



CHALMERS
UNIVERSITY OF TECHNOLOGY

Predicting sea surface wave and wind parameters from satellite radar images using machine learning

Master's thesis in Computer science and engineering

Filip Borg

Axel Brobeck

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

MASTER'S THESIS 2023

Predicting sea surface wave and wind parameters from satellite radar images using machine learning

Filip Borg
Axel Brobeck



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Predicting sea surface wave and wind parameters from satellite radar images using machine learning

Filip Borg

Axel Brobeck

© Filip Borg & Axel Brobeck, 2023.

Supervisor: Anis Elyouncha, Space, Earth and Environment

Supervisor: Adrià Amell, Space, Earth and Environment

Examiner: Leif Eriksson, Space, Earth and Environment

Master's Thesis 2023

Department of Space, Earth and Environment

Chalmers University of Technology

SE-412 96 Gothenburg

Telephone +46 31 772 1000

Typeset in L^AT_EX
Gothenburg, Sweden 2023

Predicting sea surface wave and wind parameters from satellite radar images using machine learning

Filip Borg

Axel Brobeck

Department of Space, Earth and Environment

Chalmers University of Technology

Abstract

Accurate predictions of wave and wind parameters over oceans are crucial for various marine operations. Although buoys provide accurate measurements, their deployment is limited, which necessitates the exploration of alternative data sources. Sentinel-1, a satellite mission capturing Synthetic Aperture Radar (SAR) images with high coverage, presents a promising opportunity. However, establishing the relationship between SAR images and wave/wind parameters is not straightforward. This project aims to develop a machine learning model that can effectively extract this relationship.

To accomplish this, data from all available buoys measuring significant wave height and wind speed in the year 2021 were utilized. The corresponding SAR images were located, and $2\text{ km} \times 2\text{ km}$ sub-images were extracted around each buoy. From each sub-image, a set of features were extracted. These sub-images and features served as input to train machine learning models capable of predicting buoy measurements, supplemented with model data as necessary.

The project presents two final deep learning models: one utilizing only the extracted features and another employing both the sub-images and features. These multi-class regression models simultaneously predict significant wave height and wind speed. The model using only features achieved a Root Mean Square Error (RMSE) of 0.553 m for significant wave height and 1.573 m/s for wind speed. The model incorporating both sub-images and features achieved an RMSE of 0.459 m for significant wave height and 1.658 m/s for wind speed.

The code for the project can be found on <https://github.com/SEE-GEO/sarssw>.

Keywords: Machine Learning, Computer Vision, Synthetic Aperture Radar, Significant Wave Height, Wind Speed, Radar, Master Thesis, Chalmers University of Technology

Acknowledgements

We would like to thank our supervisors Anis Elyouncha and Adrià Amell for their valuable feedback and insightful discussions.

Leif Eriksson also deserves our thanks for his dedicated service as our examiner.

We would also like to thank ESA and their Copernicus program for making the ERA5 [1] model data and the Sentinel-1 SAR data available.

The Copernicus [2] Sentinel-1 data for this project has been retrieved from NASA's Alaska Satellite Facility Distributed Active Archive Center (ASF DAAC) [3], processed by ESA.

This study has been conducted using E.U. Copernicus Marine Service Information; Global Ocean- In-Situ Near-Real-Time Observations [4] & Global Ocean Hourly Reprocessed Sea Surface Wind and Stress from Scatterometer and Model [5].

Filip Borg & Axel Brobeck, Gothenburg, 2023-09-10

Contents

List of Figures	xi
List of Tables	xiii
List of Abbreviations	xv
1 Introduction	1
1.1 Aim	2
1.2 Objectives	2
1.3 Limitations	2
2 Theory	5
2.1 Significant Wave Height and Wind Speed Interactions	5
2.2 Synthetic Aperture Radar and Sentinel-1	5
2.3 Data Sources	6
2.3.1 Buoy Data	6
2.3.2 ERA5 Significant Wave Height Model	7
2.3.3 Copernicus Wind Speed Model	7
2.4 Machine Learning	8
2.4.1 Traditional Machine Learning Algorithms for Regression	8
2.4.2 Deep Learning	9
2.4.2.1 Neural Networks and Fully Connected Layers	9
2.4.2.2 Convolutional Neural Networks	10
2.4.3 Metrics Used for Evaluation	10
3 Methods	13
3.1 Data Pipeline	13
3.1.1 Collocating SAR-images with Buoy Data	13
3.1.2 Buoy Wind Speed Measurement Height Adjustment	15
3.1.3 Model Data Assessment	16
3.1.3.1 Significant Wave Height Model	16
3.1.3.2 Wind Speed Model	16
3.1.4 Merging Buoy and Model Data	17
3.1.5 Extracting Sub-images from SAR-Images	17
3.1.6 Features	18
3.1.6.1 Calculation of the Azimuth Cutoff Wavelength	19

3.1.7	Homogeneity Filter	19
3.1.8	Data Split	25
3.1.9	Final Dataset	25
3.2	Machine Learning Models	25
3.2.1	Traditional Model Baseline	27
3.2.2	Multi-Task Learning and Loss Function	27
3.2.3	Model Structure for Only Features	28
3.2.4	Extending Model to Use Sub-images	29
4	Results	31
4.1	Metrics for Models	31
4.2	Correlation Plots from Target to Prediction	32
5	Discussion	35
5.1	Comparison with Similar Studies	35
5.2	Trustworthiness of the Results	36
5.3	Limitations in Quality of the Data	37
5.4	Improvements of the Homogeneity Filter	37
5.5	Apparent Clipping of Small Values	38
5.6	Comparison of the Models	38
5.7	Future Work	39
6	Conclusion	41
	Bibliography	43
A	Training Progress	I
B	Overlapping Histograms of Wind Speed Distributions	III

List of Figures

3.1	The four main clusters of buoys collocated with at least one SAR-image.	14
3.2	Correlation of wave height model and buoy data.	16
3.3	Overlapping histograms of the significant wave height from the buoy data separated by the interpolatability of the model. The model interpolatable data has a mean of 1.580 m and non-interpolatable data has a mean of 1.188 m	17
3.4	Correlation of wind speed model and buoy data.	18
3.5	Gaussian fit of ACF.	20
3.6	Overlapping histograms comparing the distributions homogeneity, dissimilarity, and correlation for different types of sub-images. The blue color represents sub-images which had no land in them, according to the land mask. These are similar to what sub-images looks like without the homogeneity filter, and should be mostly homogeneous. The yellow images are gathered from coordinates of wind farms. These should represent something non homogeneous that should be removed.	22
3.7	Sub-images removed by the homogeneity filter.	23
3.8	Sub-images kept by the homogeneity filter.	24
3.9	Overlapping histograms of the distributions of significant wave height values from model and buoy in the final dataset. Buoy mean is 1.47 m/s and model mean is 1.42 m/s.	26
3.10	Overlapping histograms of the distributions of wind speed values from model and buoy in the final dataset. Buoy mean is 7.14 m/s and model mean is 6.62 m/s.	26
3.11	Illustration of the features-only model. The number in each block refers to the number of neurons in that layer. In addition, batch normalisation, dropout and ReLU activation is applied after each layer.	29
3.12	Illustration of the images-features model. The sub-image is fed through ResNet50 and is reduced to a vector, which is then concatenated with the features and fed through the same architecture as for only features. This results in the two separate heads for wind and wave as during the features-only network.	30
4.1	Heatmaps for significant wave height predictions made on the test set by the two deep learning models.	33
4.2	Heatmap of wind speed predictions made on the test set data by the two deep learning models.	34

A.1	Logs of metrics during training progress. Purple is from the model using features only. Orange is from the model that use both features and sub-images. The title of each plot gives the value that is tracked, which is the y-axis. The x-axis represent the number of steps that has passed during the training progress.	I
B.1	Overlapping histograms of the distributions of wind speed values from both model and buoy in the final dataset. The same data as in Figure 3.10 but with only 30 bins.	IV

List of Tables

3.1	Distribution of SAR-image swath mode and polarisation for the collocations. IW swath and VV, VH polarisation constitute 92.7% of the data.	15
3.2	Distribution of the buoy parameters of the collocations. Note that many SAR-images were collocatable with buoys that measured multiple values and is why the collocations sum to 79 884.	15
3.3	Data sources of the final dataset.	25
4.1	Wave and wind RMSE for traditional regression models not based on deep learning.	31
4.2	Performance metrics for both the model using only features and the one using sub-images and features.	32

List of Abbreviations

ACF	Autocorrelation Function
AI	Artificial Intelligence
ASF DAAC	Alaska Satellite Facility Distributed Active Archive Center
ASM	Angular Second Moment
CMOD	Category of GMFs predicting wind speed
ECMWF	European Centre for Medium-Range Weather Forecasts
EW	Extra Wide Swath
GLCM	Gray Level Co-occurrence Matrix
GMF	Geophysical Model Function
GRD	Ground Range Detected
IW	Interferometric Wide Swath
MAE	Mean Absolute Error
MSE	Mean Square Error
PSD	Power Spectral Density
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
SAR	Synthetic Aperture Radar
SVC	Support Vector Classifier
SWH	Significant Wave Height
VAVH	Variable measuring SWH
VH	Vertical send Horizontal receive polarisation
VHM0	Variable measuring SWH
VV	Vertical send Vertical receive polarisation
WSPD	Wind Speed

List of Abbreviations

WV Wave Swath

1

Introduction

Ocean and atmospheric conditions significantly influence numerous human activities, particularly those related to the marine environment. The ability to accurately measure and predict related parameters is invaluable to several stakeholders operating in maritime spaces, from the renewable energy sector to maritime transportation, from coastal infrastructure planning to environmental conservation. The extraction of key parameters such as wave height and wind speed allows for a range of applications, including the assessment of potential locations for renewable energy extraction, maritime safety evaluation, optimisation of shipping routes, sediment movement analysis, and the planning of coastal infrastructure [6–8].

Historically, several techniques have been utilised to measure these parameters. Buoys placed in-situ, satellite-borne radio altimeters, and weather balloons have all provided useful data [9]. However, the advent of satellite instruments, particularly Synthetic Aperture Radar (SAR), offers a significant advancement. With the ability to provide global coverage, the potential to extract up-to-date wave and wind parameter data from SAR images is a promising avenue for enhancing our understanding of the marine environment.

Early research on methods to predict wind speed using data from scatterometers has been conducted and widely adopted. These studies include CMOD4 [10], CMOD5 [11], CMOD5N [12], and CMOD7 [13]. All of which formulate Geophysical Model Functions (GMFs) based on the relationship between backscatter intensity and wind speed. One after another they improve upon the earlier GMFs by incorporating new empirical observations and ideas. Similar GMFs has been developed for SAR and is benefiting from the higher spatial resolution compared to scatterometers [14]. These GMFs include C-SARMOD [15], and C-SARMOD2 [14].

The rapid advancements in Artificial Intelligence (AI) technology have further enriched the field of SAR image analysis. AI methodologies have been successfully applied to determine the optimal areas for wind energy harvesting from SAR data [7]. A notable accomplishment was made using SAR images from the Chinese satellite Gaofen-3, where the application of AI outperformed the CMOD7 GMF [16].

The application of SAR data to measure wave parameters has similarly benefited from technological progression. Early work focused on algorithms for extracting two-dimensional wave spectra [17–19]. Over time, more empirical algorithms such as CWAVE_ERS [20], CWAVE_ENV [21], and CWAVE_S1A [22] were developed,

which do not require a wave spectrum prior. AI has been pivotal in this progression as well, with machine learning methods such as Support Vector Classifiers (SVCs) [23], as well as deep learning techniques [6, 8], including Neural Networks (NNs) [22, 24], and Convolutional Neural Networks (CNNs) [25, 26]. The method developed by Quach et al. [8] was found to be so precise that it was integrated into the data suite for the Sentinel-1 satellite [27].

1.1 Aim

The aim of this project is to explore the possibility to use SAR images collected globally and features calculated from these to extract both wind and wave parameters simultaneously using a machine-learning approach. The output will be evaluated with in-situ data and the results compared with existing state-of-the-art methods.

1.2 Objectives

More specifically, during the project, the following questions and tasks will be considered:

- How to utilise the right tools to handle the large data files in regards to efficient handling, storage, and training of the models.
- Experiment with data filterings such as homogeneity filter, land, and ice masks.
- Experimenting with different approaches for the machine learning models to get a feel for what works and what does not work.
- Explore ways to combine the two prediction models of wave and wind parameters and train them together.

1.3 Limitations

The work will be limited to be focused on machine learning models and data handling for solving this problem and will not consider further development of existing GMFs.

The data utilised for this study is primarily drawn from SAR-images provided by the European Space Agency's (ESA) Sentinel-1 mission. Specifically, images from the Sentinel-1A and Sentinel-1B satellites in the Ground Range Detected (GRD) format were used. The study employs data captured in the Interferometric Wide (IW) swath mode and uses the VV and VH polarisations. An understanding of the concepts related to SAR can be found in Section 2.2. The specific reasons for choosing these settings are further explained in Section 3.1.1. Importantly, this research was limited to data from the year 2021, with Sentinel-1B ending its service at the close of that year [28].

Only data accompanying the SAR-images will be used as input in the models. No additional sources of input like sea floor depth or historical knowledge about usual

wave or wind behavior will be considered.

2

Theory

2.1 Significant Wave Height and Wind Speed Interactions

Significant Wave Height (SWH), in this report also referred to as wave height, and wind speed are key parameters for the understanding of oceanic and atmospheric interaction. SWH is a commonly used measure to describe the wave environment. It is defined as the average of the highest one-third of waves in a wave record. In other words, it is the mean wave height of the largest third of waves in a specific time period [29]. Wind Speed, on the other hand, is the speed at which air is moving horizontally past a given point, in this project assumed to be at height 10 m above the sea surface.

The interaction between wind and waves is a fundamental aspect of the Earth's oceanic and atmospheric systems. Wind generates ocean surface waves that can travel thousands of kilometres. This phenomenon occurs because wind provides an input of energy to the sea surface, causing the water to move in a circular motion and leading to the formation of waves [30]. The speed, duration, and fetch (the unobstructed distance over which the wind blows) of the wind determines the size and type of waves produced [31]. With increasing wind speed, the wave height also increases.

2.2 Synthetic Aperture Radar and Sentinel-1

Synthetic Aperture Radar (SAR) operates on the principles of radar by emitting a signal and measuring the backscattered (reflected) signal. The main advantage of SAR is its ability to produce high-resolution images irrespective of weather conditions or daylight availability [32]. The term “synthetic” in SAR refers to the practice of artificially enlarging the antenna length using data processing techniques. As a satellite with SAR moves along its path, it captures a series of radar signals over a certain period of time. These separate captures are then combined into a single large-scale observation, mimicking the output of a significantly larger antenna, hence achieving greater spatial resolution [33].

SAR sensors typically transmit linearly polarised signals, labelled as horizontal (H) or vertical (V). The combination VV, for example, signifies that the SAR sensor

transmitted a vertically polarised signal and also received the reflected signal in the same vertical polarisation. Different combinations of transmitted and received polarisations correspond to distinct scattering types: rough surface, volume, and double bounce. In ocean studies, VV polarisation is particularly informative as it is sensitive to surface roughness, enabling the capture of wave and wind patterns. Variations in these scattering signals can occur with changing signal wavelength due to differences in penetration depth [33].

The Sentinel-1 mission [34], a key component of the European Union’s Copernicus Programme, carry an all-weather, day-and-night imaging radar that utilises the C-band frequency. Offering comprehensive geographical coverage and rapid data dissemination, it has a significant operational role in marine monitoring, land monitoring, and emergency services. The mission systematically captures images of all global landmasses, coastal zones, European shipping routes, and oceans.

Among its primary acquisition modes are the Interferometric Wide swath (IW) [35] and Extra Wide swath (EW) [36]. IW mode is mostly used over land, capturing data across a 250 km swath with a 5 m×20 m resolution using the Terrain Observation with Progressive Scanning SAR (TOPSAR) technique. This technique ensures continuous coverage with uniform quality throughout the swath. EW mode, on the other hand, extends its use over sea-ice, polar zones, and maritime areas with a wider 400 km swath at a 20 m×40 m resolution.

Ground Range Detected (GRD) [37] Level-1 data represents SAR data that has been multi-looked and projected onto ground range, offering detected amplitude information with reduced speckle at the cost of slightly compromised spatial resolution.

2.3 Data Sources

This thesis primarily relies on four sources of data. SAR images are taken from Sentinel 1, as described in Section 2.2. The other three sources pertain to wave and wind measurements. These include in-situ measurements in the form of buoy data, which is described in Section 2.3.1, as well as two distinct predictive models outlined in Sections 2.3.2 and 2.3.3.

2.3.1 Buoy Data

The Global Ocean- In-Situ Near-Real-Time Observations [4] dataset from Copernicus is a comprehensive collection of in-situ measurements, encompassing a wide range of oceanic parameters. It includes data on physical properties, such as wind speed, significant wave height, wave period and direction, temperature, salinity, and current velocities, as well as biogeochemical characteristics.

This dataset is collected from seven regions together forming a dataset with data from all oceans allowing for a comprehensive analysis of different oceanic environments. The data acquisition is carried out by a handful of independent actors and is then compiled and validated to form a common dataset. The measurements are done with instruments deployed on various platforms, including research vessels, fixed

buoys, and floats. These instruments are equipped with sensors capable of measuring the targeted parameters as well as depths and location to include detailed meta-information for each datapoint. As different instruments of the same type carry different types of sensors this is also reflected in the dataset.

VAVH, VHM0, and WSPD are some of the possible variables found in the fixed buoy data. VAVH is measured as the average peak to trough of the highest 1/3 of the waves recorded during a burst [38]. VHM0 is a spectral analysis that approximates the significant wave height as four times the square root of the first moment of the wave power spectrum [39]. WSPD measures the sustained absolute speed of the wind parallel to the ground and does not contain any information regarding direction [40].

2.3.2 ERA5 Significant Wave Height Model

The ERA5 hourly data on single levels from 1940 to present [1] is a vast and comprehensive collection of climate reanalysis data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The dataset offers hourly climate-related quantities for a range of parameters for atmospheric, ocean and land related phenomena. These quantities include, among others, significant wave height, wave direction, temperature, humidity, wind speed and direction, precipitation and cloud cover. The data is supplied in a long-lat grid format with a resolution of $0.25^\circ \times 0.25^\circ$ for atmospheric and $0.5^\circ \times 0.5^\circ$ for ocean related parameters.

This dataset is generated through assimilation, a process that combines predictions from advanced atmospheric models with diverse observational data sources. This assimilation technique enhances the reliability of global predictions, particularly for parameters with limited observational coverage.

2.3.3 Copernicus Wind Speed Model

The Global Ocean Hourly Reprocessed Sea Surface Wind and Stress from Scatterometer and Model [5] consists of hourly Level-4 (L4) ocean surface wind products generated by the Royal Netherlands Meteorological Institute. These products encompass global gridded near real-time and multi-year L4 sea surface wind and stress variables at a standard height of 10 m in a long-lat spatial resolution of 0.125° .

This product was created after the observation of systematic biases in the ECMWF ERA5 models wind parameters (Section 2.3.2) as identified by Belmonte Rivas and Stoffelen [41]. The corrections are made in a method similar to Trindade *et al.* [42] where the mean biases of scatterometer measurements compared to ERA5 is calculated in vector form over several days. These biases are then added to the ERA5 model fields to generate bias-corrected parameters. The scatterometer data are the ASCAT instruments onboard Metop-A, Metop-B and Metop-C and the SeaWinds instrument carried by the QuikSCAT satellite.

2.4 Machine Learning

Machine learning is a topic with varying definitions. For the purpose of this study, in its essence machine learning is about creating algorithms that learn to solve some task by finding patterns in data. There are various kinds of machine learning depending on the task, like reinforcement learning for developing policies for agents and unsupervised learning for finding patterns in unlabeled data. This study aims to find an algorithm that can relate data from SAR images to the corresponding wave height and wind speed. The relevant class of machine learning for this is called supervised learning, which deals with data where the answer that the algorithm should predict is already known. Specifically, the task for this project is a regression task since the outputs are continuous values, as opposed to a fixed set of classes.

2.4.1 Traditional Machine Learning Algorithms for Regression

There are multiple models for solving regression tasks using machine learning. Most state of the art models today use deep learning [43]. Sometimes it is however useful to investigate more traditional models as they can be faster to train and serve as a good baseline, or even outperform more advanced models.

Linear regression is a fundamental algorithm in machine learning that assumes a linear relationship between the input variables and the output variable [44, Chapter 3]. It is simple and provides a useful prediction in various scenarios. However, in the presence of multicollinearity (i.e., when input features are highly correlated) and overfitting issues, different versions of linear regression, such as Ridge and Lasso regression, become useful [45]. Ridge Regression, with its L2 regularization, helps handle overlapping information by slightly adjusting the importance of each variable, ensuring no single variable overly dominates the prediction. Lasso, through L1 regularization, can entirely remove some variables if they're not deemed crucial, which is beneficial when working with numerous possibly redundant variables. Elastic Net combines both approaches, making it versatile when the data has many correlated variables [46].

In addition to linear regression, there are several decision tree-based regression algorithms. Decision Tree Regressor splits the data into subsets based on feature values, creating a tree that provides non-linear predictions. However, a single tree can often lead to overfitting. Random Forest Regressor, an ensemble method, overcomes this issue by aggregating predictions from a multitude of decision trees, leading to a more generalised model [47]. Gradient Boosting Regressor builds trees sequentially, with each new tree aiming to correct the errors of the previous one [48]. XGBoost Regression is an optimised distributed gradient boosting library designed to be highly efficient, flexible, and portable, providing a significant computational speed and model performance advantage [49].

2.4.2 Deep Learning

Deep learning is a class of machine learning that employs algorithms with multiple layers of computations, commonly known as hidden layers, to construct complex relationships between inputs and outputs [43]. A deep learning model is constructed by connecting these layers with non-linear functions, or activation functions, in between. Each layer has a set of trainable parameters, which can modify the relationship between its input and output.

The architecture of the model is designed such that the output layer has the same shape as the target prediction. A loss function is then applied to this output layer, quantifying the discrepancy between the model's predictions and the actual target values. The objective is to adjust the parameters to minimise the total loss, making this an optimisation problem.

The entire model is constructed to be differentiable, which allows for the application of a technique called backpropagation. Backpropagation, in tandem with an optimisation algorithm like gradient descent, calculates the gradient of the loss function with respect to the model's parameters. The parameters are then iteratively updated in the direction that minimises the loss.

While a deep learning model is theoretically capable of approximating any arbitrarily complex relationship, its practical performance is dependent on several factors. These include the availability of sufficient and representative training data, the selection of an appropriate model architecture, and the careful tuning of hyperparameters.

2.4.2.1 Neural Networks and Fully Connected Layers

Neural networks are foundational structures in deep learning, designed to process and learn from data in ways reminiscent of the human brain. At their core, neural networks consist of interconnected nodes or “neurons.” These artificial neurons do not replicate the full complexity of human brain cells, but they are inspired by the general idea of processing and passing on information.

In a typical setup, neurons receive data, process it through a weighted sum operation, and then apply an additional transformation known as an activation function. One common activation function is the Rectified Linear Unit (ReLU), which outputs the input if it is positive and zero otherwise. This helps the network capture and model nonlinear relationships in the data. As these layers of processing stack and interact, the network can recognise patterns and relationships within the input data, eventually leading to the formation of a decision boundary. This decision boundary acts as a separator, distinguishing between different classes or categories within the data.

Among various configurations of neural network layers, the “fully connected” layer is particularly noteworthy. In such a layer, each neuron is connected to every neuron from the previous layer, meaning all available information from one layer is passed to the next. This complete linkage is especially useful when no piece of data should be overlooked, making these layers common in the final stages of

many network architectures. However, while effective, they can be computationally intensive, especially when dealing with large datasets or high-resolution inputs, such as images.

2.4.2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [50] are a specialised type of neural network designed to handle grid-like data, such as images [43]. Instead of using fully connected layers, CNNs utilise kernel convolutions. This involves striding a kernel (a small matrix of weights) across the input. These kernels are not exactly the same as convolutions in other fields of engineering, but serve a similar purpose in their operation. The weights within the kernel are learnable parameters that are optimised to minimise a loss function during training. This approach enables parameter sharing across different parts of the image and reduces the number of parameters, focusing only on one neighborhood of pixels at a time [43]. Most CNNs also employ pooling, a technique that reduces the size of the input by aggregating a region of values into a single value. This can be achieved by taking the maximum value from each region, a method known as max pooling.

CNNs are a central part of modern machine learning, often employing pre-built architectures such as ResNet [51] and VGG [52], which are readily available in popular deep learning libraries. ResNet, developed by Microsoft Research, utilises residual blocks with shortcut connections to facilitate gradient flow and thus enable the training of much deeper networks. On the other hand, VGG, developed by the University of Oxford, is recognised for its simplicity, utilising multiple layers of 3×3 convolutions to achieve excellent performance in large scale image recognition tasks. Both these architectures, having been trained on massive datasets like ImageNet [53], are frequently repurposed in a technique known as transfer learning. In transfer learning, these pretrained networks are extended for specific new tasks. Instead of starting from scratch, knowledge is essentially transferred from one task (like recognising a million ImageNet images) to another (like identifying a particular set of objects or features).

2.4.3 Metrics Used for Evaluation

In these equations, n is the number of observations, y_i and \hat{y}_i are the actual and predicted values for sample number i , and \bar{y} as well as $\bar{\hat{y}}$ denote the mean of the actual and predicted values, respectively.

The Mean Absolute Error (MAE) is defined as the average of absolute differences between the predicted and actual values, and is given by:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

The Mean Squared Error (MSE) is defined as the average of the squares of the differences between the predicted and actual values, and is given by:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

The Root Mean Square Error (RMSE) is the square root of MSE, and is given by:

$$\text{RMSE} = \sqrt{\text{MSE}}$$

The bias of a model, a measure of prediction errors, is defined as:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$$

Slope is obtained by fitting a linear regression model on the predicted values as a function of the actual ones, and can be calculated as:

$$\text{Slope} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

The correlation coefficient represents the strength and direction of a linear relationship between two variables. It ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship. Given the context, it can be defined as:

$$\text{Correlation} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \times \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}$$

3

Methods

This chapter covers the two main parts of the project. Section 3.1 describes the data pipeline, and Section 3.2 explains the machine learning approaches used.

3.1 Data Pipeline

The machine learning models utilised data from four primary sources: SAR images from Sentinel-1, buoy data from 2021, and model data for both wave and wind parameters for that year. The process from raw data to input suitable for the machine learning models is termed the “data pipeline.” This section provides an overview of each component within this pipeline and the considerations for their integration. Subsequent sections will delve deeper into the specifics of each component.

The first step of the pipeline was collocating the SAR-images with the buoy data (Section 3.1.1), where for each measurement from a buoy the SAR-image closest in time was located. For each of these SAR-images, a small sub-image was extracted around each buoy measurement (Section 3.1.5). From these sub-images, a number of additional features were extracted (Section 3.1.6). A filter was then applied to all sub-images that flags for if there was any inhomogeneity in the image, like a boat or ice (Section 3.1.7). After this, the final dataset for the machine learning could be compiled (Section 3.1.9), which included the sub-images as well as a table of features and the corresponding labels.

Special care had to be taken for the labels since some buoys only measure wind or wave. Buoys offer the most precise data, but in order to enable predicting both parameters at the same time, the model values were used to supplement missing buoy measurements (Section 3.1.3).

3.1.1 Collocating SAR-images with Buoy Data

From the world-wide buoy data described in Section 2.3.1, the data from the year 2021 was extracted. Specifically, the parameters VHM0 and VAVH for significant wave height and the parameter WSPD for wind speed were extracted. The measurements were also filtered to exclude data points with a minimum euclidean distance to land of 0.01° .

This data was then collocated with the SAR-images from 2021 found on Alaska

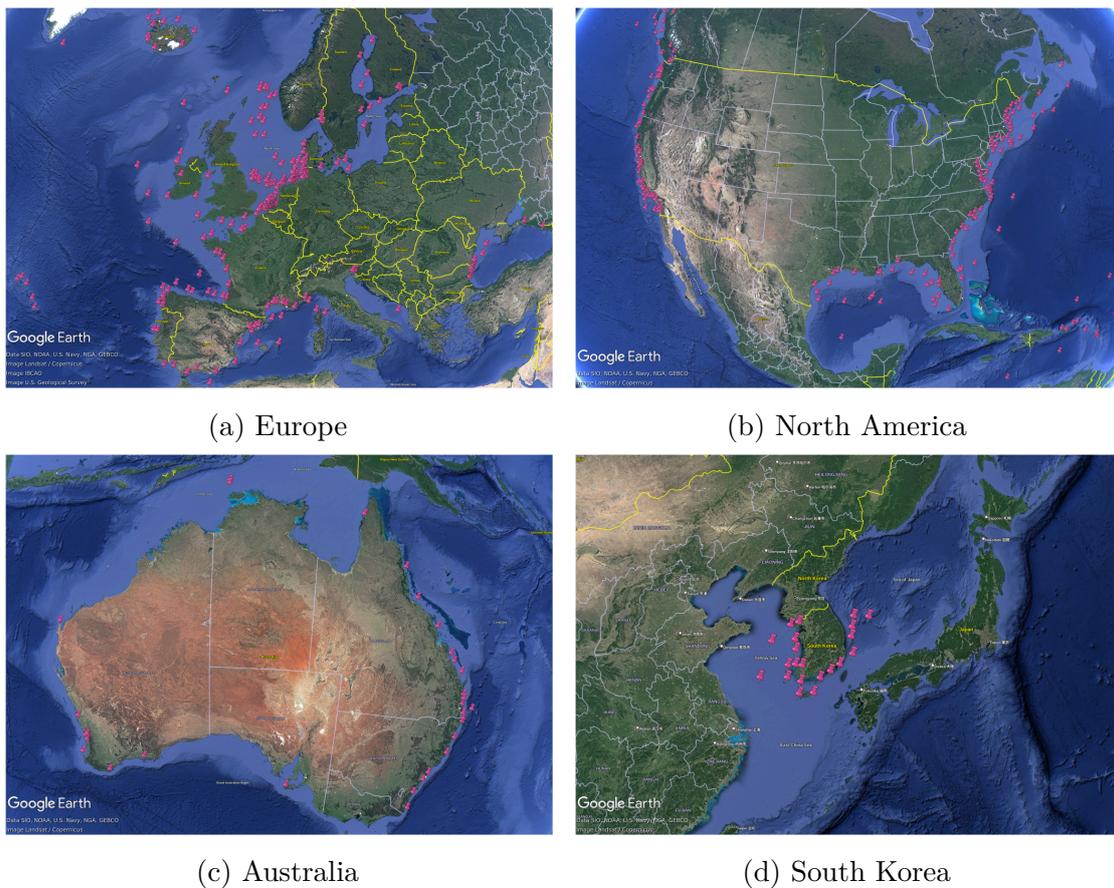


Figure 3.1: The four main clusters of buoys collocated with at least one SAR-image.

Satellite Facility’s Distributed Active Archive Center (ASF DAAC) [3] using their Python module *asf_search* [54] for accessing the Sentinel-1 data. The SAR-images were saved in GRD format for swaths IW and EW and all possible polarisations. The collocation process then saved, for each spatially overlapping SAR-image and buoy, the image itself and the measurement from the buoy temporally closest to the midpoint of the SAR-image acquisition up to a maximum difference of one hour. This collocation technique would, for SAR-images containing multiple buoys, possibly save one collocation for each buoy.

Figure 3.1 shows the geographical locations of the buoys collocated with at least one SAR-image with the four main clusters of buoys. There are some buoys not shown scattered across the rest of the globe.

For the SAR-images from 2021 22 798 were collocatable with buoy data and resulted in a total of 53 851 collocations since some SAR-images contained multiple buoys. The distribution of the collocations in regards to swath and polarisation is shown in Table 3.1 and in regards to the parameter types in Table 3.2.

The imbalance in regards to the distribution of swath and polarisation as seen in Figure 3.1 was the motivation for the limitation to only use IW swath and VV, VH polarisation since the addition of the other swath, polarisation combinations would

Swath	Polarisation	Number of collocations
IW	VV, VH	49 893
EW	HH, HV	3331
EW	VV, VH	333
IW	HH, HV	246
IW	HH	48

Table 3.1: Distribution of SAR-image swath mode and polarisation for the collocations. IW swath and VV, VH polarisation constitute 92.7% of the data.

Buoy variable name	Number of collocations
VHM0	36 844
WSPD	22 137
VAVH	20 903

Table 3.2: Distribution of the buoy parameters of the collocations. Note that many SAR-images were collocatable with buoys that measured multiple values and is why the collocations sum to 79 884.

make the data more inhomogeneous while only adding an insignificant amount of data.

3.1.2 Buoy Wind Speed Measurement Height Adjustment

Height adjustment of wind speed parameters is an essential process when working with buoy data measured at different heights as wind speed can vary significantly as a function for height due to factors like earth’s surface structure and friction [55, p. 40]. The wind profile power law [56, 57] is a well-known empirical formula used to adjust wind speed measurements to a common reference height. This adjustment enables more accurate comparisons and integrations of wind data collected from different sources and heights, thereby improving the quality of subsequent analyses.

The Wind profile power law assumes that the wind speed increases with height following a power-law distribution as it relates the wind speed at one height to the wind speed at another height via the Equation (3.1), where u_2 is the wind speed at height z_2 , u_1 is the wind speed at height z_1 , and α is the wind profile exponent, in this case 0.11 taken from [57] to reflect the open ocean condition. By using this law, wind speed measurements from buoys have been adjusted to a standardised height of 10 m, the same as the reference height of the wind speed model described in Section 2.3.3.

$$u_2 = u_1 \left(\frac{z_2}{z_1} \right)^\alpha \quad (3.1)$$

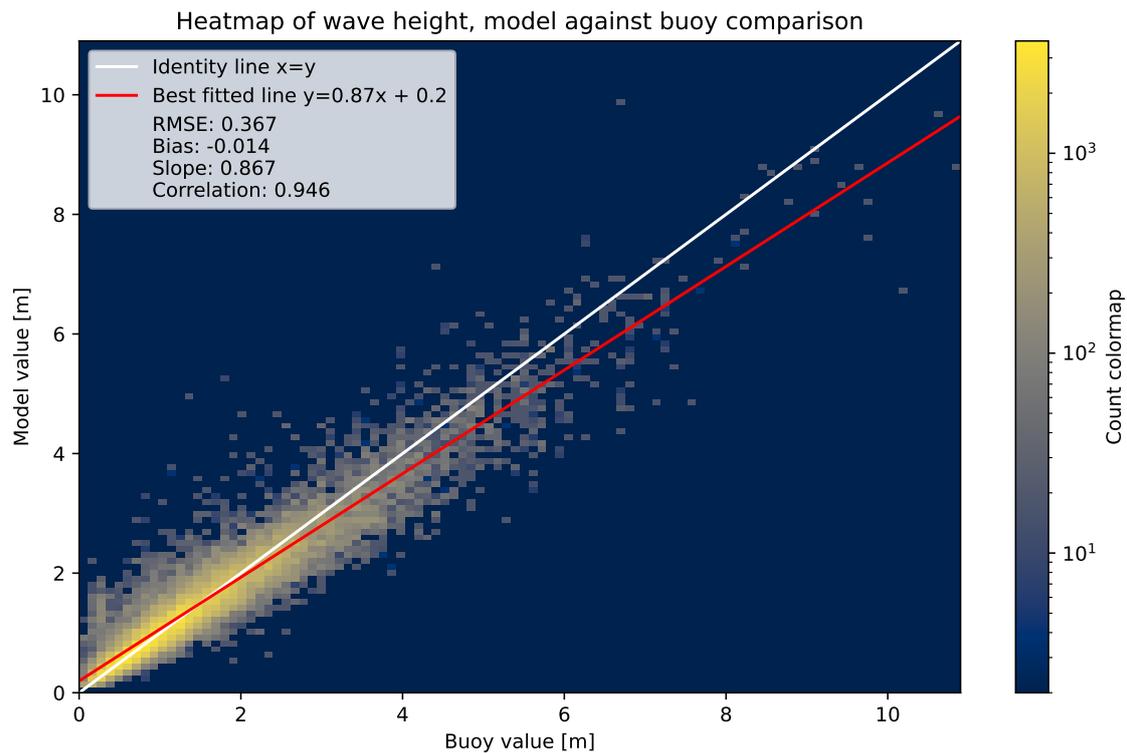


Figure 3.2: Correlation of wave height model and buoy data.

3.1.3 Model Data Assessment

This chapter shows the data analysis done for the models providing significant wave height and wind speed prediction in preparation to extending the existing buoy measurements from the collocation to form a complete dataset.

3.1.3.1 Significant Wave Height Model

For all the buoy measurements for the variables VHM0 and VAVH found during the collocation step detailed in Section 3.1.1, overlapping predictions from the wave height model described in Section 2.3.2 were extracted and compared. The heatmap of the correlation between model and buoy values is shown in Figure 3.2. Between the model and buoy data the RMSE is 0.367 m and the bias is -0.014 m. Note that the model lacks full geographical coverage near the shore, as 46.7% of the buoy data points could not be linearly interpolated due to undefined values in the model. The distribution of these two separate datasets, interpolatable and non-interpolatable buoy data, can be seen in Figure 3.3 where the model interpolatable buoy data has a mean of 1.580 m while non-interpolatable data has a mean of 1.188 m, 24.8% lower than interpolatable values.

3.1.3.2 Wind Speed Model

The wind speed model used to complement the wave height buoy data where wind speed measurements are missing is described in Section 2.3.3. In Figure 3.4 the

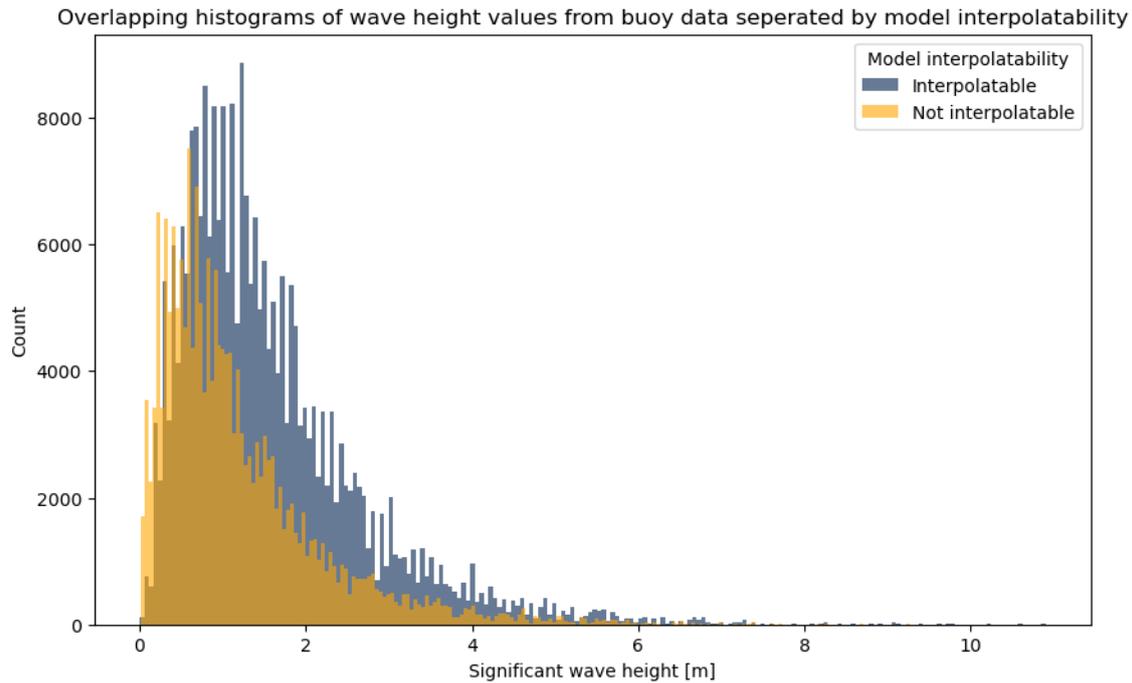


Figure 3.3: Overlapping histograms of the significant wave height from the buoy data separated by the interpolatability of the model. The model interpolatable data has a mean of 1.580 m and non-interpolatable data has a mean of 1.188 m

interpolated model data is plotted against height adjusted buoy data from the WSPD parameter. The RMSE is 1.862 m/s and bias is -0.025 m/s.

3.1.4 Merging Buoy and Model Data

From the collocation step each SAR-image has been paired with a value for at least one of the buoy variables VAVH, VHM0, and WSPD, sometimes multiple. Since the subsequent multi-task machine learning algorithm required complete labels with values for both wind speed and wave height the models were used to complete the labels for missing buoy values. In case the WSPD value was missing in the buoy data the corresponding value from the wind speed model was used. For VAVH and VHM0 the situation was a bit different since they both measure wave height. If the buoy only has a measurement for one of the variables that one is used as the label and if both variables are found VAVH is used, finally if none is found the label is taken from the wave height model.

3.1.5 Extracting Sub-images from SAR-Images

The next step in the data pipeline was extracting smaller sub-images from the larger SAR-images, centered specifically on the buoy's location. To achieve this, the buoy's coordinates were mapped to specific locations on the SAR-image. A square patch, measuring 2 km on each side, was then taken from the SAR-image, corresponding to a resolution of 200 pixels by 200 pixels.

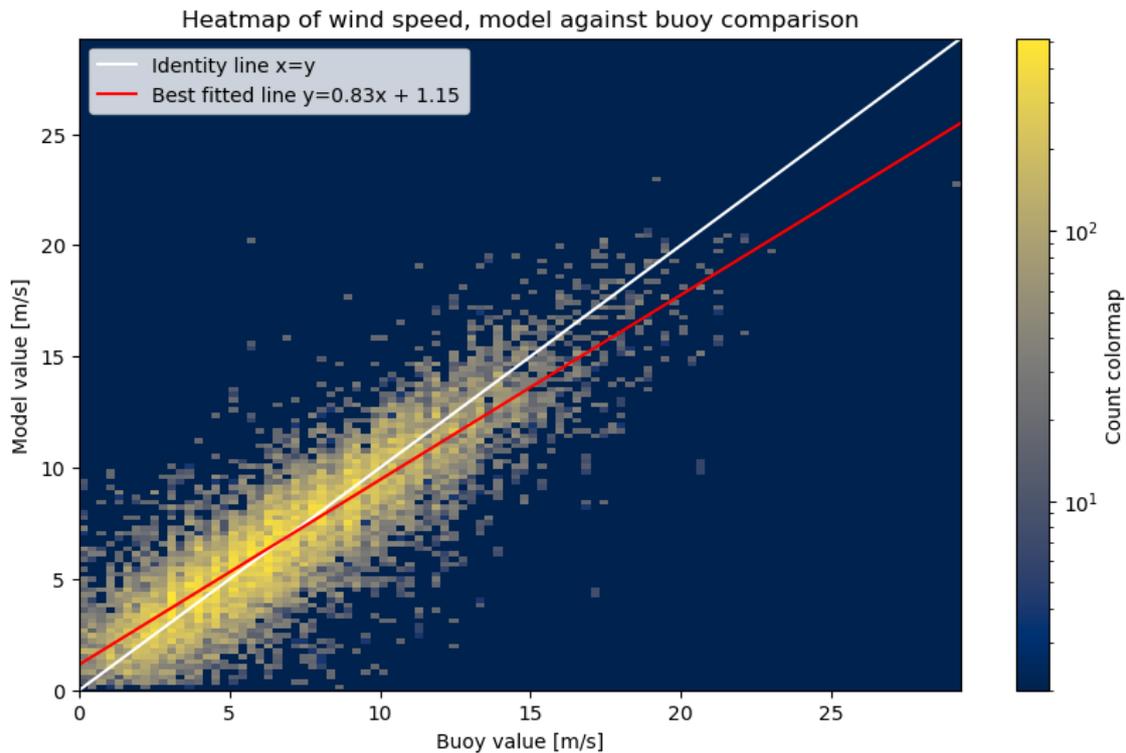


Figure 3.4: Correlation of wind speed model and buoy data.

To augment and increase the amount of data, a shifting technique was employed. Starting from the buoy’s image position, a distance of 500 m was either added, subtracted, or kept unchanged, both horizontally and vertically. This procedure yielded a 3×3 grid, creating nine distinct sub-images from each primary SAR-image.

Only those sub-images fully within the larger SAR-image and free from land or undefined values were kept. For each retained sub-image, several attributes were recorded, the most important being the backscatter σ^0 and incidence angle. Additional information like land masks, swath, time, and coordinates were stored to aid in potential troubleshooting later. These refined sub-images were then stored using 32-bit precision in the NetCDF-4 format. Examples of what these sub-images look like can be found in Figure 3.7 and Figure 3.8.

3.1.6 Features

For each sub-image, several features were calculated. A number of features were extracted using the Gray Level Co-occurrence Matrix (GLCM) [58]. These features included contrast, dissimilarity, homogeneity, energy, correlation, and Angular Second Moment (ASM). Additionally features based on statistics from the backscatter were used. These were the mean, variance, mean divided by variance, minimum, maximum and difference between minimum and maximum value. In addition, four variations of azimuth cutoff wavelength were included. Finally, the incidence angle from the centre of the image crop was included. Since each image used has two polarisations, each of these features except incidence angle are doubled: once for the VV polarisation and

once for the VH.

3.1.6.1 Calculation of the Azimuth Cutoff Wavelength

Another series of features that were calculated were related to the azimuth cutoff wavelength. The azimuth cutoff wavelength is a property of SAR-images that has been shown to be related both to significant wave height and wind speed [59]. It is possible to make an estimation when exact calculation is not necessary.

The procedure for estimating the azimuth cutoff wavelength that was followed was described by Corcione *et al.* [60]. The procedure involves first calculating the autocorrelation function (ACF) in the azimuth direction. This can be done through the Wiener Khinchin theorem. The theorem states how to estimate the ACF from the backscatter σ^0 , which is done the following way. First, a 2-D power spectral density (PSD) is computed from σ^0 :

$$\text{PSD} = \left| \text{FFT2}(\sigma^0) \right|^2 \quad (3.2)$$

From the resulting 2-D PSD, a 1-D azimuth PSD was then derived by averaging the 2-D PSD along the range direction. Subsequently, the ACF was obtained by using the inverse Fourier transform on the 1-D azimuth PSD. This ACF was then min-max normalised, and a 7×1 median filter was finally applied, which served as a speckle filter and removed the 0-lag contribution.

Once the ACF is calculated, the Gaussian function in Equation 3.3 can be fitted to it:

$$C(x) \sim e^{-\pi^2 \frac{x^2}{\lambda_c^2}} \quad (3.3)$$

$C(x)$ represents the Gaussian function's value at a given point x . The λ_c that makes the Gaussian fit closest to the ACF is the azimuth cutoff wavelength [61]. The fitted Gaussian is shown in Figure 3.5.

It was not certain how the median filter should be best applied, and if the backscatter σ^0 should be given in linear or decibel units. The feature was extracted for every combination of these to allow for future feature selection. Finally, the fitting of a curve was slow and unreliable for the data of this project. A workaround was to instead take the standard deviation of the AACF. This should give a number closely related to the azimuth cutoff wavelength. From the perspective of a machine learning model, these should give the same information.

3.1.7 Homogeneity Filter

The sub-images that were extracted required a filtration of undesired objects that sometimes appeared. The land mask included in the SAR-images was too coarse to accurately filter out all land at the resolution required. The mask also did not account for things like piers and harbors, as well as offshore constructions like wind farms. Transient features like ships were also sometimes present in the sub-images.

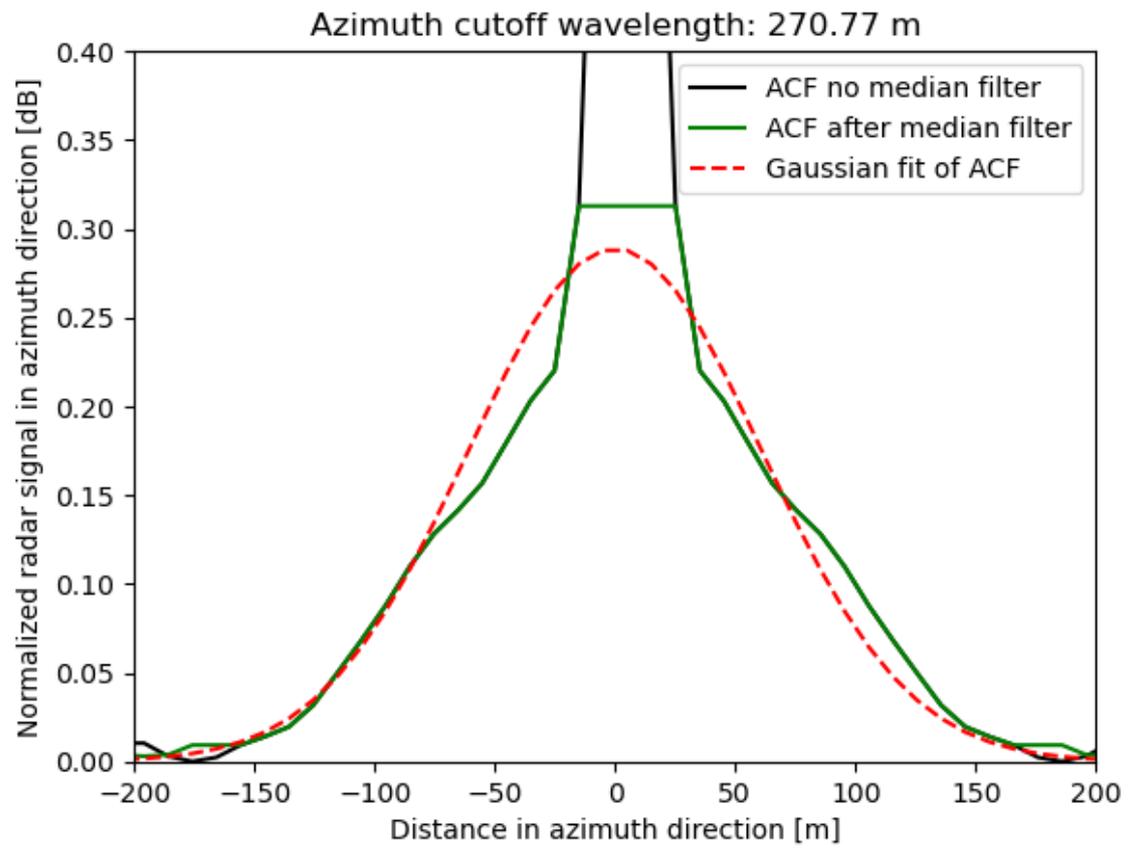


Figure 3.5: Gaussian fit of ACF.

These are the hardest to remove since their location is not fixed, and therefore can not be masked away like land.

These irregularities within the data could have potentially introduced confusion into the model and it was decided that the implementation of a homogeneity filter was necessary. The concept for the homogeneity filter involved collecting a set of example sub-images that represented both homogeneous and non homogeneous conditions. Different features could then be calculated between these two datasets which could be used to differentiate the two classes.

The homogeneous dataset was assembled by slicing out sub-images from randomly selected coordinates in SAR-images where land was absent. These sub-images align closely with those the model is expected to encounter and provide a useful baseline. However, the resulting dataset was not completely homogeneous, which is also the motivation of needing one of these filters. This process was used to extract 875 sub-images.

For the non homogeneous dataset, wind farms were found to best represent the non homogeneous conditions this filter should detect. Sub-images depicting wind farms were obtained using a list of coordinates for wind farms in the Baltic, North, and Irish seas [62]. Since the coordinates only indicated the center point of the wind farms, which can span several kilometers, a random offset was applied to capture more images from each farm. This method led to the collection of 203 sub-images of wind farms.

After assembling the dataset, features were extracted using the GLCM [58]. These were homogeneity, dissimilarity, and correlation. Figure 3.6 shows histograms comparing these features across the various sources of sub-images. Upon analysing these plots, the decision was made to separate the classes by using a Support Vector Classifier (SVC) [63].

A SVC was trained using the features calculated with the GLCM. The SVC was specifically trained to differentiate between ocean and wind farm images, and achieved an accuracy of 93%. Since the dataset was not manually labeled, perfect accuracy was not the goal; rather, the aim was to find a classifier capable of establishing the best threshold between these two types of images. Instead, a visual inspection was employed to judge the separation. Some sub-images that were filtered out are shown in Figure 3.7 and some that were kept in Figure 3.8. These images are from sub-images never seen by the model during training. Furthermore, they are the first 25 sub-images that appear alphabetically in the final dataset (Section 3.1.9), and are not hand picked.

Visually, it appears as the filter not only have learned to filter out wind farms like it was trained on, but extends to what looks like ships and some land/harbour structure. However, some images that were filtered out arguably are homogeneous and could have been used. Moreover, the images that were marked homogeneous are all visually homogeneous. This is more important since these images are the ones that were kept. The homogeneity filter removed 24% of the images in the final dataset.

3. Methods

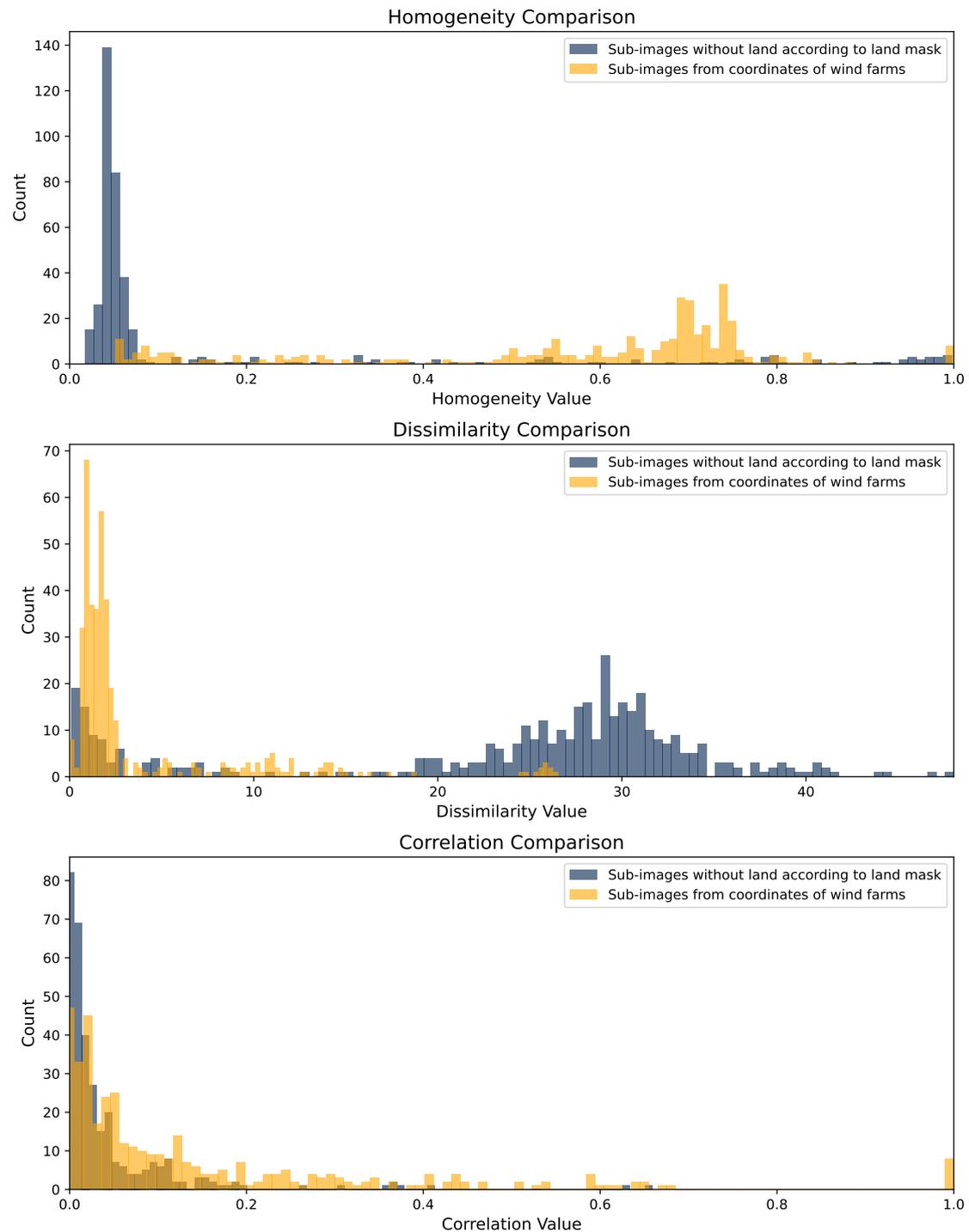


Figure 3.6: Overlapping histograms comparing the distributions homogeneity, dissimilarity, and correlation for different types of sub-images. The blue color represents sub-images which had no land in them, according to the land mask. These are similar to what sub-images looks like without the homogeneity filter, and should be mostly homogeneous. The yellow images are gathered from coordinates of wind farms. These should represent something non homogeneous that should be removed.

Sub-images marked as non homogeneous by the homogeneity filter

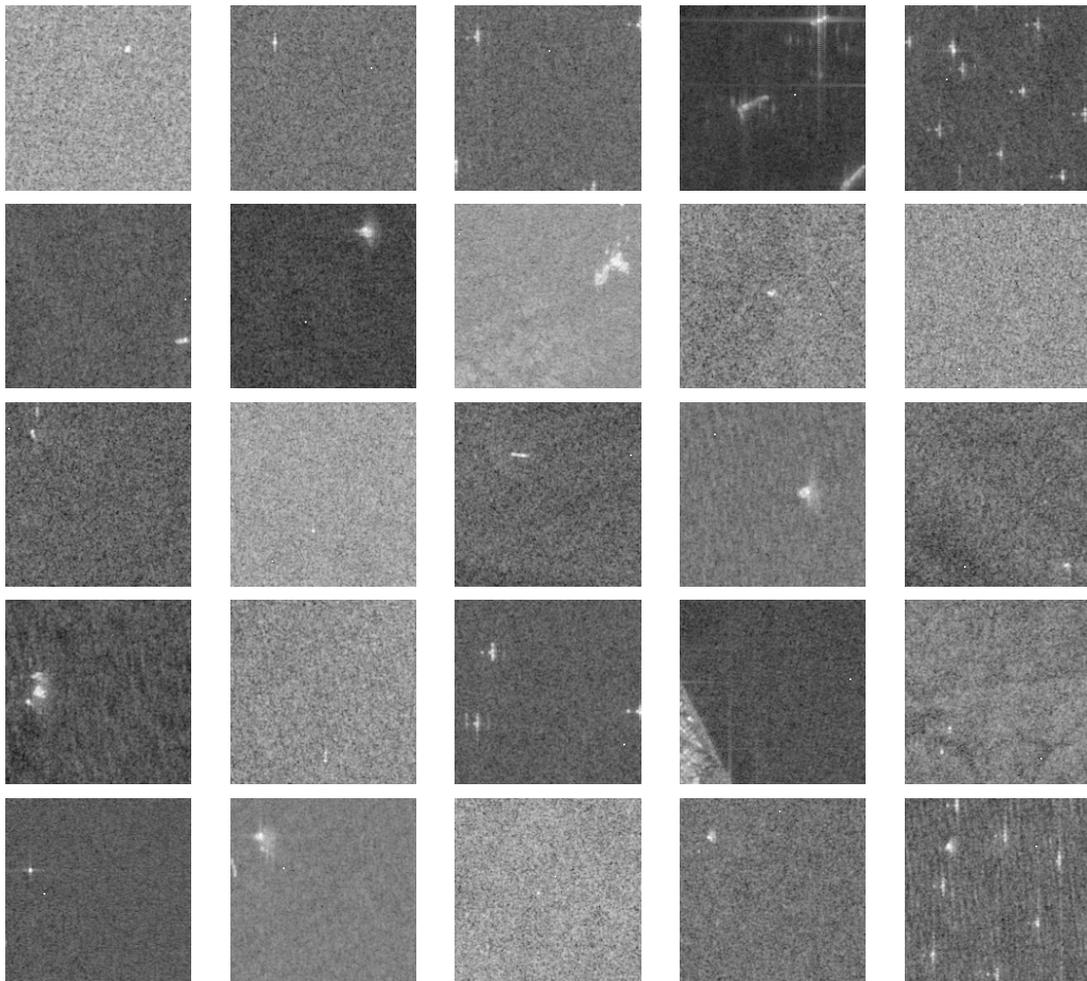


Figure 3.7: Sub-images removed by the homogeneity filter.

Sub-images marked as homogeneous by the homogeneity filter

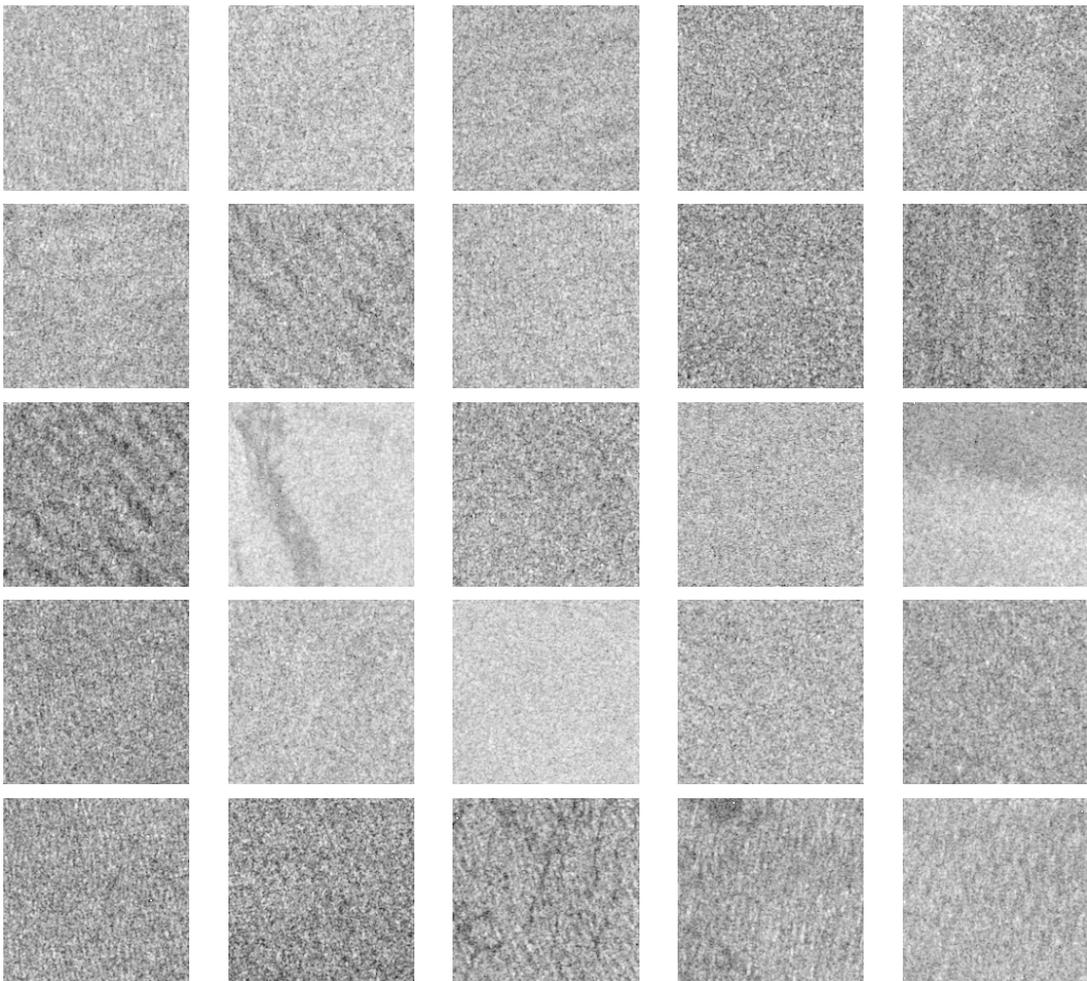


Figure 3.8: Sub-images kept by the homogeneity filter.

Wave source	Wind source	Nr. images	Proportion of the data
Buoy	Model	146 651	0.56
Buoy	Buoy	92 094	0.35
Model	Buoy	22 488	0.09

Table 3.3: Data sources of the final dataset.

3.1.8 Data Split

In the process of developing a machine learning model, the data splitting strategy plays a crucial role in ensuring the robustness and generalisability of the model. A common method of partitioning the data is a 60/20/20 split, where 60% of the data is used for training, 20% for validation during model development, and the remaining 20% for testing the final model’s performance. This strategy ensures that the model is not overfitted to the training data, as well as enabling effectively hyper-parameter tuning using the validation set, and provides an unbiased evaluation of the final model on unseen data from the test set [44, Chapter 7].

Since there is data augmentation in the SAR-image extraction step in the form of overlapping sub-image offsets, as described in Section 3.1.5, a regular independent random split cannot be used since it would introduce data leakage [64]. Instead the split uses a hash-based approach that combines the original SAR-image name and buoy name to create a unique identifier for each SAR-image and buoy name pair. The hash of this identifier is then used to decide which set (training, validation, or testing) the sub-image data point should belong to. This method ensures that all sub-images from the original SAR-image cut around each buoy is placed in the same set. Another benefit with this method is ensuring a consistent split regardless of changes or additions to other parts of the dataset and is beneficial since maintaining the same split throughout the project further mitigates data leakage.

3.1.9 Final Dataset

The pipeline yielded a final dataset of 261 233 sub-images, each with corresponding features and metadata. The combined size of these images was 122 GB where each sub-image consisted of two layers representing the VV and VH polarisations. Of the final dataset the training set included 157 229 sub-images, while the validation and test sets contained 51 635 and 52 369 sub-images respectively. The breakdown of data sources used for labels can be found in Table 3.3 while the distributions of wave height and wind speed values are depicted in Figures 3.9 and 3.10 respectively.

3.2 Machine Learning Models

After extracting the data, the next step involved discerning the patterns between the SAR data and the associated wave and wind parameters. A baseline was first established using models not based on deep learning, because of ease and speed of using these. Once the features had been evaluated and a baseline had been

3. Methods

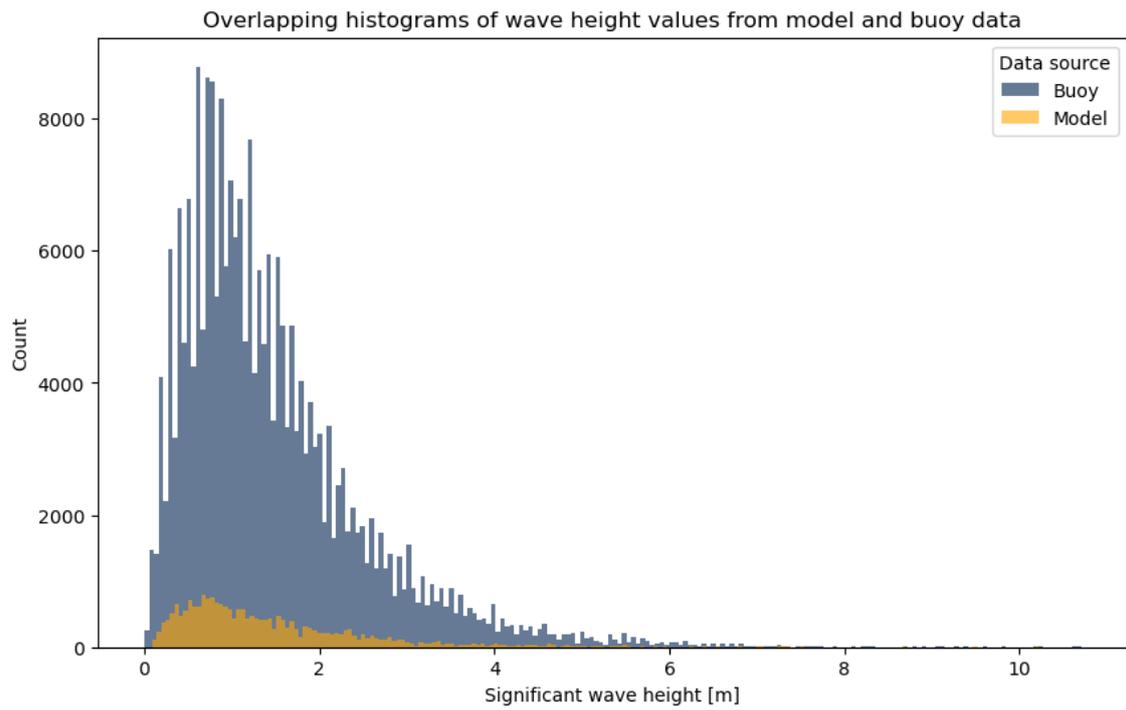


Figure 3.9: Overlapping histograms of the distributions of significant wave height values from model and buoy in the final dataset. Buoy mean is 1.47 m/s and model mean is 1.42 m/s.

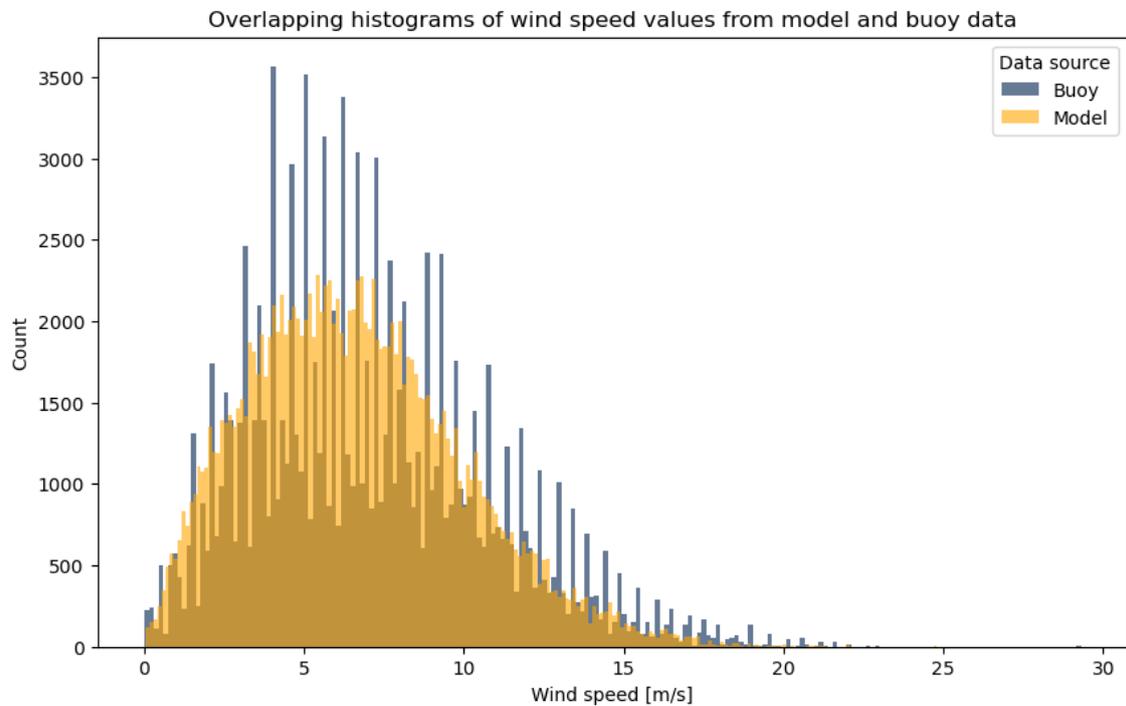


Figure 3.10: Overlapping histograms of the distributions of wind speed values from model and buoy in the final dataset. Buoy mean is 7.14 m/s and model mean is 6.62 m/s.

established, different types of deep learning models were developed to achieve the best results possible.

3.2.1 Traditional Model Baseline

To establish a preliminary understanding of the potential performance, a selection of models not based on deep learning was employed and trained. All these models were only trained on the features extracted from the sub-images, and not the sub-images themselves. The models were also trained on wind and wave independently, and not in a combined model. The models were trained on the training set and evaluated on the test set using RMSE.

This assortment of models included various linear algorithms, namely, Linear Regression, Ridge Regression, Lasso Regression, and Elastic Net. Additionally, the Decision Tree Regressor, a non-linear model, was also tested. The Linear Regression was also applied to each feature individually to evaluate how useful each feature was.

Subsequently, a series of ensemble methods were utilised to potentially enhance model performance. These methods incorporated the Random Forest Regressor, Gradient Boosting Regressor, and XGBoost. These ensemble methods combine multiple models to achieve better predictive performance than could be obtained from any of the constituent models individually. Since the intention of this experiment was only to investigate a baseline model to compare the final model against as well as making sure the features had sufficient predictive power, no hyperparameter tuning was done except for the XGBoost model. Here, the hyperparameter tuning library Optuna [65] was used. The results are presented in Section 4.1.

3.2.2 Multi-Task Learning and Loss Function

The project’s objective included an exploration of the potential performance of a deep learning model for this prediction task. Several different architectures were tested. The first critical design decision involved creating a model that would concurrently predict both target variables, wave and wind. This decision was motivated by the observed correlation between these two variables as explained in Section 2.1. A model designed in such a manner could utilise this overlapping information to share useful intermediate features within its hidden layers. Another reason for this choice was the project’s ultimate goal – predicting both values. Hence, it was deemed most efficient to bundle these values into a single model.

To facilitate this, a custom loss function was adopted that could account for both variables, enabling the simultaneous update of parameters for wave and wind during backpropagation. Importantly, the goal was to minimise the RMSE for both targets. However, as RMSE is dependent on the scale of the target, the loss function inputs were normalised by the mean value of each target precomputed on the training set. This adjustment ensured that both wave and wind would equally contribute to the loss. The RMSEs of the normalised targets were then combined using root mean square. The custom loss function, which incorporates these RMSEs, is then given by (3.4) where $\mathbf{y}_{\text{parameter}}$ and $\hat{\mathbf{y}}_{\text{parameter}}$ are the target and prediction vectors for each

parameter, and $\mu_{\text{parameter}}$ is the mean value of each parameter calculated from the training set. In this case $\mu_{\text{wave}} = 1.46$ and $\mu_{\text{wind}} = 6.85$.

$$\mathcal{L} = \sqrt{\frac{\text{RMSE}\left(\frac{\mathbf{y}_{\text{wave}}}{\mu_{\text{wave}}}, \frac{\hat{\mathbf{y}}_{\text{wave}}}{\mu_{\text{wave}}}\right)^2 + \text{RMSE}\left(\frac{\mathbf{y}_{\text{wind}}}{\mu_{\text{wind}}}, \frac{\hat{\mathbf{y}}_{\text{wind}}}{\mu_{\text{wind}}}\right)^2}{2}} \quad (3.4)$$

The aggregation was accomplished using root mean square rather than the normal arithmetic mean, motivated by the fact that root mean square gravitates more towards the larger value. This aspect implies that the gradient step will be more attuned to the output that is more significantly incorrect.

In addition to the loss, the RMSE and MAE is tracked for both wave and wind independently for the training and validation set during the training of the model. Whenever these metrics are calculated for the validation set, it is only done on data from buoys and not model, to give a more truthful evaluation.

3.2.3 Model Structure for Only Features

Once the structure of multi-task learning had been decided, designing the network was initiated. This included finding a structure of layers that worked well, as well as tuning several hyperparameters. Trying every combination of hyperparameters and settings is impossible due to the combinatorial explosion of possible choices. If a hyperparameter with a number of choices is added to the search, the size of the search space is multiplied by that number. Instead of a complete search, smaller subsets of the search space were explored at a time. Once an intuition was built up for which hyperparameters were important and which were not, successive experiments honed in on the final network structure.

Hyperparameters and design choices that were included in the search for the best network will be detailed here. Learning rate was found to be very influential and was subsequently present in all the searches. Moreover, network depth and width, dropout rate, optimiser, standardisation, and which features to include were all at different times explored. The hyperparameters were chosen as to minimise the validation loss. The hyperparameter tuning was done using the Optuna library [65], which hones in on the region of the search space which hones in on the region that gives the best results and prunes trials that are not fruitful.

The final feature-only network uses all features, which are normalised using standard scaling. The features based on statistics from the backscatter were given in linear scale. These features serve as inputs to a series of five fully connected layers, each with 1024 neurons. Batch normalisation, dropout, and ReLU activation functions are consistently applied after each of these layers. The network then diverges into separate branches for wave and wind predictions. Each branch has five fully connected layers with a successively decreasing number of neurons: 1024, 512, 256, 128, and 64. Like the shared layers, each layer in these branches also uses batch normalisation, dropout, and ReLU activation function. Both branches ultimately terminate in a

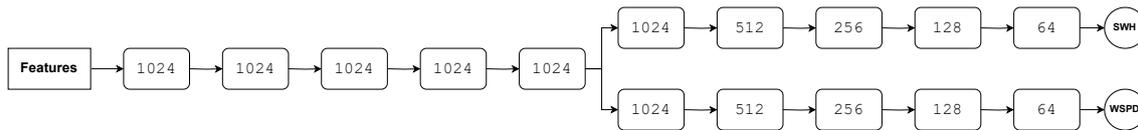


Figure 3.11: Illustration of the features-only model. The number in each block refers to the number of neurons in that layer. In addition, batch normalisation, dropout and ReLU activation is applied after each layer.

single neuron each, serving as the prediction for wave and wind, respectively. The network structure is illustrated in Figure 3.11.

The network is configured with a learning rate of 0.0087 and trained using the Adam optimiser. A dropout probability of 0.42 is used at each dropout step. Every random seed was set to 0. It was trained for 100 epochs using all the available data and a batch size of 128, which took 1.71 hours on four A100 graphics cards. The model that achieved the best validation loss is the model presented in the results.

3.2.4 Extending Model to Use Sub-images

An extended model using the extracted sub-images from the SAR-images as well was developed. This network takes the sub-images and feeds them through a CNN to reduce the images to a vector. This vector is concatenated with the feature vector and the concatenation of these two is then fed through the exact same network architecture as when using only the features.

Here a number of additional hyperparameters contributed to the search space. Whether the image should be in linear or decibel was tried. A number of CNN architectures were tried, namely ResNet- 18, 34 and 50 [51], as well as VGG- 16 and 19 [52]. These networks were taken from the built in models in PyTorch, but the first layer was swapped with a kernel filter with two channels, one for each polarisation in the SAR-image. It was tested both if this network should be initialised with random weights or pre-trained on ImageNet.

For the final image-feature network, the images are normalised using standard scaling and then augmented with mirrorings and multiples of 90° rotations. The final CNN architecture was ResNet50 with weights pretrained on ImageNet. The images are given in linear scale as opposed to decibel. The 512 neuron output of ResNet50 is concatenated with the features and then goes through the exact same architecture as the feature-only network. An illustration of the network is shown in Figure 3.12

The network is configured with a learning rate of 0.0005 and trained using the Adam optimiser. A dropout probability of 0.2 is used at each dropout step. Every random seed was set to 0. It was trained for 57 epochs using all the available data and a batch size of 128, which took 1.92 hours on four A100 graphics cards. The model that achieved the best validation loss is the model presented in the results.

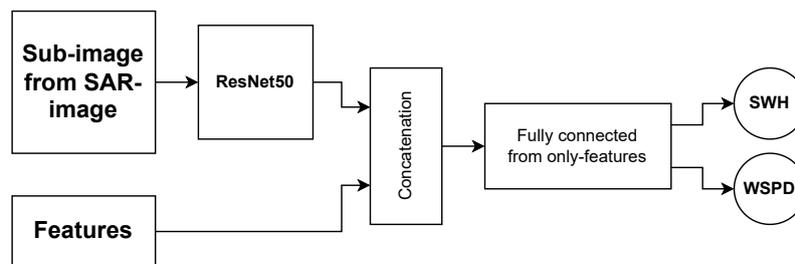


Figure 3.12: Illustration of the images-features model. The sub-image is fed through ResNet50 and is reduced to a vector, which is then concatenated with the features and fed through the same architecture as for only features. This results in the two separate heads for wind and wave as during the features-only network.

4

Results

In this chapter, the results of the project is presented. As a point of comparison, a baseline of traditional models is first presented in Section 4.1. Then, the results from the final deep learning models using only features as well as sub-images and features are presented in Section 4.1 through Section 4.2. All these evaluations are done on the test set, which only contain datapoints from in-situ buoy measurements. This test set for wind speed contains 22 677 datapoints and for wave height 47 796 datapoints.

4.1 Metrics for Models

Traditional Models as Baseline

In Table 4.1 the RMSE of the traditional models are presented. These results serve as a point of comparison for the other deep learning models.

Model	Wave RMSE	Wind RMSE
Linear Regression	0.644	2.100
Ridge Regression	0.644	2.100
Lasso Regression	1.065	2.771
Elastic Net	1.018	2.538
Decision Tree Regressor	0.794	2.364
Random Forest Regressor	0.576	1.698
Gradient Boosting Regressor	0.592	1.778
XGBoost	0.580	1.600

Table 4.1: Wave and wind RMSE for traditional regression models not based on deep learning.

Deep Learning Models

The metrics for the final models are shown in Table 4.2. The metrics include loss, as well as RMSE, bias, slope and correlation for both significant wave height and wind speed individually. How training and validation loss, as well as validation RMSE for both wind and wave developed during the training can be found in Appendix A.

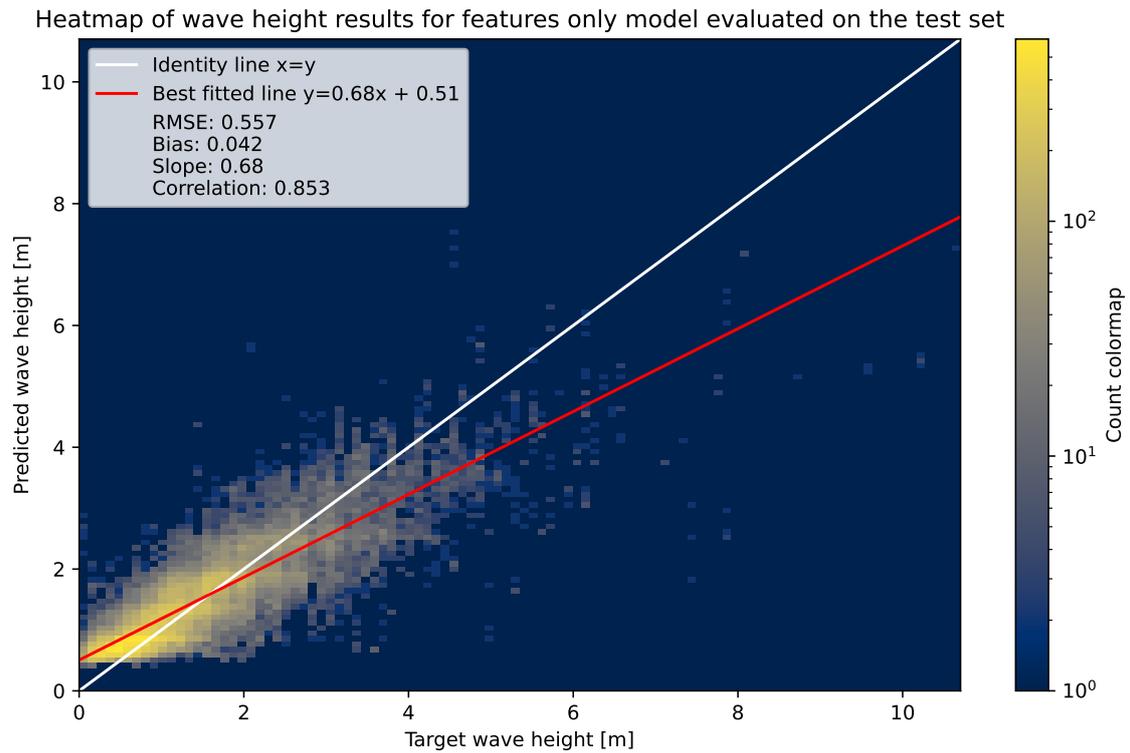
Model	Parameter	Loss	RMSE	Bias	Slope	Correlation
Features-only	Wave	0.294	0.553	-0.020	0.662	0.857
	Wind		1.573	-0.225	0.721	0.916
Images-features	Wave	0.267	0.459	-0.046	0.749	0.906
	Wind		1.658	-0.425	0.767	0.903

Table 4.2: Performance metrics for both the model using only features and the one using sub-images and features.

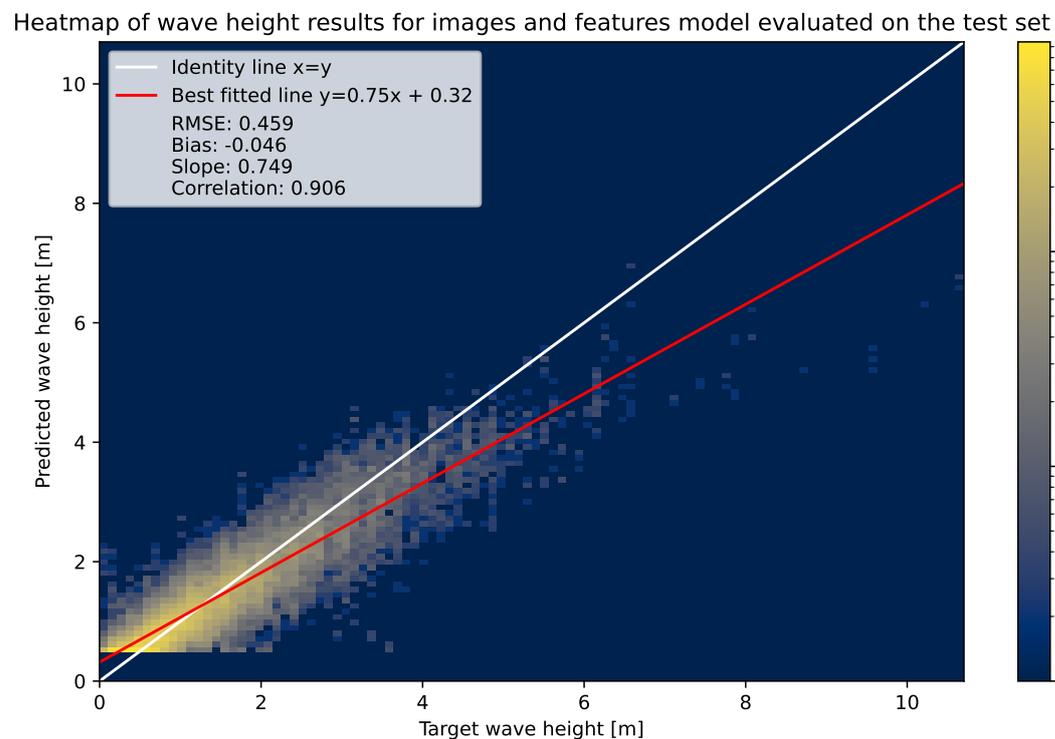
4.2 Correlation Plots from Target to Prediction

A heatmap of target significant wave height against predicted significant wave height is shown in Figure 4.1. The model that uses only features is shown in Figure 4.1a and the one that uses both sub-images and features is shown in Figure 4.1b.

A heatmap of target wind speed against predicted wind speed height is shown in Figure 4.2. The model that uses only features is shown in Figure 4.2a and the one that uses both sub-images and features is shown in Figure 4.2b.



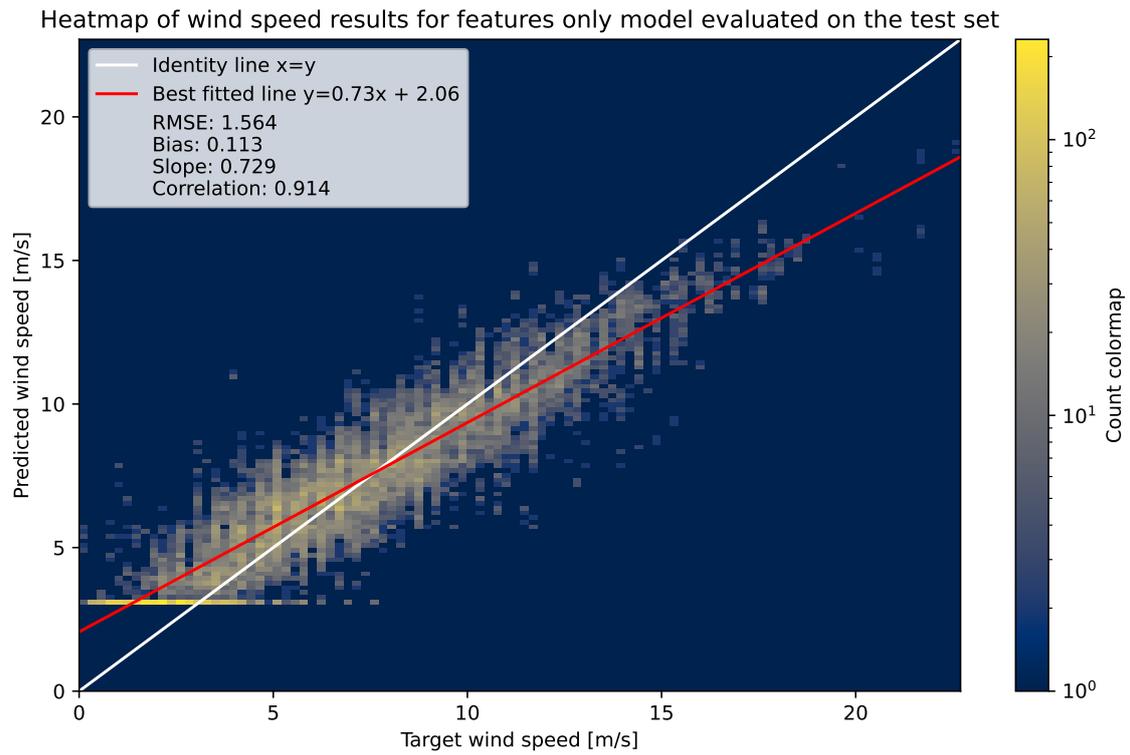
(a) Features only model results.



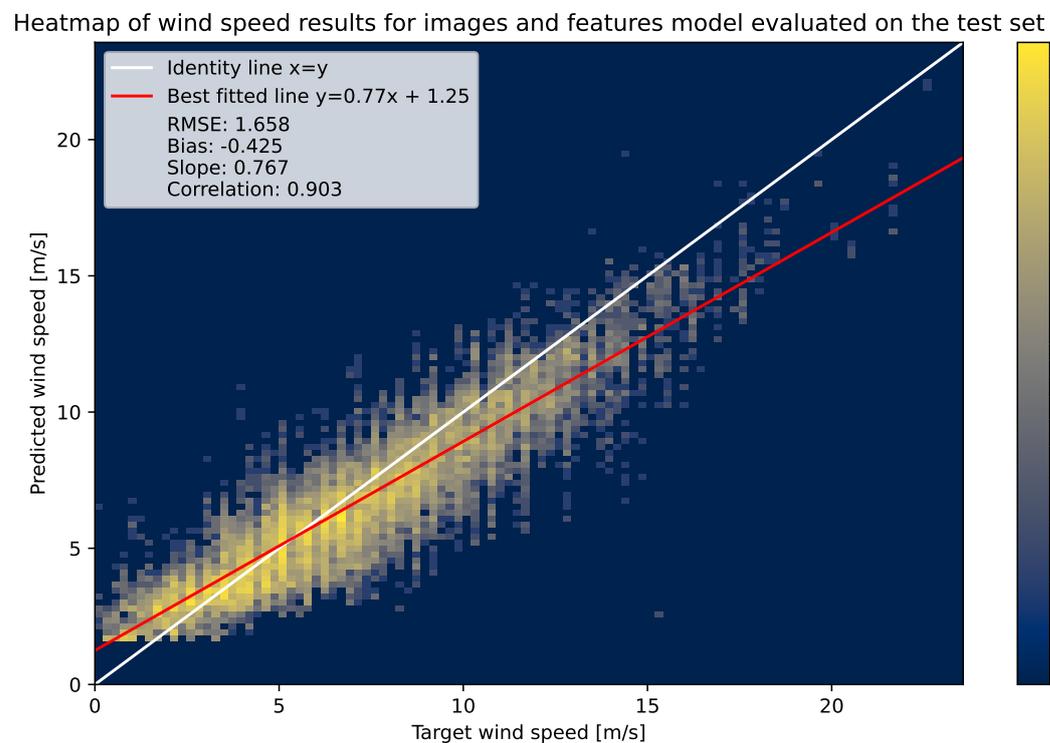
(b) Images and features model results.

Figure 4.1: Heatmaps for significant wave height predictions made on the test set by the two deep learning models.

4. Results



(a) Features only model results.



(b) Images and features model results.

Figure 4.2: Heatmap of wind speed predictions made on the test set data by the two deep learning models.

5

Discussion

5.1 Comparison with Similar Studies

The resulting RMSE of the image and feature model of 0.459 m for significant wave height is comparable with other state of the art methods using similar technique and data. For example Quach *et al.* [8] that uses Sentinel-1 data acquired from wave mode collocated with altimeter data achieve a RMSE of 0.53 m on their buoy test set and 0.307 m compared to altimeter data. Although their dataset is bigger their collocation step allows for up to a three hour difference between the SAR and altimeter measurement.

Another similar result is from Xue *et al.* [25] that attain a RMSE of 0.45 m and 0.52 m for their model trained on VV polarised data and VV and VH data respectively. Their network is trained on a dataset where buoy measurements are collocated with 3330 SAR-images. Although their data source is significantly smaller they still achieve similar results with a network of comparable size.

Yet another study with comparable results is conducted by Wu *et al.* [24]. They use SAR images from EW swath collocated with radar altimeter data from the Arctic sea and achieves a RMSE of 0.71 m on their test set constituted of radar altimeter data.

These studies reveal that contemporary methods for predicting significant wave height predominantly rely on machine learning. More specifically, these approaches are grounded in neural network architectures and utilise SAR data, which is typically paired with various other data sources. The model presented in this project also use the same methodologies.

As for wind speed predictions according to Lu *et al.* [14] all state of the art GMF achieve RMSE around 2 m/s on their test set of 1452 SAR images from coastal regions collocated with in-situ buoy measurements. More specifically the results are RMSE of 2.22 m/s for CMOD4 [10], 1.98 m/s for CMOD5 [11], 1.86 m/s for CMOD5.N [12], 1.93 m/s for CMOD7 [13], 1.92 m/s for C_SARMOD [15] and 1.84 m/s for their own GMF C_SARMOD2 [14].

Regarding wind speed predictions, the model trained on both features and images has demonstrated an RMSE of 1.658 m/s, thereby outperforming the other GMFs mentioned. However, it is worth noting that Shao *et al.* [16] achieved an RMSE of

0.74 m/s using the Gaofen-3 satellite. This satellite captures images in wave mode with quad-polarizations: VV, HH, VH, and HV. Despite the unparalleled precision, the results of this study are still unmatched.

5.2 Trustworthiness of the Results

While the results section evaluates the performance of the model on the available dataset, the applicability of these results in real-world scenarios remains to be investigated. This section discusses the measures taken to ensure the model's ability to handle new, unseen data and assesses the efforts to simulate realistic testing conditions.

To enhance the model's ability to generalise, four major strategies were implemented.

1. **Diverse Dataset:** The model was trained on a rich and diverse dataset encompassing 22 798 SAR-images. These images were collected over a full year from various global locations, ensuring the model was exposed to patterns across different oceans and seasons.
2. **Early Split of Test Set:** A fixed test set was carved out early in the project to ensure consistent evaluation. This dataset was not involved in any part of the model's training or hyperparameter tuning process, providing a reliable, unbiased evaluation of the model's performance on completely new data.
3. **Data Leakage Avoidance:** The dataset split was conducted considering the SAR-image names and buoy names, which ensured that images captured from the same ocean area at the same time were placed in the same subset. This approach effectively minimised data leakage, preventing artificially inflated performance metrics that could occur if closely related data points were present across different data subsets.
4. **High-Quality Validation Data:** Even though model data was used for training, the validation and testing stages exclusively used buoy data, ensuring the evaluation was based on the highest quality data available.

By implementing these four strategies, the risk was minimised that the model might learn patterns not applicable to future data. As a result, although the model's reported performance metrics might appear more modest, these strategies ensure the model is better equipped to handle new data, enhancing its real-world applicability.

One way our final model might be untrustworthy comes from the dataset's sparse coverage of open oceans. Since the data was gathered from buoys, which are primarily situated near coastlines and smaller oceans, the model's performance in vast open oceans like the Atlantic or Pacific might be different.

5.3 Limitations in Quality of the Data

Since the data is crucial to the success of a machine learning project, it is important to be aware of limitations of the data that may produce unwanted results. Not only to be better prepared for designing the model but also to guide future research towards more accurate and reliable data collection and interpretation with similar techniques. In this section, two such possible limitations are discussed: biases arising from incomplete coverage in the wave model, and discrepancies between buoy measurements and model predictions.

In the analysis of the significant wave height model in Section 3.1.3.1 it is noted that the model does not have perfect geographical coverage of the buoy data as 46.7% of the data is missing in the heatmap in Figure 3.2. This is due to the relatively low resolution of the model (Section 2.3.2) and that it does not contain wave parameters over land, making the shoreline only crudely approximated resulting in buoy data close to shore being excluded from the analysis. This phenomenon can be seen in Figure 3.3, as the buoy values with interpolatable values in the model have a wave height 33.0% higher than the non-interpolatable thereby introducing some bias to the dataset. But note also that there is an imbalance in the final dataset where model values for wave height only make up 9% of the data as seen in Table 3.3 mitigating this bias.

The wind model provides comprehensive coverage of all geographical areas represented in the buoy data. However, the buoy measurements display a pattern where certain values are more frequently recorded, as depicted by spikes in Figure 3.10. This pattern is likely due to limitations in the resolution of some sensors responsible for measuring wave height. When reducing the number of bins in the histogram, the distributions between the model and buoy data start to align more closely, as shown in Appendix B. This suggests a general consistency between the two datasets. A similar, albeit less pronounced, trend can be observed in the wind speed data presented in Figure 3.9. It is essential to consider these data characteristics when interpreting model predictions, as a model's outputs are influenced by its input data.

5.4 Improvements of the Homogeneity Filter

The homogeneity filter is one of the parts of the project that might benefit from being revised. The main source of uncertainty about the results of the homogeneity filter is that it is purposefully trained on imperfect data. There were both images in the homogeneous dataset that had objects or land in them, and images that were only of the ocean in the wind farm dataset. This leads to some instability where the model is trained hoping that the images misclassified also were mislabeled.

An alternative method to develop this filter would be to manually label the images the correct classes first. This would take some manual work, but would be doable within reasonable time given the size of the dataset. Also, since a machine learning model was trained to separate the classes, more features might as well have been extracted since these might help find a better separation. A more thorough evaluation

would also help to prove the usefulness of the filter, where more images would be inspected. Another evaluation would be to train machine learning models both with and without using the homogeneity filter. For this project, this test would be the ultimate judge of how helpful the homogeneity filter is.

5.5 Apparent Clipping of Small Values

In the heatmaps showing the relationship between the target and predicted values for the deep learning models in Figure 4.1 and 4.2, there is a clipping phenomenon where no predictions are made below a certain threshold.

By examining the model's intermediate layers preceding the output, one can trace the origins of the observed behavior. The network employs multiple ReLU activation functions throughout its architecture. Inherently, a ReLU function sets all negative values to zero, which can be perceived as 'clamping' numerous values at the 0 mark. If the weights in the last layer are positive or not too large, this behaviour gets carried through to the output where many values bunch up close to the bias term of the last layer. This is algebraically why the behaviour arises, the question is though why the weights end up being tuned this way.

This phenomenon is thought to have been learned by the model due to the imbalance of the dataset. This imbalance in the distribution of the two output parameters can be seen in the histograms in Figure 3.9 and 3.10. As the model gets exposed to training samples in proportion to the frequency of their occurrence, the extremities of the distribution are rarely seen. To adjust for this, the training samples can be oversampled proportionally to the inverse of their frequency. Alternatively, the loss can be adjusted to account for the underrepresented samples.

5.6 Comparison of the Models

As several models were developed during this project, it is interesting to compare these and discuss the benefits of the different models. Depending on the needs of a future application of this project, different models could serve the purpose best. First, given how simple the model is, it is noteworthy that a linear regression was not terrible, and is near instantaneous to train. Further, a model like Random Forest Regressor or XGBoost showed some performance given the time these could be trained compared to the deep learning models. Although they offered worse performance than the deep learning models, they are noteworthy in a drafting stage of a project to provide quick evaluation of the features, and could still be a viable option in a project where quick training and predictions as well as simple and small models are valuable.

In evaluating the models, it is evident that the deep learning models outperform the traditional ones in terms of metrics. While the feature-only model exhibits faster training times, the fact that training is only done once makes this advantage marginal. If investing in training, it would be prudent to select a model that produces superior

results, even if it requires slightly more time.

Moreover, while the feature-only model does offer quicker prediction times and a more compact storage footprint, these factors are not important in the project's context. The primary aim of the project did not put limitations on speed or storage efficiency as critical constraints. Therefore, given the project's objectives, the model that integrates both sub-images and features is deemed the most appropriate because of its superior loss where both wave and wind predictions are weighed together.

5.7 Future Work

To conclude this discussion, several avenues for future research emerge. A primary concern to address would be the observed clipping of values below a certain threshold. It remains undetermined if this is a result of an error within the implemented program (Section 5.5). If the model were to be updated so that the correct predictions were made for these values, it would make a direct improvement to the results.

Refining the homogeneity filter, as detailed in Section 5.4, could result in significant improvements to the outcomes. Beyond modifications to the filter, expanding the dataset is worth consideration. This could involve adding data from subsequent years or implementing more extensive data augmentation strategies. Nonetheless, given the comprehensive dataset already employed in this study, it remains undetermined whether additional data would lead to enhanced model performance.

One interesting extension would be to extend the study to also include data from the open ocean to adapt the model for these conditions. Two such sources would be SAR-images in EW and wave (WV) swath since these acquisition modes are more common over open ocean than IW [36, 66]. Another interesting data source to explore could be radar altimeters as done by Quach *et al.* [8] and Wu *et al.* [24].

The potential performance benefits from the current model structure appear to have reached their peak. This conclusion is drawn from two observations: firstly, the validation loss remained steady throughout training, as detailed in Appendix A. Secondly, because of extensive hyperparameter tuning, as outlined in Section 3.2.3, further adjustments seemed unlikely to yield significant improvements. To achieve a notable performance boost, a new method may have to be tried, such as incorporating an additional feature. An alternative strategy could be to train separate networks for wave and wind. Even though this diverges from the project's initial aim, observations during training indicated the model often had to prioritise optimising for either wave or wind, rather than both (see Appendix A). Thus, distinct networks for each might enhance the overall results.

6

Conclusion

This project ventured to investigate the potential of harnessing globally collected SAR-images and their derived features to determine wind and wave parameters through machine learning. The models have been thoroughly evaluated with in-situ data, both using several metrics as well as a visual investigation of the prediction patterns. The resulting algorithms capture the relationship between input and output, although some potential errors and improvements have been highlighted. Thought has been put into making sure that the results are honest and applicable to diverse data.

A large portion of the project was devoted to developing tools for handling large amounts of data, and in the end all of the data for the given year from all buoys available could be utilised. This effort of thorough data handling have led the algorithm to reflect the patterns in the data as accurately as possible.

The project has also investigated removing images based on land and ice masks. The final project uses a land mask and a homogeneity filter. Some improvements that could be made to these are highlighted in the discussion, but the mask and filter helps to make the data more uniform.

Initially, it was not decided which polarisations and swath modes should be used. Eventually, a model was however only developed for the IW swath mode. Although the reasons for this design choice have been presented in the report, it still remains interesting to investigate if the project could be extended to work for EW swath mode as well.

The two deep learning models developed predict both the wave and wind parameters simultaneously. This has shown that it is possible to design a model this way, and comes with the benefits of ease of use and speed of training. However, the best results were not necessarily achieved for both output values simultaneously. The network being able to predict one did not mean it could also predict the other. This suggests it might be possible to achieve even better results by training two separate networks for each parameter. One could even utilise the prediction from the other. The way the models were structured aligned the most with the aim of the project. However, in the future different design choices could be made to potentially get different results.

Bibliography

- [1] H. Hersbach *et al.*, *ERA5 hourly data on single levels from 1940 to present*, 2023. DOI: 10.24381/cds.adbb2d47. (visited on 05/16/2023).
- [2] ESA, *ESA - Copernicus*, 2022. [Online]. Available: https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus (visited on 08/08/2023).
- [3] ASF, *ASF Home | Alaska Satellite Facility*, [Online; accessed 08-August-2023], 2022. [Online]. Available: <https://asf.alaska.edu/>.
- [4] *Global Ocean- In-Situ Near-Real-Time Observations*. DOI: 10.48670/moi-00036.
- [5] European Union-Copernicus Marine Service, *Global Ocean Hourly Reprocessed Sea Surface Wind and Stress from Scatterometer and Model*, 2019. DOI: 10.48670/MOI-00185.
- [6] G. Atteia, M. J. Collins, A. D. Algarni, and N. A. Samee, “Deep-Learning-Based Feature Extraction Approach for Significant Wave Height Prediction in SAR Mode Altimeter Data,” *Remote Sensing*, vol. 14, no. 21, 2022. DOI: 10.3390/rs14215569.
- [7] M. Majidi Nezhad, A. Heydari, E. Pirshayan, D. Groppi, and D. Astiaso Garcia, “A novel forecasting model for wind speed assessment using sentinel family satellites images and machine learning method,” *Renewable Energy*, vol. 179, pp. 2198–2211, 2021. DOI: 10.1016/j.renene.2021.08.013.
- [8] B. Quach, Y. Glaser, J. E. Stopa, A. A. Mouche, and P. Sadowski, “Deep Learning for Predicting Significant Wave Height From Synthetic Aperture Radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1859–1867, 2021. DOI: 10.1109/TGRS.2020.3003839.
- [9] A. Ribal and I. R. Young, “33 years of globally calibrated wave height and wind speed data based on altimeter observations,” *Sci. Data*, vol. 6, no. 1, p. 77, 2019. DOI: 10.1038/s41597-019-0083-9.
- [10] A. Stoffelen and D. Anderson, “Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4,” *Journal of Geophysical Research: Oceans*, vol. 102, no. C3, pp. 5767–5780, 1997. DOI: 10.1029/96JC02860.
- [11] H. Hersbach, A. Stoffelen, and S. de Haan, “An improved C-band scatterometer ocean geophysical model function: CMOD5,” *Journal of Geophysical Research: Oceans*, vol. 112, no. C3, 2007. DOI: 10.1029/2006JC003743.
- [12] H. Hersbach, *CMOD5.N: A C-band geophysical model function for equivalent neutral wind*. Apr. 2008. DOI: 10.21957/mzcfm6jfl.

- [13] A. Stoffelen, J. A. Verspeek, J. Vogelzang, and A. Verhoef, “The CMOD7 Geophysical Model Function for ASCAT and ERS Wind Retrievals,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2123–2134, 2017. DOI: 10.1109/JSTARS.2017.2681806.
- [14] Y. Lu, B. Zhang, W. Perrie, A. A. Mouche, X. Li, and H. Wang, “A C-Band Geophysical Model Function for Determining Coastal Wind Speed Using Synthetic Aperture Radar,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 7, pp. 2417–2428, 2018. DOI: 10.1109/JSTARS.2018.2836661.
- [15] A. Mouche and B. Chapron, “Global C-Band Envisat, RADARSAT-2 and Sentinel-1 SAR measurements in copolarization and cross-polarization,” *Journal of Geophysical Research: Oceans*, vol. 120, no. 11, pp. 7195–7207, 2015. DOI: 10.1002/2015JC011149.
- [16] W. Shao *et al.*, “Intelligent Wind Retrieval from Chinese Gaofen-3 SAR Imagery in Quad Polarization,” *Journal of Atmospheric and Oceanic Technology*, vol. 36, no. 11, pp. 2121–2138, 2019. DOI: 10.1175/JTECH-D-19-0048.1.
- [17] S. Hasselmann, C. Brüning, K. Hasselmann, and P. Heimbach, “An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra,” *Journal of Geophysical Research: Oceans*, vol. 101, no. C7, pp. 16 615–16 629, 1996. DOI: 10.1029/96JC00798.
- [18] J. Schulz-Stellenfleth, S. Lehner, and D. Hoja, “A parametric scheme for the retrieval of two-dimensional ocean wave spectra from synthetic aperture radar look cross spectra,” *Journal of Geophysical Research: Oceans*, vol. 110, no. C5, 2005. DOI: 10.1029/2004JC002822.
- [19] J. Sun and H. Kawamura, “Retrieval of surface wave parameters from SAR images and their validation in the coastal seas around Japan,” *Journal of oceanography*, vol. 65, no. 4, pp. 567–577, 2009. DOI: 10.1007/s10872-009-0048-2.
- [20] J. Schulz-Stellenfleth, T. König, and S. Lehner, “An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data,” *Journal of Geophysical Research: Oceans*, vol. 112, no. C3, 2007. DOI: 10.1029/2006JC003970.
- [21] X.-M. Li, S. Lehner, and T. Bruns, “Ocean Wave Integral Parameter Measurements Using Envisat ASAR Wave Mode Data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 155–174, 2011. DOI: 10.1109/TGRS.2010.2052364.
- [22] J. E. Stopa and A. Mouche, “Significant wave heights from Sentinel-1 SAR: Validation and applications,” *Journal of Geophysical Research: Oceans*, vol. 122, no. 3, pp. 1827–1848, 2017. DOI: 10.1002/2016JC012364.
- [23] D. Gao, Y. Liu, J. Meng, Y. Jia, and C. Fan, “Estimating significant wave height from SAR imagery based on an SVM regression model,” *Acta Oceanologica Sinica*, vol. 37, pp. 103–110, 2018. DOI: 10.1007/s13131-018-1203-7.
- [24] K. Wu, X.-M. Li, and B. Huang, “Retrieval of Ocean Wave Heights From Spaceborne SAR in the Arctic Ocean With a Neural Network,” *Journal of Geophysical Research: Oceans*, vol. 126, no. 3, 2021. DOI: 10.1029/2020JC016946.

-
- [25] S. Xue, X. Geng, X.-H. Yan, T. Xie, and Q. Yu, "Significant wave height retrieval from Sentinel-1 SAR imagery by convolutional neural network," *Journal of Oceanography*, vol. 76, no. 6, pp. 465–477, 2020. DOI: 10.1007/s10872-020-00557-3.
- [26] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A new algorithm for SAR image target recognition based on an improved deep convolutional neural network," *Cognitive Computation*, vol. 11, pp. 809–824, 2019. DOI: 10.1007/s12559-018-9563-z.
- [27] European Space Agency, *Total significant wave height is now provided for the Sentinel-1 Wave Mode acquisitions*, 2022. [Online]. Available: <https://sentinel.esa.int/web/sentinel/-/total-significant-wave-height-is-now-provided-for-the-sentinel-1-wave-mode-acquisitions/1.1> (visited on 01/26/2023).
- [28] European Space Agency, *Mission ends for Copernicus Sentinel-1B satellite*, Aug. 3, 2022. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite (visited on 06/15/2023).
- [29] L. H. Holthuijsen, *Waves in Oceanic and Coastal Waters*. Cambridge University Press, Jan. 2007. DOI: 10.1017/cbo9780511618536.
- [30] A. George, *What causes waves in the ocean?* [Online]. Available: <https://www.newscientist.com/question/causes-waves-ocean/>.
- [31] *When Do Ocean Waves Become 'Significant'? A Closer Look at Wave Forecasts*. [Online]. Available: https://www.vos.noaa.gov/MWL/apr_06/waves.shtml.
- [32] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6–43, 2013. DOI: 10.1109/MGRS.2013.2248301.
- [33] Earth Science Data Systems, NASA, *What is synthetic aperture radar?* Apr. 2020. [Online]. Available: <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar>.
- [34] European Space Agency, *User Guides - Sentinel-1 SAR - Overview - Sentinel Online*. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/overview>.
- [35] European Space Agency, *User Guides - Sentinel-1 SAR - Interferometric Wide Swath - Sentinel Online*. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath>.
- [36] European Space Agency, *User Guides - Sentinel-1 SAR - Extra Wide Swath - Sentinel Online*. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/extra-wide-swath>.
- [37] European Space Agency, *User Guides - Sentinel-1 SAR - Level-1 - Sentinel Online*. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/product-types-processing-levels/level-1>.

- [38] Natural Environment Research Council, *Concept: Wave height mean of waves (highest one third) on the water body*. [Online]. Available: <http://vocab.nerc.ac.uk/collection/P01/current/GAVHZZ01/3/> (visited on 08/2023).
- [39] Natural Environment Research Council, *Concept: Spectral significant wave height of waves on the water body*. [Online]. Available: <https://vocab.nerc.ac.uk/collection/P01/current/HMZEZZ01/2/> (visited on 08/2023).
- [40] Natural Environment Research Council, *Concept: Speed of wind in the atmosphere*. [Online]. Available: <https://vocab.nerc.ac.uk/collection/P01/current/EWSBZZ01/2/> (visited on 08/2023).
- [41] M. Belmonte Rivas and A. Stoffelen, “Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT,” *Ocean Science*, vol. 15, no. 3, pp. 831–852, 2019. DOI: 10.5194/os-15-831-2019.
- [42] A. Trindade, M. Portabella, A. Stoffelen, W. Lin, and A. Verhoef, “Erastar: A high-resolution ocean forcing product,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1337–1347, 2020. DOI: 10.1109/TGRS.2019.2946019.
- [43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [44] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
- [45] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. DOI: 10.1111/j.1467-9868.2011.00771.x.
- [46] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005. DOI: 10.1111/j.1467-9868.2005.00527.x.
- [47] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324.
- [48] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of statistics*, pp. 1189–1232, 2001. DOI: 10.1214/aos/1013203451.
- [49] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” ACM, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.
- [50] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [52] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014. DOI: 10.48550/arXiv.1409.1556.
- [53] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *CVPR09*, 2009. DOI: 10.1109/CVPR.2009.5206848.

-
- [54] ASF, *Basics - asf sar data search manual*, [accessed April-2023]. [Online]. Available: https://docs.asf.alaska.edu/asf_search/basics/.
- [55] T. Wizelius, *Developing wind power projects: theory and practice*. Earthscan, 2007, p. 40.
- [56] E. W. Peterson and J. P. Hennessey, "On the Use of Power Laws for Estimates of Wind Power Potential," *Journal of Applied Meteorology and Climatology*, vol. 17, no. 3, pp. 390–394, 1978. DOI: 10.1175/1520-0450(1978)017<0390:OTUOPL>2.0.CO;2.
- [57] S. A. Hsu, E. A. Meindl, and D. B. Gilhousen, "Determining the Power-Law Wind-Profile Exponent under Near-Neutral Stability Conditions at Sea," *Journal of Applied Meteorology and Climatology*, vol. 33, no. 6, pp. 757–765, 1994. DOI: 10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2.
- [58] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973. DOI: 10.1109/tsmc.1973.4309314.
- [59] G. Grieco, W. Lin, M. Migliaccio, F. Nirchio, and M. Portabella, "Dependency of the Sentinel-1 azimuth wavelength cut-off on significant wave height and wind speed," *International Journal of Remote Sensing*, vol. 37, no. 21, pp. 5086–5104, Sep. 2016. DOI: 10.1080/01431161.2016.1226525.
- [60] V. Corcione, A. Buono, F. Nunziata, and M. Migliaccio, "A Sensitivity Analysis on the Spectral Signatures of Low-Backscattering Sea Areas in Sentinel-1 SAR Images," *Remote Sensing*, vol. 13, no. 6, p. 1183, Mar. 2021. DOI: 10.3390/rs13061183.
- [61] V. Kerbaol, B. Chapron, and P. W. Vachon, "Analysis of ERS-1/2 synthetic aperture radar wave mode imagerettes," *Journal of Geophysical Research: Oceans*, vol. 103, no. C4, pp. 7833–7846, Apr. 1998. DOI: 10.1029/97jc01579.
- [62] *Lists of offshore wind farms*, Jun. 2022. [Online]. Available: https://en.wikipedia.org/wiki/Lists_of_offshore_wind_farms.
- [63] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, S2CID 206787478. DOI: 10.1007/BF00994018.
- [64] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 4, Dec. 2012. DOI: 10.1145/2382577.2382579.
- [65] [Online]. Available: <https://optuna.org/>.
- [66] European Space Agency, *User guides - sentinel-1 sar - wave - sentinel online*. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/wave> (visited on 08/2023).

A

Training Progress

In Figure A.1 it is shown how the metrics changed during training of the deep learning models. The metrics shown include training loss, validation loss and validation RMSE for significant wave height and wind speed individually.

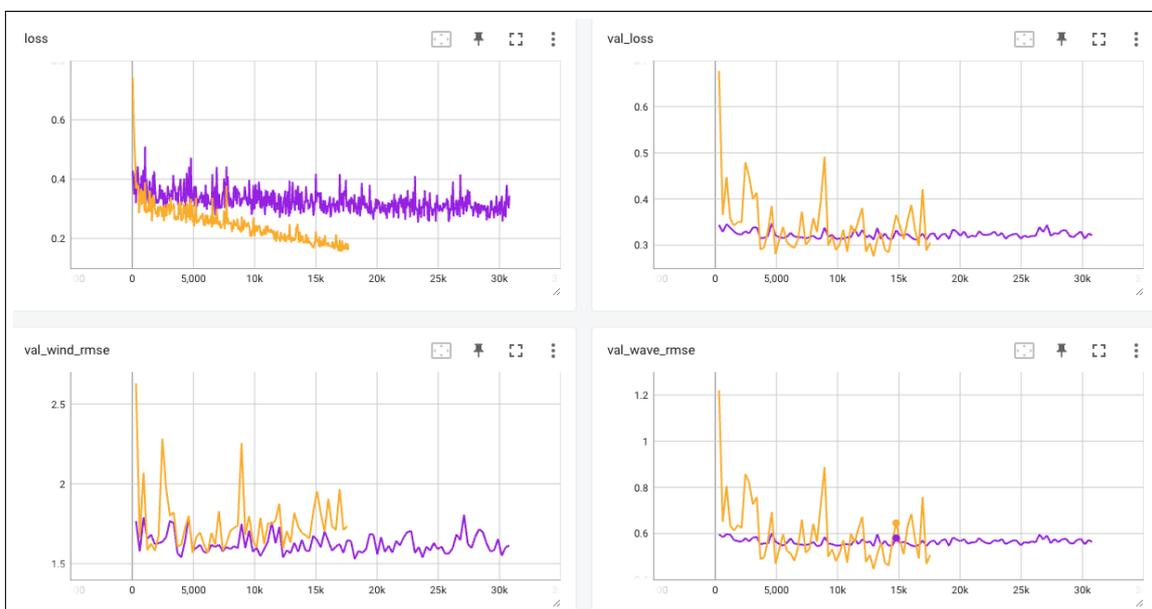


Figure A.1: Logs of metrics during training progress. Purple is from the model using features only. Orange is from the model that use both features and sub-images. The title of each plot gives the value that is tracked, which is the y-axis. The x-axis represent the number of steps that has passed during the training progress.

B

Overlapping Histograms of Wind Speed Distributions

In this appendix, the overlapping histograms of distributions of wind speeds are presented in Figure B.1. This plots the same data found in Figure 3.10 but has fewer bins to show that the spiking behaviour disappears as the number of bins is decreased.

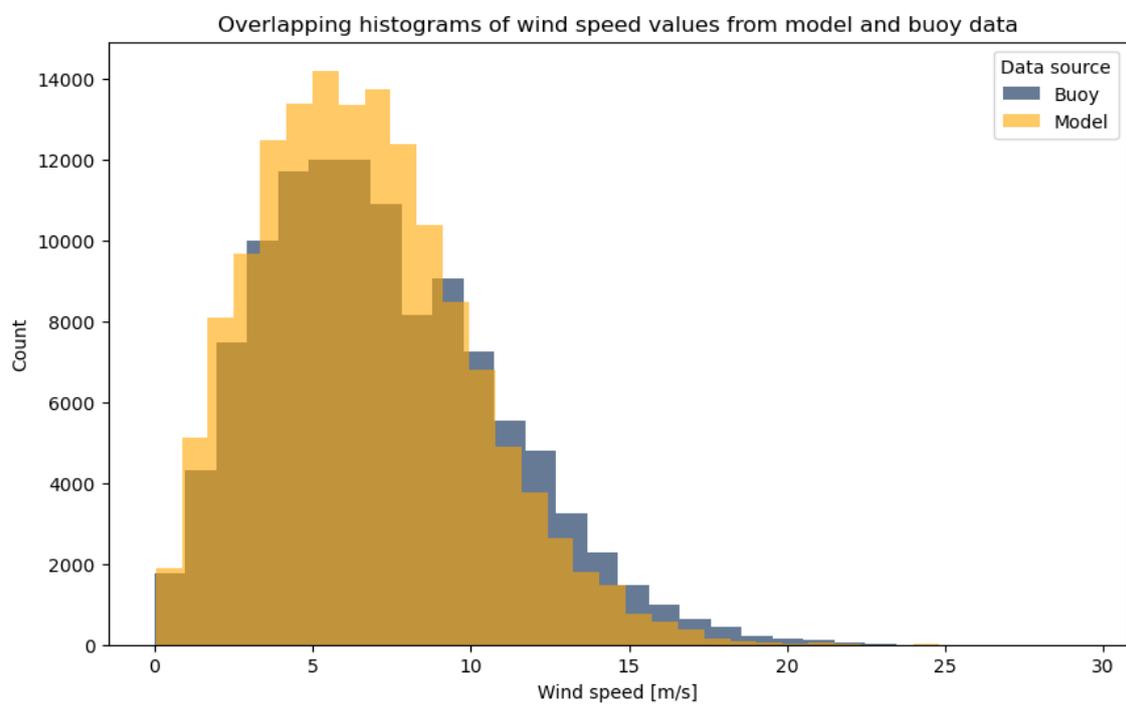


Figure B.1: Overlapping histograms of the distributions of wind speed values from both model and buoy in the final dataset. The same data as in Figure 3.10 but with only 30 bins.