is nothing more than doing what has been already done, but instead of using the past data, using the forecast data provided by global databases. In this way, we would have both the historical study of the surfing conditions in the area chosen worldwide, and the wave predictions in the next few days, thus completing the existing temporal casuistry.

Bibliography

Albuquerque, João et al. (2018). "Directional correction of modeled sea and swell wave heights using satellite altimeter data". In: *Ocean Modelling* 131, pp. 103 –114. ISSN: 1463-5003. DOI: https://doi.org/10.1016/j.ocemod.2018.09.001. URL: http://www.sciencedirect.com/science/article/pii/S1463500318301379.

Bernabeu, Ana, Raúl Medina, and Cesar Vidal (June 2003). "A morphological model of the beach profile integrating wave and tidal influences". In: *Marine Geology - MAR GEOLOGY* 197, pp. 95–116. DOI: 10.1016/S0025-3227(03)00087-2.

Black, Kerry et al. (June 2003). "The Science of Surfing Waves and Surfing Breaks – A Review". In:

Caires, Sofia and Andreas Sterl (Mar. 2003). "Validation of ocean and wind data using triple collocation". In: *Journal of Geophysical Research* 108. DOI: 10.1029/2002JC001491.

Camus, Paula, Fernando J. Mendez, and Raul Medina (2011). "A hybrid efficient method to downscale wave climate to coastal areas". In: *Coastal Engineering* 58.9, pp. 851 –862. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2011.05.007. URL: http://www.sciencedirect.com/science/article/pii/S0378383911000676.

Camus, Paula et al. (2011a). "Analysis of clustering and selection algorithms for the study of multivariate wave climate". In: *Coastal Engineering* 58.6, pp. 453 –462. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2011.02.003. URL: http://www.sciencedirect.com/science/article/pii/S0378383911000354.

Camus, Paula et al. (Nov. 2011b). "Multivariate Wave Climate Using Self-Organizing Maps". In: *Journal of Atmospheric and Oceanic Technology* 28.11, pp. 1554–1568. ISSN: 0739-0572. DOI: 10.1175/JTECH-D-11-00027.1. eprint: https://journals.ametsoc.org/jtech/article-pdf/28/11/1554/3345564/jtech-d-11-00027_1.pdf. URL: https://doi.org/10.1175/JTECH-D-11-00027.1.

Camus, Paula et al. (2013). "High resolution downscaled ocean waves (DOW) reanalysis in coastal areas". In: *Coastal Engineering* 72, pp. 56 –68. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2012.09.002. URL: http://www.sciencedirect.com/science/article/pii/S0378383912001457.

Durrant, Thomas et al. "CAWCR Wave Hindcast 1979-2010". In: (). DOI: https://doi.org/10.4225/08/523168703DCC5.

Espejo Hermosa, Antonio (Oct. 2011). "Variabilidad espacial y temporal del recurso surf: metodología y resultados". PhD thesis. URL: http://hdl.handle.net/10902/1326.

González, Mauricio et al. (July 2007). "An integrated coastal modeling system for analyzing beach processes and beach restoration projects, SMC". In: *Computers Geosciences* 33, pp. 916–931. DOI: 10.1016/j.cageo.2006.12.005.

González Trueba, J.J. (2012). "Manifiesto para la protección de las olas". In: *Ed. Surf Nature Alliance, Santander*.

Holthuijsen, Leo H. (2007). *Waves in Oceanic and Coastal Waters*. Cambridge University Press. DOI: 10.1017/CBO9780511618536.

Kennard, R. and LA Stone (Apr. 2012). "Computer Aided Design of Experiments". In: *Technometrics* 11, pp. 137–148. DOI: 10.1080/00401706.1969.10490666.

Kohonen, Teuvo, Manfred Schroeder, and Thomas Huang (Jan. 2001). *Self-Organizing Maps*. ISBN: 3540679219.

Kumar, Nirnimesh et al. (Apr. 2017). "Bulk vs. Spectral Wave Parameters: Implications on Stokes Drift Estimates, Regional Wave Modeling, and HF Radars Applications". In: *Journal of Physical Oceanography* 47. DOI: 10.1175/JPO-D-16-0203.1.

Moragues, María Victoria, María Clavero, and Miguel Losada (Apr. 2020). "Wave Breaker Types on a Smooth and Impermeable 1:10 Slope". In: *Journal of Marine Science and Engineering* 8, p. 296. DOI: 10.3390/jmse8040296.

Mínguez, R. et al. (Nov. 2011). "Directional Calibration of Wave Reanalysis Databases Using Instrumental Data". In: *Journal of Atmospheric and Oceanic Technology* 28.11, pp. 1466–1485. ISSN: 0739-0572. DOI: 10.1175/JTECH-D-11-00008.1. eprint: https://journals.ametsoc.org/jtech/article-pdf/28/11/1466/3345390/jtech-d-11-00008_1.pdf. URL: https://doi.org/10.1175/JTECH-D-11-00008.1.

Narra, Pedro, Carlos Coelho, and Jorge Fonseca (May 2015). "Sediment grain size variation along a cross- shore profile – representative d50". In: *Journal of Coastal Conservation* 19. DOI: 10.1007/s11852-015-0392-x.

Oqielat, Moa'Ath (Oct. 2017). "Radial Basis Function Method For Modelling Leaf Surface from Real Leaf Data". In: *Australian Journal of Basic and Applied Sciences*.

Perez, Jorge, Melisa Menendez, and Inigo J. Losada (2017). "GOW2: A global wave hindcast for coastal applications". In: *Coastal Engineering* 124, pp. 1 –11. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2017.03.005. URL: http://www.sciencedirect.com/science/article/pii/S0378383917300443.

Polinsky, A. et al. (July 1996). *Software for automated design of exploratory and targeted combinatorial libraries*, pp. 219 –232.

Qian, Chengcheng et al. (Nov. 2019). "Climatology of Wind-Seas and Swells in the China Seas from Wave Hindcast". In: *Journal of Ocean University of China* 19.1, pp. 90–100. DOI: 10.1007/s11802-020-3924-4.

Reguero, B.G. et al. (2012). "A Global Ocean Wave (GOW) calibrated reanalysis from 1948 onwards". In: *Coastal Engineering* 65, pp. 38 –55. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2012.03.003. URL: http://www.sciencedirect.com/science/article/pii/S0378383912000452.

Rippa, Shmuel (Nov. 1999). "An algorithm for selecting a good parameter c in radial basis function interpolation". In: *Advances in Computational Mathematics* 11, pp. 193–210. DOI: 10.1023/A:1018975909870.

Sola, J. and Joaquin Sevilla (July 1997). "Importance of input data normalization for the application of neural networks to complex industrial problems". In: *Nuclear Science, IEEE Transactions on* 44, pp. 1464 –1468. DOI: 10.1109/23.589532.

Vos, Kilian et al. (2019). "CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery". In: *Environmental Modelling Software* 122, p. 104528. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2019.104528. URL: http://www.sciencedirect.com/science/article/pii/S1364815219300490.

Wickham, Hadley (Sept. 2014). "Tidy data". In: *The American Statistician* 14. DOI: 10.18637/jss.v059.i10.

Wright, L.D and A.D Short (1984). "Morphodynamic variability of surf zones and beaches: A synthesis". In: *Marine Geology* 56.1, pp. 93 –118. ISSN: 0025-3227. DOI: https://doi.org/10.1016/0025-3227(84)90008-2. URL: http://www.sciencedirect.com/science/article/pii/0025322784900082.

Appendix A

Preliminary concepts

The concepts that will be described in this section are indispensable variables and ideas that are used in almost all the coastal engineering studies. These concepts can include the variables and the spectrum that define a sea state, which encompasses the different conditions that are present in the sea at a certain moment.

A.0.1 Sea state

When a surfer reaches the beach and observes the sea, different things can be observed, and this landscape is the result of a lot of different circumstances that are happening at the same time. In oceanography, sea state is the general condition of the free surface on a large body of water with respect to wind waves and swells at a certain location and moment. A sea state is characterized by statistics, including the wave height (significant wave height, H_S), period (peak period, T_P), direction (mean wave direction, θ_m) and power spectrum, $S(f, \theta)$. The sea state varies with time, as the wind conditions or swell conditions change. The sea state can either be assessed by an experienced observer, like a trained mariner, or through instruments like weather buoys, wave radar or remote sensing satellites.

In case of buoy measurements, the statistics are determined for a time interval in which the sea state can be considered to be constant. This duration has to be much longer than the individual wave period, but smaller than the period in which the wind and swell conditions vary significantly. Typically, records of one hundred to one thousand wave periods are used to determine the wave statistics.

A lot of new concepts have been introduced in this brief explanation that must be commented. Then, the variables that define a sea state, which are the wave height and the period and direction of the waves are defined as the vertical distance between the highest and the lowest surface elevation in a wave and the time interval between the start and the end of the wave and its direction, respectively. These are the main variables that define a sea state, but other important ones can be mentioned and will be explained in detail below.

Significant wave height, H_S

One of the parameters, if not the most important one when giving an assessment about waves is precisely the wave height, more specifically the breaking wave height, which depends directly on the significant wave height, which is explained next. In coastal engineering, the height of the

waves has always been one of the variables par excellence, and that more headaches and more innovations and research has contributed to the engineers, both in the oceanic branch and in the more properly engineering one, which can include the construction of breakwaters, movement of sandbanks or evenly the study and building of entire ports. This wave height is defined as the vertical distance that the sea surface suffers when a wave advances freely. Thus, if we refer to a buoy in order to understand this variable better and we put in the case that this buoy is being measuring for a period of time close to one hour, then the significant wave height will be the average of the highest third of individual heights measured:

$$H_S \mid H_{1/3} = \frac{3}{N} \sum_{i=1}^{N/3} H_i \tag{A.1}$$

being this one of the main parameters in almost all coastal studies. In the case of surfing it is the break what matters, so the wave height at break will be explained below (refer to the section Parameters that define surfing at breaker scale).

Wave period, T_m and T_P

The period is another key point in this type of studies, and it is just the frequency with which a wave presents consecutive peaks or valleys. In the field of surfing, this translates into the time a surfer who sits waiting for the next wave to come in has to wait.

However, understanding this variable is not that simple, as in real life it is rarely the case that beaches present a monochromatic train of waves (waves defined by a single period, frequency and wavelength) incident, but there are usually several waves that come from different areas and have been generated at different times that finally coincide in their arrival to coast. This is why the actual period observed theoretically is not so easy to see. In the ideal case of finding ourselves on a beach where a single differentiated swell is arriving, then the crests of the waves, as long as the wave height is large enough, can be observed spaced in space and time, see Figure A.1. But this does not end here, because even in this case where there is only a monochromatic wave on this ideal beach, there will still exist frequencial and directional dispersions. In the second one we will focus later, but the first one is in charge of making the waves arrive in order and in similar groups, with clear wave periods, but that can change from one group to another.

This is where the peak period comes into play, which will be the actual variable used in the study of the surfing variability, and which is only the period associated with the biggest waves, as they are usually the ones that attract the most intention on the part of the surfing community and those that generate the best waves, remember that bigger periods support bigger refractions. In this case, and again having in mind a buoy that is measuring for an enough interval of time, the peak period can be defined as:

$$T_P = \frac{3}{N} \sum_{i=1}^{N/3} T_i \cdot 1.05 \tag{A.2}$$

FIGURE A.1: Clear waves with a large period at Snapper Rocks, Gold Coast, Australia, that allow the surfers to catch consecutive crests and surf together at the same time (the length of the wave is also very large and allows this sharing). This image is also good to see how waves increase its height when they are close to break and breaking, as the waves appear smaller in open sea. Source: https://stabmag.com/.

where the left side of the equation refers, as in the previous case with the significant wave height, to the mean of the third biggest periods. The mean period has then an obvious formulation, as it will not be more than the standard mean of all the periods measured during that certain interval of time:

$$T_m = \frac{1}{N} \sum_{i=1}^{N} T_i \tag{A.3}$$

Mean wave direction, θ_m

Another remaining principal variable that must be included to correctly characterize a sea state is the direction of the waves, the mean wave direction if a unique bulk parameter must be presented (refer to Fig. B.1 to see how waves reach the coast with a direction, defined here as θ_1). In this way, the mean wave direction will represent the mean direction in which waves are propagating. Notice that this is a very important variable, as a change in the direction of the waves could drastically vary the location of good waves. Here in the cantabric sea, more precisely in Cantabria, the majority of the large swells come from the W-NW, and it is then when good waves appear. Windseas arriving from the E-NE present little fetchs and also little distances to propagate and order before they reach the coast, so worse waves appear (the difference between windseas and swells is explained below). For example, surfbreaks as El Sardinero (in this case it could be also seen as a beach) that present an orientation towards the E will not have good surfing conditions when medium energetic swells arrive from the W-NW part, having all this a very important significance in the morphodinamics of the beach too.

The oceanography wind fetch, also known as fetch length or simply fetch, is the length of water over which a given wind has blown without obstruction. Fetch is used in geography and

meteorology and its effects are usually associated with sea state and when it reaches shore it is the main factor that creates storm surge which leads to coastal erosion and flooding. It also plays a large part in longshore drift as well.

Fetch length, along with the wind speed (wind strength), determines the size (sea state) of waves produced. The wind direction is considered constant. The longer the fetch and the faster the wind speed, the more wind energy is imparted to the water surface and the larger the resulting sea state will be.

Returning to the direction, just mention the angular and radial dispersion that waves suffer through its path to the shore. It will be explained in detail below, but waves that travel long distances suffer larger dispersions, ending up in ground swells that can generate very good surfing conditions.

A.0.2 Spectrum and free surface elevation

But the spectrum has been also mentioned when defining a sea state. The aim of describing ocean waves with a spectrum is not so much to describe in detail one observation of the sea surface (i.e., one time record), but rather to describe the sea surface as a stochastic process and to characterise all possible observations (time records) that could have been made under the conditions of the actual observation. An observation is thus formally treated as one realisation of a stochastic process. Here, we base this treatment on the random-phase/amplitude model (Holthuijsen, 2007), which leads to the wave spectrum, which is the most important form in which ocean waves are described. The basic concept of the wave spectrum is simple, but its many aspects make it seem rather complicated. To distinguish the essence from these additional aspects, consider first a wave record, the surface elevation $\eta(t)$ at one location as a function of time, with duration D, obtained at sea with a wave buoy or a wave pole (see Fig. A.2).
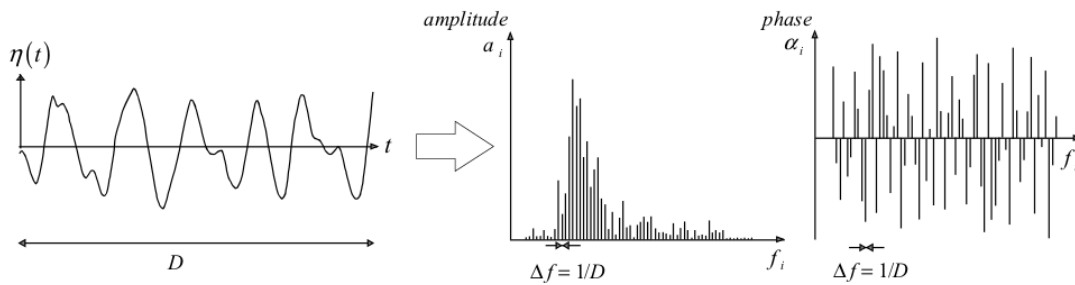

FIGURE A.2: The observed surface elevation and its amplitude and phase spectrum. Source: (Holthuijsen, 2007).

We can exactly reproduce that record as the sum of a large number of harmonic wave components (a Fourier series):

$$\eta(t) = \sum_{i=1}^{N} a_i \cos(2\pi f_i t + \alpha_i) \tag{A.4}$$

where a_i and α_i are the amplitude and phase, respectively, of each frequency $f_i = i/D$ (i = 1, 2, 3, ...; the frequency interval is therefore $\Delta f = 1/D$). With a Fourier analysis, we can

determine the values of the amplitude and phase for each frequency and this would give us the amplitude and phase spectrum for this record (see Fig. A.2). By substituting these computed amplitudes and phases into Eq. A.4, we exactly reproduce the record.

For most wave records, the phases turn out to have any value between 0 and 2π without any preference for any one value. Since this is almost always the case in deep water (not for very steep waves), we will ignore the phase spectrum (just keep this uniform distribution in mind and apply that knowledge when called for).

The above one-dimensional variance density spectrum characterises the stationary, Gaussian surface elevation as a function of time (at one geographic location). To describe the actual, three-dimensional, moving waves, the horizontal dimension has to be added. To that end we expand the random phase-amplitude model by considering a harmonic wave that propagates in x, y-space, in direction θ relative to the positive x-axis (we use w instead of f for the sake of brevity in the notation):

$$\eta\left(x,y,t\right) = a\cos\left(\omega t - k\,x\cos\theta - k\,y\sin\theta + \alpha\right) \tag{A.5}$$

where the wave number $k = 2\pi/L$ (where L is the wave length of the harmonic wave) and θ is the direction of wave propagation (i.e., normal to the wave crest of each individual component). Analogously to the one-dimensional model, the corresponding three-dimensional random phase-amplitude model (in x, y and t-space) is the sum of a large number of such propagating harmonic waves (see Fig. A.3):

$$\eta\left(x,y,t\right) = \sum_{i=1}^{N}\sum_{j=1}^{M} a_{i,j}\cos\left(\omega_i t - k_i\,x\cos\theta_j - k_i\,y\sin\theta_j + \alpha_{i,j}\right) \tag{A.6}$$

where every individual wave component in this three-dimensional model has a random amplitude $a_{i,j}$ (Rayleigh distributed) and a random phase $\alpha_{i,j}$ (uniformly distributed). Furthermore, this two-dimensional random phase-amplitude model represents a Gaussian process that is stationary in time and homogeneous in x, y-space: a spatial pattern of chaotically moving surface elevations, seen as the sum of many wave components propagating with various amplitudes, phases and frequencies (or wave lengths) in various directions across the ocean surface. The effect is a realistic representation of random, short-crested waves (see Fig. A.3).

The amplitude spectrum provides enough information to describe the sea-surface elevation realistically as a stationary, Gaussian process. However, for several reasons, it is more relevant to present the information in this spectrum in a different way: consider the variance $E\left\{\frac{1}{2}a_i^2\right\}$ rather than the expectation of the amplitude $E\left\{a_i\right\}$ (Holthuijsen, 2007). In other words, consider the variance spectrum instead of the amplitude spectrum. This seems trivial and also enough to characterise the sea-surface elevation. However, both the amplitude and the variance spectrum are based on discrete frequencies and directions, whereas Nature does not select such discrete quantities. All frequencies and directions are present at sea. The random phase-amplitude model needs therefore to be modified. This is done by distributing the variance $E\left\{\frac{1}{2}a_i^2\right\}$ over the frequency interval Δf_i at frequency f_i and the direction interval $\Delta\theta_i$ at direction θ_i. This spectrum is defined for all frequencies and directions, but it still varies discontinuously from

FIGURE A.3: Free surface reconstruction graphically explained as a sum of the different frequencies and directions existent in the analyzed spectrum. Source: (Holthuijsen, 2007).

one frequency or direction band to the next. A continuous version is obtained by having both widths bands approaching zero, see Fig. A.4:

$$E(f,\theta) = \lim_{\Delta f \to 0} \lim_{\Delta \theta \to 0} \frac{1}{\Delta f \Delta \theta} E \left\{ \frac{1}{2} a^2 \right\} \tag{A.7}$$

where the spectrum is calculated in terms of the frequency f and the direction θ. The variance density spectrum gives a complete description of the surface elevation of ocean waves in a statistical sense, provided that the surface elevation can be seen as a stationary, Gaussian process. This implies that all statistical characteristics of the wave field can be expressed in terms of this spectrum. The dimension and S.I. unit of the variance density $E(f,\theta)$ follow directly from its definition: the dimension of the amplitude a is [length] and its S.I. unit is [m]; the dimension of the frequency band Δf is [time]$^{-1}$ and its S.I. unit is [s^{-1}] or rather [Hz]; and the dimension of the direction θ is [degree] and its S.I. unit is [radians]. The dimension of $E(f,\theta)$ is therefore [length^2 / (time $\cdot$ degree)] and its unit is [m^2 / (Hz $\cdot$ radians)].

The standard way to calculate the spectrum has been presented, as once the free surface elevation is obtained, this variance density spectrum can be easily obtained, but in this work, the spectrum calculation has been performed in coast, where partitioned sea state variables (partitions will be explained below) and not free surface elevations exist. In this way, this spectrum reconstruction is performed using the variables propagated H_S, T_P, T_m and θ_m, where $S(f,\theta) = S(f) \cdot D(\theta)$, being $S(f)$ the spectral energy associated to the frequency and $D(\theta)$ the spectral energy associated to the direction, is the total wave energy of the spectrum. These two energy variables grouped represent the total spectral energy associated to each frequency and direction in a sea state and are calculated using Equations A.8 and A.11, which are also very well explained in (Kumar et al., 2017). First, the frequency distributed spectrum is calculated
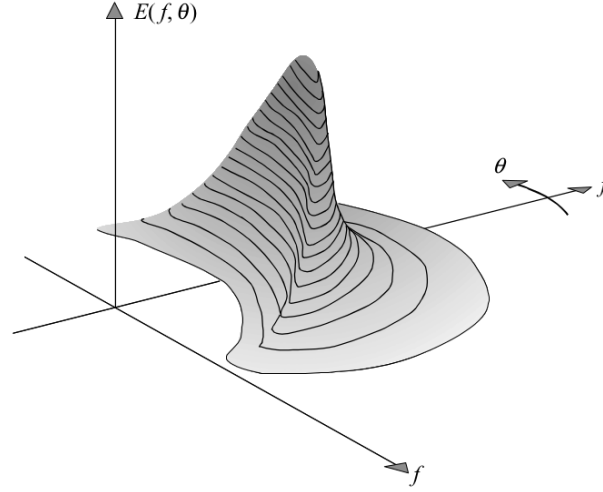
FIGURE A.4: The two-dimensional spectrum of wind-generated waves (shown in polar coordinates) at time t. Source: (Holthuijsen, 2007).

for each frequency as:

$$S(f) = \alpha \cdot H_S^2 \cdot T_P^{-4} \cdot f^{-5} \cdot \exp\left(-1.25(T_P \cdot f)^{-4}\right) \cdot \gamma^{\exp \frac{-(T_P \cdot f - 1)^2}{2\sigma^2}} \tag{A.8}$$

where γ gives an idea about the shape of the spectrum, being high for very disordered spectrums and lower for narrower and more powerful swells, and α and σ are calculated in the next way:

$$\alpha = \frac{0.06238}{0.23 + 0.0336 \cdot \gamma - 0.185 \left(1.9 + \gamma\right)^{-1}} \left(1.0940.01915 \cdot \ln \gamma\right) \tag{A.9}$$

$$\sigma = 0.07 \ \text{if} \ f < f_P \ | \ \sigma = 0.09 \ \text{if} \ f \geq f_P \tag{A.10}$$

being $f_P = T_P^{-1}$ the peak frequency of the spectrum. For the directional spectrum, the spectral energy is also calculated for each direction:

$$D(\theta) = \frac{2^{2s-1}}{\pi} \frac{\Gamma^2(s+1)}{\Gamma(2s+1)} \cos \frac{\theta - \theta_S}{2}^{2s} \tag{A.11}$$

where $s = \frac{2}{\sigma_\theta^2} - 1$ is a shape parameter, θ_S is the direction of the peak frequency and $\int_0^{2\pi} D(\theta) d\theta = 1$. Once these spectra are reconstructed we just have to calculate the aggregated parameters (as our surfing index will use bulk parameters), which can be done using some formulas that are explained in (Espejo Hermosa, 2011) and also in (Kumar et al., 2017), but in this case we will use a computational software called wavespectra. This software uses those equations, but it is already validated by different people, and this accompanied with the results obtained in Figures 3.12 and 3.11 prove the methodology used. Notice that in this work, we have reconstructed the spectrums in each of the surfbreaks of interest, thus having a historical record of totally explained surfing conditions in those spots. Finally, just comment here the reduction performed in the data in this step as the calculation of the spectrums was a very

expensive task at a computational level, so the data was reduced to just daily hours where the sport can be done, which was also one of the ideas that had to be taken into account in the study. This reduces the amount of data to more than a half, but ending up with a total number of observations of about 50000 for each surfbreak, so the subsequently explained neural network has been able to work correctly.

Wave partitions: Windseas and Swells

Last, having this spectrum explanation (refer to Figs. A.4 and A.5) it is possible to differentiate between the two main types of waves that a surfer can encounter when approaching the beach. If the spectrum is narrow and energetic, high wave heights form energetic spectrums, thus the train of waves reaching the shore can be considered a swell while broader and less energetic spectrums represent the existence of sea states called windseas, as these waves did not have the time enough to correctly abandon the location where they were formed and did not have time enough to get ordered both in frequencies and directions. These differences are also seen in the way the data from the wave reanalysis that will be explained below is obtained. These numerical models calculate different partitions that are present at the same time in a sea state, thus giving information about a windsea and different swells that can be reaching the shore at the same time. In our case (CSIRO wave reanalysis), one windsea will be always present if wind exists, and a total amount of three swells would be calculated.

To summarize, spectrums are the best possible option for the study of the wave climate in our days. They actually provide all the necessary information to define a sea state correctly and can be calculated by two different ways. First, based on the free measured surface and second, using the previously described equations and calculating the energy associated to each frequency and direction based on the sea states variables existent in that moment, which is the case we have followed, as the partitioned data we had in the surfbreaks had total, H_S, T_P, T_m and θ_m, (bulk is not used here as it is preferred for the total sea state) parameters for each of the different windsea and swell partitions.
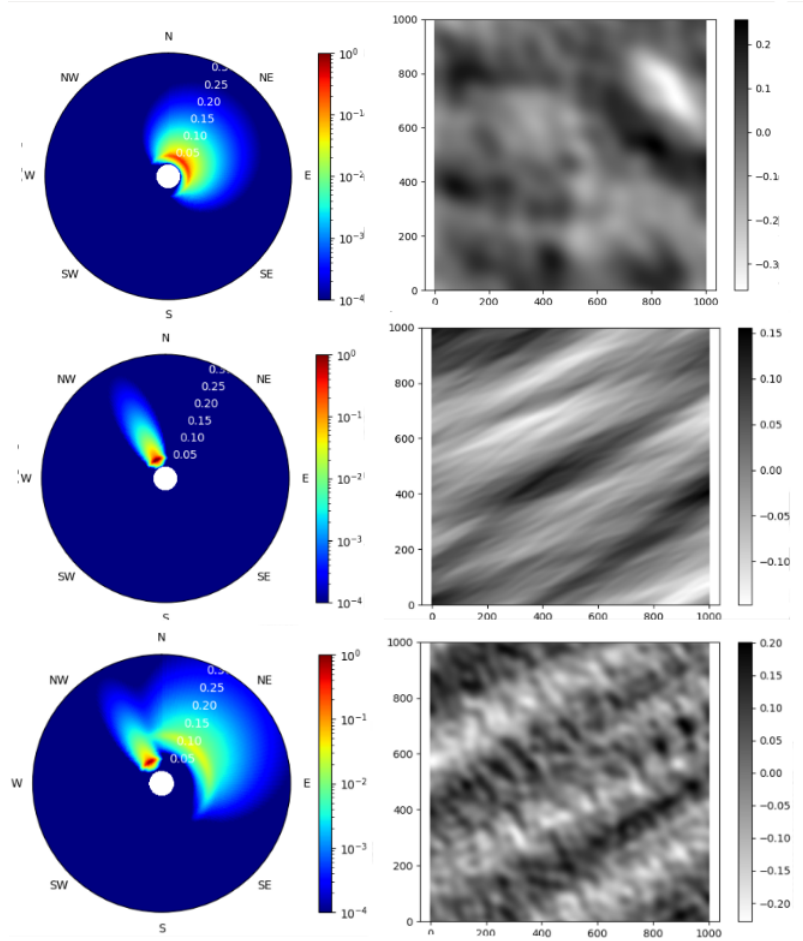
FIGURE A.5: Spectrum and free surface reconstructions are shown. The axis are frequency [Hz], direction [rad] and energy [m^2 / (Hz · rad)] in the figures in the right and space [m] and elevation [m] in the other ones. As it can be seen, the different subplots show three different sea states. The first one (top of the image) represents a windsea, very spread in frequencies and directions that come from the NE, which is the normal case in the cantabric sea. The free surface associated to it shows a disturbed see state where mean waves appear, generating bad surfing conditions. Second, a very well defined swell coming from the NW is shown. In this case, waves seem clean and the conditions are perfect though. Lastly, in the bottom image, one sea state with a swell coming from the NW but also a windsea approching with NE direction is shown. These are the cases where intelligent surfers can catch good waves, but it is not that easy as in the previous case. Notice that the majority of the swells come always from the W-NW in the cantabric sea, as it is in this direction where the major portion of sea is encountered ("fetch", this was previously mentioned). Finally, the surface reconstruction shows very different surfing sessions. In the left, a disperse and untidy summer day is shown and in the right, an ordered and epic session is waiting for us.
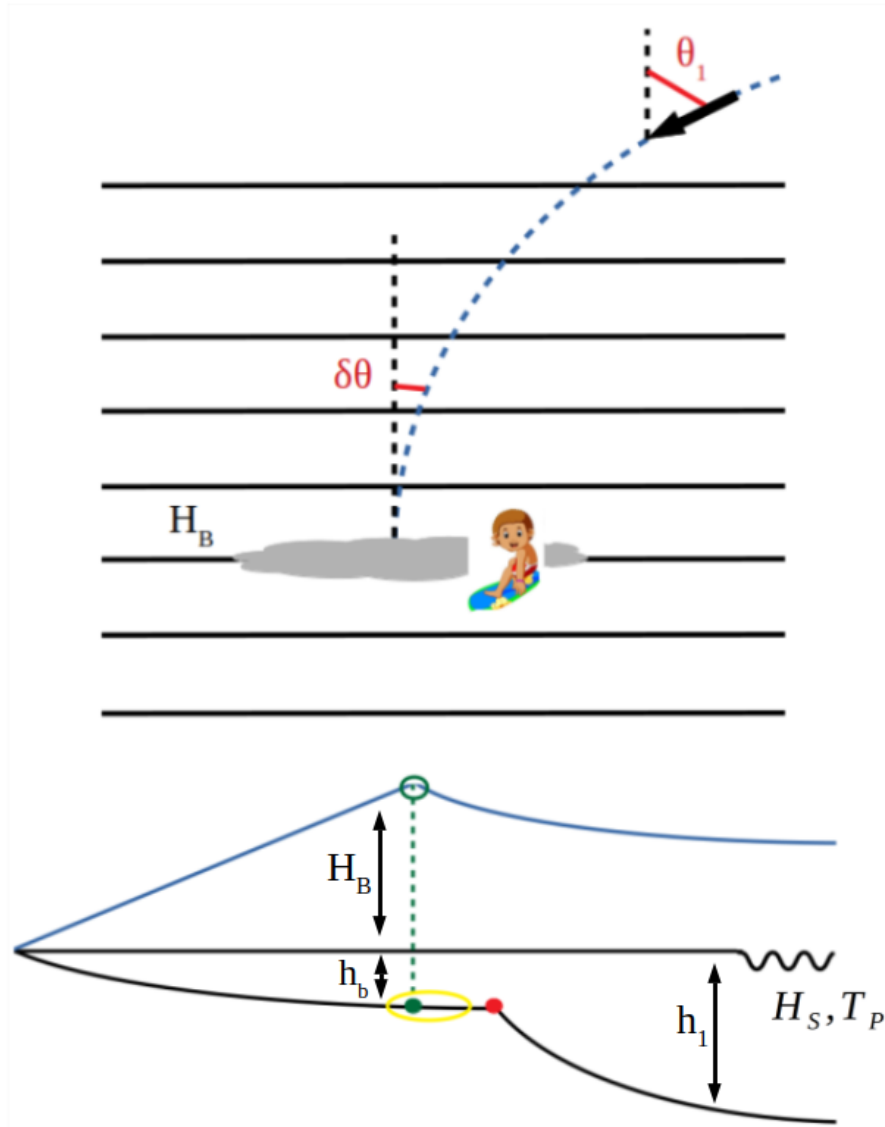
Appendix B

Images

FIGURE B.1: This figure shows the two previously mentioned phenomenons that the wave suffers when approximating the shore. First, refraction (snell law refraction) appears changing the value of the relative mean direction to the coast but also changing the values for the wave height, the period and the dispersion or cleanliness of the waves, and finally, the effect of the bottom is seen by the wave, which increases its height until it breaks, generating the surfable wave. All the existent variables have already been explained, just notice that the variable θ_1 is already the realtive direction to the coast, and not just the mean wave direction, as the same criteria must be applicable in the study to all the surfbreaks.

FIGURE B.2: These two pictures show how the principal surfing websites, windguru on the left and surf forecast in the right, show their predictions to the surfing community. Both predictions have similarities, but there are also differences that must be commented. On the left, the windguru website shows a more quantitative prediction, as it just shows numbers and arrows for the direction of the waves and the wind, sixth and third lines respectively. The other lines represent the peak period, the significant wave height... everything already explained and that can be consulted in the web. Notice that in this case, the windguru surfing index rates the existence of strong winds as good (the range of quality goes from 1 to 3 stars), and this is not what has been explained here. On the other hand, the surf forecast prediction is more of our style, as it is more focused on standard surfing and it shows other very important qualitative factors as the wave height map of the region of interest, the existence of different swells approximating coast in the moment of the prediction (what we called partitions previously in this chapter), the tide level and a rating range between 1 and 5 that appears for every single prediction, and which is focused, as we already said, in standard surfing, and not into windy sports such as windsurfing, kitesurfing or sailing, which is not our objective in this project. Sources: https://www.windguru.cz/48699 and https://es.surf-forecast.com/breaks/Liencres/forecasts/latest/six$_d$ay, respectively.
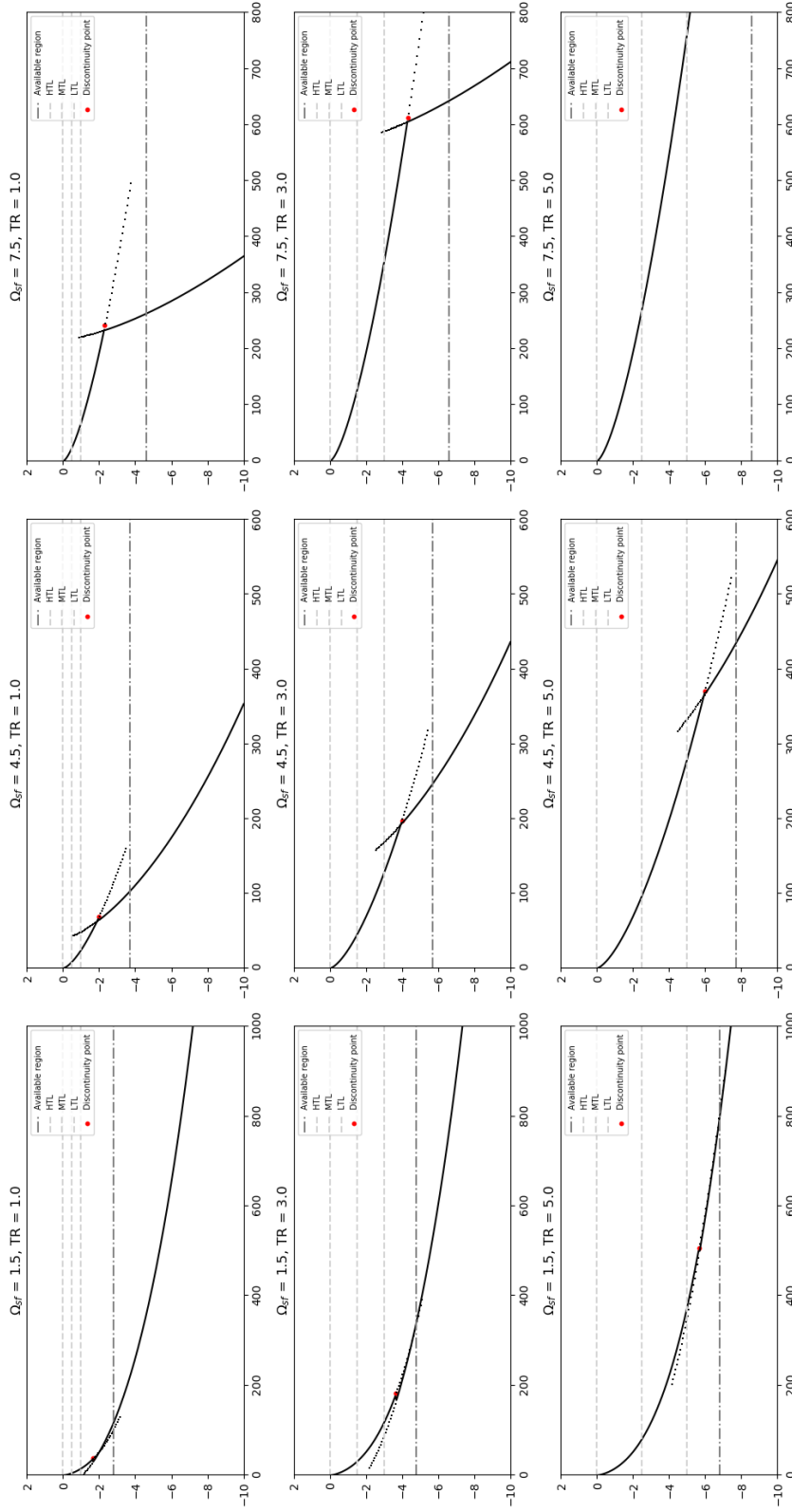
FIGURE B.3: Different types of biparabolic beach profiles for each value set for the $\Omega_{sf} = \Omega$ in Eq. 2.3 are shown. These profiles will be used for the calculation of the slope the wave sees when reaching its breakage. The profile in the right-down part of the figure does not show its discontinuity point (shoaling-surfing), as the tidal range is too large that the profile changes its shape far away from the shore.
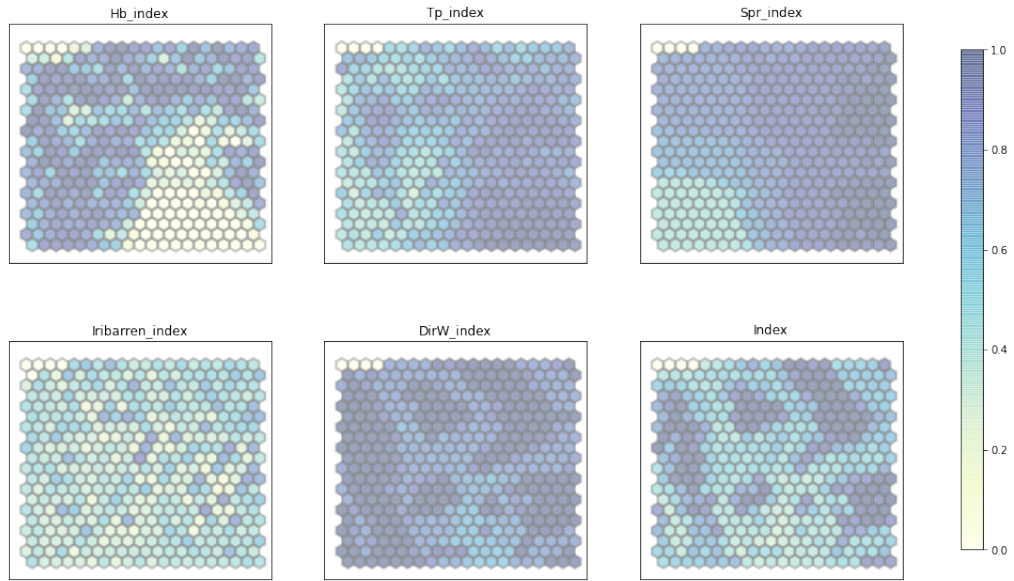
FIGURE B.4: The different individual and the global index calculated are shown.

FIGURE B.5: Probabilities for each cluster in the different beaches are exhibited.
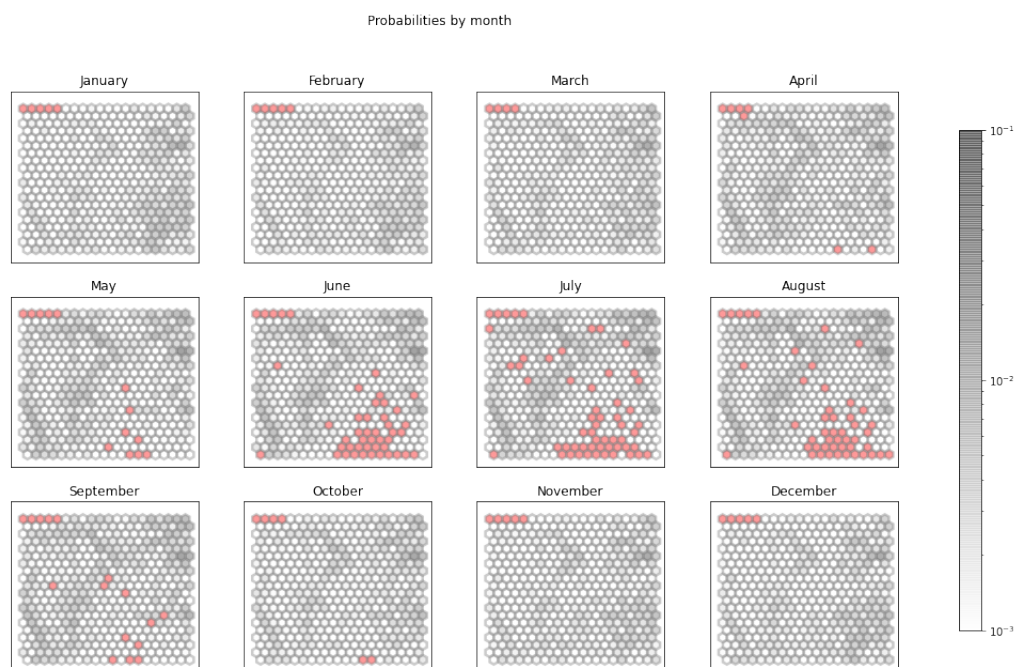
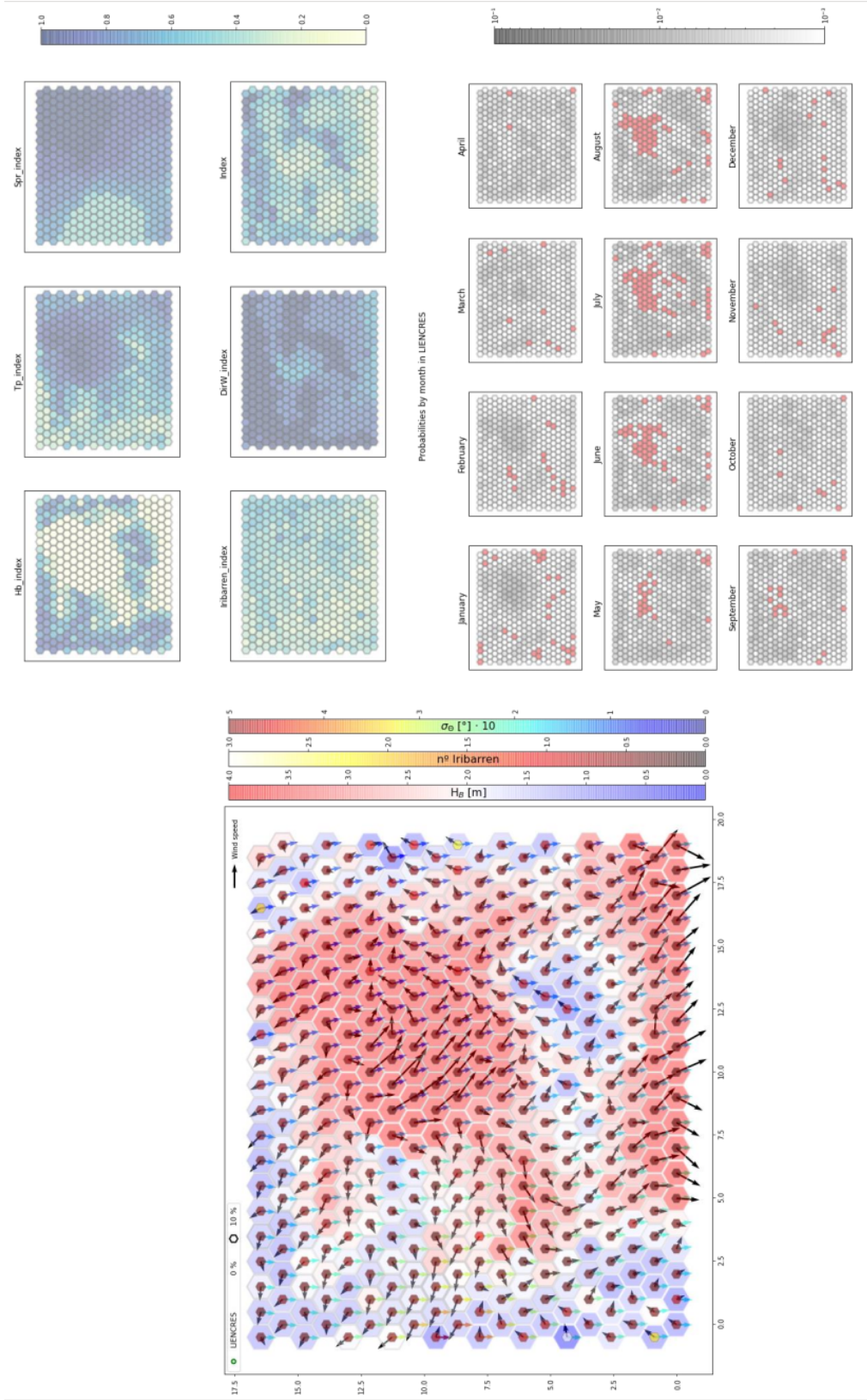FIGURE B.6: Probabilities for each cluster in the different months are exhibited.

FIGURE B.7: The results at Liencres beach are very close to the observed reality and are just a good summary of everything mentioned so far in the study. The temporal variability is clear, in winter months very big waves flatten the profile while strong winds appear, generating wave breaks that do not give rise to good surfing, even when good conditions of wave height and period appear. On the other hand, in the summer months is when the less energetic waves appear, generating good surfing conditions for the enjoyment of the sportsmen. In months like September - October, which has been previously mentioned, very good conditions appear, and this is due to the state of the steeper profile due to the passage of summer. The number of clusters is very high, and it is possible that many of them present surfing conditions that cannot be clearly framed in reality, but it is always possible to reduce the number of them to make a more rough study. Finally, simply emphasize the similarity between the results obtained and the observed reality, since the temporal variability on this beach is traced perfectly.
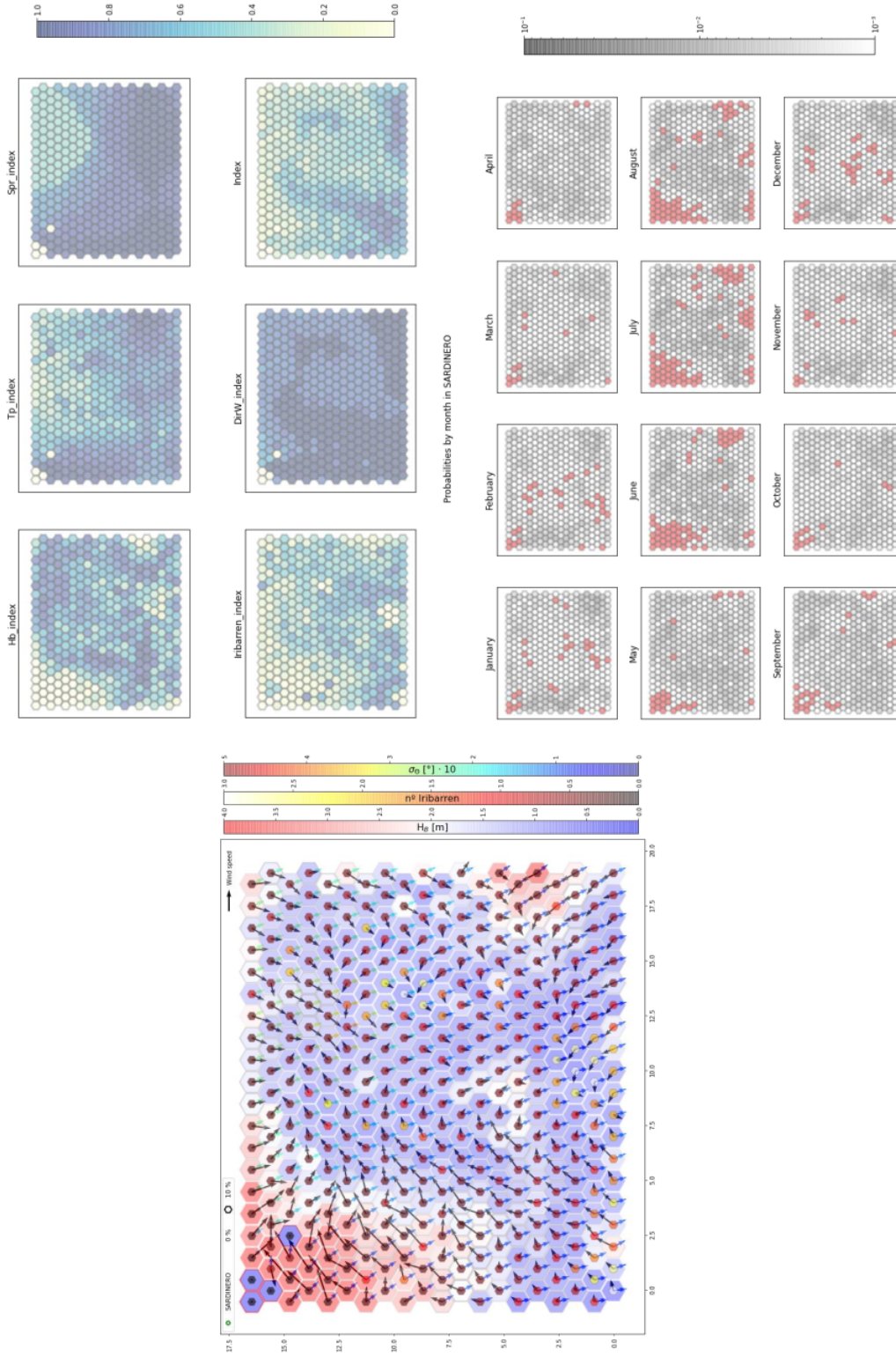
FIGURE B.8: As far as El Sardinero beach is concerned, the way in which more and less energetic waves appear is obviously the same as far as the time is concerned, but the analysis is very different. The beach of El Sardinero is oriented to the East, and this makes only those swells with great wave heights and periods reach the shore to break (remember that the main direction of the swells on the Cantabrian coast is W-NW). This way, good surfing conditions will appear in those winter months where the waves are bigger, and accompanied by high periods and offshore winds, as the so-called "galician" is the most common wind in these months, generating surfable waves. Another aspect to mention in this beach is its reflective character, this is due to the constant legacy of weaker seas to the coast during a large part of the year, which makes the profile steeper ("pindio" in spanish) than in other beaches of the coast. This profile, and more specifically its abrupt change from high to low tide, is capable of generating tube waves (plunging) in certain particulars. Again, it should be noted that the number of clusters is very high, perhaps even exceeding the number of observable circumstances, but the temporal variability fits with the observed, as does the spatial variability if we put the two beaches mentioned together.