Escuela Técnica Superior de
Ingeniería de Caminos, Canales y Puertos
UNIVERSIDAD DE CANTABRIA

AUTOMATIC CALIBRATION OF HYDRODYNAMIC MODELS FOR FRINGING REEFS

Trabajo realizado por:
Pablo Zubía Palazuelos

Dirigido:
Fernando Méndez Incera
Beatriz Pérez Diaz

Titulación:
Máster Universitario en Ingeniería de Caminos, Canales y Puertos

Santander, Octubre 2024.

Automatic Calibration of Hydrodynamic models for Fringing Reefs

Pablo Zubía Palazuelos

Santander, October 2024

Acknowledgements

I would like to take this opportunity to express my heartfelt gratitude to everyone who has supported me throughout my university journey:

First and foremost, I want to thank my parents for their unwavering support and unconditional trust throughout my studies. Their encouragement and respect for every decision I have made have been invaluable.

I am also deeply grateful to my tutor, Fernando Méndez, for giving me the incredible opportunity to work on such an engaging project. I have learned so much from this experience. My thanks extend to Beatriz and Alba for their endless patience and dedication whenever I needed their help.

Additionally, I appreciate the chance to pursue a double degree with the École des Ponts. This experience has been truly unforgettable, offering me incredible memories and valuable lessons that will undoubtedly benefit me throughout my life.

Lastly, I want to acknowledge all my friends—both from outside the university and those from within, including my classmates and colleagues from the department. To everyone who has shared this journey with me over these seven years of university, thank you very much.

Abstract

Title. Automatic Calibration of Hydrodynamic Models for Fringing Reefs
Author. Pablo Zubía Palazuelos.
Directors. Fernando Méndez Incera, Beatriz Pérez Diaz and Alba Ricondo Cueva.
Call. October 2021.
Degree. Master's in Civil Engineering.
Specialization. Water, Energy, and Environment.
Keywords. Nearshore hydrodynamics, Hybrid modelling, Fringing Coral Reefs, Wave Breaking Coefficient, Friction Coefficient.

As global sea levels rise and storm frequency and intensity shift due to climate change, tropical coral reef-lined coasts are becoming increasingly susceptible to wave-driven flooding. In addition, coral reefs, which are essential for coastal protection, are deteriorating due to ocean acidification and other environmental stressors, thereby weakening their ability to dissipate wave energy. In this scenario, accurate downscaled predictions of nearshore wave processes are vital for reducing the susceptibilities of these environments and developing adequate adaptation strategies.

To achieve these accurate predictions, traditionally, forecasting tools relied on modeling coastal dynamics using high-fidelity numerical models. While these models provide detailed simulations, they are costly in terms of computational resources when applied under a dynamic downscaling approach. To overcome this limitation, hybrid approaches, also known as metamodels, have been developed. These metamodels combine numerical models with statistical techniques, aiming to reduce computational costs by predicting wave behavior with fewer exhaustive simulations.

However, the accuracy achieved by these approaches is highly dependent on the correct estimation of certain calibration coefficients of the numerical models. A calibration process commonly involves comparing real measurements with the numerical model outputs to identify the most accurate set of calibration coefficients. The longer the observations and the greater the number of coefficients to be calibrated, the more the number of required simulations increases exponentially, which can lead to significant computational effort and make it difficult to obtain accurate results.

To address these challenges, this project introduces CHySwash, an advanced methodology that builds upon the foundation of its predecessor, HySwash (Ricondo et al., 2024). CHySwash is designed to streamline the calibration of numerical models, significantly reducing the time and computational resources typically required. It achieves this by integrating advanced techniques, including sampling, clustering, and interpolation, alongside an automatic calibration process powered by the Shuffled Complex Evolution optimization algorithm, renowned for its effectiveness in parameter optimization. The proposed methodology is applied in a monitored coral reef-lined coast, Molokai, Hawaii. Specifically, we aim to predict the optimal wave breaking and friction coefficients, which govern the wave breaking process and the dissipation of wave energy.

Resumen

Título. Calibración automática de modelos hidrodinámicos para arrecifes en franja.
Autor. Pablo Zubía Palazuelos.
Directores. Fernando Méndez Incera, Beatriz Pérez Diaz y Alba Ricondo Cueva.
Convocatoria. Octubre 2024.
Titulación. Máster en Ingeniería de Caminos, Canales y Puertos.
Especialidad. Agua, Energía y Medioambiente.
Palabras clave. Hidrodinámica costera, Modelado híbrido, Arrecifes de coral de franja, Coeficiente de rotura de olas, Coeficiente de fricción.

A medida que el nivel global del mar aumenta y la frecuencia e intensidad de las tormentas varían debido al cambio climático, las costas tropicales protegidas por arrecifes de coral se vuelven cada vez más vulnerables a las inundaciones. Además, estos arrecifes de coral, que son esenciales para la protección costera, se están deteriorando debido a factores como la acidificación del océano y otro tipo de impactos ambientales, debilitando así su capacidad para disipar la energía de las olas. En este contexto, las predicciones precisas de los procesos de oleaje cercanos a la costa son cruciales para mitigar las vulnerabilidades de estos ecosistemas y desarrollar estrategias de adaptación adecuadas.

Con el fin de obtener dichas predicciones precisas, tradicionalmente, se ha dependido de la modelización de la dinámica costera a través de modelos numéricos de alta fidelidad. Si bien estos modelos proporcionan simulaciones muy detalladas, son costosos en términos de recursos computacionales. Para superar esta limitación, se han desarrollado enfoques híbridos, también conocidos como metamodelos. Estos metamodelos combinan modelos numéricos con técnicas estadísticas, con el objetivo de reducir los costos computacionales al predecir el comportamiento de las olas con menos simulaciones necesarias.

Sin embargo, la precisión que se alcanza con estos enfoques depende en gran medida de la correcta estimación de ciertos coeficientes de calibración de los modelos numéricos. El proceso de calibración comúnmente implica comparar mediciones reales con los resultados obtenidos del modelo numérico para identificar la combinación más precisa de dichos coeficientes de calibración. Cuanto más largas sean las observaciones y mayor sea el número de coeficientes a calibrar, mayor será el número de simulaciones necesarias, lo que puede conllevar un esfuerzo computacional significativo y dificultar la obtención de resultados precisos.

Para abordar estos desafíos, este proyecto introduce CHySwash, una metodología avanzada basada en los fundamentos de su predecesor, HySwash (Ricondo et al., 2024). CHySwash está diseñado para agilizar la calibración de modelos numéricos, reduciendo significativamente el tiempo y los recursos computacionales típicamente necesarios. Lo logra integrando técnicas avanzadas, como el muestreo, la agrupación y la interpolación, junto con un proceso de calibración automático impulsado por el algoritmo de optimización *Shuffled Complex Evolution*, conocido por su efectividad en la optimización de parámetros. La metodología propuesta se implementa en una costa rodeada por arrecifes de coral en Molokai, Hawái, donde se ha llevado a cabo un monitoreo exhaustivo. Nuestro objetivo es estimar los coeficientes óptimos de rotura de olas y fricción del coral, los cuales regulan el proceso de rotura y la disipación de la energía de las olas.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Hs	Significant Wave Height
WL	Water Level
Hs/Lo	Wave Steepness
Tp	Peak Period
Cr	Wave Breaking Coefficient
Cf	Coral Friction Coefficient
Hrms	Root Mean Square Wave Height
HsIG	Infragravity Significant Wave Height
LHS	Latin Hypercube Sampling
MDA	Maximum Dissimilarity Algorithm
PCA	Principal Component Analysis
RBF	Radial Basis Function
SCE	Shuffled Complex Evolution
RMSE	Root Mean Square Error

1. Introduction

1.1. Background and motivation

As sea level and storm frequency and intensity are expected to rise along numerous coastlines throughout the world (Oppenheimer et al., 2022), tropical coral reef-lined coasts face increasing exposure to storm wave-driven flooding. In addition, ocean warming and acidification pose an increasing threat to the structural integrity of reefs compromising their ability to mitigate wave energy and protect coastlines. This deteriorating condition of coral reefs poses a significant challenge, as these bio-geomorphic systems play a crucial role in coastal protection (Ferrario et al., 2014).

Given the ecological importance and inherent vulnerability of coral reef-lined coasts (Winter et al., 2020), addressing the escalating risks associated with storm wave-driven flooding is of paramount importance. This challenge necessitates accurate predictions of coastal wave processes, which are essential for the effective design and implementation of nature-based shoreline protection measures and coastal zone management strategies (Clark, 1997). By integrating predictive tools into management practices, decision-making can be more informed, thereby enhancing the resilience of coastal ecosystems to wave impacts.

To effectively address these challenges, it is crucial to enhance our understanding of how reef morphology, vegetation characteristics, and hydrodynamic conditions influence wave behavior. Fringing coral reefs, with their varied topographies, from steep slopes to shallow platforms, have the potential to dissipate the majority of the incoming wave energy (Brander et al., 2004 ; Péquignet et al., 2011 ; Pomeroy et al., 2012). The roughness of the reef surface and wave breaking at the crest are vital for energy dissipation and the formation of nearshore currents. Therefore, accurately estimating wave breaking (Cr) and friction (Cf) coefficients is essential for modeling these processes effectively. Precision in these estimates is crucial for generating realistic outcomes in coastal management and improving risk assessments (Lee et al., 2021).

In order to reduce the computational effort of classical wave downscaling, hybrid approaches, also referred to as metamodels, have emerged as potential alternatives to transfer nearshore waves to coastal areas (e.g., Camus et al., 2011a, Ricondo et al., 2023), or nearshore waves to the evolution of the surf-zone hydrodynamics (Ricondo et al., 2024, Pearson et al., 2017). These methods incorporate high-fidelity numerical models with statistical techniques (i.e., data mining, selection and interpolation) to create transfer functions of minimum computational cost. However, when applied to real coral reef morphologies, these metamodels often require separate calibration of the numerical models, which can significantly increase computational demands. Calibrating hydrodynamic models for nearshore wave processes typically involves an iterative process of adjusting model input coefficients and comparing simulations with observed data. As the number of coefficients and the length of observations increase, the number of required simulations grows exponentially, making calibration more computationally intensive.

To address these challenges, we propose CHySwash, an innovative metamodel that builds upon the foundation of its predecessor, HySwash (Ricondo et al., 2024). CHySwash seamlessly integrates advanced techniques such as sampling, clustering, and interpolation with the Shuffled Complex Evolution (SCE) optimization algorithm (Duan et al., 1994), known for its effectiveness in parameter optimization. By streamlining the calibration process of numerical models, CHySwash significantly

reduces the time and computational resources typically required, while enhancing prediction precision. This approach not only accelerates simulations but also improves accuracy by adjusting key model parameters, such as wave breaking and friction coefficients, based on observational data. The proposed methodology is applied to the coral reef-lined coast of Molokai, Hawaii, ultimately supporting more informed decision-making for coastal protection and adaptation strategies.

1.2. Objectives

The primary objective of this project is to develop a robust and efficient methodology for the automatic calibration of hydrodynamic models. To achieve this, the project focuses on the development, application, and validation of a hybrid statistical-numerical metamodel, CHySwash, that integrates an optimization technique for finding the optimum calibration coefficients. This approach aims to enhance the efficiency and accuracy of model calibration processes.

The specific objectives are:

1. **Develop the CHySwash metamodel,** based on its predecessor HySwash (Ricondo et al., 2024)**.** The primary objective is to create the CHySwash metamodel, a hybrid tool that combines statistical and numerical methods to simulate nearshore wave dynamics with high precision. This metamodel will integrate advanced techniques such as sampling, clustering, and interpolation with the established 1D hydrodynamic model SWASH (Simulating WAves till SHore) (Zijlema et al., 2011). The goal is to develop a model that efficiently predicts wave behavior in coastal areas, offering both accuracy and reduced computational time. This involves creating a response function through interpolation methods after simulating representative cases. This function approximates the behavior of the original model, enabling predictions for parameter combinations that were not directly simulated. As a result, the CHySwash metamodel will allow for the reconstruction of extensive datasets without the need for time-consuming direct simulations, thus optimizing the model's performance and usability.

2. **Apply the CHySwash metamodel to the Molokai study area.** The second objective is to apply the CHySwash metamodel to a real-world coastal setting, specifically the southern coastal area of Molokai Island, Hawaii, where extensive sensor data is available. The goal is to use the metamodel to accurately reconstruct wave characteristics, such as significant wave height (Hs) and infragravity wave height (Hs_{IG}) , based on this data. This application involves calibrating key coefficients—namely, the wave-breaking coefficient and the coral friction coefficient—to ensure that the model reflects observed hydrodynamic conditions accurately. The calibration process will be validated by comparing the metamodel's predictions with actual sensor measurements, aiming to minimize prediction errors. By utilizing real sensor data from Molokai, the metamodel will reconstruct wave evolution in the nearshore zone using five defined parameters: significant wave height (Hs), wave steepness (Hs/Lo), water level (WL), friction coefficient (Cf), and wave-breaking coefficient (Cr). Special emphasis will be placed on calibrating the last two coefficients to align the model outputs with real-world observations.

3. **To implement an optimization algorithm that automates the calibration process.** The third objective is to enhance the calibration process by implementing an optimization algorithm that automates the search for the best combination of calibration coefficients. This will be achieved using the SCE algorithm. Specifically, the SCE will be programmed to systematically search for the combination of calibration coefficients that results in the lowest error, ensuring the model best aligns with the observed data. By automating this process, we aim to streamline calibration and improve the accuracy of the model's predictions.

1.3. Project Outline

The project is structured to ensure a comprehensive understanding of the motivation and development of CHySwash. The 1.1. Background and Motivation section provides essential context about the increasing exposure of tropical coral reef-lined coasts to wave-driven flooding. This section underscores the critical role of coral reefs in coastal protection and highlights the urgent need for effective decision-support tools for disaster managers and coastal planners. Following this, the 1.2. Objectives are clearly defined, focusing on creating a tool capable of predicting the optimal combination of wave breaking and coral roughness coefficients. The study specifically aims to characterize these coefficients using the innovative one-dimensional hybrid hydrodynamic model, CHySwash, calibrated with real data from sensors deployed on Molokai Island, Hawaii.

The next major section, 2. Study Area and Data, provides a comprehensive overview of the Molokai Reef tract. The 2.2. Field Data subsection describes the deployment of 19 strategically positioned sensors along Molokai's southern coast. These sensors captured various hydrodynamic parameters crucial for accurately modeling the processes impacting the reef. Additionally, the 2.3. Bathymetric Profile subsection presents a detailed bathymetric profile of the study area, which is essential for the accurate simulation of hydrodynamic processes.

In the 3. CHySwash Methodology section, the project delves into the technical aspects of the study. The 3.1. Parameterization subsection explains the careful selection process for hydrodynamic parameters, wave breaking coefficients, and friction coefficients, all of which are critical for accurate modeling of wave dynamics on coral reefs. Following this, the 3.2. Sampling and Selection subsection describes the application of Latin Hypercube Sampling (LHS) and the Maximum Dissimilarity Algorithm (MDA). These techniques are used to create a representative dataset for simulation. The 3.3. Numerical Model SWASH subsection provides detailed information on the setup and configuration of the numerical model used to simulate wave dynamics and their impact on the reef. The 3.4. Reconstruction subsection outlines methodologies for creating the interpolation surface using Principal Component Analysis (PCA) and Radial Basis Function (RBF) interpolation. This allows for the estimation of output variables for input combinations that were not modeled. To ensure the model's accuracy and reliability, the 3.5. Numerical Validation subsection describes the application of k-fold cross-validation. Finally, the 3.6. Automatic Calibration subsection details the process of refining the model's accuracy. This is achieved through hyperparameter optimization, using real sensor data to fine-tune the model.

The 4. Results section provides a comprehensive overview of the numerical simulations, focusing on the calibration and accuracy of the hydrodynamic model. It highlights the identification of optimal calibration coefficients, which significantly improved the precision of reconstructed variables.

The section also addresses the model's limitations in handling bidimensional effects, particularly at the reef crest, where wave dynamics are more complex. Additionally, other applications of the metamodel are presented, being the flood proxy concept introduced, demonstrating the model's capability to assess coastal flood risks under various future climate scenarios.

Following this, the 5. Environmental Self-Assessment and Reflection on the Project's Contribution to Transitions section evaluates the project's impact according to Green Budget Criteria, focusing on areas such as climate change mitigation, adaptation to climate change, water resources management, and biodiversity preservation.

The project concludes with the 6. Conclusions and Future Research Lines section, summarizing the key findings and their implications for future research. This section offers recommendations for further studies aimed at enhancing the accuracy and applicability of the developed metamodel, thereby contributing to improved coastal management and protection strategies.

2. Study area and data

This section offers a comprehensive overview of the study area, focusing on the south coast of Molokai for the calibration of our metamodel. We delve into a detailed description of the Molokai Reef tract, highlighting its unique characteristics. Additionally, we describe the field data, which involved the strategic deployment of 19 sensors along Molokai's southern coast to capture various hydrodynamic parameters. A detailed bathymetric profile of the study area is also presented, crucial for accurately modeling the hydrodynamic processes impacting the reef.

2.1. Study Area

The coral reefs of Maui Nui (Figure 1), encompassing the islands of Maui, Molokai, Lanai, and Kahoolawe in the Hawaiian Islands, are crucial to the local ecology, culture, and economy. Despite their importance, the health of these reefs has been steadily declining over the past few decades. Currently, they face numerous threats, including overfishing, land-based pollution, and climate change (Storlazzi, 2019). Our research is concentrated on the south coast of Molokai, selected due to the presence of extensive coral reefs and the availability of field data.

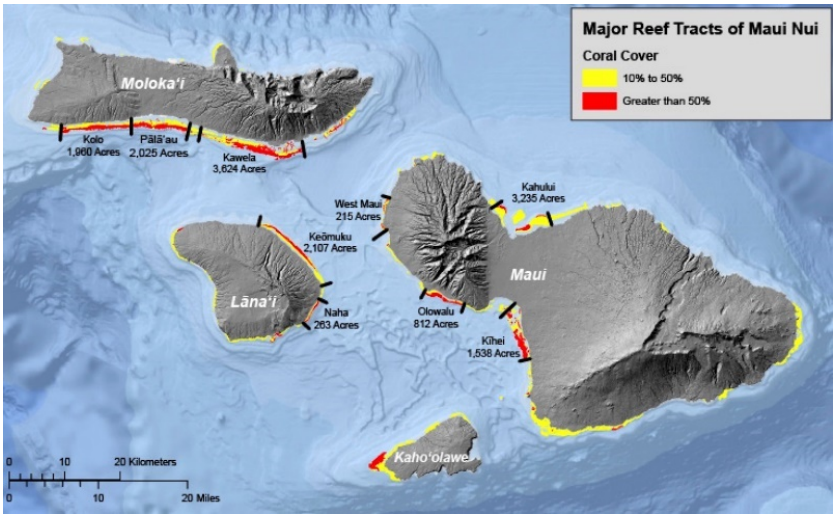


Figure 1. Map of nine major shallow reef tracts of Maui Nui, Hawaii, showing percentage of coral cover. (Source: Storlazzi, 2019).

2.1.1. Molokai Reef tract

The island of Molokai is located approximately at 21°N, 157°W in the north-central Pacific, situated between the islands of Oahu and Maui in the Hawaiian Archipelago (Figure 2). The island is 62 km long from east to west and averages 13 km wide from north to south. It is comprised of two basaltic shield volcanoes that formed roughly between 1.90 and 1.76 million years ago (Clague and Dalrymple, 1989), being the East Molokai volcano 1515 meters high, and the west one 420 meters high.

The south shore of Molokai boasts a 53 km long fringing coral reef, stretching across the Kalohi Channel between Molokai and Lanai, and the Pailolo Channel between Molokai and Maui. This reef holds significant ecological importance, serving as habitat for a diverse range of marine species while also acting as a natural barrier against erosion and wave impact. Its vital role in the marine environment makes it indispensable for both the local ecosystem and the cultural and economic activities of surrounding communities.

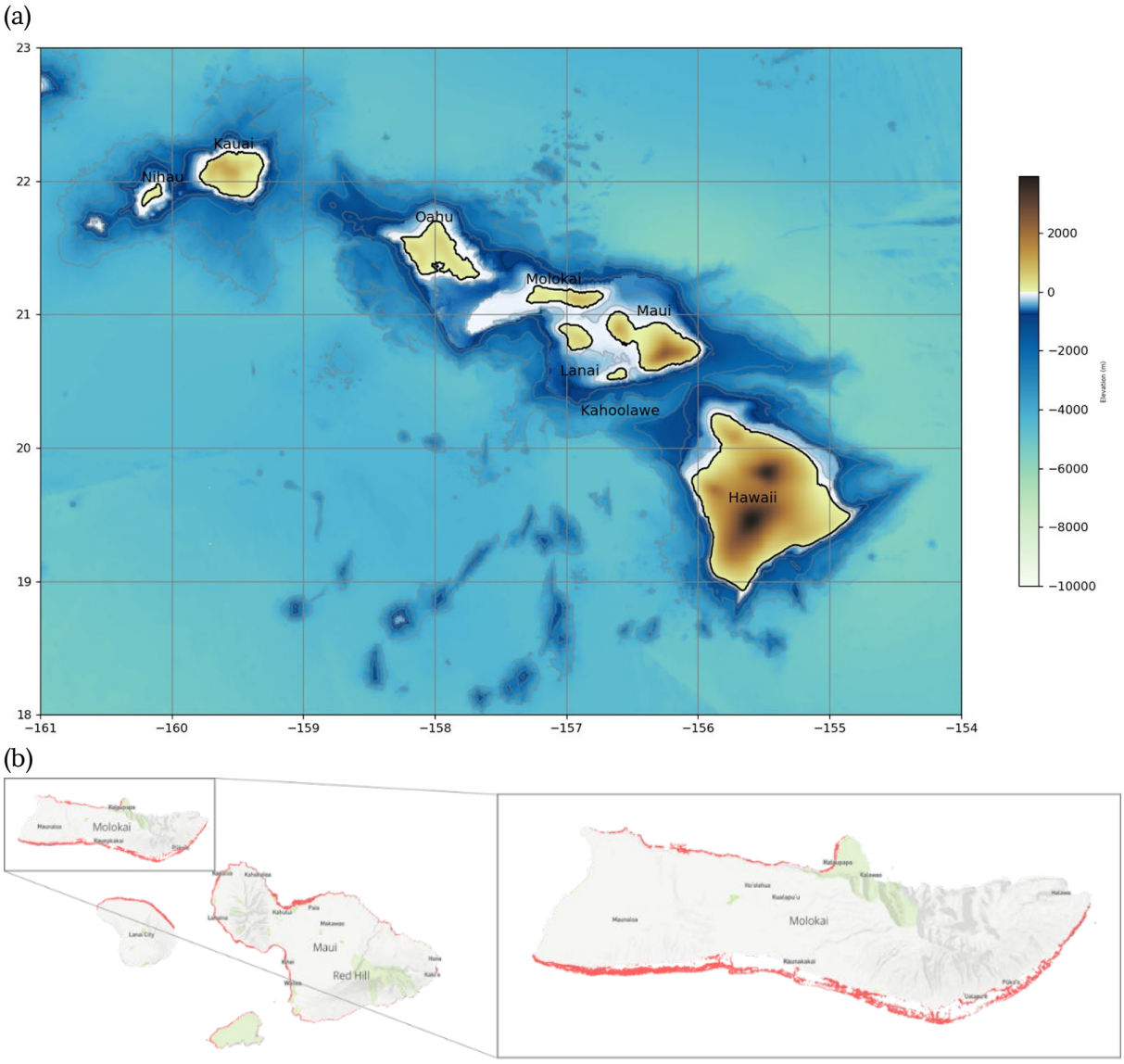


Figure 2. Map of the study area. (a) Map of the Hawaiian Islands showing the location of Molokai (Source : Created by the author) ; (b) Location of coral zones in the islands of Maui, Molokai, Lanai, and Kahoolawe (Source: Allen Coral Atlas).

The Molokai Reef is characterized by a complex and varied structure, featuring distinct morphological sections such as the reef flat, reef crest, and fore reef (Figure 3). Below, we explore the detailed morphology of these sections, highlighting the specific features and sediment compositions that define each part of the reef.

- **Forereef.** To begin with, at depths between 3 to 30 meters, we can find the forereef, which is the deepest part of the reef and faces the open ocean. The steep slope in this area helps dissipate some of the wave energy while also directing water flow toward the upper parts of the reef. The turbulence generated here initiates the process of energy dissipation. The steep incline of the forereef forces waves to disperse their energy vertically. This area usually has the highest coral cover (Storlazzi et al., 2004) and features 1- to 3-meter-high ridges and grooves running perpendicular to the shore.

- **Reef Crest**. Furthermore, the reef crest, the highest part of the reef, is where most waves start the process of breaking. This area is clearly defined along much of the southern Molokai reef. Here, wave energy is significantly dissipated due to the abrupt reduction in depth and the increased friction. Surface roughness is particularly high in this zone because of the dense presence of corals and other biological structures that act as physical barriers against waves. The reef crest is mostly covered with tough coralline algae and encrusting corals (Storlazzi et al., 2004).

- **The Reef Flat.** Finally, the reef flat is a mostly flat area under shallow water, between 0.3 and 2.0 meters deep. It extends from the shore out to about 1.5 km into the sea. The bottom of the reef is mostly made up of calcareous marine sediment, with large grains making up 58–65% of it (Calhoun and Field, 2002). The finer particles include both marine sediment and material from the erosion of the island's volcanic rocks. The inner part of the reef flat is covered by a layer of muddy sand, which is mostly made up of silt and fine grains from the land. This muddy sand extends about 200–300 meters from the shore. Beyond this area, up to about 500 meters from the shore, the ancient reef surface is exposed and usually covered with algae. Additionally, from 500 meters to about 1000 meters offshore, the reef flat has a pattern of ridges and troughs. The ridges have live coral, and the troughs are filled with calcareous sediment (Storlazzi et al., 2004). This flatter and shallower area allows the remaining wave energy to dissipate further through friction.

(a) (b)

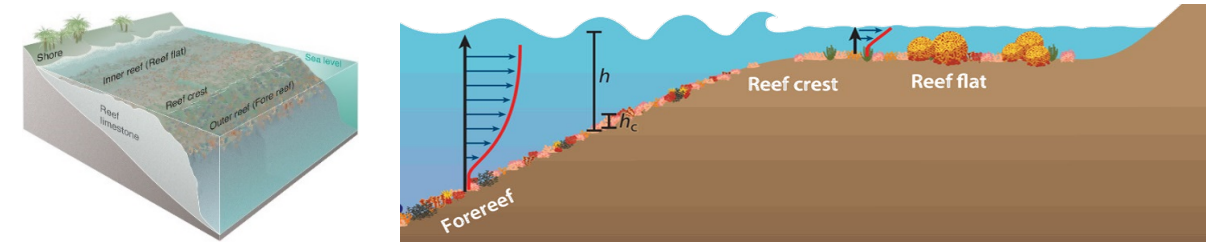


Figure 3. General coral reef morphology highlighting the key components. (a) Tridimensional morphology of a general reef morphology (Source: https://www.usgs.gov/media/images/coral-reef-block-diagram).
(b) Cross-shore profile showing the evolution of waves interacting with the reef.
(Source: https://www.annualreviews.org/content/journals/10.1146/annurev-marine-042120-071823)

2.1.2. Oceanography and Meteorology

The waves arriving at Molokai encompass: North Pacific swells, Northeast Trade wind waves, Southern Ocean swells, and Kona storm waves (Moberly and Chaimberlin, 1964). As analyzed by Moberly and Chaimberlin, the North Pacific swell doesn't affect the central part of the south shore reef much because the island is oriented east-west. Northeast Trade wind waves happen all year but are strongest from April to November when the Trade winds blow the hardest, usually between 10 and 20 meters per second. These waves are about 1–4 meters high but have short intervals of 5–8 seconds between them. Southern swells come from storms in the Southern Ocean during the Southern Hemisphere's winter. These waves are smaller, about 1–2 meters high, but have long intervals of 14–25 seconds between them. Finally, Kona storm waves come from the south during local storms or fronts. They are not very frequent or regular and are usually about 3–5 meters high with intervals of 8–12 seconds.

The island's rainfall is mainly influenced by the Northeast Trade winds and Kona storms. The Northeast Trade winds hit the northeast side of the island, get directed around the East Molokai volcano, and usually approach the reef near Kamiloloa from the southeast (Fletcher et al., 2002). Because of the high East Molokai volcano, most rain falls on the northeast part of the island, around 200–400 cm per year, while the south-central and west parts get much less rain, under 60 cm per year on average (State of Hawaii, 2001).

2.2. Field Data

As part of a coastal monitoring initiative (U.S. Geological Survey, 2020), aimed at advancing the understanding of coastal dynamics and supporting the preservation of coral reef ecosystems, 25 sensors were deployed along the southern coastal zone of Molokai Island, Hawaii. These sensors were strategically arranged in a linear formation perpendicular to the coastline, as illustrated in Figure 4, to collect crucial data on various oceanographic parameters over a period of 86 days. Positioned directly on the seafloor, the sensors were designed to measure the pressure exerted by the water column above them. This data is vital for understanding nearshore circulation and the variability of hydrodynamic properties over the coral reefs in this region.

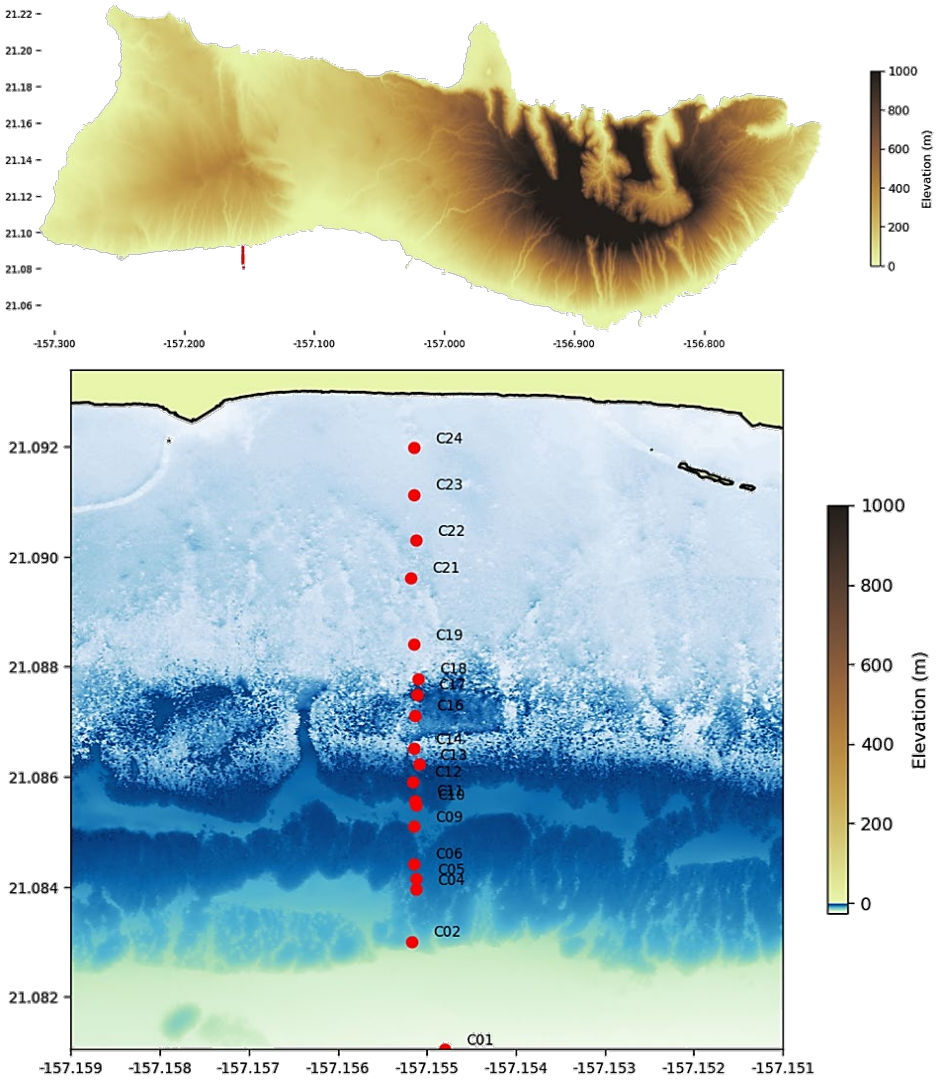


Figure 4. Sensor deployment area on the southern coast of Molokai Island. Horizontal and vertical axes with longitude and latitude coordinates (Source: Created by the author).

It's important to highlight that although 25 sensors were installed, only the data from 19 sensors were deemed valid for further analysis. The data from the remaining sensors were discarded due to incorrect pressure measurements or invalid readings caused by sediment accumulation, which hindered accurate water column pressure measurements.

The 19 valid sensors began recording measurements on June 24, 2018, at 21:00 hours. However, the total duration of data collection varied among the sensors. Some sensors recorded measurements for a full period of 2064 hours, until September 18, 2018, at 21:00 hours, while others recorded for shorter periods, with the shortest being 1125 hours.

The parameters inferred from the water column pressure include WL, tide, *Hs*, *Hrms*, dominant period, *Tp*, energy, as well as infragravity and low-frequency components, both represented also by their *Hs*, *Hrms*, *Tp*, and energy. Data for all these parameters were collected on an hourly basis.

Relevant information from the sensors is displayed in Table 1. It can be observed that the position is shown in coordinates of longitude and latitude, as well as the elevation, in meters, referenced to the local mean sea level. Additionally, the duration of measurements for each sensor is provided. It can be noted that there are certain sensors whose duration does not correspond to the total of 2064 hours.

Sensor	Longitude	Latitude	Elevation	Start Time	End Time	Duration
C01	-157.154799	21.081046	-24.786127	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C02	-157.155172	21.083003	-16.747430	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C04	-157.155119	21.083952	-8.576710	2018-06-24 21:00:00	2018-09-03 02:00:00	1685.0
C05	-157.155121	21.084145	-6.933797	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C06	-157.155143	21.084421	-5.184982	2018-06-24 21:00:00	2018-08-16 10:00:00	1261.0
C09	-157.155150	21.085098	-7.386092	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C10	-157.155121	21.085492	-4.578356	2018-06-24 21:00:00	2018-08-15 14:00:00	1241.0
C11	-157.155137	21.085569	-3.787843	2018-06-24 21:00:00	2018-08-10 18:00:00	1125.0
C12	-157.155162	21.085913	-2.816314	2018-06-24 21:00:00	2018-08-11 02:00:00	1133.0
C13	-157.155092	21.086227	-2.040946	2018-06-24 21:00:00	2018-08-11 10:00:00	1141.0
C14	-157.155144	21.086507	-1.439283	2018-06-24 21:00:00	2018-08-14 11:00:00	1214.0
C16	-157.155136	21.087112	-1.873183	2018-06-24 21:00:00	2018-08-20 08:00:00	1355.0
C17	-157.155108	21.087495	-2.340674	2018-06-24 21:00:00	2018-08-15 06:00:00	1233.0
C18	-157.155093	21.087781	-1.398484	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C19	-157.155150	21.088405	-0.960255	2018-06-24 21:00:00	2018-08-14 13:00:00	1216.0
C21	-157.155187	21.089616	-0.958671	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C22	-157.155119	21.090295	-1.222854	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C23	-157.155147	21.091120	-0.932724	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0
C24	-157.155149	21.091969	-0.875177	2018-06-24 21:00:00	2018-09-18 21:00:00	2064.0

Table 1. Sensor Data Overview. This table provides an overview of sensor data collected including sensor ID, position (lon, lat), elevation (meters), start time, end time, and duration (hours).

We can examine the *Hrms* measurements from all sensors throughout the entire duration to get an understanding of the wave height evolution at the study site. This data is depicted in Figure 5. As mentioned earlier, some sensors don't cover the full 2064-hour period. We observe that as we move towards the shore, *Hrms* decrease, starting from the values recorded by the offshore sensor, C01, and declining as we approach the coast, reaching their lowest point at sensor C24.

The *Hrms* recorded by the offshore sensor reaches magnitudes of up to 1.06 meters in the month of July, with the average offshore sensor *Hrms* around 0.34 meters in height. By contrast, the maximum *Hrms* recorded by the sensor closest to the coast is around 0.14 meters in height, with the *Hrms* average in this zone of 0.048 meters. This highlights the decrease in wave height as we move towards the coast.

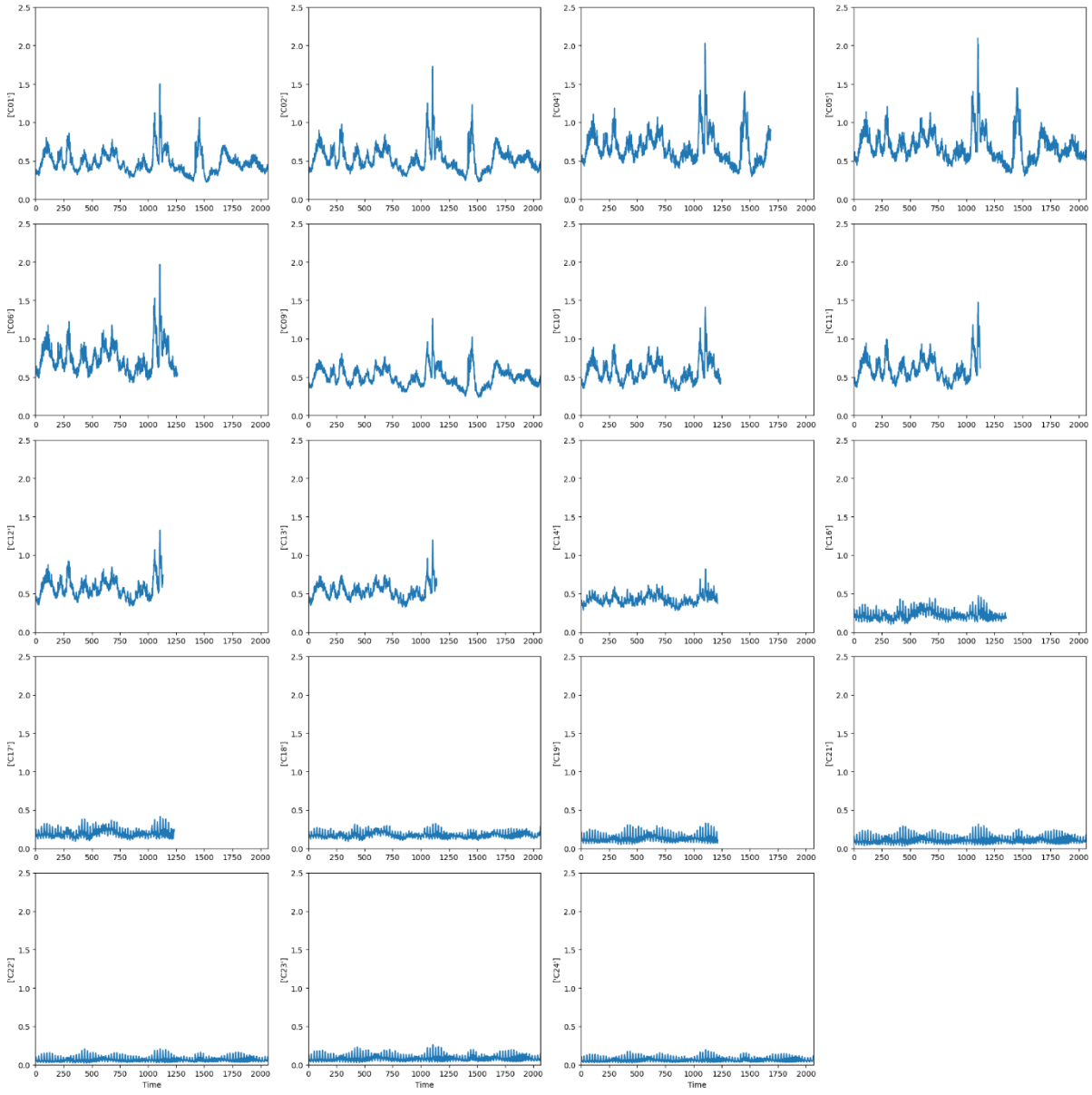


Figure 5. Temporal evolution of Hrms measured by each sensor deployed in our study area. The horizontal axis represents time, in hours, and the vertical one reflects Hrms in meters. (Source: Created by the author).

Furthermore, it's worth mentioning that sensors C04, C05 and C06 record the highest *Hrms* values, peaking at up to 1.48 meters. This increase in height is due to the fact that, as evidenced in the following section where the bathymetric profile and detailed sensor positions on the profile are shown; they are situated in the reef crest area. Therefore, as waves approach, the decrease in bathymetry causes the waves to feel the bottom, leading to wave steepening, initiating the wave-breaking process in this zone, and waves progressively decrease in height as they approach the coast.

2.3. Bathymetric profile

Following the presentation of the study area and the spatial distribution of the sensors from which we obtained real measurements for calibrating our metamodel, we proceed to obtain the bathymetric profile of our area. The bathymetric data corresponds to the 2013 USACE NCMP Topobathy Lidar DEM of the southern coast of Molokai, with a resolution of 1.5 meters (National Oceanic and Atmospheric Administration, 2013).

Using this dataset in QGIS, we plotted three transects, as shown in Figure 6. These transects provide crucial bathymetric data regarding the sensor group, with the exception of the first sensor, C01, which is slightly offset to the right of the transects. However, this discrepancy does not notably impact the results due to minimal horizontal depth variation at this point. Following the plotting of these three transects, an average profile was derived. This profile serves to define a comprehensive representation encompassing all sensors, achieved by averaging measurements taken along transects perpendicular to the coastline.

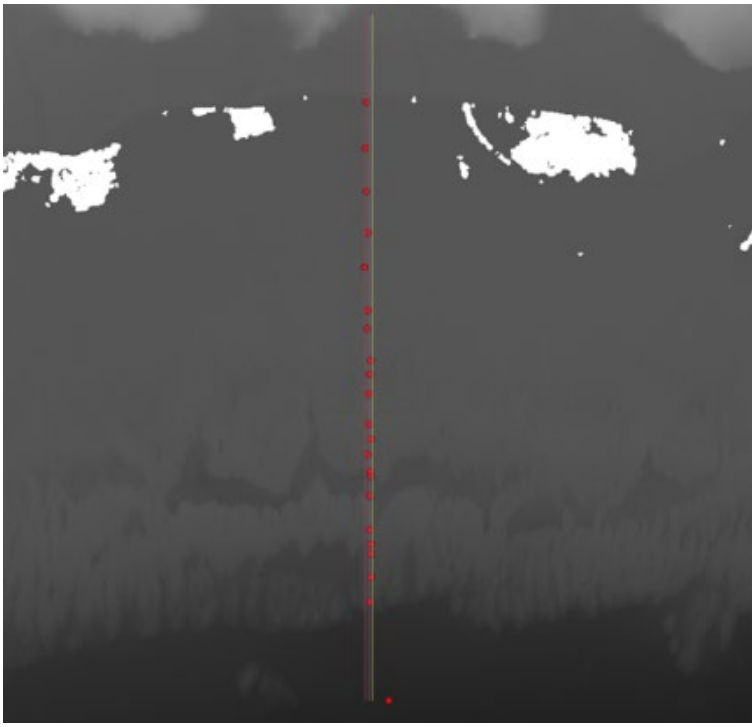


Figure 6. Bathymetric surface of the southern coast of Molokai, Hawaii, based on the 2013 USACE NCMP Topobathy Lidar DEM with a 1.5-meter resolution from NOAA. Three transects plotted in QGIS provide crucial data for the sensor group, excluding sensor C01, which is slightly offset. (Source: Created by the author).

Thus, following the aforementioned process, we obtain a bathymetric elevation profile in the study area, where we can locate the sensors, with elevation information at every 1.5-meter point. The one-dimensional bathymetric profile is presented in Figure 7, showing the position of each sensor.

Over the 2000 meters of the profile, several distinctive geomorphological characteristics explained earlier in Section 2.1.1 can be identified. Starting from the coast, in our study area, the reef flat extends for about 500 meters, featuring a relatively flat, horizontal rocky platform where the water depth varies around 0.7 meters. This is followed by a much rougher area filled with grooves extending for about 250 meters, leading to a significantly deeper and wider groove, and then to the reef crest.

After this, the fore reef extends for approximately 250 meters, transitioning into the sandy seabed where the furthest sensor from the coast, sensor C01, is located. This detailed profile provides a comprehensive understanding of the coastal environment under study, which is crucial for our research on local hydrodynamic processes.

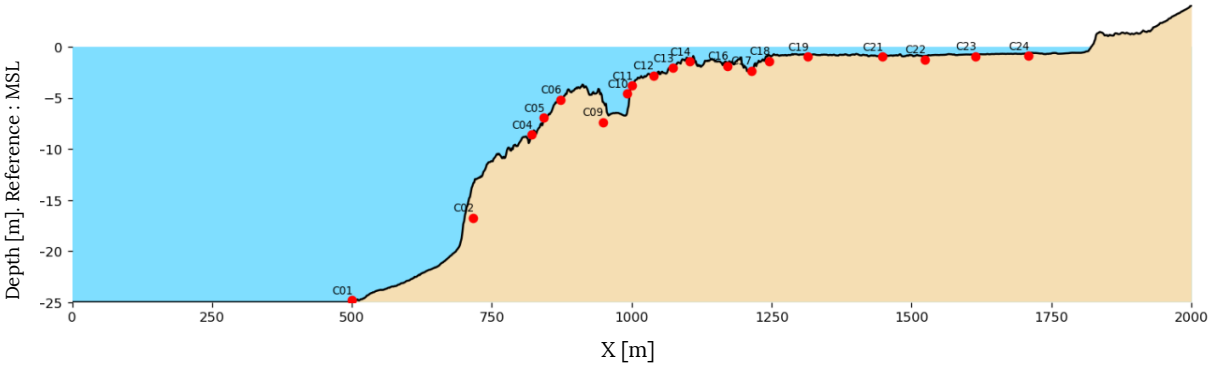


Figure 7. One-dimensional bathymetric elevation profile of the study area with sensor positions indicated. Elevation information is provided at every 1.5-meter interval. The reference is the local mean sea level. (Source: Created by the author).

When conducting our study, we must carefully consider various factors related to the bathymetric profile and sensor placement.

1.- Elongation of the Bathymetric Profile. Initially, the profile obtained from the average of the three transects was 1500 meters long. However, to position the numerical wave generator outside the area of interest, a flat region 500 meters in length at a depth of 25 meters was added before the fore reef, extending beyond the first sensor, C01. Consequently, the final profile used for our study and the numerical swash model extends over a total of 2000 meters, ranging from the coast to a total depth of 25 meters. It's worth mentioning, however, that this elongation will not affect the wave evolution in the original profile; it is merely for numerical purposes. This extension facilitates stabilization of waves by the time they reach the position corresponding to sensor C01.

2.- Sensor Positioning Discrepancies. It can be observed that sensors like C02 and C09 are slightly positioned below the actual bathymetric surface. This discrepancy arises from the fact that the represented profile is the average of three transects. As a result, there are slight differences, on the order of centimeters, between the bathymetry considered in these points and the actual bathymetry where the sensors were placed on the seafloor. This averaging process smooths out the bathymetric variations, which explains the minor offsets for these particular sensors. It should be highlighted that these differences are centimetric and will not significantly impact the overall results.

3.- One-dimensionality of the Profile. As we have seen, the profile where we will launch the numerical simulation is one-dimensional. Therefore, it would be logical to consider some wave refraction effect for the waves that approach the reef profile obliquely. This means knowing the direction from which the waves are coming to determine the H_s that we should launch perpendicular to the coast, following the direction of our one-dimensional profile. However, the sensors did not record information about the wave direction. This makes such a transformation impossible. Given this and considering that the wave heights are generally not very large except for some specific moments during the recording period, we make the hypothesis of not transforming them according to its wave direction. Instead, we directly use the H_s values measured by the offshore C01 sensor to establish the limits of the actual wave conditions.

3. CHySwash Methodology

As previously explained, the methodology to be used for characterizing the friction and breaking coefficients is based on the one-dimensional hybrid hydrodynamic metamodel named HySwash (Ricondo et al., 2024). The methods presented in this metamodel have been accordingly modified to incorporate two new parameters into the numerical model for subsequent calibration. Therefore, the new metamodel is called CHySwash, where "C" stands for calibration and "HySwash" stands for Hybrid Swash, reflecting the fusion of hybrid techniques with the SWASH numerical model to estimate surf-zone hydrodynamics.

A metamodel, often referred to as a "model of models," functions as a streamlined representation of intricate systems, providing a manageable and computationally efficient approach to understanding their behavior (Figure 8). At its core, the process begins with defining a problem using multiple parameters X = {x_1, x_2, x_3, ..., x_n} which encapsulate the various inputs influencing the system under study. These inputs could encompass a wide array of factors depending on the nature of the problem. The heart of the metamodeling process lies in a model, which can take the form of an analytical model based on mathematical equations or a numerical model relying on computational simulations. This model takes the defined parameters as input and generates corresponding outputs, representing the system's response, Y, to different combinations of input parameters.

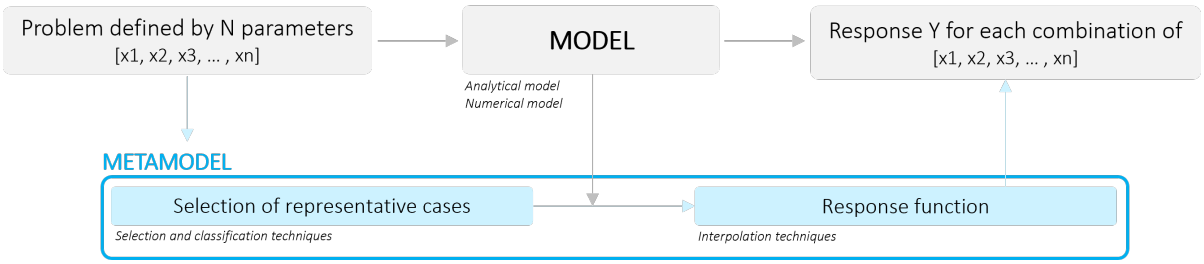


Figure 8. Schematic overview of metamodel functionality. (Source: Created by the author).

Generating responses involves running simulations or calculations for each unique combination of input parameters. This step can be resource-intensive and time-consuming, especially for complex systems with numerous variables. To streamline this process, a metamodel is constructed. This involves selecting a subset of representative cases instead of simulating every possible combination of parameters. These cases are chosen strategically to provide a comprehensive yet concise portrayal of the system's behavior. Various selection and classification techniques are employed to ensure the chosen cases accurately represent the system's dynamics. Once the representative cases are identified, a response function is constructed using interpolation methods. This function approximates the behavior of the original model, enabling predictions for combinations of input parameters that were not directly simulated by the numerical model.

With the metamodel in place, predicting the system's response for any new combination of input parameters becomes straightforward. The metamodel utilizes the response function to estimate the response Y efficiently, bypassing the need for extensive computations required by the detailed model. This significantly reduces computational time and resources, making it particularly valuable in scenarios where running full-scale simulations for every possible input combination is impractical due to time or resource constraints.

In this context, the proposed CHySwash methodology aims to automate the search for the optimal combination of calibration coefficients extending the capabilities of the HySwash metamodel. The process begins with case selection algorithms, utilizing the LHS technique to create a synthetic database of nearshore wave conditions. Subsequently, the MDA (Camus et al., 2011b) is employed to select a reduced number of representative cases. These selected conditions are dynamically downscaled using SWASH. The output variables then undergo PCA to reduce their dimensionality while preserving spatial structure. Finally, RBFs are used to create an interpolation function that models new nearshore conditions and predicts their hydrodynamics evolution along the cross-shore profile. Numerical validation is conducted through k-fold cross-validation.

Once we have the interpolation surface created from a synthetic case library, we can use it to obtain the response evolution of various variables along our profile for a non-modeled combination of input variables. Utilizing this feature, we will reconstruct the sea states measured by the offshore sensor C01 with the objective of reproducing and comparing the numerical evolution with the actual sensor measurements to optimize the combination of calibration coefficients. This process ensures precise calibration, systematically adjusting coefficients to accurately reflect observed conditions on the reef. This general methodology of the proposed metamodel is shown in Figure 9.

The metamodel initiates with synthetic generation and case selection in Section 3.2, focusing on the 5 variables chosen in Section 3.1. Section 3.3 follows, detailing the numerical model employed for simulating wave-reef interactions. Subsequently, Section 3.4 conducts comprehensive analysis and results reconstruction using techniques such as PCA and RBF interpolation. These methods reduce data dimensionality and establish an interpolation surface to accurately predict the evolution of output variables like *Hrms* and Hs_{IG} along the reef profile. Next, Section 3.5 employs K-fold cross-validation to evaluate the metamodel's generalization and reliability across diverse input scenarios. Finally, Section 3.6 focuses on model calibration using offshore sensor data to optimize coefficients such as *Cf* and *Cr*. This step aims to improve the precision of predicting H_{rms} and Hs_{IG} along the reef profile.

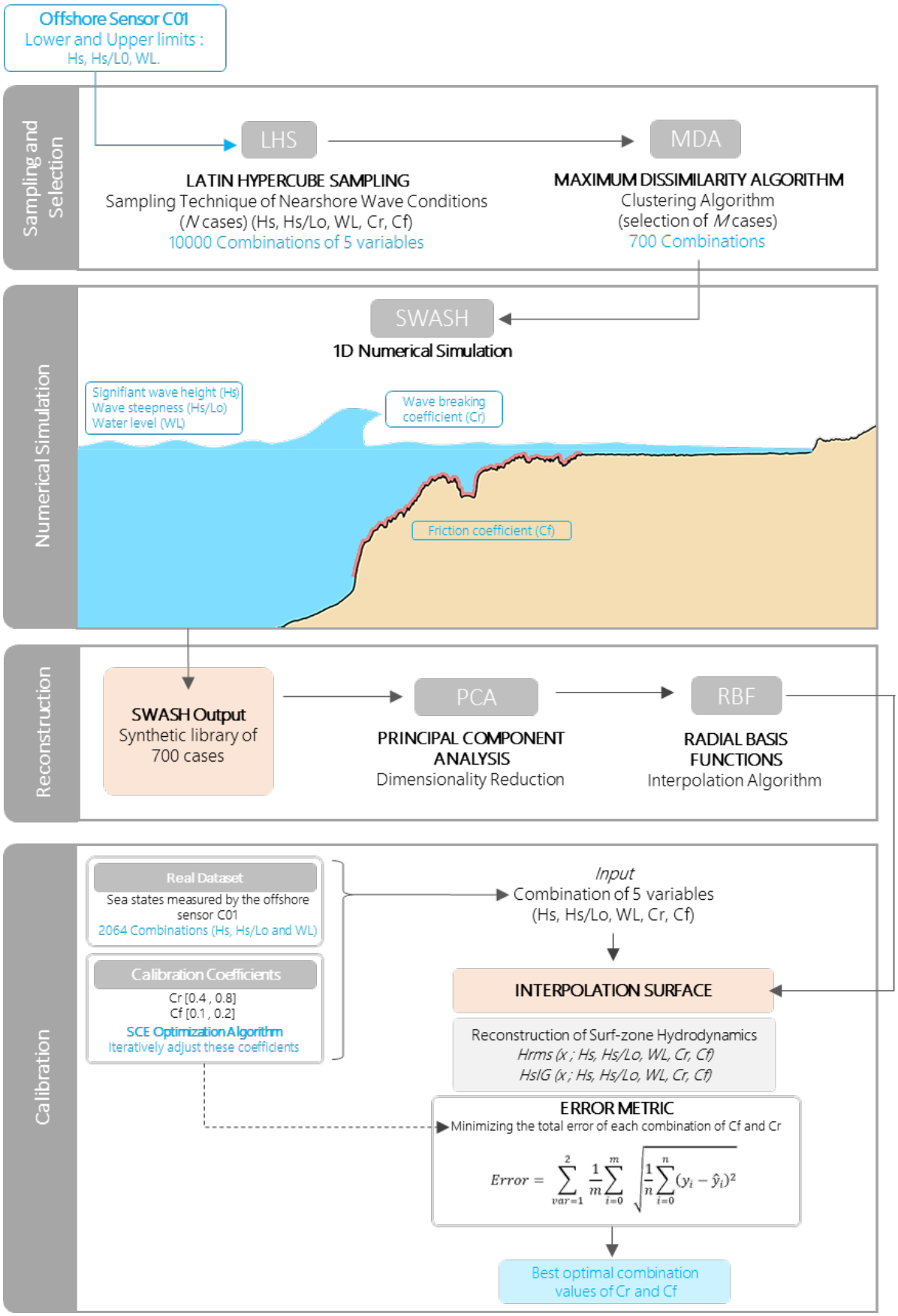


Figure 9. Flow chart of the proposed CHySwash metamodel. (Source: Created by the author).

3.1. Parameterization

It is essential to carefully select the parameters that will feed into the presented metamodel. The numerical model SWASH is designed to solve the Navier-Stokes equations numerically, relying on key hydrodynamic boundary conditions. Drawing from previous research on hydrodynamic modeling, it has been proven that main boundary conditions to solve wave processes encompass *Hs*, *Hs/L0*, and offshore *WL*. These variables are essential for the initial setup of the model.

However, when it comes to simulating the complex process of wave breaking in the nearshore zone, SWASH requires additional inputs to accurately reflect real-world conditions. The model alone cannot determine these parameters, necessitating user-defined inputs. Therefore, we also include the two critical coefficients responsible for governing wave breaking and energy dissipation: the *Cr* and the *Cf*. Proper calibration of these coefficients is crucial for the model to realistically simulate wave behavior and energy dissipation in coral reef environments.

3.1.1. Hydrodynamic parameters

Waves and tidal oscillations are the most relevant predictive offshore hydrodynamic variables for flooding (Pearson et al., 2017). These variables include short-term (i.e., local wind waves, distant-source swells) and long-term (i.e., tides, surges, mean sea level, increasing sea level) processes. The metamodel parameters that define the offshore wave and water level conditions are the *Hs*, *Hs/L0*, and *WL*. In this manner, to commence the algorithm execution, it is pivotal to acquire representative parameter combinations of the hydrodynamic model.

- **Significant Wave Height.** This is defined as the average height of the highest one-third of waves during a specific time period. It provides a measure of the energy and intensity of the wave conditions.
- **Water Level.** This refers to the elevation of the water offshore relative to an established reference point, in this case, the local mean sea level.
- **Wave steepness**. The wave slope is calculated by dividing the significant wave height (*Hs*) by the dominant wavelength (*Lo*). This parameter is related to the wave period.

The reason for choosing wave steepness instead of directly using the wave period is to produce physically plausible combinations of wave heights and periods. By sampling the wave steepness, we avoid generating unrealistic scenarios such as very high waves with short periods or very low waves with long periods. This approach ensures controlled and realistic combinations of wave height and period. The relationship between wave steepness and the wave period is demonstrated through the following formula:

$$T = \sqrt{\frac{2\pi \cdot Hs}{g \cdot (Hs/Lo)}} \quad \text{(Eq. 1)}$$

T : wave period
Hs : significant wave height
g : gravitational constant
Hs/Lo : wave steepness

3.1.2. Calibration parameter 1: Wave Breaking Coefficient

Once the hydrodynamic variables are defined, it is necessary to establish the wave breaking coefficient for the SWASH model. This coefficient helps determine when wave breaking starts. In this numerical model, it is described as follows: If the rate of change of the free surface elevation with respect to time exceeds a certain value related to this coefficient, the waves are considered to be breaking (Delft University of Technology, n.d.). The mathematical expression for this is:

$$\frac{\partial \zeta}{\partial t} > Cr \cdot gh \quad \text{(Eq. 2)}$$

$\partial\zeta/\partial t$: rate of change of the free surface elevation.
Cr : the wave breaking coefficient.
g : the acceleration due to gravity.
h : the water depth.

By default, SWASH uses a value of 0.6 for this coefficient (Delft University of Technology, n.d.), which is generally considered valid for most simulations. However, in our study, we will treat it as a calibration parameter, varying it from 0.4 to 0.8. A higher Cr value means that waves require a steeper slope on the water surface to break. Conversely, a lower Cr value indicates that waves can break even with a gentler slope.

3.1.3. Calibration parameter 2: Friction coefficient

The Cf describes the resistance to water flow due to the seabed. The rougher the seabed surface, the higher the friction, which consequently reduces the flow velocity. This coefficient is vital for predicting water flow behavior, velocity distribution, and water depth under various conditions. Playing a significant role in energy dissipation, this coefficient affects the wave breaking process. Higher friction results in greater energy loss as waves travel, leading to changes in wave height and potentially altering the location and intensity of wave breaking.

In SWASH, the default value for this coefficient is 0.002. However, we will modify this default value for the area within our study profile that is covered by coral structures (Section 3.3.1).

3.2. Sampling and selection

The first step in implementing our metamodel involves generating and selecting specific representative cases that combine the variables previously identified. These combinations will accurately simulate realistic sea states in SWASH. Therefore, our task is to create realistic combinations of these variables to feed into the numerical model later on.

3.2.1. Sampling technique: Latin Hypercube Sampling (LHS)

Firstly, to generate all these potential realistic cases from which we will subsequently feed Swash, the initial algorithm we are going to employ is LHS. This algorithm is a statistical technique designed to efficiently select parameter values across multiple dimensions while ensuring controlled randomness in the sampled data. To implement this technique, we start by establishing the limits for each variable. Once each variable has its limits defined, the Latin Hypercube Sampling method is applied to generate a set of synthetic cases. This sampling method is a type of stratified Monte Carlo technique, which demands less computational effort and time consumption than Monte Carlo simulations.

LHS operates in a manner that produces a sample size N from the n variables by dividing their correspondent ranges into N non-overlapping, equally probable intervals. From each interval, a value is picked attending to the probability density that defines it. This procedure is carried out N times for each variable, thus having n sets (one per variable) of N values (one per interval). These N values of a variable are paired with the N values of another variable randomly, and so forth, until reaching the n^{th} variable. In this way, a set of $n \cdot N$ tuples are constructed.

Below in Figure 10 is a two-dimensional example of how LHS works. The grid shows the number of regions for each variable, in this case, we have x and y. A random selection is made for the first case, represented by the green point. Any combination involving the same values of x or y is then prevented from being selected again. Subsequently, another case is selected, represented by the red point, and the corresponding values of x and y are blocked again from being chosen in future iterations. This process continues consecutively until the entire space is covered. Therefore, the sample size will match the number of partitions of the variables.

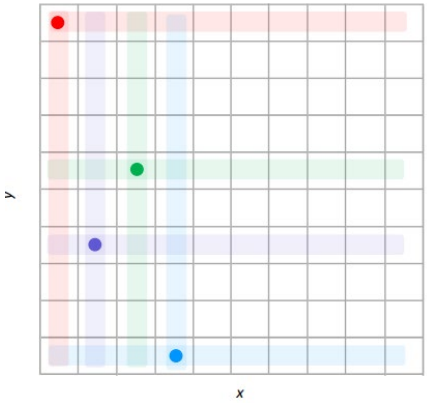


Figure 10. Two-dimensional example of the Latin Hypercube Sampling. Axes of the graph without units. (Source: Méndez, 2024).

As mentioned, it is necessary to establish the limits of the different variables to begin generating the cases. In the case of hydrodynamic variables, aiming to set realistic maximum and minimum limits, data from sensor C01 is utilized, as it is located offshore, thus providing measurements closer to offshore reality. Accessing the recorded data from this sensor, including Hs, Tp, and WL, allows us to identify the maximum and minimum limits registered during the recording period. Subsequently, the minimum and maximum values of Tp are transformed into wave steepness with their corresponding Hs values to generate physically realistic cases, as explained earlier.

On the other hand, to establish the limits of Cr, referring to the Swash manual, it sets the default Cr value at 0.6 (Delft University of Technology, n.d.). However, to broaden the range of study for this coefficient, we will widen the range from 0.4 to 0.8. Finally, concerning the Manning friction coefficient, after consulting various sources, an interval is established ranging from 0.01 to 0.25 (Karim and Nandasena, 2023 ; Liu et al., 2023). The maximum and minimum limits for each of the variables are shown in Table 2.

Variable	Boundary Values
Hs	[0.15, 1.60] m
Hs/Lo	[0.0005, 0.009] m/m
WL	[-0.6, 0.356] m
Cf	[0.025, 0.20]
Cr	[0.40, 0.80]

Table 2. Upper and lower boundary values used for each variable.

Once the limits for each variable are defined, the LHS is applied to generate a set of synthetic cases. Initial samples generated for the study area consist of N = 10000 hourly sea states.

3.2.2. Selection technique: Maximum Dissimilarity Algorithm (MDA)

The high computational cost of propagating the entire dataset of N = 10000 cases produced by LHS necessitates the use of statistical tools to reduce the data set to a manageable number of representative cases for hybrid downscaling. For this purpose, the MDA is implemented.

Given a data sample X = $\{x_1, x_2, \dots, x_N\}$ consisting of N n-dimensional vectors, a subset of M vectors $\{v_1, v_2, \dots, v_M\}$ 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 remaining M-1 elements are selected iteratively by calculating the dissimilarity between each data point left in the database and the elements already in the subset. The most dissimilar data point is then transferred to the subset. Dissimilarities between points are computed using Euclidean distance (Eq. 3). The process finishes when the algorithm reaches M iterations.

$$D_i = \sum_{j=1}^{N-1} \|x_i - x_j\|; i = 1, \dots, N \rightarrow \max\{D_i ; i = 1, \dots, N\} \quad \text{(Eq. 3)}$$

Additionally, the algorithm prioritizes exploring the edges of this space, where combinations of extreme wave parameters are located. By focusing on the edges of the multidimensional space, MDA effectively represents the parameter population on the boundary, which ensures that the selected subset is suitable for accurate interpolation, minimizing the need for extrapolation, which tends to be less reliable.

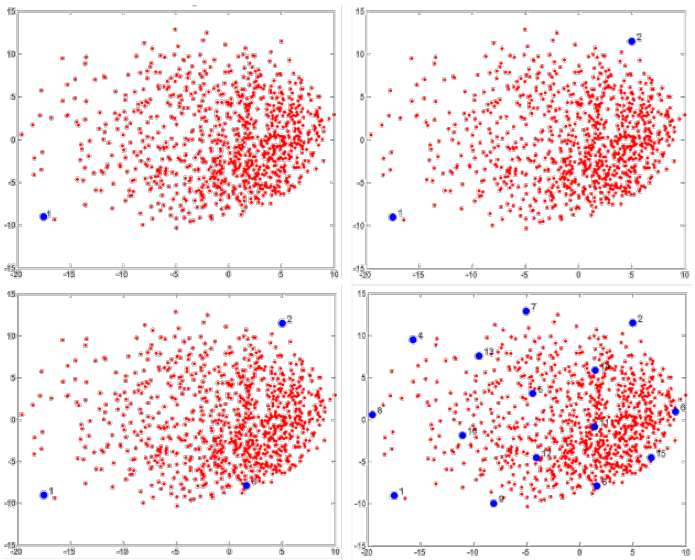


Figure 11. Example of a two-dimensional MDA algorithm showing the selection of points (blue) inside the big population defined by two variables (red). (Source: Méndez, 2024).

Figure 11 illustrates an example of the MDA algorithm in action. In this scenario, from a total population of N red points, we aim to select M representative points, which in this case is 16. The selection process begins with an initial subset M containing just one vector, v_1. From this starting point, the algorithm calculates the most dissimilar point in the population based on Euclidean distances. This new vector, point 2, is added to the subset, which now consists of v_1 and v_2.

The algorithm then identifies the most dissimilar point relative to the two points in the subset, resulting in the selection of point 3, which is subsequently added to the subset. This process is repeated so that we obtain the desired M number of points.

In our case, given the data sample generated by LHS consisting of N (= 10000) n-dimensional (= 5) vectors, MDA is used to obtain a subset of M = 700 vectors that represent the diversity of the data, as we can see in Figure 12.

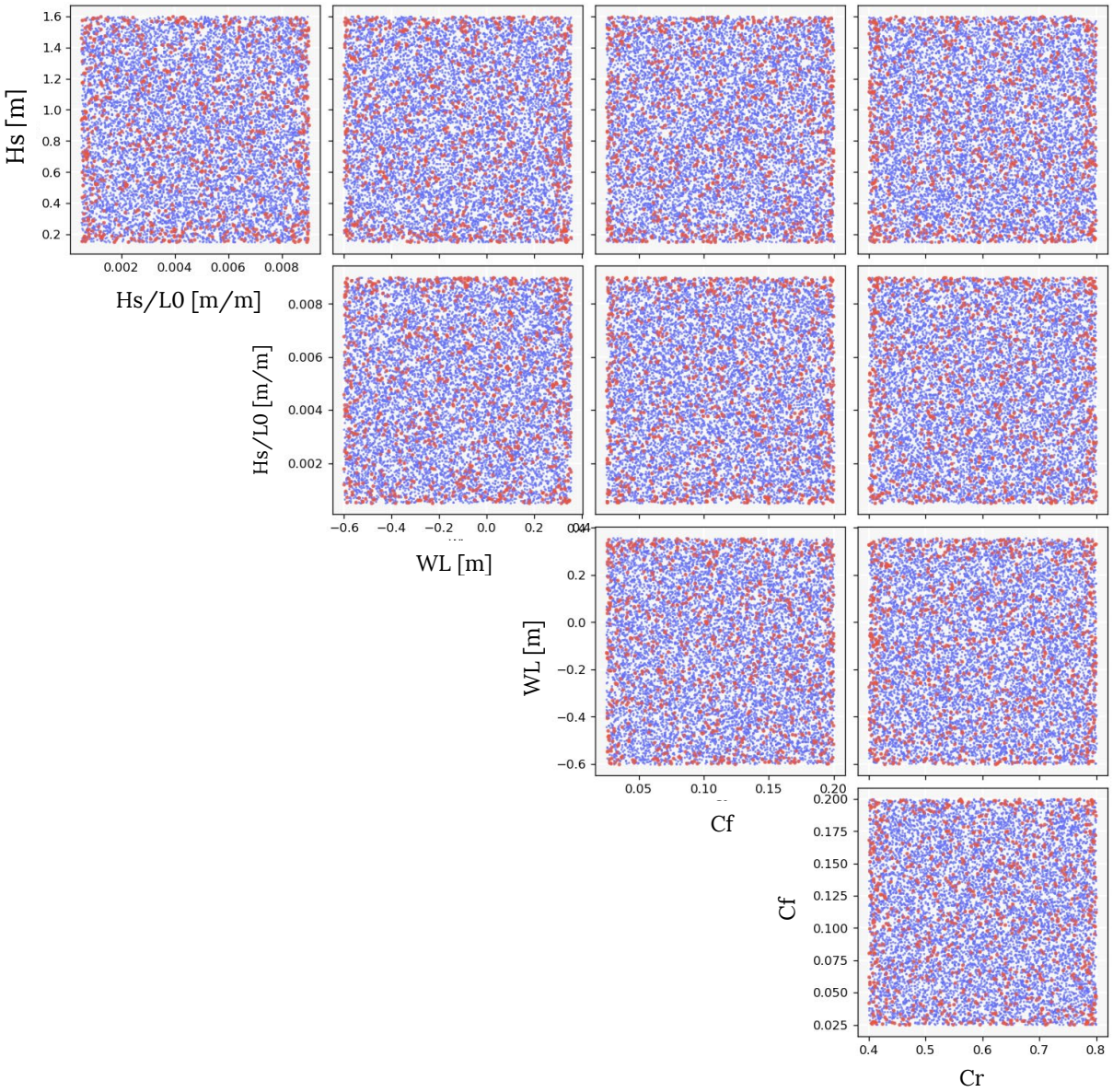


Figure 12. Multidimensional distribution of the LHS synthetically generated dataset (Blue) and the MDA-selected cases (Red). In this case for a total of 5 variables: Hs, Hs/Lo, WL, Cr and Cf. (Source: Created by the author).

3.3. Numerical Model SWASH

Once the hydrodynamic cases to be numerically simulated have been selected through the MDA algorithm, we proceed to launch these cases in SWASH. It is a versatile numerical tool designed to simulate a non-hydrostatic, phase-resolving wave model capable of simulating waves from deep waters to the shoreline, modeling wave breaking, bottom friction, wave- induced setup and runup, and the generation and propagation of infragravity waves (Delft University of Technology, n. d.).

The accuracy of the model is comparable to that of lower order Boussinesq models and has been widely used for modeling wave dynamics and inundation on sandy beaches and coral reefs (Liu et al., 2021). Leveraging this powerful tool, the wave transformation along the coral reef was simulated using one-dimensional SWASH simulations.

It's important to note that while the discussion regarding SWASH and its specific applications is presented in a general context, the main focus of this work does not delve into specific details of SWASH. For a deeper and more detailed understanding, it is recommended to consult relevant literature, including the SWASH user manual. This explanation remains concise to avoid unnecessary elaboration, as the primary focus is directed towards other aspects of the metamodel.

3.3.1. Preliminary Setup and Configuration

The following section presents the key and relevant information that must be considered and introduced before launching SWASH. This information is crucial for our model.

1.- Cross-shore profile. The propagation of sea states will occur from the offshore position to the coastal zone, corresponding with the bathymetry presented in Section 2.3, Figure 7.

2.- Bottom Friction. As we know by this stage of the project, the primary goal is to characterize the surface of the reef covered by coral and determine its *Cf*. SWASH allows users to activate bottom friction, and in the default option, the Manning's coefficient is considered. Our next step is to assign the friction coefficient to the specific zones within the computational grid of SWASH.

The friction coefficient values selected by the MDA, representing the coral friction coefficient for each case, need to be applied accurately. To do this, we must first identify the areas of the reef profile covered by coral structures. Using the Allen Coral Atlas (Allen Coral Atlas, n.d.), we can access the benthic map of the study area, which provides detailed information on the seabed's distribution and characteristics. This data enables us to establish the start and end points of the coral structures within our cross-shore profile. The Figure 13 illustrates the designated coral region, marked in pink, where the selected *Cf* values of the 800 selected cases will be applied to the numerical SWASH grid.

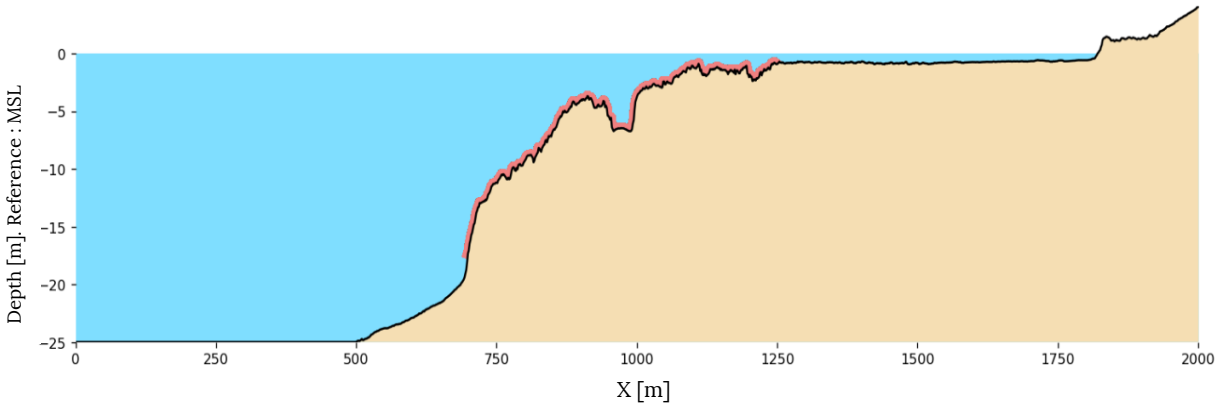


Figure 13. Coral region (shown in pink) delineated in the cross-shore profile used for the numerical simulations in SWASH. Reference is the local mean sea level. (Source: Created by the author).

The Allen Coral Atlas benthic map also offers information about the seabed for the remainder of the profile, identifying sandy areas in the offshore zone and rocky areas in the reef flat zone. However, our focus is not on characterizing the friction coefficient for these regions. Therefore, we will apply the *Cf* only to the coral-covered region and use the SWASH default bottom friction value of 0.002 (Delft University of Technology, n.d.) for the rest of the grid profile.

3.- Boundary Conditions. The boundaries of the computational grid in SWASH are either land, beach or water. In our case, the wave condition is imposed on the west boundary of the computational domain, so that the wave propagation is pointing eastward. To simulate entering waves without some reflections at the wavemaker boundary, a weakly-reflective boundary condition allowing outgoing waves is adopted. Here, the boundary conditions are defined as irregular unidirectional waves by means of 1D spectrum. Both the initial water level and velocity components are set to zero.

4.- Sea States. Once the representative *M* cases defining our model are selected, we proceed to calculate the different sea states of the waves at the start of the profile using a single-peaked unidirectional TMA spectrum. The TMA spectrum (Bouws et al., 1985), is a generalized version of the JONSWAP spectrum for offshore waves and is valid for waters of any depth. It was designed to modify the spectral shape as water depth decreases, limiting the energy of low-frequency waves in shallower waters.

To create the TMA spectrum, we multiply the JONSWAP spectrum by a function that depends on wave frequency and water depth. For our model, considering an initial depth of 25 meters, we transform the JONSWAP spectrum into a TMA spectrum and then rescale it to maintain constant energy. This process provides a wave spectrum that is more suitable for simulating the specific conditions of our hydrodynamic model in the study area.

5.- Wind and turbulence. The influence of wind action and turbulence are not considered.

6.- Simulation Configuration. In terms of certain parameters that indicate the resolution of the simulation, we considered several aspects:

- **Simulation Time.** The simulation time is set to 2 hours to ensure that the generation of infragravity energy is well modeled.
- **Horizontal Resolution.** The number of nodes per wavelength dictates the horizontal resolution of the numerical model's mesh. For instance, if we consider a value of 60 nodes

per wavelength, and the modeled wavelength is 200 meters, the mesh resolution is 3.3 meters. The resolution therefore varies for every simulation with the case wavelength.

- **Vertical Resolution.** Another important parameter is the number of vertical numerical layers into which the water column is divided, over which the equations will be solved. These layers are in sigma coordinates, so along the length of our profile, the number of layers remains constant, adapting to changes in bathymetry while maintaining a constant layer thickness.

Based on these three aspects, a sensitivity analysis was conducted to see how they affect the numerical simulation using SWASH. Below, Table 3 summarizes the parameters used for each simulation, along with graphs showing the output of the numerical model. In this case, the simulation corresponds to the conditions measured by the sensors at hour 450. The selection of the hour was completely random.

Case	Simulation time (sec)	Horizontal Mesh Resolution: number of nodes per wavelength	Number of vertical layers in the numerical model	Case resolution time (sec)	Graph Color
1	7200	30	2	182.36	Blue
2	7200	60	2	520.37	Yellow
3	7200	30	3	280.08	Green
4	7200	60	3	794.78	Red
5	7200	30	5	441.92	Purple
6	7200	60	5	1374.43	Brown

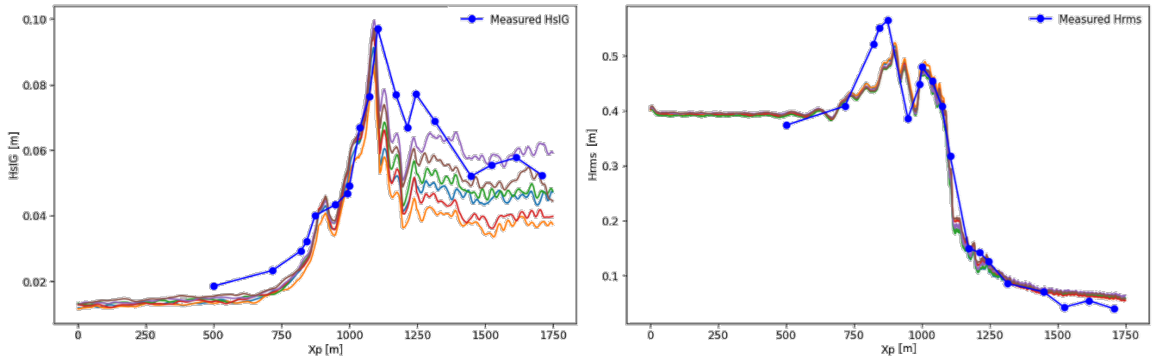


Table 3. Summary of simulation parameters and their corresponding outputs. Each case varies by simulation time, horizontal mesh resolution (nodes per wavelength), and the number of vertical layers in the numerical model. Graph colors represent different parameter combinations. (Source: Created by the author).

Based on the graph shown, we can observe that, taking into consideration that the vertical scales are not of the same order, the variations in *Hrms* are not significantly different from each other. This magnitude is expressed in centimeters (ranging from 5 to 50 centimeters), and the differences between simulations are in millimeters. Therefore, *Hrms* does not provide much clarity on the ideal numerical parameters to use. If we now focus on the SWASH output for Hs_{IG}, we can see that, with an order of magnitude in millimeters (ranging from 1 mm to 10 mm), there are more noticeable differences between cases, particularly in the area closer to the coast, after the wave breaks. Case 6 (brown) best matches the sensor measurements.

As a result, the selected combination of parameters includes a simulation time of 2 hours, a horizontal resolution of 60 nodes per wavelength, and a vertical resolution of 5 layers. We consider it not worthwhile to increase the number of vertical layers to 10, for example, because it would significantly increase the simulation time without a corresponding significant improvement in the results.

3.3.2. Analysis of SWASH Output Variables for Model Calibration

The primary output variable generated by SWASH include the time series of water level along the profile. This output serves as crucial indicator of the hydrodynamic processes impacting the reef. From this SWASH output, we compute various other key variables, particularly the three variables chosen for model calibration in our study: *Hrms*, *HsIG*, and the *Setup* (initially chosen for calibration but subsequently discarded due to issues that will be discussed later). These variables are essential for capturing both low-frequency and high-frequency wave dynamics across the spectrum.

1.- The root-mean-square wave height (*Hrms*). It provides an average measure of wave heights, essential for understanding the overall wave energy. It is calculated spectrally using the formula:

$$Hrms = \sqrt{8\, mo} = \sqrt{8\, \sigma^2} \qquad \text{(Eq. 4)}$$

mo : zero moment of the wave spectrum.
σ^2 : the standard deviation of the water surface.

2.- The significant infra-gravity wave height (*HsIG*). It is crucial to understand the impact of low-frequency waves on flooding along reef-lined coasts, influenced by resonance modes (Gawehn et al., 2016). To calibrate *HsIG*, we include additional frequency band components extracted from the power spectra of the water level signal. In this case, we use a Python library to apply the Welch function to estimate the power density of the water level time series, specifically targeting the frequency band of 0.004 – 0.04 Hz.

3.- The mean wave setup ($\bar{\eta}$). It is calculated by averaging the water level and subtracting the offshore water level. Initially intended for model calibration, it has posed challenges due to inconsistent bathymetric data alignment with sensor depth readings, preventing accurate computation of the residual (water level minus tide). While the model yields reasonable wave setup results (Stockdon et al., 2006), significant differences in sensor depths and the uncertainties in our bathymetric network prevent us from using this variable for calibration, as these differences can induce errors greater than the setup value itself for the analyzed sea states. Moreover, considering that the sensors themselves have a measurement uncertainty of 1 cm (Nortek, 2024), we will present the setup output later, acknowledging the model's output as good and acceptable but not suitable for calibration due to the aforementioned issues.

3.4. Reconstruction : Principal Component Analysis (PCA) and Radial Basis Function (RBF)

Based on the M-selected inputs and their M-transformed SWASH-1D outputs, we can approximate the spatial relationship between input and output using an interpolation surface. This surface allows for estimating output variables for input cases combinations that have not been modeled. We employ RBF interpolation, which is particularly effective for high dimensional and irregularly distributed data (Alfeld, 1989). For a new vector of input parameters V, the RBF interpolation returns a univariate metamodel output y(V).

This applies to any SWASH output evaluated at a specific location in the profile. However, to reconstruct the evolution of spatial variables, such as *Hrms* or Hs_{IG} in our case, we would need to fit as many RBFs as there are target locations in the profile. This can significantly increase computational time and may lead to inconsistencies between adjacent points.

Therefore, we use PCA before RBF interpolation to reduce the dataset's dimensionality while preserving spatial patterns. PCA projects the original data onto a new space, capturing the maximum variance in the sample data. This new space consists of eigenvectors (empirical orthogonal functions, EOFs) and the transformed components of the original data over the EOFs, known as Principal Components (PCs). The EOFs capture the major oscillation patterns, while the PCs represent their variability across the M cases. By selecting the number of EOFs that define a desired percentage of explained variance, we achieve dimensionality reduction.

For our project, we decided to retain the principal modes that explain 99.8% of the variance. After applying PCA to the SWASH output quantity $y1(x; Hrms)$ and $y2(x; Hs_{IG})$ of dimensions P x M, where P is the number of spatial locations (x) along the studied reef profile, the originally modeled output for each case can be obtained as a linear combination of EOFs and PCs following:

$$y(x; Hrms) = \bar{y}(x) + \sum_{n=1}^{M} EOF_n(x) \cdot PC_n(Hrms) \qquad \text{(Eq. 5)}$$

where $\bar{y}(x)$ is the mean of the output variable, in this case $Hrms$, along the M cases. As the interpolated metamodel output must be univariate, we use RBF interpolation to reconstruct the reduced set of PCs, one by one.

The general RBF interpolation surface is of the form:

$$RBF(V) = p(V) + \sum_{i=1}^{M} a_i \, \Phi \left(\|V - V_i\| \right) \qquad \text{(Eq. 6)}$$

where p(V) is a linear polynomial with coefficients $b = \{b_0, b_1, \dots, b_m\}$, m is the dimension of the input parameters, $\| \bullet \|$ is the Euclidean norm, and ϕ is a Gaussian function with a user-specified parameter (ε) that modifies the shape of the distributions. This is a commonly used function that decreases exponentially with distance, expressed by the following formula:

$$\phi(r) = e^{-(\epsilon \cdot r^2)} \qquad \text{(Eq. 7)}$$

As mentioned, ε controls the rate of decay of the influence of each center. As shown in the Figure 14, we observe how the shape of the function changes based on the value of ε applied to the Gaussian function.

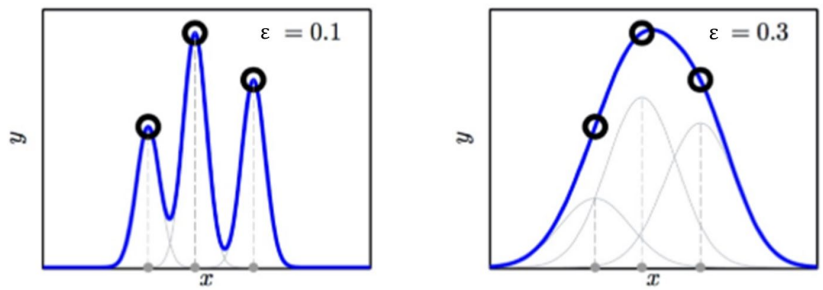


Figure 14. Effect of the value epsilon in the shape of the interpolation surface. (Source: Méndez, 2024).

Here, we adopt the algorithm proposed by Rippa (1999) to obtain the optimal ε parameter. The coefficients a_i and a_i need to be found for C principal components that explain the defined variance ($Pc_1, PC_2, \dots , Pc_c$). Therefore, the estimated spatial field of the response function is:

$$y(x; Hrms) \approx \bar{y}(x) + EOF_1(x) \bullet PC_1(Hrms) + \cdots + EOF_c(x) \bullet PC_c(Hrms) \qquad \text{(Eq. 8)}$$

Once the coefficients are obtained, multidimensional RBF interpolation can be used to replace the SWASH model, enabling the estimation of the principal components of various output quantities across the domain for unmodeled forcing conditions.

What we will do next in Section 3.6 is use this interpolation surface subsequently to reconstruct all the sea states recorded by the sensors, in order to carry out the corresponding calibration. The computational time saved is enormous, as we can instantly reconstruct a specific combination of the 5 variables (*Hs*, *Hs/Lo*, *WL*, *Cf*, and *Cr*) and obtain the spatial evolution of the selected output variables for calibration: *Hrms* and Hs_{IG}.

3.5. Numerical validation : K-fold cross validation

K-fold cross-validation is a widely used technique to assess the performance of a machine learning model. This method is particularly useful when we have a limited amount of data, as it provides insights into how well the model will perform on unseen data.

In k-fold cross-validation, the dataset is divided into '*k*' equally sized folds or subsets. For instance, if we have 100 data points and set *k*=5, the data is split into 5 folds, each containing 20 data points. The process involves '*k*' iterations of training and testing. In each iteration, '*k*-1' folds are used to train the model, and the remaining fold is used to test it. For example, in the first iteration, the model is trained on folds 1 to 4 and tested on fold 5. In the second iteration, the training is done on folds 1, 2, 3, and 5, and testing on fold 4. This process continues until each fold has been used exactly once as the test set.

3.5.1. Assessment of Modeled versus Reconstructed Data Accuracy

Using our dataset of 700 cases, we implement the k-fold algorithm to evaluate our model's performance. The 700 cases are divided into 5 groups, each containing 140 cases. Each group is then reconstructed with the other 4 remaining groups, meaning that the 140 cases are reconstructed using the remaining 560 cases. This process is repeated 5 times, ensuring that all cases are used for reconstruction. This allows us to create two datasets: one corresponding to cases modeled directly by the numerical model and another corresponding to cases reconstructed using k-fold cross-validation. We can then compare the evolution of the variables studied in both the modeled and reconstructed cases.

In Figure 15, it is represented the evolution of *Hrms*, Hs_{IG} and *Setup* for the first 20 cases selected by the MDA algorithm, comparing the modeled results to the reconstructed ones. As we can observe, the reconstruction of the cases is extremely well executed and almost perfectly matches the modeled results. This indicates that our model performs exceptionally well, and the reconstructions using our model provide highly accurate solutions for unseen data.

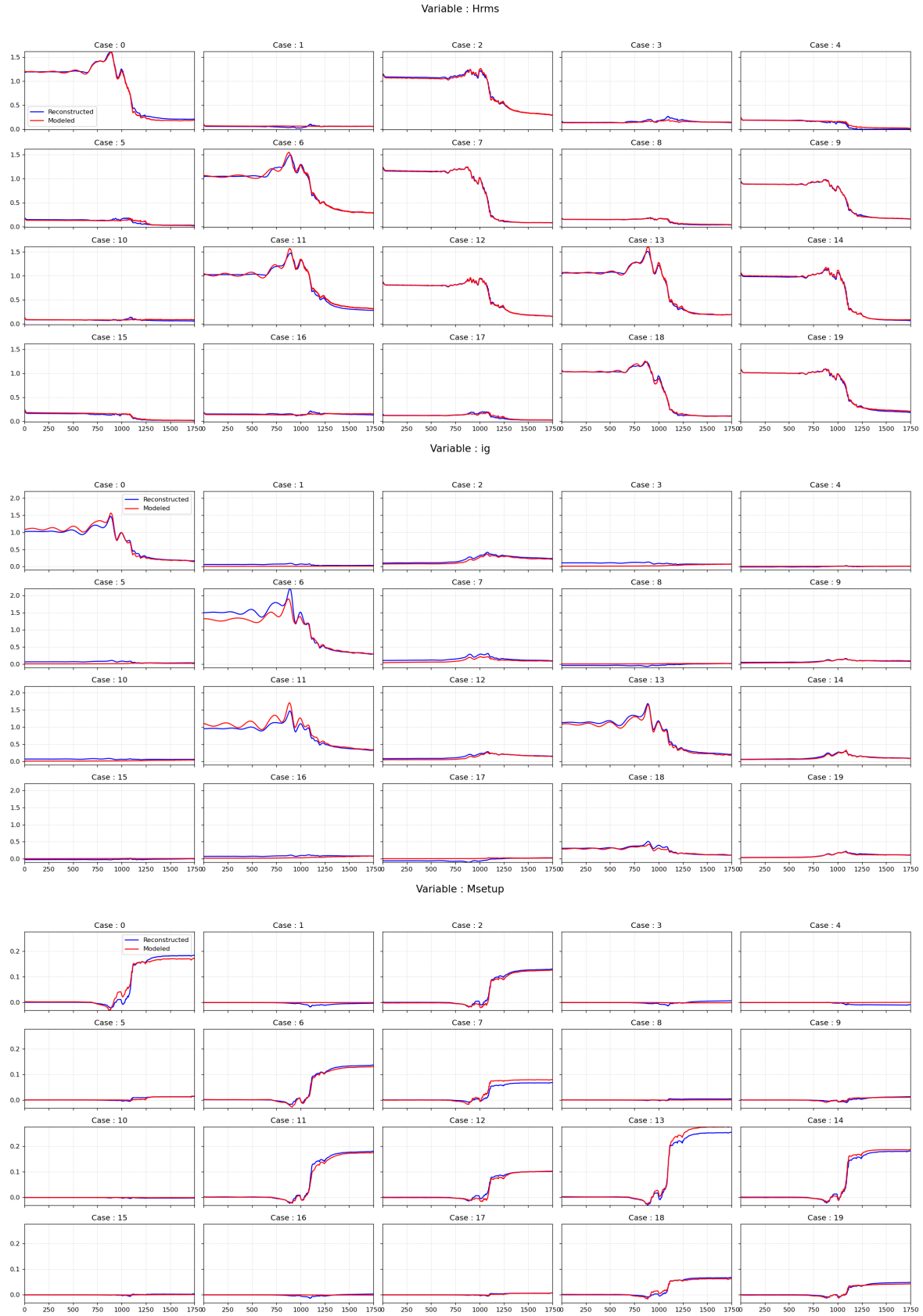


Figure 15. Comparison of modeled (red line) versus reconstructed (blue line) for both output variables *Hrms* and H_{SIG}, for the first 20 cases of the MDA.

3.5.2. Minimum Number of Cases: RMSE Analysis Based on Case Count

To determine the minimum number of cases required to establish an effective metamodel, we employ again k-fold cross-validation with k = 5. We conduct multiple iterations with varying case counts. Initially, we use 50 cases, dividing them into five folds and performing k-fold validation to obtain the mean RMSE. We then incrementally increase the number of cases by 50, repeating the process for each case count until we reach 700 cases. This iterative approach allows us to derive mean RMSE values corresponding to each case quantity used for reconstruction.

Naturally, as the number of cases used for reconstruction increases, the interpolation surface becomes more accurate, resulting in a decrease in RMSE. This trend continues until further increments produce minimal reductions in error. This analysis helps us identify the minimum number of MDA-selected cases needed for optimal interpolation.

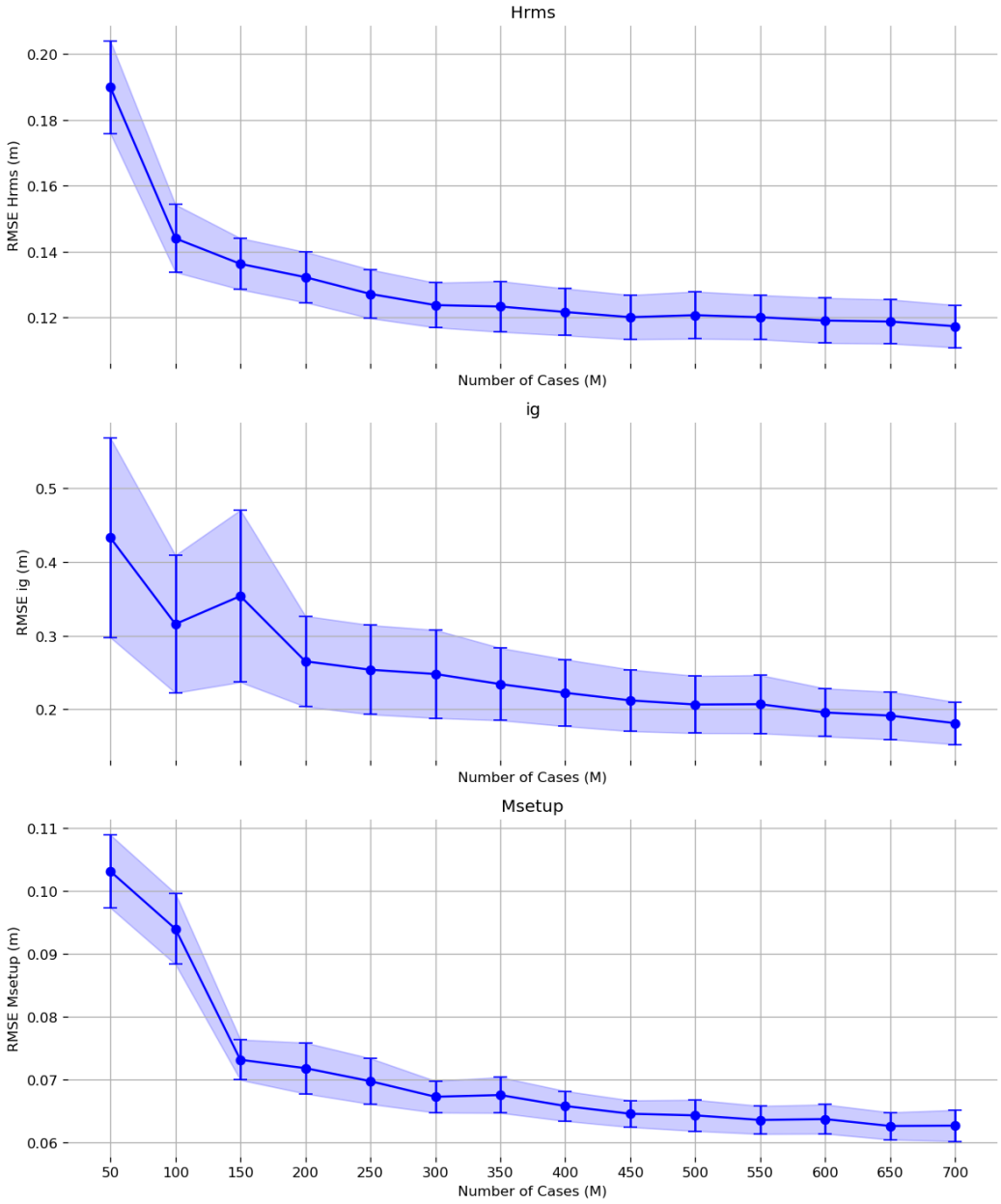


Figure 16. K-fold cross validation on Hrms, HsIG and Setup, as a function of the number of cases that set up the metamodel. The point is the mean RMSE value, and the error bars are the standard deviation of the k-folds.

Figure 16 illustrates the RMSE as a function of the number of cases for the three variables studied in this project (*Hrms*, *HsIG*, and *Setup*). Our observations reveal that the accuracy of the metamodel improves with an increasing number of cases, as evidenced by the decreasing RMSE and associated uncertainty. Additionally, we see that after approximately 300 cases, the reduction in error becomes progressively smaller, indicating that a good interpolation surface can be established with just the initial 300 MDA cases.

However, given that we have 700 MDA cases available, we will use them to ensure the highest accuracy, as both the mean and standard deviation are minimized at this point. We chose to work with 700 cases due to computational resource constraints, as executing a larger number of cases would require significant memory and the available storage space was limited. Consequently, we have determined that using 700 MDA cases is optimal for applying the hybrid method across all output magnitudes.

3.6. Automatic Calibration

Once our model has been validated and we have access to the interpolation surface, which allows us to obtain the spatial evolution of the variables used for calibration (*Hrms* and *HsIG*) for any combination of the 5 input parameters (*Hs*, *Hs/Lo*, *WL*, *Cf*, and *Cr*), our next step is to reconstruct the sea states measured by the sensors using this interpolation function.

The offshore sensor C01 has recorded a total of 2064 hours, during which values for *Hs*, *WL*, *Tp* and numerous other variables were obtained. Our approach involves constructing a dataset comprising these 2064 sea states measured by the offshore sensor. These sea states will be input into the interpolation function along with different combinations of *Cf* and *Cr* to obtain the evolution of *Hrms* and *HsIG* along the profile. Afterwards, the reconstructed evolution using RBF interpolation will be compared with the actual evolution measured by the sensors to determine the optimal combination of *Cf* and *Cr*. The goal is to identify the combination that minimizes the error metric that will be presented next (Eq. 12), ensuring that the modeled evolution of both, *Hrms* and *HsIG*, closely matches the observed data.

For this calibration, we will employ two methods. The first method is a simple approach known as Grid Search, where the user selects the values of the coefficients *Cf* and *Cr* for each reconstruction. On the other hand, we will also use an automatic alternative method using the SCE optimization algorithm, which offers significant improvements in accuracy and efficiency over traditional techniques. In this automatic method, the SCE algorithm iteratively searches for the optimal combination of coefficients without requiring the user to provide different values. By specifying the parameter space to explore, the algorithm adjusts these combinations automatically until it finds the optimal solution.

3.6.1. Error metric

It is clear that we are engaged in hyperparameter optimization, focusing specifically on identifying the optimal combination of parameters – namely, *Cf* and *Cr* – within our modeling framework. Our goal is to systematically explore a defined subset of potential *Cf* and *Cr* values to enhance the simulation accuracy of variables such as *Hrms* and *HsIG*. This process entails iterating through various combinations of these coefficients and evaluating how well they replicate real-world observations within our reef ecosystem.

To assess the fidelity of each combination in replicating the actual evolution of these variables, we calculate the error associated with each combination. The primary objective is to minimize this error, thereby achieving the most accurate representation of variable evolution along the profile. Here are the steps followed to compute the total error associated with each combination.

Initially, we compute the Root Mean Square Error (RMSE) for each sensor and variable :

$$RMSE_{(m\,;\,var)} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \qquad \text{(Eq. 9)}$$

m : denotes the specific sensor for which RMSE is being calculated (total of 19 sensors).
var: represents the specific variable for which RMSE is being calculated: Hrms, HsIG, and *Mean Setup*.
n : is the total number of measurements by each sensor (ranging between xxx and 2064 measurements).
y_i : stands for the real measurement of the studied variable at instant i.
$\hat{y}_i$: represents the modeled value of the studied variable at instant i.

After obtaining the RMSE for each sensor, the next step is to average the RMSE values from all sensors along the profile to get a representative total error for the specific combination.

$$Error_{(var)} = \frac{1}{m}\sum_{i=0}^{m}(RMSE_i) \qquad \text{(Eq. 10)}$$

m : is the total number of sensors distributed along the profile
var : represents the specific variable for which RMSE is being calculated: Hrms, HsIG, and mean setup.

In summary, the process involves first calculating the RMSE for each sensor, then, averaging these errors across all sensors to obtain a final error value along the profile. Next, the error corresponding to each variable is summed, resulting in a unique total error value associated with each combination of coefficients.

$$Total\ combination\ error = \sum Error_{(var)} \qquad \text{(Eq. 11)}$$

Having broken down the total error function for a specific combination into parts, we now present the entire function to be optimized.

$$Total\ Combination\ Error = \sum_{var=1}^{2}\frac{1}{m}\sum_{i=0}^{m}\sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \qquad \text{(Eq. 12)}$$

3.6.2. Grid Search automatic calibration

Grid Search involves defining a grid of potential values for each parameter. The performance of each model is typically evaluated using a predefined assessment metric. In our case, we will be using the error metric discussed earlier (Eq. 12). During this calibration process, we tested various realistic combinations of Cf and Cr to observe how the reconstructions of the variables subject to calibration would appear. In total, we explored 189 points in the parameter space, with Cf ranging from 0.1 to 0.2 in increments of 0.05, and Cr ranging from 0.4 to 0.8 in increments of 0.05. Figure 17 displays all the combinations of Cf and Cr for which reconstructions were performed.

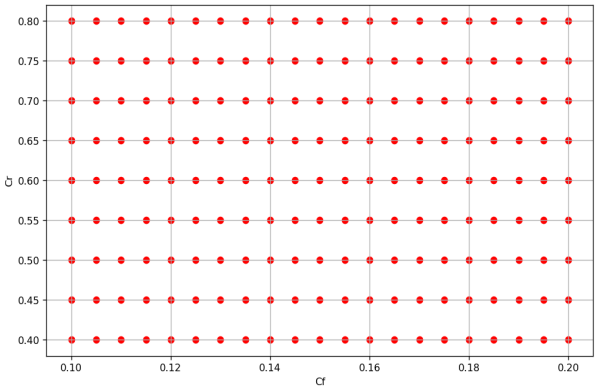


Figure 17. Exploration of various combinations of Cf and Cr in the study space. Each point in the space corresponds to a combination for which the corresponding reconstruction of the variables to be calibrated has been performed, followed by the calculation of the error. (Source: Created by the author)

This approach yields output values for *Hrms* and Hs_{IG} at each point along our profile for every combination. In this way, by knowing the reconstructed evolutions of both of these variables, along with the actual measured evolutions, the error is computed following the process shown in Figure 18. This leads to the error surfaces shown in Figure 20, which are detailed in the following section.

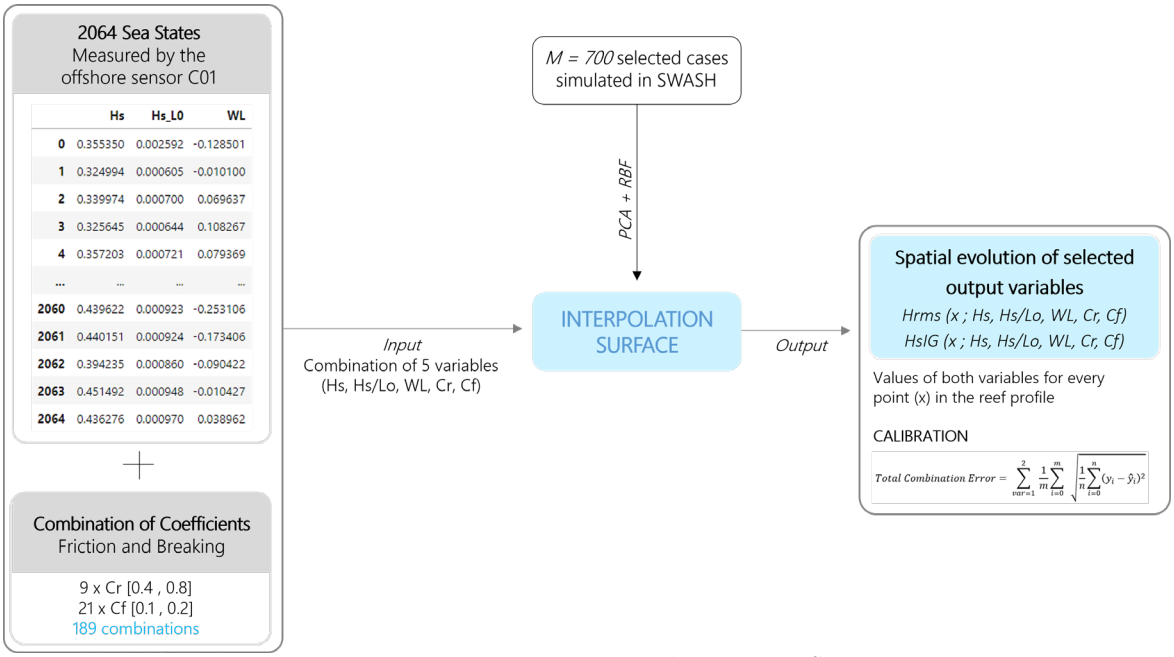


Figure 18. Grid Search automatic calibration workflow.

3.6.3. SCE-based automatic calibration

Metaheuristic algorithms provide an efficient means to search a large solution space and identify near-optimal solutions for problems with multiple variables, nonlinear constraints, and complex objective functions (Amini et al., 2024). In the present project, a key innovation is the application of a metaheuristic optimization algorithm for calibrating the friction and breaking coefficients. This enables efficient automated searches to identify optimal values that align with measurements. We found that the optimization algorithms rapidly converge to precise coefficients, enhancing accuracy and overcoming the limitations of Grid Search calibration, which can be laborious and inconsistent.

In a Grid Search, the search space is defined by establishing the maximum and minimum limits of each variable and the interval between these limits to perform a point sweep. However, with this algorithm, there is no need to specify the interval; it will directly search for the optimum with decimal precision, which would be very costly and time-consuming to achieve using a Grid Search calibration due to the vast number of combinations that would need to be tested.

In this way, there are two fundamental reasons for utilizing this algorithm. First, it allows for rapid identification of the optimal solution with greater precision than Grid Search methods. Second, it can be adapted for the calibration of additional variables, where the dimensionality of the calibration space increases significantly. As the number of coefficients grows, the calibration space expands, and Grid Search methods would require immense time and effort. In contrast, this algorithm efficiently handles high-dimensional spaces, swiftly offering optimal solutions and proving to be a powerful tool for future, more complex calibrations.

SCE Algorithm

The SCE algorithm is a global optimization method designed to solve complex model calibration problems (Duan et al., 1994). This algorithm combines several techniques to efficiently explore the parameter space and find the optimal solution. Here is a simple explanation of how this algorithm works:

1. **Sample Generation.** First, the algorithm generates an initial set of 's' points (parameters) randomly within the allowable parameter space. These points are generated using a uniform probability distribution if there is no prior information about the location of the global optimum.
2. **Point Ordering.** The generated points are ordered according to the value of the objective function, so that the first point corresponds to the lowest value (in minimization problems).
3. **Partition into Complexes.** The set of points is divided into several 'p' complexes. Each complex is a subset of 'm' points selected to maintain good coverage of the parameter space. This partitioning is done following a specific pattern to ensure diversity.
4. **Complex Evolution.** Each complex evolves independently through a process called Competitive Complex Evolution (CCE). In this process, the points within each complex are iteratively adjusted to improve the value of the objective function. This adjustment is based on a combination of deterministic and probabilistic methods.
5. **Shuffling**. Periodically, the complexes are shuffled and reorganized. This reordering introduces new points to each complex, helping to prevent the algorithm from getting stuck in local optima and promoting better exploration of the parameter space.

In Figure 19, the algorithm's operation is illustrated. Initially, we have a population of 10 points divided into 2 complexes of 5 points each. Throughout each iteration, these points are reordered and adjusted in an effort to identify the minima. By the second panel, it is evident that each complex is nearing a minimum. While the asterisks indicate a relative minimum, the dots represent the absolute minimum within the studied space. Despite this, the algorithm does not stop at the relative minimum; instead, it persists in its search for the global optimum. This iterative process, involving constant reordering and adjustment, ensures a comprehensive exploration of the parameter space and convergence to the global minimum.

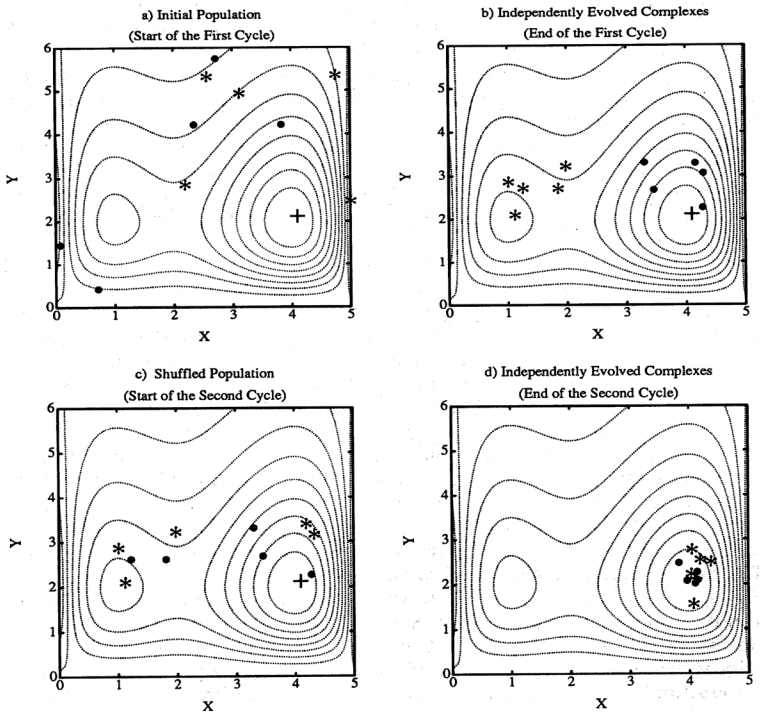


Figure 19. Illustration of the shuffled complex evolution (SCE) method. (Source : Duan et al., 1994)

This algorithm is designed to address complex optimization problems with several key features. One of its main strengths is the ability to converge globally, even when multiple regions of attraction exist in the parameter space, allowing it to find the best possible solution rather than being limited to a local optimum. Additionally, this algorithm is also robust against differences in parameter sensitivity and interdependence, making it suitable for problems where parameters have varying degrees of influence on the outcome. Finally, the SCE efficiently handles high dimensionality in parameters, ensuring thorough exploration of the search space even in cases with many parameters.

The effectiveness of the SCE algorithm depends on the appropriate selection of certain parameters, the most important being:

- ***bl and bu***: These are the lower and upper boundaries of each parameter. For the friction coefficient (Cf) and the wave breaking coefficient (Cr), the boundaries are between 0.1 and 0.2, and 0.4 and 0.8, respectively.
- ***ngs***: This is the number of subgroups into which the initial population will be divided. Each subgroup independently seeks to minimize the RMSE error, increasing the probability of finding the global minimum error. In our case, we have divided the population into 5 subgroups.
- ***peps***: This defines the stopping criterion based on the relative improvement of the objective function. If the improvement in the objective function (in our case, the RMSE error metric) is less than *peps*, the algorithm considers it has converged. For the current case, a value of 0.001 is used, meaning the algorithm will consider it has converged if the improvement in RMSE error is less than 0.1% between consecutive iterations.
- ***maxn***: This is the number of objective function evaluations to be performed. In our case, it is set to 100. Once this limit is reached, the algorithm will stop even if it has not yet converged.

- ***kstop***: This defines the maximum number of evolution cycles without significant improvement before the algorithm is considered to have converged. With *kstop* = 10, if no significant improvement in RMSE error is observed after 10 evolution cycles, the algorithm will stop, assuming it has reached convergence.
- ***pcento***: This specifies the allowable percentage change in the objective function during *kstop* evolution cycles before the algorithm is considered to have converged. In the current case, a value of *pcento* = 0.01 is used, meaning the algorithm will consider it has converged if the percentage change in RMSE error is less than 1% over 10 consecutive cycles. This ensures the RMSE error has stabilized before the algorithm stops.

For the reasons presented above, the algorithm has been selected for our work as it addresses the complexity of our problem effectively, being a robust and efficient tool that combines various optimization approaches to explore and exploit the parameter space, ensuring a high probability of finding the global optimum in our complex calibration problem.

In the following section, we will explore the results provided by this algorithm in terms of obtaining the optimal combination of coefficients. Through detailed analysis of the SCE algorithm's output, we aim to determine the set of coefficients that most accurately mirrors the observed data, ensuring the model's reliability and precision in capturing the actual physical phenomena.

4. Results

Following the detailed explanation of the CHySwash metamodel's functionality, we now present the results for our case study in Molokai where this methodology has been applied. As outlined in the previous section, the interpolation surface is utilized to perform reconstructions based on the five input parameters (*Hs*, *Hs/Lo*, *WL*, *Cf*, and *Cr*). The first three parameters are measured by sensors over a period of 2064 hours (86 days), while the latter two are the calibration coefficients. These two coefficients are tested by reconstructing the selected output variables, *Hrms* and Hs_{IG}, which are then compared with measurements to finally obtain the optimal values of both *Cf* and *Cr*.

4.1. Error Surfaces and Optimal Combination

After defining the points representing the 189 different combinations (*Cf* ranging from 0.1 to 0.2 in increments of 0.05, and *Cr* ranging from 0.4 to 0.8 in increments of 0.05) (Figure 17), we calculated the error for each combination concerning the two output variables, resulting in the corresponding error surfaces. To individualize the error for each coefficient combination, the total error for a given combination was determined by summing the errors of the *Hrms* and the Hs_{IG} variables. The error combination is based on a sum because the errors for both output variables are of the same order of magnitude, allowing direct summation without the need for dimensional adjustment. Figure 20 shows the error surfaces for the two studied variables and the total error.

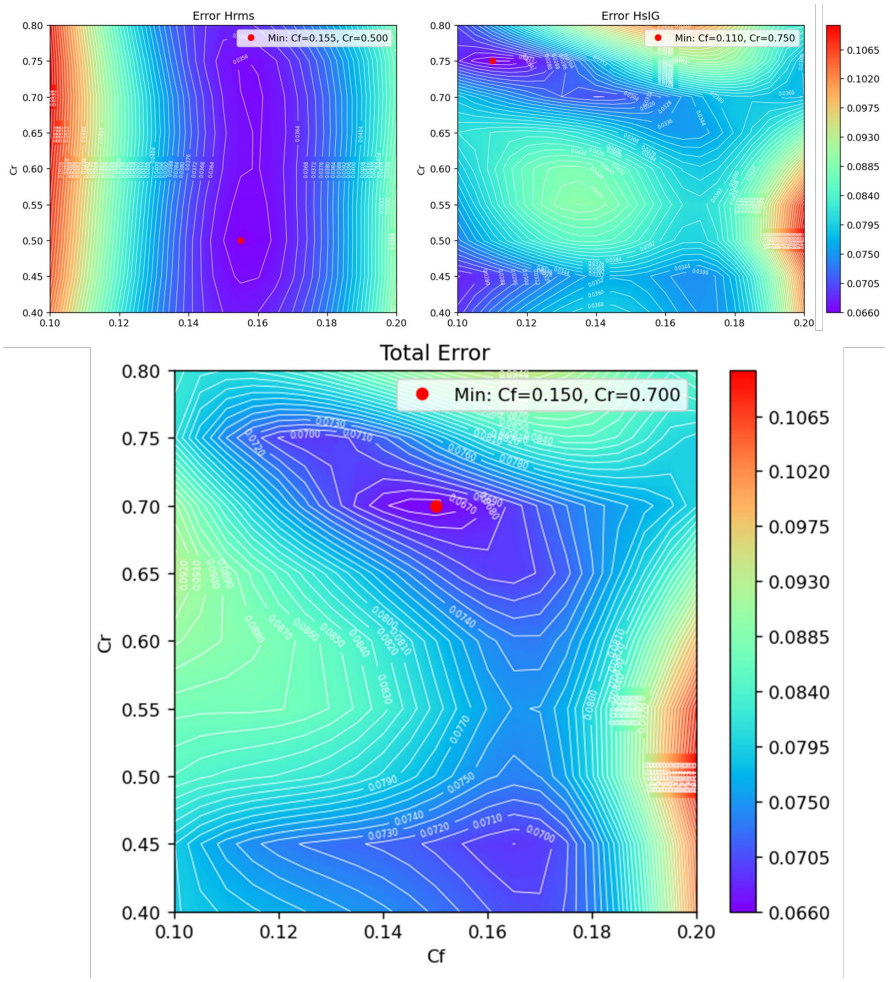


Figure 20. Error surfaces for the variables *Hrms* and Hs_{IG}, and the total error. (Source: created by the author).

As observed, the error surfaces indicate which combinations of the wave breaking coefficient and coral friction coefficient yield the most accurate reproduction of hydrodynamic evolution in the surf zone. Based on Grid Search calibration, the optimal combination identified is Cf = 0.15 and Cr = 0.70. However, one of the project's objectives was also to implement the SCE algorithm for the automatic search of the optimal combination that minimizes the error without testing all the combinations shown earlier. The algorithm would directly provide the optimal combination of coefficients, even with greater decimal precision. Thus, by applying the SCE algorithm to find the optimal point, we find that the solution it offers is a Cf of 0.145 and a Cr of 0.708, resulting in a total error of 0.066 m. Given the greater accuracy of the latter solution, the finally adopted optimal combination of coefficients is:

- A coral-covered zone friction coefficient of 0.145.
- A wave-breaking coefficient of 0.708.

This combination best reproduces the evolution of *Hrms* and Hs_{IG} along our one-dimensional coastal profile of Molokai, concluding that these values accurately reflect the real conditions controlling the area's hydrodynamics.

4.2. Spatial evolution of output variables

Using the adopted optimal combination, we can observe how our metamodel reconstructs the different variables. The following sections present the evolutions of *Hrms*, Hs_{IG}, and *Setup* along our profile.

4.2.1. Variable 1: Hrms

Figure 21 presents the reconstructions of the *Hrms* variable for all tested coefficient combinations. The light gray lines represent the full set of reconstructions for all coefficient combinations evaluated, while the dark gray lines show the reconstruction for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along our studied 1D profile. The reconstructed variable is displayed alongside the actual sensor measurements (red points). For illustration, we present four randomly selected hours from the total of 2064 measured hours. The reconstructions are highly accurate and closely align with the sensor measurements.

4.2.2. Variable 2: HsIG

Figure 22 presents the reconstructions of the Hs_{IG} variable for all tested coefficient combinations. The light gray lines represent the full set of reconstructions for all coefficient combinations evaluated, while the dark gray lines show the reconstruction for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along our studied 1D profile. The reconstructed variable is displayed alongside the actual sensor measurements (red points). For illustration, we present four randomly selected hours from the total of 2064 measured hours. At first glance, the reconstruction may seem less accurate than for *Hrms*, but we must consider that the values of this variable are centimetric. The reconstruction is quite accurate, closely matching the sensor measurements with errors of less than 1 centimeter, as seen in the vertical scale of the graph.

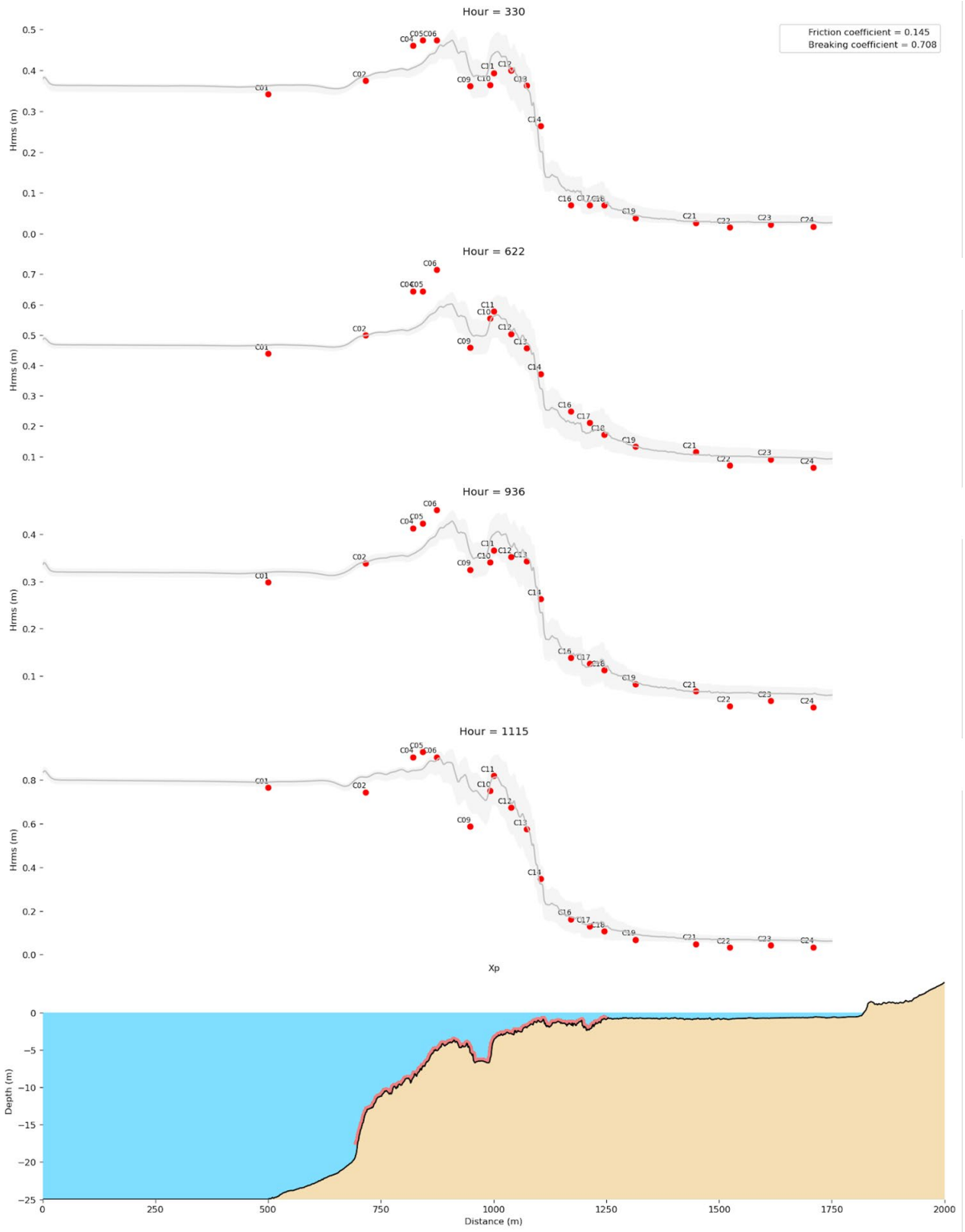


Figure 21. Reconstruction of the *Hrms* variable for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along our studied 1D profile. The reconstructed variable (gray line) is shown along with the actual sensor measurements (red points). Four random hours are displayed. (Source: Created by the author).

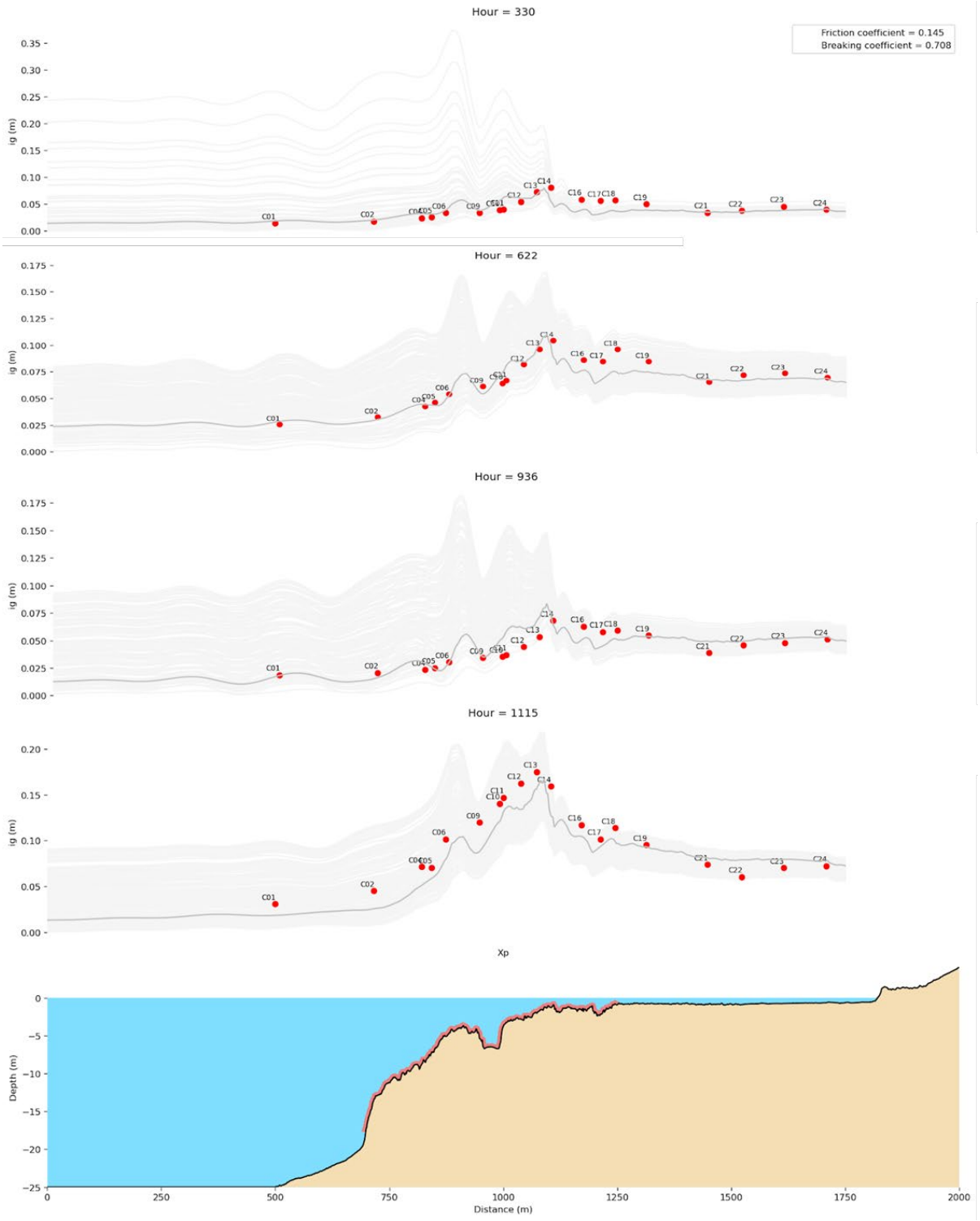


Figure 22. Reconstruction of the Hs_{IG} variable for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along our studied 1D profile. The reconstructed variable (gray line) is shown alongside the actual sensor measurements (red points). Four random hours are displayed.

4.2.3. Variable 3: Setup

As thoroughly explained in Section 3.3.2, we initially intended to use the *Setup* (η) for model calibration, but ultimately, this was not possible. Nevertheless, we believe that the model provides setup outputs that are indeed accurate and considered valid.

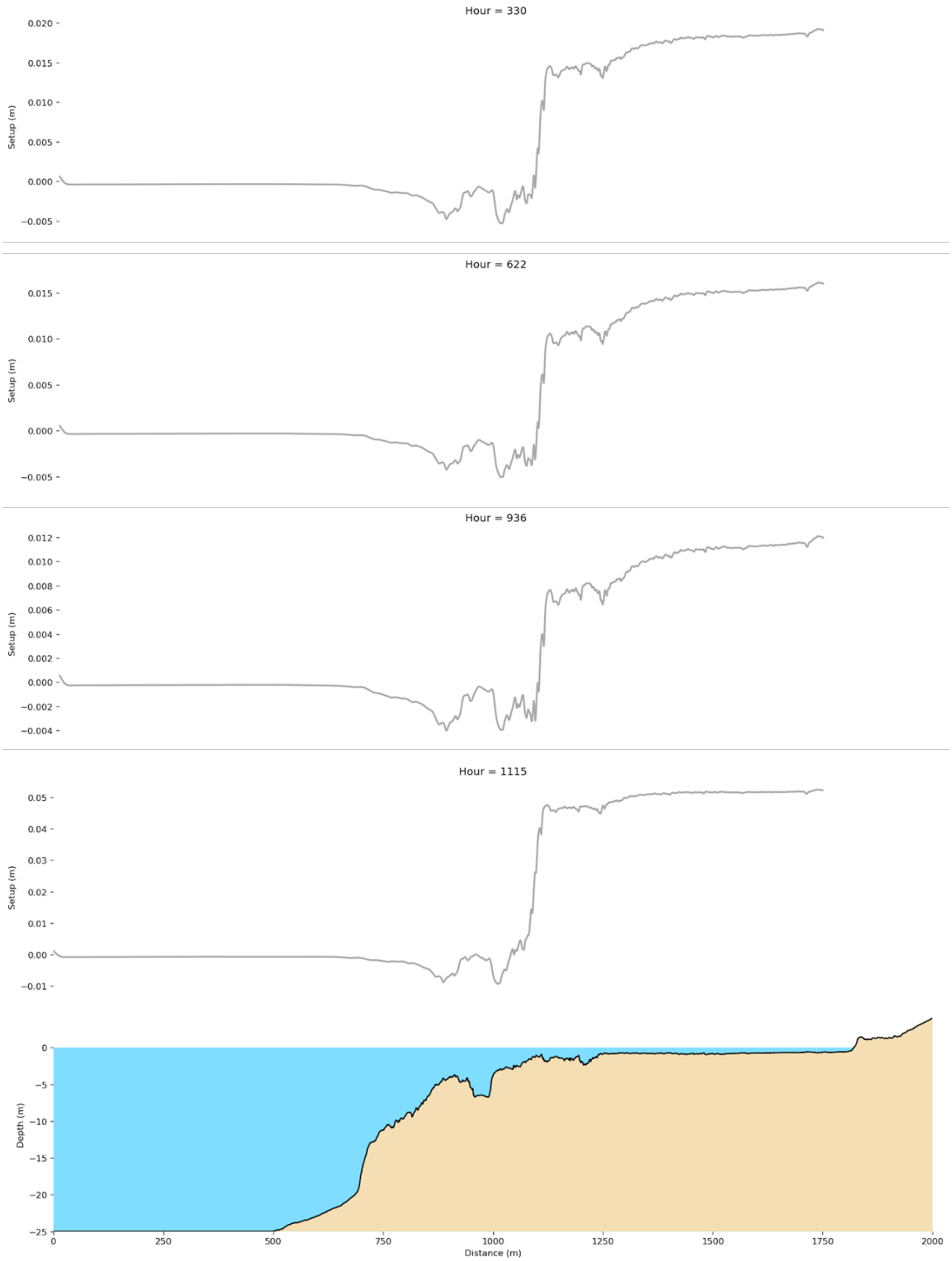


Figure 23. Reconstruction of the Setup variable for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along our studied 1D profile, where the reconstructed variable is depicted by the gray line. Four random hours are displayed.

In Figure 23, we can observe how, for the hours 330, 622, 936 and 1115 recorded by the sensors, the evolution of the *setup*, along the profile follows a coherent trend. In all three cases, the setdown begins precisely when the coral friction profile starts, reaching its maximum at the fore reef zone.

As previously mentioned, given the small wave magnitudes in this studied region, the *setups* are on the order of centimeters. Therefore, due to the lack of precise data, calibrating this variable is nearly impossible due to their small values. Despite the lack of calibration, we conclude that the outputs of our metamodel for the *setup* variable are entirely valid and produce excellent results.

4.3. Two-dimensional Analysis

There is a specific area where our model fails to accurately reconstruct the variable *Hrms*. This area is precisely the reef crest, where sensors C04, C05, and C06 are located. As shown in Figure 24, the reconstruction for hour 977 is accurate along the entire profile length, including the reef crest area. However, for hour 3, the reconstruction is also accurate except for the problematic zone where the three mentioned sensors are located. This issue with reproducing *Hrms* has been observed for several sensor hours. While some hours, like hour 977, are perfectly reconstructed, others fall short, with no tested *Cf* and *Cr* combinations being able to accurately simulate the measurements in this zone.

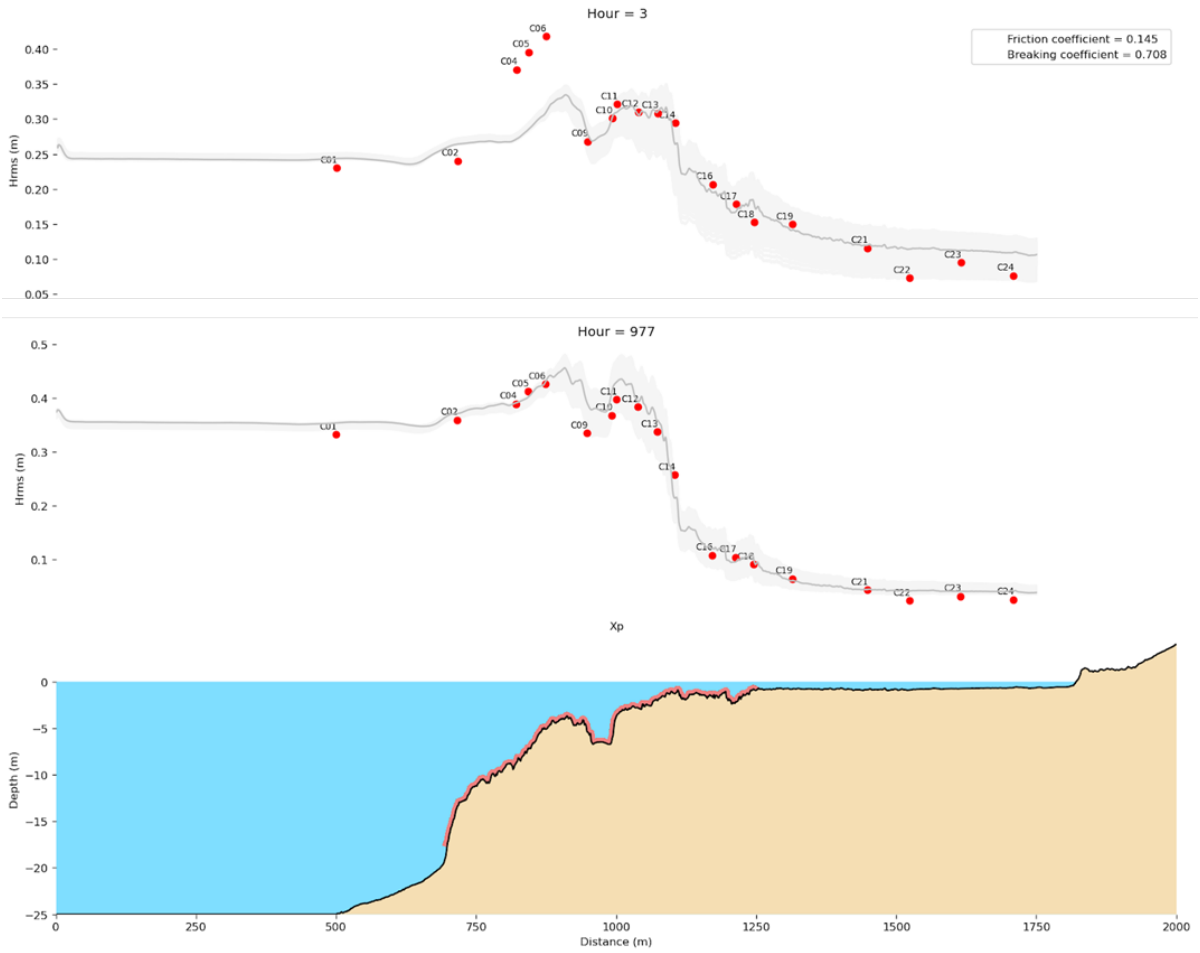


Figure 24. Comparison of the reconstructions of the variable *Hrms* for the hours 3 and 977 recorded by the sensors.

To address these discrepancies, we investigated the challenging nature of the reef crest area, where wave dynamics are more complex due to the higher interaction with the seafloor and the initiation of breaking processes. We believe that bidimensional effects are a primary reason for the inaccuracies. Our simulations are one-dimensional and may not account for transverse factors. Given that we did not have data on wave direction, we assumed waves were perpendicular to the coast, following the direction of our profile. Thus, it is crucial to consider:

- **Transverse Effects and Energy Concentration.** The placement of sensors in an upwelling zone, as indicated by the bathymetric map, suggests the formation of a wave front that concentrates wave energy. This phenomenon could lead to local increases in wave heights measured in situ, which may not be fully captured by the current model that does not explicitly account for complex transverse effects.

- **Wave Obliquity.** The lack of detailed information on the exact direction of incident waves in the study area is another factor to consider. Wave obliquity can cause significant variations in wave heights measured in situ, which may be higher than those predicted by the model assuming simplified perpendicular incidence.

To explore these bidimensional effects, a series of cases were simulated using the XBeach model in non-hydrostatic mode (Roelvink et al., 2009 ; Roelvink et al., 2015), considering different wave directions to see their impact on wave propagation at the sensor locations. Figure 25 shows the wave height propagation coefficient maps ($Kp = Hrms/Hrms0$) corresponding to a specific instant of the sea state corresponding to hour 3 (case shown in Figure 24) for different directions (ranging from 160 to 190). Additionally, Figure 26 presents the evolution of this wave propagation coefficient Kp along the transect. Significant changes can be observed for small differences in wave direction, particularly at sensors C4, C5 and C6, where the Kp is amplified for wave directions less than 180. It is important to emphasize that this was solely a qualitative sensitivity analysis regarding wave direction. Therefore, the XBeach model was not calibrated, i.e. all simulations were conducted using the default calibration coefficients.

Therefore, the differences observed between some reconstructed sea states using the CHySwash calibration metamodel and the in-situ measurements can likely be attributed to several factors. First, the absence of wave direction information in the boundary conditions may contribute to the differences. Second, measurement errors and the accuracy of the sensors could also play a role. Lastly, the methodological limitations of using a profile model that cannot account for transverse effects may further explain these inconsistencies.

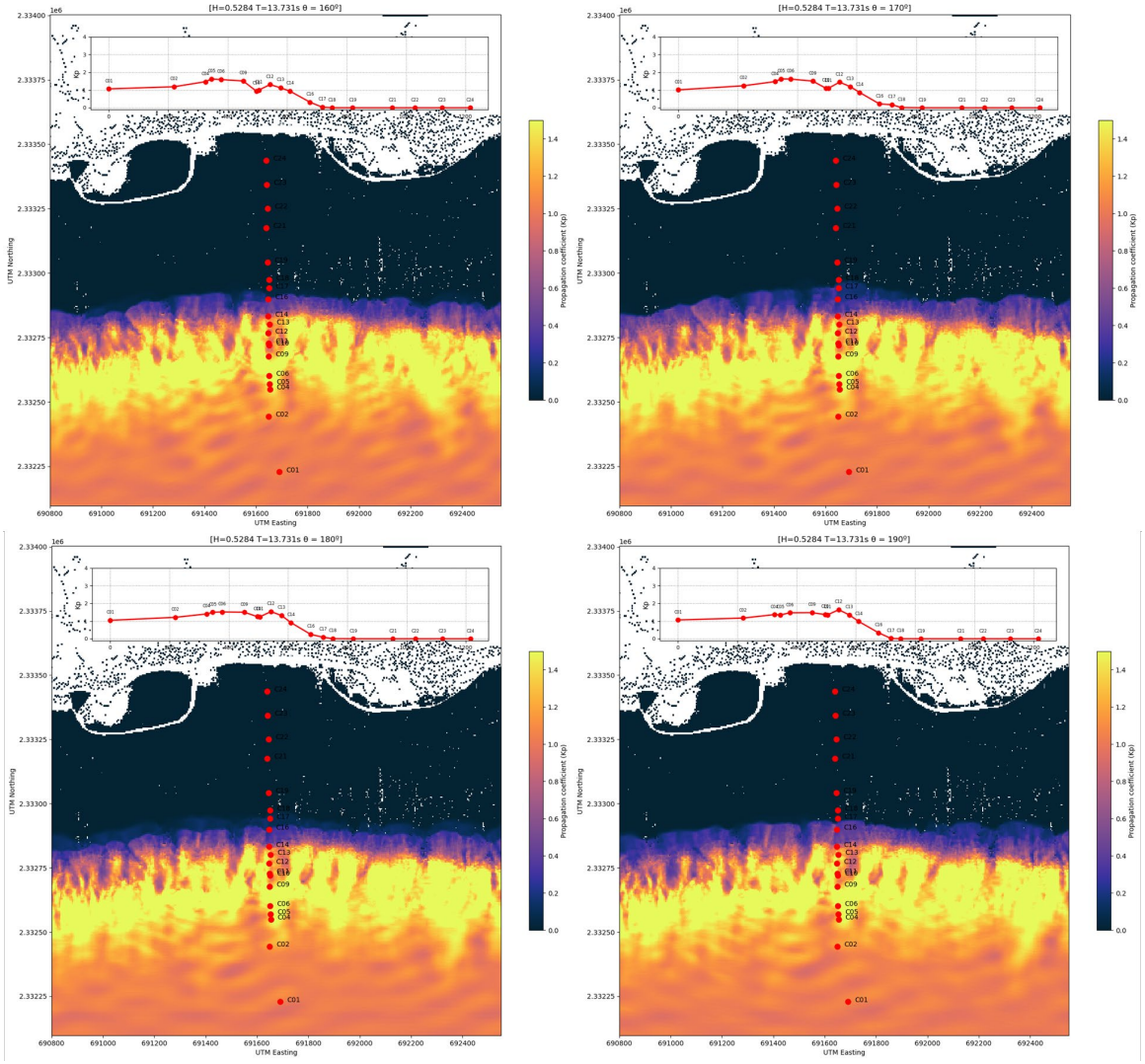


Figure 25. Maps of the *Hrms* propagation coefficient (*Kp*) for the same instant of the sea state corresponding to hour 3, considering different wave directions.

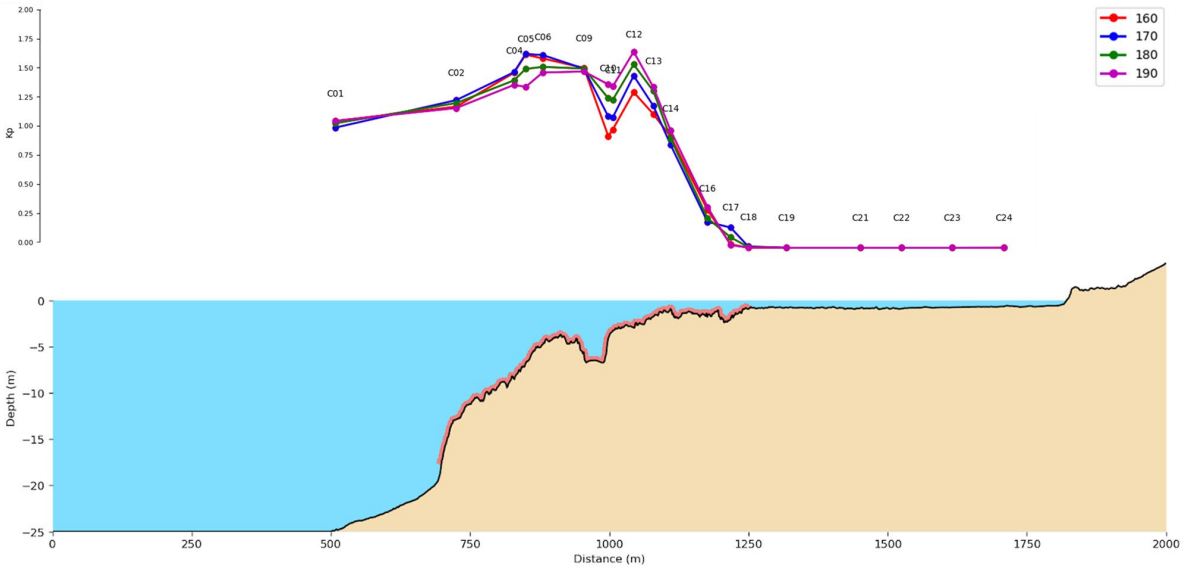


Figure 26. Evolution of the *Hrms* propagation coefficient (*Kp*) along the transect for the same instant and sea state as in Figure 25 and for different wave directions.

4.4. Additional applications of CHySwash

The potential of CHySwash methodology extends to different types of studies. Here, we showcase a potential application for assessing risks of coastal flooding under different climate scenarios, including the consideration of adaptation strategies involving nature-based solutions. To demonstrate this application, we define a flood proxy, a variable used to estimate or infer the likelihood and extent of flooding in a given area. This flood proxy can be calculated by the following expression:

$$Flood\ Proxy = WL + Setup + Hs_{IG} \quad \text{(Eq. 13)}$$

By combining water level, setup, and infragravity waves, this measure provides a comprehensive estimation of total water level along our bathymetric profile. This combined metric (Eq. 13) enhances flood risk assessment by incorporating both immediate water levels and additional contributions from waves and infragravity effects.

Having defined the flood proxy, we identify the hour during the sensor-recorded period when this indicator reached its highest value, signifying the most intense flood risk. The hour with the highest flood proxy value is 1108, with a flood proxy value of 0.449 meters, using the optimal coefficient combination (Cf = 0.145 and Cr = 0.708).

Considering the negative impacts of climate change, such as sea level rise and coral degradation due to issues like coral bleaching, we project future conditions. Sea levels are expected to rise by 30 centimeters by 2050 (Nunez, 2017). In this scenario, we simulate the reconstruction of sensor data over 2064 hours, assuming a sea level rise of 30 cm and a change in the coral friction coefficient to Cf = 0.10, due to coral bleaching, while maintaining Cr = 0.708. Using our fast and robust metamodel, which enables us to reconstruct a sea state in seconds, we reconstruct the variables under these new input conditions and observe the changes in the flood proxy compared to the current scenario. The flood proxy value for the hypothetical 2050 conditions is found to be 0.679 meters, representing an increase of 0.230 meters compared to the current conditions.

As observed in Figure 27, the flood proxy has increased, and the changes in the three output variables of our model are evident. Firstly, for *Hrms*, the values remain roughly the same in the initial part of the profile. However, upon entering the coral friction profile, the heights do not attenuate as much, and at the end of the profile, the heights increase from approximately 0.2m to 0.3m.

In the case of infragravity waves, under the hypothetical 2050 scenario with degraded coral structure and 0.3 meters of SLR, the intensity does not decrease once the coral friction profile ends, maintaining values around 0.14 meters. In contrast, in the current situation, after reaching values of 0.16, the wave decreases to 0.09 meters. Lastly, for *setup*, there is a noticeable increase. Near the coast, the setup value is 0.02 meters in the current situation, whereas, considering the increased WL and reduced friction, it reaches 0.035 meters.

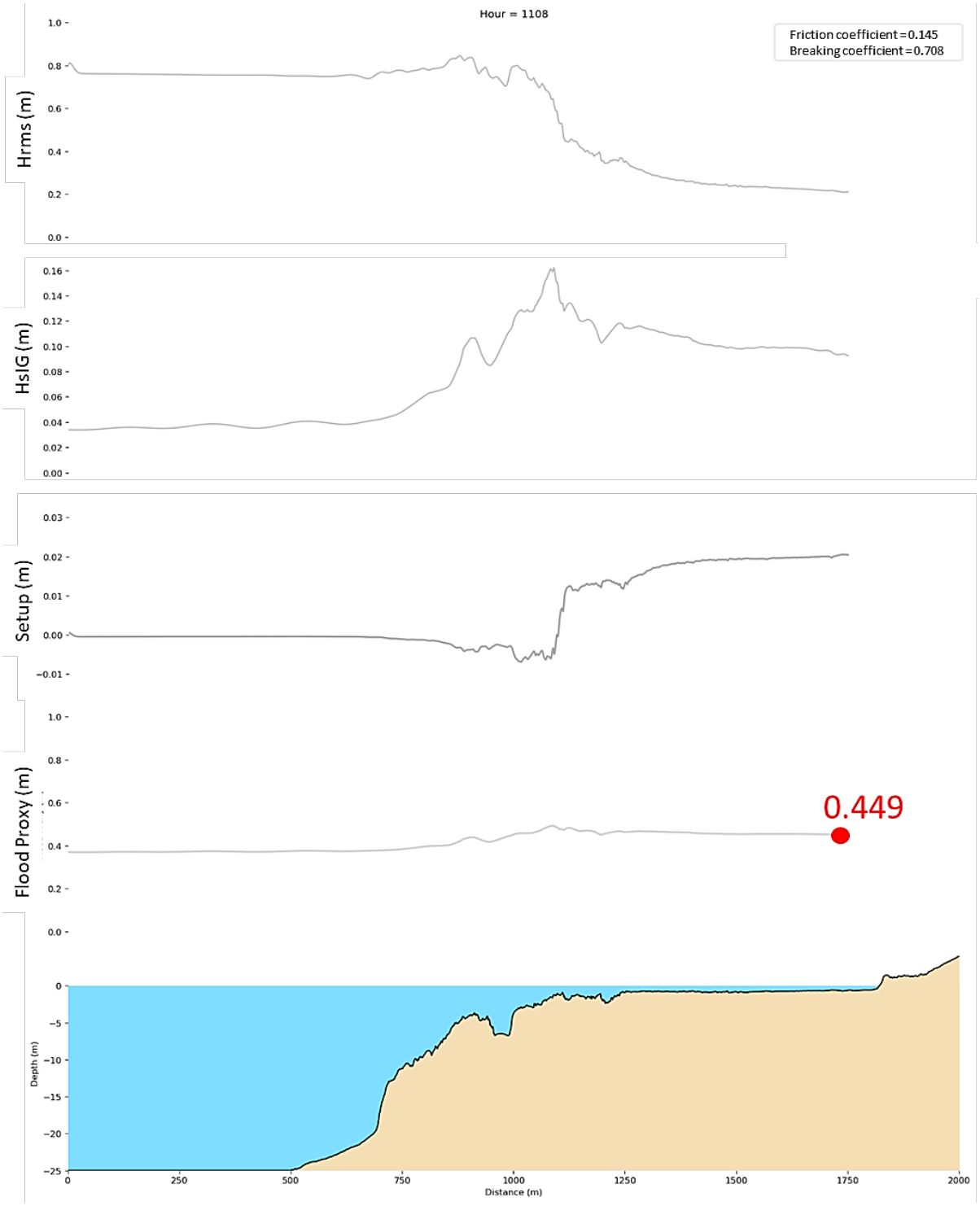


Figure 27. Variables defining the flood proxy for the current situation (*Cf* = 0. 145) of the reef. The breaking coefficient *Cr* is considered to be 0.708. Analysis made for the hour 1108.

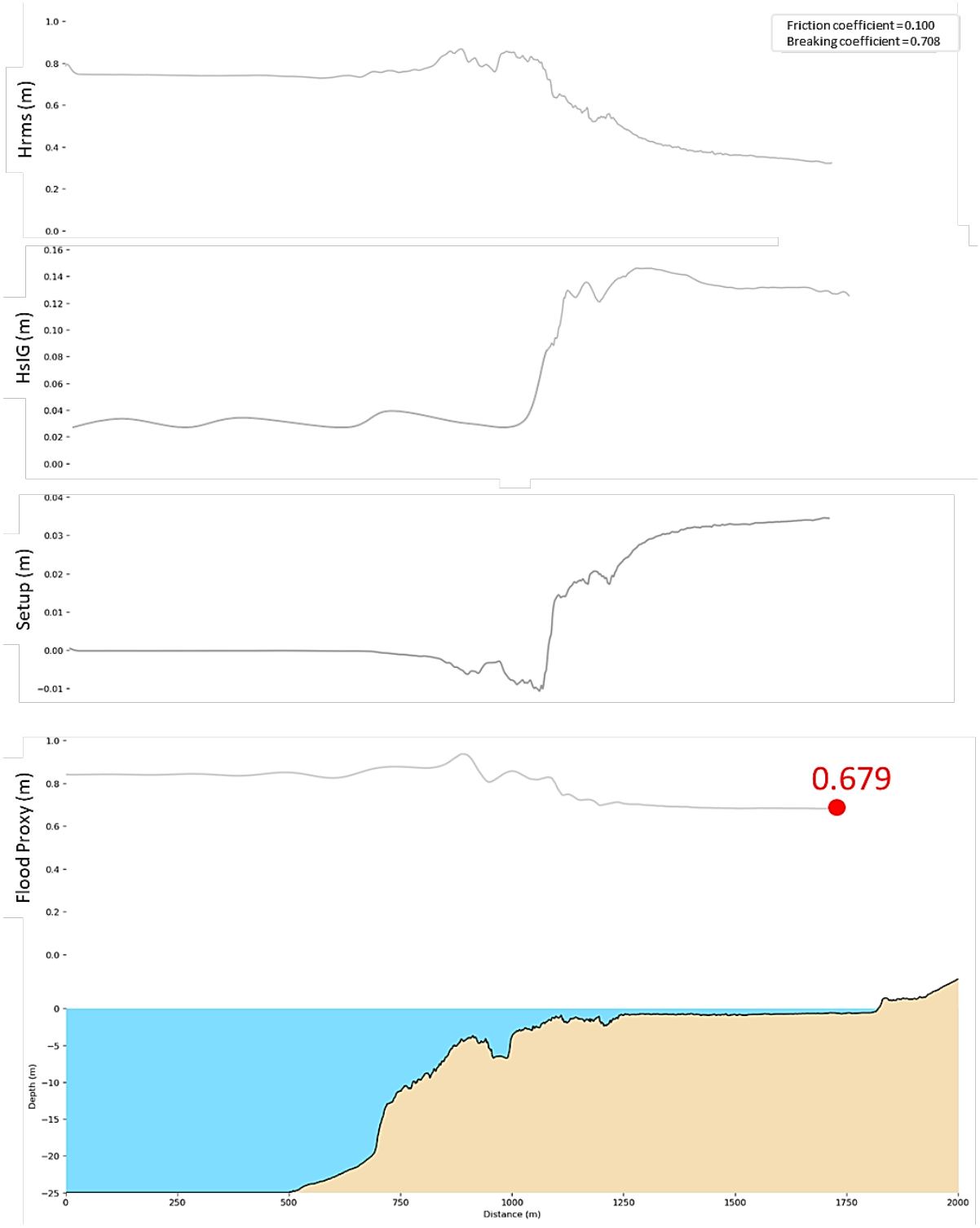


Figure 28. Variables defining the flood proxy for the hypothetical future situation considering a 30 cm increase in WL, along with coral degradation reaching a *Cf* value of 0.10. The breaking coefficient *Cr* is considered to be 0.708. Analysis made for the hour 1108.

This demonstrates our model's capability to predict and reconstruct future scenarios based on anticipated changes in input parameters. However, this is not the end of our analysis, because utilizing our fast metamodel, we can also explore additional solutions. In the context of an extreme future scenario driven by climate change, one potential mitigation strategy is to restore the coral reef, which, in this context, translate into increase in the friction in the coral-covered area. Our study determined that in the extreme 2050 scenario, returning the flood proxy to the current value (i.e. no changes) would require increasing the friction coefficient from the degraded value of 0.1 to 0.30. Although this increase may seem excessive and significantly high, it is important to note that this is merely a simple demonstration of another application of CHySwash, and achieving the same flood proxy value as the current situation would likely require a combination of various adaptation measures, not just coral reef restoration.

Nonetheless, this is an illustration of an application with CHySwash methodology. Naturally, the metamodel would need to be tailored to meet the specific requirements of the application, considering other parameters that It is true that may play a significant role in the study.

5. Environmental Self-Assessment and Reflection on the Project's Contribution to Transitions

This chapter provides a thorough evaluation of the project's environmental impact and its contributions to various transitions. First, we will examine the project's alignment with sustainable practices and its adherence to green budget principles. Finally, we will conclude with a critical reflection on the project's contributions in the context of transitions, offering an in-depth analysis of its influence on environmental, digital, and social transitions.

5.1. Environmental Self-Assessment of the Project According to Green Budget Criteria

In the context of the current ecological transition, the "green budget" outlines the environmental impact of budgetary credits and tax expenditures included in the Finance Bill (PLF). This innovative classification system evaluates fiscal expenditures based on their environmental impact and identifies public resources dedicated to environmental initiatives. France is the first country in the world to have implemented, as early as 2020, a tool for analyzing the environmental impact of its budget. The goal is to better integrate environmental issues into the management of public policies. The French Parliament made it a legal requirement through the Finance Law of December 28, 2019, effective from 2020 (Ministry of Economy, Finance, Industrial and Digital Sovereignty, 2023).

Expenditures in the green budget are classified into three types based on their impact: favorable, neutral, and unfavorable. Favorable expenditures include those with a primary environmental objective or that directly contribute to the production of an environmental good or service, those with no environmental objective but with a proven indirect impact, and favorable expenditures with a controversial impact due to short-term benefits. Neutral expenditures have no significant effect on the environment. Unfavorable expenditures directly harm the environment or encourage environmentally harmful behaviors. Expenditures are evaluated according to six environmental objectives, both at an aggregated level, including state expenditure norms and total tax expenditures, and at the budgetary mission level corresponding to parliamentary votes:

- Combating climate change.
- Adapting to climate change and preventing natural risks.
- Managing water resources.
- Transitioning to a circular economy, managing waste, and preventing technological risks.
- Combating pollution.
- Preserving biodiversity and protecting natural, agricultural, and forest areas.

In this section, the impact of the present project will be evaluated in each of these areas to determine if the studied methodology and the surrounding context affect each category positively, negatively, or neutrally. The following is a detailed analysis for each category, explaining how the concepts presented in the work can be applied to each one.

5.1.1. Climate change mitigation

As previously mentioned, forecasting coastal dynamics has traditionally relied on high-fidelity numerical models, which, although detailed, are computationally expensive. To reduce these costs, metamodels have been developed, combining numerical models with statistical techniques to predict wave behavior more efficiently. However, these metamodels require complex calibration to ensure accuracy, involving the adjustment of parameters to match observed data. As the number of parameters and the duration of observations increase, the calibration process becomes exponentially more demanding, leading to significant computational challenges and potential difficulties in achieving precise results.

The efficient metamodel developed in this project addresses these challenges by allowing the reproduction of any case with minimal simulations, drastically reducing computational costs. Without this metamodel, conducting the same calibration exercise dynamically would require significantly more computational resources and time. This reduction in computational demand directly lowers the carbon footprint by reducing the number of computers needed and their operational time. Thus, the methodology presented in this project not only enhances tools for predicting coastal scenarios but also contributes to climate change mitigation by reducing the energy consumption and carbon emissions associated with extensive computational modeling. For these reasons, the impact of this project in this category should be considered **FAVORABLE**.

5.1.2. Adapting to climate change and preventing natural risks

The increasing frequency and intensity of storms and rising sea levels underscore the need for effective strategies. Thus, the methodology developed in this project provides precise parameters to accurately characterize coastal hydrodynamic spaces and enhance coastal management tools. This facilitates informed decision-making and the implementation of proactive measures to protect vulnerable coastal communities.

Moreover, coral reef surface roughness is a key indicator of ecosystem health, linked to important metrics such as fish biomass, carbonate production, and live coral cover (Price et al., 2019). Our methodology allows us to obtain optimal values for reef surface roughness, enabling effective monitoring and correlation with coral health.

Being informed about local conditions empowers us to devise flood control strategies and take actions to preserve coral health. Understanding the ideal roughness of coral structures also facilitates the implementation of alternative approaches. For instance, in scenarios of significant coral degradation with suboptimal roughness values, tactics such as installing artificial corals can be utilized to mitigate flooding and erosion, thereby ensuring sustainable coastal protection (Kim et al., 2022). The overall assessment for this category is **FAVORABLE**.

5.1.3. Water resources management

The methodology developed in this project plays a crucial role in water resources management, particularly in the context of coastal protection and flood control. By accurately calibrating parameters such as friction coefficients and wave breaking over coral reefs, this methodology enhances our understanding of coastal hydrodynamics, contributing directly to effective management of coastal water resources by mitigating impacts from storm-driven flooding and reducing coastal erosion.

Additionally, coral reefs in nearshore waters degrade more severely compared to offshore reefs due to the combined effects of thermal stress caused by global environmental stressors and increased sediment and nutrient concentrations resulting from local disturbances (Baumann et al., 2021). This decrease in water quality further accelerates coral reef degradation (Nalley et al., 2021). Thus, if we have the roughness coefficient as an indicator of coral structure health, as mentioned earlier, utilizing this methodology to monitor roughness in the area allows us to link water quality with roughness.

Regarding water filtration, coral reefs play a crucial role in keeping nearshore waters clean. Many corals and sponges are filter feeders, meaning they consume particulate matter, pollutants that do not dissolve in water. This prevents these particles from settling on the ocean floor and polluting the sea with harmful substances (United Nations Environment Programme, 2020)

In conclusion, the project has a clearly **FAVORABLE** impact in this category. While the methodology primarily focuses on coastal hydrodynamics, its implications for water resources management are profound. By enhancing our ability to predict and manage coastal flooding and erosion, this project offers a comprehensive tool to support sustainable water resource management in coastal areas. Monitoring the role of coral reefs as natural barriers that absorb wave energy helps protect coastlines from erosion and flooding. Moreover, by preserving and restoring these reefs, the project indirectly contributes to maintaining water quality. Healthier reefs reduce the influx of sediments and pollutants into coastal waters, which is crucial for overall marine ecosystem health.

5.1.4. Transitioning to a circular economy, managing waste, and preventing technological risks

The methodology developed in this project initially appeared to not directly address these areas, with its main contribution focused on improving coastal management through better calibration of hydrodynamic models and monitoring coral reef health. However, upon further consideration, the impact in this category should be viewed as **FAVORABLE**.

By providing an efficient metamodel, this project has the potential to significantly reduce technological risks, particularly in developing communities with limited computational resources. For instance, by offering the metamodel with a pre-established database, these communities can make accurate predictions without needing to run complex numerical models. This approach has already proven beneficial in places like Pacific Island Countries (PICs) (Cagigal et al., 2024 ; Ricondo et al., 2023) where resources are scarce. By enabling these communities to perform high-quality predictions without extensive computational demands, the project not only enhances their capacity to manage coastal risks but also prevents potential technological challenges associated with resource limitations. Therefore, the methodology contributes positively to technological risk prevention and supports the transition toward a more sustainable and resilient coastal management framework.

5.1.5. Pollution abatement

As previously mentioned, coral reefs play a crucial role in preserving the cleanliness of nearshore waters by acting as natural filters. This prevents particles from settling on the ocean floor and thus avoids ocean contamination with harmful substances. While the methodology does not directly address pollution reduction, its contributions to coral reef health and stability have significant secondary benefits.

By fostering healthier reef ecosystems, the project indirectly supports efforts to combat marine pollution. However, because the direct application of the methodology does not primarily focus on pollution control, the impact in this category is considered **NEUTRAL**.

5.1.6. Biodiversity and sustainable land use

As previously mentioned in earlier categories, there is a direct correlation between coral reef surface roughness and its biodiversity (Tucker, 2021). Surface roughness of coral reefs serves as a key indicator of ecosystem health and is linked to other important ecosystem metrics such as high fish biomass, carbonate production, and live coral cover. However, coral reef surface roughness, in terms of benthic structural or architectural complexity, has declined globally over recent decades (Alvarez-Filip et al., 2013). A decline in coral reef surface roughness may indicate that the coral reef ecosystem is approaching functional collapse (Harvey et al., 2018). Therefore, monitoring coral reef ecosystems may include surface roughness as a metric to track ecosystem dynamics and enable detailed descriptions of coral reef habitat zones. Thus, the present methodology provides valuable insights into the actual roughness values of coral structure in the area.

Furthermore, coral reefs are critical habitats for a diverse range of marine species. The current methodology, which enhances our understanding and management of local hydrodynamics, can aid in favoring coral preservation and, consequently, biodiversity conservation. As emphasized throughout various points in the project, healthy coral reefs are pivotal for multiple ecological functions, including acting as natural barriers that mitigate wave energy and protect coastlines from erosion. Additionally, they significantly contribute to improving water quality. Reefs can trap and stabilize sediments and pollutants, preventing these substances from dispersing and causing further environmental damage. This natural filtration process is crucial for maintaining the clarity and quality of coastal waters.

In conclusion, our project is considered to have a **FAVORABLE** impact in this category. The calibration of friction and wave breaking coefficients in the proposed methodology helps understand and predict the behavior of coastal waters. By doing so, it supports the conservation and restoration of coral reefs.

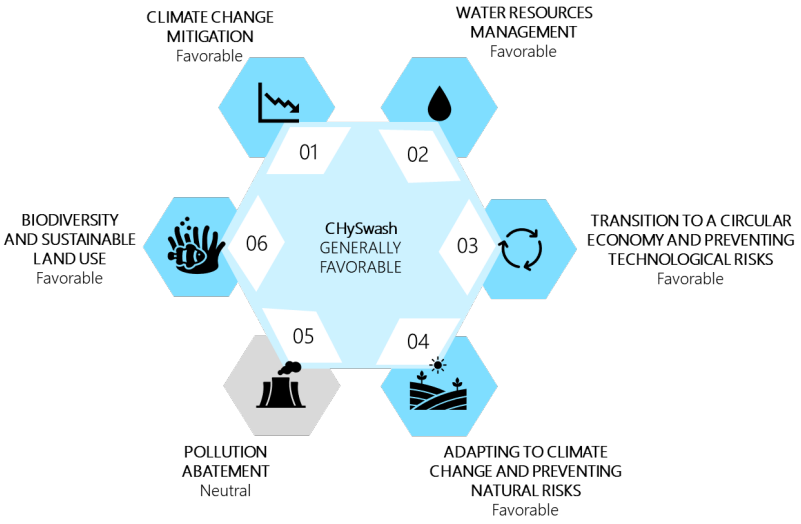


Figure 29. Overall environmental score of the project based on the six impact categories. (Source: Created by the author)

The project's methodology significantly enhances our capacity to manage coastal hydrodynamics, offering critical insights for the effective management of coastal flooding and erosion. By providing an efficient metamodel, the project not only reduces computational demands—thereby lowering the carbon footprint and contributing to climate change mitigation—but also empowers resource-constrained communities to perform accurate predictions without the need for extensive computational resources, reducing technological risks.

Moreover, the methodology provides essential parameters, such as coral roughness, which serve as indicators of the health and stability of reef ecosystems. These parameters aid in preserving coral reefs and indirectly support efforts to improve water quality and biodiversity. By facilitating the conservation and restoration of coral reefs, the project contributes to broader ecological sustainability and resilience, which is crucial for protecting coastlines, maintaining marine biodiversity, and ensuring the long-term health of coastal environments.

Given the favorable impact across multiple environmental categories—including climate change mitigation, adaptation to natural risks, water resource management, technological risk prevention, and biodiversity conservation— the project is considered **GENERALLY FAVORABLE** overall, underscoring its significant contributions to sustainable coastal management and ecosystem conservation, as summarized in Figure 28.

5.2. Critical Reflection on the Project's Contribution to Transitions

Contribution to the Ecological Transition

This project, centered on the automatic calibration of hydrodynamic models, plays a vital role in advancing the ecological transition towards more sustainable coastal management. By improving our ability to predict and manage coral reefs impact on coastal protection and biodiversity, the project aids in the conservation of these crucial ecosystems. The use of coral roughness coefficients to assess reef health links water quality with reef structure, addressing degradation issues caused by factors like thermal stress and increased sediments. Additionally, the project enhances sustainability by reducing computational resource needs, which lowers the overall carbon footprint of the modeling process.

Contribution to the Digital Transition

The project incorporates advanced digital technologies and modeling techniques that are fundamental to the digital transition in natural resource management. The application of this methodology represents a significant advancement in the use of predictive analysis in environmental sciences. These techniques allow for greater precision in the calibration of hydrodynamic models, facilitating more efficient and effective management of coral reefs. Additionally, the development of these digital tools not only enhances researchers' ability to model and predict changes in ecosystems but also promotes the adoption of emerging technologies in the field of environmental management. This can lead to greater integration of geographic information systems (GIS), remote sensors, and real-time data analysis, driving the digitization of monitoring and conservation of coastal ecosystems.

Contribution to the Social Transition

From a social perspective, the project has the potential to involve local communities in reef conservation, promoting sustainable practices and raising awareness about the importance of coral reefs. Implementing these practices not only protects marine ecosystems but also supports local economies that depend on the health of these ecosystems, such as fishing and tourism. Collaboration with local communities is crucial for the continuous monitoring and protection of reefs. These communities can provide invaluable data and be at the forefront of conservation efforts. Additionally, the project can foster education and training in conservation and monitoring techniques, empowering communities to play an active role in protecting their natural environment.

Future Perspectives and Recommendations

Based on the project's findings, several recommendations can be proposed to enhance coral reef conservation:

- **Implementation of Conservation Policies.** It is essential for local and national authorities to implement specific policies for coral reef conservation, based on scientific data obtained from projects like this. Integrating this methodology into reef restoration programs could significantly improve their effectiveness.

- **Collaboration with Local Communities.** Collaboration with local communities and scientists is crucial for continuous monitoring and protection of reefs. These communities can provide invaluable data and be at the forefront of conservation efforts.

- **Nature-Based Solutions (NBS).** Nature-Based Solutions refer to approaches that work with nature to address various environmental challenges while providing benefits to people and ecosystems. These solutions leverage natural processes and ecosystems to enhance resilience and mitigate adverse impacts from environmental changes. In the context of coral reef conservation, NBS involve integrating natural processes and habitats to support reef health and coastal protection.

    1. **Use of artificial reefs.** Artificial reefs are human-made structures placed in marine environments to simulate natural reef habitats, offering several benefits for reef conservation and coastal protection. These reefs act as barriers that absorb and dissipate wave energy, which helps to reduce coastal erosion by shielding shorelines from the full force of incoming waves. Additionally, artificial reefs create new habitats for marine species, providing refuge, breeding grounds, and feeding areas, which supports and enhances biodiversity. By introducing these structures into the marine environment, artificial reefs attract a variety of species that might not otherwise be present, leading to increased biodiversity as they become colonized by diverse marine organisms such as fish, invertebrates, and algae.

    2. **Exploring Additional Measures.** While artificial reefs are a valuable tool, they should be considered as part of a broader strategy that includes other nature-based solutions.

        - **Mangroves.** Planting and preserving mangrove forests can provide critical coastal protection by stabilizing sediments and reducing wave energy. Mangroves also serve as important breeding and nursery grounds for many marine species, contributing to the health of coral reef ecosystems.

- **Seagrass Meadows.** Protecting and restoring seagrass meadows can enhance water quality and provide habitat for marine life. Seagrasses help to trap sediments, reduce turbidity, and support diverse marine communities, all of which benefit coral reefs.

- **Grey Solutions.** In addition to natural approaches, integrating 'grey solutions' (such as coastal engineering techniques) with nature-based solutions can offer a comprehensive approach to coastal management. Combining these strategies can provide more robust and adaptable solutions for managing and protecting coastal ecosystems.

Incorporating artificial reefs into conservation efforts represents a proactive step towards enhancing coastal resilience and biodiversity. However, their effectiveness can be further amplified when combined with other nature-based solutions. This holistic approach ensures that multiple layers of protection and support are provided to coral reef ecosystems, contributing to their long-term sustainability and health.

In conclusion, the project's contribution to environmental, digital, and social transitions is multifaceted. Environmentally, it strengthens sustainable coastal management by providing advanced tools for hydrodynamic model calibration. Digitally, the project employs advanced modeling and data analysis techniques, representing a significant advancement in scientific methodology for ecosystem management. Socially, the project has the potential to involve local communities in reef conservation, promoting sustainable practices and raising awareness about the importance of coral reefs. The project not only offers a valuable tool for the management and conservation of coral reefs but also lays the foundation for future research and policies that promote the sustainability and resilience of coastal ecosystems.

6. Conclusions and future research lines

The development of this project marks a significant advancement in the field of hydrodynamic modeling, particularly for coastal regions with fringing coral reefs. The primary objective of this study was to create a hybrid metamodel capable of automatic calibration of hydrodynamic models, which has been successfully achieved.

We highlight the versatility of the presented methodology, which combines sampling, clustering, interpolation, and optimization tools within a numerical framework, creating a robust system adaptable to any numerical model or parameter requiring calibration. It is not confined to SWASH model but can be extended to numerical models or fields beyond coastal engineering. By defining input variables for numerical simulations and customizing the model to achieve desired outputs, this methodology can effectively address a wide range of engineering challenges.

In the specific case study carried out in this project, the methodology has yielded satisfactory results when applied to coastal engineering. The tool has proven effectiveness in predicting the optimal combination of coefficients that govern wave-breaking behavior and coral roughness in coastal areas with fringing corals. Through a rigorous calibration process, the most effective combination of these coefficients was determined, accurately simulating the hydrodynamic conditions observed in the field data collected from the study area. Finally, the optimal combination of coefficients found consists of a *Cf* of 0.145 and a *Cr* of 0.708.

With the optimized coefficients, coastal risk assessments can incorporate more reliable predictions of wave impacts, inundation levels, and potential damage to coastal infrastructure and ecosystems, which is particularly valuable for areas with fringing corals where wave dynamics are complex and highly influenced by coral morphology.

Another major innovation of this project is the hybrid technology's capacity to drastically reduce computation time. By constructing a response function through interpolation, the model approximates the behavior of the original simulations, enabling efficient prediction and reconstruction of hydrodynamic variables. This approach has proven particularly beneficial in processing extensive datasets, such as the 2064 hours of sensor data used in this study. The goal was to reconstruct this data using our metamodel, enabling comparison of simulated data with real measurements. Without this methodology, finding the optimal combination of friction and wave breaking coefficients would require performing 2064 sea states multiplied by the 189 different combinations of coefficients that have been used for Grid Search calibration, resulting in 390,096 SWASH simulations. Given that each simulation took 30 min to run in an Ubuntu (x86-64) PC, using up to eight 3 GHz Intel i7-9700 processors and 32 GB of RAM, this would total 195,048 hours, or 22.26 years of simulation time, which is computationally impossible.

However, our method of selecting representative cases and reconstructing the interpolation surface allows us to build a comprehensive case library with only 700 simulations. This translates to a total of 350 hours, or approximately 15 days of simulation. From this library, we can almost instantaneously reconstruct the evolution of *Hrms* and Hs_{IG} along the entire profile based on different combinations of the five selected input variables:

$$Hrms\ (x\ ;\ Hs,\ Hs/Lo,\ WL,\ Cr,\ Cf)$$
$$Hs_{IG}\ (x\ ;\ Hs,\ Hs/Lo,\ WL,\ Cr,\ Cf)$$

The impact of this project extends far beyond its technical innovations, making significant contributions across multiple areas. Environmentally, it advances sustainable coastal management by providing sophisticated calibration tools. Digitally, it represents a major methodological advancement in ecosystem management. Socially, it has the potential to engage local communities in reef conservation, fostering sustainability and raising awareness about the crucial role of coral reefs.

Looking ahead, the successful development of the hybrid metamodel opens up several promising avenues for future research:

- **Friction Mapping.** In this study, the focus has been on calibrating the friction coefficient specifically within coral-covered reef areas. However, it is also highly beneficial to extend the analysis to include the entire bathymetric profile. By analyzing friction coefficients across the complete bathymetric profile, researchers can generate a comprehensive friction map.
- **Application to Diverse Coastal Areas.** Expanding the application of the calibration methodology to other coastal environments beyond coral reefs. Adapting the metamodel parameters to different coastal settings with varying challenges will enhance its versatility and utility. This includes incorporating diverse environmental conditions and coastal features to generate robust and adaptable calibration frameworks.
- **2D Analysis and Enhanced Parameterization.** Future research should focus on conducting 2D studies in additional coral reef regions where detailed data on reef morphology and friction coefficients are available. This will facilitate the development of more comprehensive parameterizations, such as friction curves, that capture the nuances of different reef types and conditions.
- **Incorporating Climate Change and Adaptation Measures.** Redefining the metamodel to account for climate change impacts and various nature-based solutions (NBS) is essential. Future studies could explore how different adaptation strategies, including mangrove restoration or flood defenses, affect coastal dynamics. This will enhance the metamodel's ability to predict and manage coastal resilience under changing climate conditions.
- **Addressing Other Coastal Phenomena.** Adapting the metamodel to tackle other coastal phenomena, such as saltwater intrusion or sediment transport issues, will broaden its application. This includes developing calibration parameters specific to these phenomena and exploring its relevance in other fields that require numerical model calibration, such as river systems or urban flood management.

In summary, this project not only advances hydrodynamic model calibration but also lays a strong foundation for future research and applications. The blend of methodological versatility, computational efficiency, and potential for real-time use establishes a solid base for further studies aimed at improving the understanding and management of coastal and hydrodynamic systems.

7. Bibliography

Alfeld, P., 1989. Scattered data interpolation in three or more variables. In: Mathematical Methods in Computer Aided Geometric Design, vol. 5. Academic Press, Inc. https://doi.org/10.1016/B978-0-12-460515-2.50005-6.

Allen Coral Atlas. https://allencoralatlas.org/.

Alvarez-Filip, L., Carricart-Ganivet, J. P., Horta-Puga, G., & Iglesias-Prieto, R., 2013. Shifts in coral-assemblage composition do not ensure persistence of reef functionality. Scientific Reports, 3(1). https://doi.org/10.1038/srep03486.

Amini, E., Marsooli, R., & Neshat, M., 2024. A Multi-faceted Methodology for Calibration of Coastal Vegetation Drag Coefficient. Ocean Modelling, 102391. https://doi.org/10.1016/j.ocemod.2024.102391.

Baumann, J. H., Bove, C. B., Carne, L., Gutierrez, I., & Castillo, K. D., 2021. Two offshore coral species show greater acclimatization capacity to environmental variation than nearshore counterparts in southern Belize. Coral Reefs, 40(4), 1181-1194. https://doi.org/10.1007/s00338-021-02124-8.

Bouws, E., Günther, H., Rosenthal, W., & Vincent, C. L., 1985. Similarity of the wind wave spectrum in finite depth water: 1. Spectral form. Journal Of Geophysical Research, 90(C1), 975-986. https://doi.org/10.1029/jc090ic01p00975.

Brander, R. W., Kench, P. S., & Hart, D., 2004. Spatial and temporal variations in wave characteristics across a reef platform, Warraber Island, Torres Strait, Australia. Marine Geology, 207(1-4), 169-184. https://doi.org/10.1016/j.margeo.2004.03.014

Cagigal, L., Méndez, F. J., Ricondo, A., Gutiérrez-Barceló, D., Bosserelle, C., & Hoeke, R., 2024. BinWaves: An additive hybrid method to downscale directional wave spectra to nearshore areas. Ocean Modelling, 102346. https://doi.org/10.1016/j.ocemod.2024.102346

Calhoun, R.S., Field, M.E., 2002. Beach and reef-flat sediments along the south shore of Molokai, Hawaii. Proc. Carbonate Beaches 2000, American Society of Civil Engineers, pp. 163–171.

Camus, P., Méndez, F. J., & Medina, R.., 2011a. A hybrid efficient method to downscale wave climate to coastal areas. Coastal Engineering, 58(9), 851-862. https://doi.org/10.1016/j.coastaleng.2011.05.007

Camus, P., Mendez, F.J., Medina, R., Cofiño, A.S., 2011b. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coast. Eng. 58 (6), 453 – 462. https://doi.org/10.1016/j.coastaleng.2011.02.003.

Clague, D.A., Dalrymple, G.B., 1989. Tectonics, geochronology, and origin of the Hawaiian-Emperor volcanic chain. In: E.L. Winterer, D.M. Hussong, R.W. Decker (Eds.), The Eastern Pacific Ocean and Hawaii, pp. 188–217. In: The Geology of North Amercia, vol. N, Geological Society of America, Boulder, Colorado.

Clark, J. R.., 1997. Coastal zone management for the new century. Ocean & Coastal Management, 37(2), 191–216. https://doi.org/10.1016/s0964-5691(97)00052-5.

Delft University of Technology [SWASH Team]. SWASH User Manual 7.01 Version. https://swash.sourceforge.io/online_doc/swashuse/swashuse.html.

Duan, Q., Sorooshian, S., & Gupta, V. K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal Of Hydrology, 158(3-4), 265-284. https://doi.org/10.1016/0022-1694(94)90057-4.

Ferrario, F., Beck, M. W., Storlazzi, C. D., Micheli, F., Shepard, C. C., & Airoldi, L., 2014. The effectiveness of coral reefs for coastal hazard risk reduction and adaptation. Nature Communications, 5(1), 1–9. https://doi.org/10.1038/ncomms4794.

Fletcher, C.H., Richmond, B.M., Grossman, E.E., 2002. Atlas of Natural Hazards in the Hawaiian Coastal Zone. USGS Geologic Investigations Series I-2716.

Gawehn, M., van Dongeren, A., van Rooijen, A., Storlazzi, C.D., Cheriton, O.M., Reniers, A., 2016. Identification and classification of very low frequency waves on a coral reef flat. J. Geophys. Res.: Oceans 121 (10), 7560 – 7574. https://doi.org/10.1002/2016JC011834.

Harvey, B. J., Nash, K. L., Blanchard, J. L., & Edwards, D. P., 2018. Ecosystem-based management of coral reefs under climate change. Ecology and Evolution, 8(12), 6354–6368. https://doi.org/10.1002/ece3.4146.

Karim, F., & Nandasena, N., 2023. Experimental and numerical assessment of marine flood reduction by coral reefs. Natural Hazards Research, 3(1), 35-41. https://doi.org/10.1016/j.nhres.2022.12.003

Kim, T., Kwon, Y., Lee, J., Lee, E., & Kwon, S., 2022. Wave attenuation prediction of artificial coral reef using machine-learning integrated with hydraulic experiment. Ocean Engineering, 248, 110324. https://doi.org/10.1016/j.oceaneng.2021.110324.

Lee, W. K., Tay, S. H. X., Ooi, S. K., & Friess, D. A., 2021. Potential short wave attenuation function of disturbed mangroves. *Estuarine, Coastal and Shelf Science*, 248(106747), 106747. https://doi.org/10.1016/j.ecss.2020.106747.

Liu, Y., Liao, Z., Fang, K., & Li, S., 2021. Uncertainty of wave runup prediction on coral reef-fringed coasts using SWASH model. *Ocean Engineering*, 242, 110094. https://doi.org/10.1016/j.oceaneng.2021.110094

Liu, Y., Li, S., Liao, Z., Liu, Q., Zou, Q., & Liu, W., 2023. Explicit wave-runup formula for beaches fronted by coral reefs using tree-based models. *Coastal Engineering*, 183, 104308. https://doi.org/10.1016/j.coastaleng.2023.104308

Méndez, F., 2024. Teoría de Metamodelos [PowerPoint slides]. ETS Ingenieros de Caminos, Canales y Puertos, Universidad de Cantabria.

Ministry of Economy, Finance, Industrial and Digital Sovereignty, 2023. Le budget vert : Qu'est-ce que c'est ? economie.gouv.fr. https://www.economie.gouv.fr/actualites/budget-vert#.

Moberly, R.M., Chaimberlin, T., 1964. Hawaiian Beach Systems, University of Hawaii.

Nalley, E. M., Tuttle, L. J., Barkman, A. L., Conklin, E. E., Wulstein, D. M., Richmond, R. H., & Donahue, M. J., 2021. Water quality thresholds for coastal contaminant impacts on corals: A systematic review and meta-analysis. Science Of The Total Environment, 794, 148632. https://doi.org/10.1016/j.scitotenv.2021.148632.

National Oceanic and Atmospheric Administration., 2013. Topobathy Lidar DEM : Molokai. https://www.fisheries.noaa.gov/inport/item/64983.

Nortek. 2024. Technical Specifications : AWAC 600 kHz - 300 m, Generation 2. https://www.nortekgroup.com/products/awac2-600-khz.

Nunez, C., 2017. ¿Qué es el aumento del nivel del mar? National Geographic. Online. 2017. https://www.nationalgeographic.es/medio-ambiente/que-es-el-aumento-del-nivel-del-mar

Oppenheimer, M., Glavovic, B.C., Hinkel, J., van de Wal, R., Magnan, A.K., AbdElgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R.M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., Sebesvari, Z., 2022. Sea level rise and implications for low-lying islands, coasts and communities. In: Portner, ¨ H.-O., Roberts, D.C., MassonDelmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M. (Eds.), The Ocean and Cryosphere in a Changing Climate. Cambridge University Press, pp. 321–446. https://doi.org/10.1017/9781009157964.006.

Pearson, S. G., Storlazzi, C. D., Van Dongeren, A. R., Tissier, M. F. S., & Reniers, A. J. H. M., 2017. A Bayesian-Based System to Assess Wave-Driven Flooding Hazards on Coral Reef-Lined Coasts. Journal Of Geophysical Research Oceans, 122(12), 10099-10117. https://doi.org/10.1002/2017jc013204

Péquignet, A., Becker, J. M., Merrifield, M. A., & Boc, S. J. (2011). The dissipation of wind wave energy across a fringing reef at Ipan, Guam. Coral Reefs, 30(S1), 71-82. https://doi.org/10.1007/s00338-011-0719-5

Pomeroy, A., Lowe, R., Symonds, G., Van Dongeren, A., & Moore, C. (2012). The dynamics of infragravity wave transformation over a fringing reef. Journal Of Geophysical Research Atmospheres, 117(C11). https://doi.org/10.1029/2012jc008310

Price, D. M., Robert, K., Callaway, A., Lo Lacono, C., Hall, R. A., & Huvenne, V. A. I., 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. *Coral Reefs*, 38(5), 1007-1021. https://doi.org/10.1007/s00338-019-01827-3.

Ricondo, A., Cagigal, L., Rueda, A., Hoeke, R., Storlazzi, C. D., & Méndez, F. J., 2023. HyWaves: Hybrid downscaling of multimodal wave spectra to nearshore areas. Ocean Modelling, 184, 102210. https://doi.org/10.1016/j.ocemod.2023.102210

Ricondo, A., Cagigal, L., Pérez-Díaz, B., & Méndez, F. J., 2024. HySwash: A hybrid model for nearshore wave processes. Ocean Engineering, 291, 116419. https://doi.org/10.1016/j.oceaneng.2023.116419.

Rippa, S., 1999. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Adv. Comput. Math. 11 (2), 193 – 210. https://doi.org/10.1023/A:1018975909870.

Roelvink, D., Reniers, A., Van Dongeren, A., Van Thiel de Vries, J., McCall, R., & Lescinski, J., 2009. Modelling storm impacts on beaches, dunes and barrier islands. Coastal Engineering, 56(11-12), 1133-1152. https://doi.org/10.1016/j.coastaleng.2009.08.006

Roelvink, D., Dastgheib, A., Spencer, T., Möller, I., Christie, E., Berenguer, M., & Sempere-Torres, D., 2015. Resilience-increasing Strategies for Coasts–Toolkit. Improvement of Physical Processes. XBeach Improvement & Validation; Wave Dissipation over Vegetated Marshes and Flash Flood Module. Tech. Rep.(Deltares, UNESCO-IHE, UCAM, UPC-CRAHI, SURFSARA, 2015).

State of Hawaii, 2001. Map of median annual rainfall in millimeters and inches: Molokai, Hawaii. Department of Land and Natural Resources.

Stockdon, H. F., Holman, R. A., Howd, P. A., & Sallenger, A. H., 2006. Empirical parameterization of setup, swash, and runup. *Coastal Engineering*, 53(7), 573-588. https://doi.org/10.1016/j.coastaleng.2005.12.005

Storlazzi, C., 2019. Maui-area coral reefs described in detail, including threats | U.S. Geological Survey. https://www.usgs.gov/news/maui-area-coral-reefs-described-detail-including-threats.

Storlazzi, C., Ogston, A., Bothner, M., Field, M., & Presto, M., 2004. Wave- and tidally-driven flow and sediment flux across a fringing coral reef: Southern Molokai, Hawaii. Continental Shelf Research, 24(12), 1397-1419. https://doi.org/10.1016/j.csr.2004.02.010.

Tucker, D. T., 2021. Researchers test physics of coral as an indicator of reef health. Phys.org. https://phys.org/news/2021-12-physics-coral-indicator-reef-health.html.

United Nations Environment Programme, 2020. Seven ways you're connected to coral reefs. UNEP. https://www.unep.org/news-and-stories/story/seven-ways-youre-connected-coral-reefs.

U.S. Geological Survey. 2020. Waiakane, Molokai, Hawaiian Islands, wave and water level data, 2018. Science Base. https://www.sciencebase.gov/catalog/item/5d7ab280e4b0c4f70d02a48a

Winter, G., Storlazzi, C., Vitousek, S., van Dongeren, A., McCall, R., Hoeke, R., Skirving, W., Marra, J., Reyns, J., Aucan, J., Widlansky, M., Becker, J., Perry, C., Masselink, G., Lowe, R., Ford, M., Pomeroy, A., Mendez, F., Rueda, A., & Wandres, M., 2020. Steps to develop early warning systems and future scenarios of storm wave-driven flooding along coral reef-lined coasts. Frontiers in marine science, 7. https://doi.org/10.3389/fmars.2020.00199.

Zijlema, M., Stelling, G., & Smit, P., 2011. SWASH: An operational public domain code for simulating wave fields and rapidly varied flows in coastal waters. *Coastal Engineering*, 58(10), 992-1012. https://doi.org/10.1016/j.coastaleng.2011.05.015

CHySwash

Automatic Calibration of Hydrodynamic models for Fringing Reefs

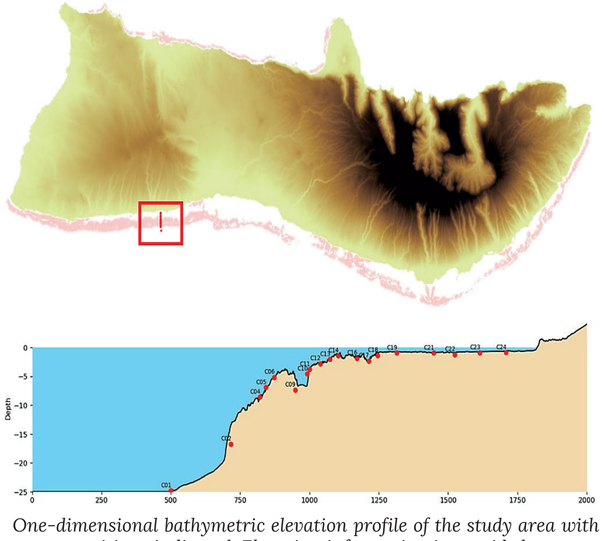
Pablo Zubía Palazuelos

Objectives

- Develop the **CHySwash metamodel**, which integrates statistical techniques with numerical simulations to efficiently predict nearshore wave dynamics while **reducing computational time**.
- Apply the CHySwash metamodel to the Molokai study area, using real sensor data to accurately reconstruct wave characteristics and **calibrate key parameters** such as **wave-breaking (*Cr*) and coral friction coefficients (*Cf*)**.
- Implement the **Shuffled Complex Evolution (SCE) optimization algorithm** to automate the calibration process, ensuring that the model finds the best combination of coefficients to match observed data and improve prediction accuracy.

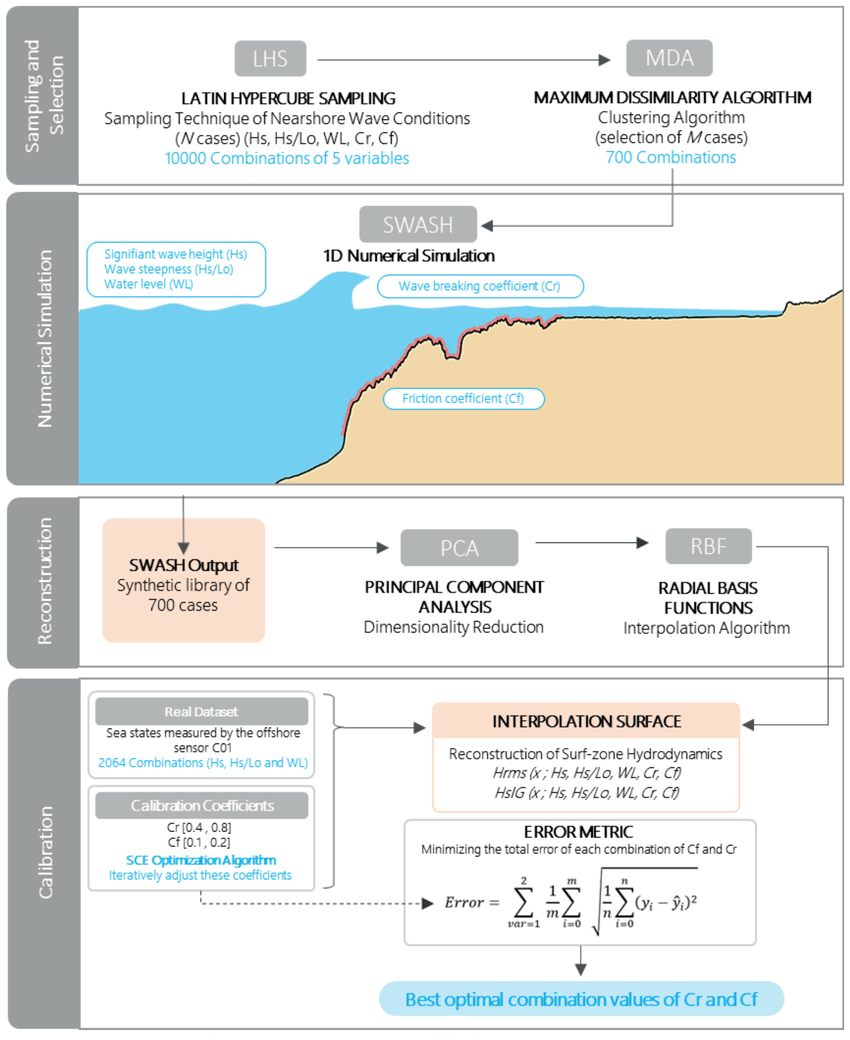
Study Area and Data

The project is located on the southern coast of **Molokai (Hawaii)**, an area with a 53 km coral reef that serves as a barrier against erosion and provides marine habitat. For calibration, data from **19 sensors** (U.S. Geological Survey, 2020) were used, which collected measurements over a total period of 2064 hours. Additionally, a **one-dimensional bathymetric profile** (National Oceanic and Atmospheric Administration, 2013) of the study area was created, essential for subsequent simulations with the SWASH numerical model.



One-dimensional bathymetric elevation profile of the study area with sensor positions indicated. Elevation information is provided at every 1.5-meter interval. The reference is the local mean sea level.

CHySwash Methodology



Flow chart of the proposed CHySwash metamodel.

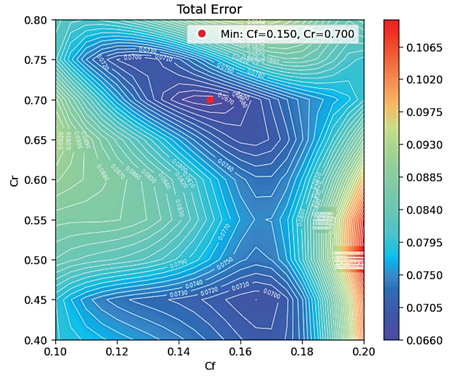
Results

Optimal combination

The optimal combination of the numerical model calibration coefficients is found :

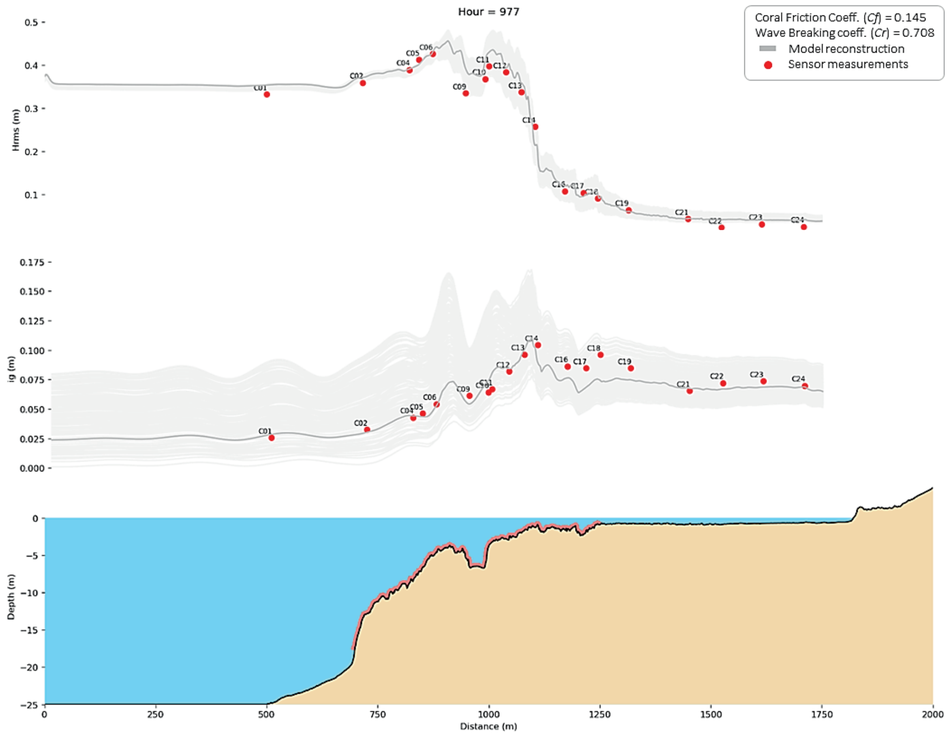
$Cf = 0.145$
$Cr = 0.708$

which minimizes the defined error metric. This combination of coefficients **best represents the hydrodynamic conditions** in Molokai.



Spatial Evolution of Output Variables

The metamodel reconstructions for *Hrms and HsIG*, using the optimal combination of coefficients, closely align with sensor measurements, demonstrating **high accuracy** in the results.



Reconstruction of the Hrms variable for the optimal coefficient combination (Cf = 0.145, Cr = 0.708) along the studied 1D profile. The reconstructed variable (dark gray line) is shown along with the actual sensor measurements (red points).

Conclusions

- The project **successfully developed a hybrid metamodel** capable of automatically calibrating hydrodynamic models, particularly for coastal areas with fringing coral reefs.
- The methodology combines sampling, clustering, interpolation, and optimization tools, making it **adaptable** to various numerical models and engineering challenges.
- The metamodel proved effective in predicting the **optimal friction and wave-breaking coefficients** for the Molokai coastal area, accurately reflecting real-world conditions.
- The hybrid technology **drastically reduces simulation time**, enabling the reconstruction of large datasets.
- Future studies should expand the calibration to other coastal areas, conduct 2D analyses, integrate climate change impacts, and apply the metamodel to other coastal phenomena.

References

- *U.S. Geological Survey. 2020. Waiakane, Molokai, Hawaiian Islands, wave and water level data, 2018. Science Base.*
- *National Oceanic and Atmospheric Administration. 2013. Topobathy Lidar DEM : Molokai.*