AN ABSTRACT OF THE DISSERTATION OF

Ashley Ellenson for the degree of Doctor of Philosophy in Civil Engineering presented on August 12, 2021.

Title: The Application of Machine Learning Techniques to Marine Science Models

Abstract approved: _____

Gregory Wilson

A variety of stakeholders require information about marine systems. In the open ocean, pilots of marine vessels require knowledge about environmental conditions for safe passage and route planning. On the coastline, communities rely on information about nearshore dynamics to increase safety from coastal hazards such as nearshore pollutants, coastal erosion, or dangerous recreational conditions (e.g., rip currents). Models provide information for environmental health and safety in the form of forecasts or general knowledge of the marine science systems.

Large volumes of data from a variety of marine sensors are now available thanks to progress in computer processing and data storage. These data should be leveraged to advance the boundaries of marine science knowledge. Herein, Machine Learning (ML) techniques are applied to improve different types of marine science models and increase the knowledge of marine science systems. Two different types of ML techniques are considered; traditional machine learning and deep learning. The techniques are applied in a transparent way, ensuring that the ML routine has made predictions with appropriate reasoning. Also, the transferability of the ML routines is assessed to determine the limits of ML routine generalizability. The thesis is organized in a manuscript format, where the first and last chapters serve as overall Introduction and Conclusions, respectively. The central three chapters are individual manuscripts.

The second chapter applies a traditional ML technique called a decision tree to numerical wave model output. The decision tree predicts corrections of 24-hour time horizon significant wave height forecasts generated by a numerical wave model. The wave model output was located at buoy locations offshore of the United States Pacific Northwest coastline. The application of the decision tree increased wave model skill more for winter than for summer. The decision tree also made accurate predictions in a geospatial transfer experiment, where the decision tree predicted error for a location that was not used in training data. However, the decision tree predictions were less accurate when it was applied to a different time period. The transparent nature of the algorithm allowed for inspection of the algorithm's architecture, finding consistent underestimations of significant wave height for data points associated with mid wave periods (6-12s).

The third chapter develops an automated technique to recognize morphological shapes within coastal imagery using a Convolutional Neural Network (CNN). The morphological shapes are morphological patterns that occur frequently in the nearshore called beach states. The input to the CNN was coastal imagery from two different study sites and the output was beach state labels. The two different study sites were Narrabeen, Sydney, Australia and Duck, North Carolina, United States. Three ensembles of CNNs were trained; two single-site CNNs (trained at individual locations) and one multi-site CNN (trained at both locations). The CNNs were applied to both locations to determine skill at the location it was trained (the original location) in a self-test and skill at the location where it was not trained (the alternate location) in a transfer-test. For the self-tests, the CNN skill was comparable to inter-labeller agreement, with skill at Duck higher than skill at Narrabeen (F-scores of 0.8 for Duck and 0.59 for Narrabeen). The CNN skill was reduced in the transfer tests. However, if at least 25% of the training data came from the alternate location, the skill increased to within 10% of the skill at the original location. A visualization technique (Guided Grad-CAM) revealed areas of importance within input images for CNN decision making, and confirmed that the CNN identified the appropriate morphological characteristics

(e.g., terraces or rip currents) for each classification.

The fourth chapter builds off the third, and applies a CNN to a long (>20 years) dataset to detect alongshore variability of beach state quantified as a beach probability simplex, thereby advancing the beach state framework from discrete space to continuous space. The approach from the third chapter is modified to detect alongshore differences in beach state using a windowing technique. The CNN produced beach probability simplices from a 28-year dataset of images from Duck, NC, and results showed that most (67%) of the resulting beach probability simplices encompassed more than one state. The 28-year time series was dominated by an annual cycle, where simplices that encompassed onshore states occurred in summer, offshore states in winter, and intermediate states in fall or spring. The mean value of the beach probability simplex exhibited a strong relationship with significant wave height (28-year daily average R=0.77) and mean wave direction (28-year daily average R=0.84). The simplices that encompassed the highest number of states (three) were most likely to occur in fall, specifically the month of September.

The Application of Machine Learning Techniques to Marine Science Models

by

Ashley Ellenson

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented August 12, 2021
Commencement June 2022

Doctor of Philosophy manuscript of Ashley Ellenson presented on August 12, 2021.

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Ashley Ellenson, Author

ACKNOWLEDGEMENTS

I would like to acknowledge the graduate students (Maggie Exton, Alex Simpson, Kai Parker, Harri Ko, Dylan Anderson, Annika O'Dea, Liz Holzenthal, Chris O'Day, Peter Davidson, Paige Hovenga, Meredith Leung, Nick Mathews, Kyle Niezgoda, Dean Henze, Kayla Delventhal, Daniel Pierdrahita) within Coastal and Ocean engineering, CEOAS, and Civil Engineering who were the motivation for staying for a PhD, and have made my time as a graduate student spectacular and filled with friendship. Thanks to Tuba Özkan-Haller for believing in me and inspiring me to continue on towards my PhD. Thanks to Greg Wilson who supported my pursuit of individual thinking, taught me how to carefully quantify relationships, and understood when I felt burned out. Thanks to the students at the Water Research Laboratory in Sydney, who showed me how incredibly collaborative and supportive research groups can be. Thanks to Kristen Splinter, Joshua Simmons and Ty Hesser, who helped me sort through and communicate my ideas, and for supporting me from afar. Thanks to Sean Fleming who answered random for statistics questions and gave me his time for brainstorming sessions. Thanks to David Reinert who made maps for me. Thanks to other IT support staff - Tom Leach and John Stanley - for helping me install an NVIDIA driver on my machine and take my computer home with me during the pandemic. Thanks to my committee for reviewing my work and being available throughout the journey. Finally, thanks to the beaches of Newport that provided sanctuary during the hard times and joy during the good times.

CONTRIBUTION OF AUTHORS

Xiaoli Fern provided the idea for implementing the decision tree from Chapter 2. Yuanli Pei assisted in implementation by providing machine learning expertise.

Tuba Özkan-Haller provided ideas, analysis and discussion for Chapters 2 and 3.

Greg Wilson provided ideas, analysis, discussion, and writing assistance for Chapters 2, 3, and 4.

Joshua Simmons provided data and code for Chapter 3.

Joshua Simmons, Kristen Splinter and Ty Hesser provided ideas, analysis, discussion and writing assistance for Chapters 3 and 4.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure | Page

LIST OF TABLES

LIST OF APPENDIX FIGURES

Chapter 1: General Introduction

Coastal zones provide ecological habitat and resources, recreational opportunity, and are home to a third of the US population. Nearshore physical dynamics directly influence coastal hazards such as coastal erosion or recreational safety and coastal environmental health such as the relative abundance of ecological fauna and transport and fate of pollutants (Benedet, Finkl, & Klein, 2006; Hughes, Aagaard, Baldock, & Power, 2014; Shanks et al., 2018; Winckler, Liu, & Mei, 2013). Coastal scientists seek to to increase our predictive power and understanding of nearshore systems, ultimately contributing to society's well-being.

The digital era has changed the way we live our lives. Powerful pattern recognition technologies aid decisions about travel, media consumption, or product purchasing. New technologies should not be confined to the commercial sector, but should also be applied towards scientific discovery in that they can uncover patterns within data that are otherwise seemingly unextractable. The following dissertation encompasses three examples of applying cutting-edge pattern recognition techniques to improve marine science models.

The scientific method requires the development and testing of models. Models that encapsulate the most up-to-date theory can take on physical, conceptual, theoretical, or numerical forms, and are either derived from or validated by observations. Observations were first made by visually noting processes or other natural phenomena. In the modern era, scientists are often removed from their object of study, as digital sensors, such as optical imagery, wave buoys, LiDar, or radar, are used in place of or to augment visual observations Malde, Handegard, Eikvil, and Salberg (2020). Thanks to a global network of digital sensors, marine data are growing by the petrabytes. The nature of the data, both in terms of volume and form (raw data from sensors are not interpretable) are no longer interpreted with computational programming languages such as FORTRAN or MATLAB. The

acquisition of the massive amounts of data has ushered in the "fourth paradigm" of science, wherein protocols for data management, sharing and accessibility are necessary to fuel data-derived scientific discovery (Hey, Tansley, Tolle, et al., 2009). With advances in computational resources and open source software, scientists can now leverage powerful pattern recognition tools to uncover relationships within data and expand the boundaries of scientific knowledge (Bergen, Johnson, Hoop, & Beroza, 2019).

Pattern recognition tools, including Machine Learning (ML) algorithms were born from a cross-fertilization of ideas between computer science and statistics and have been applied in a variety of ways within marine science. As a few examples, ML has been used to determine relevant components of nearshore systemic dynamics (Beuzen et al., 2018b; Biel, Hacker, & Ruggiero, 2019; Bulteau et al., 2015; Eadi Stringari, 2020), make predictions of morphology or runup statistics (S.-T. Chen, 2019; Nieves, Radin, & Camps-Valls, 2021; Pape, Plant, & Ruessink, 2010), or emulate flooding, storm surge or wave models (James, Zhang, & O'Donncha, 2018; Kyprioti, Taflanidis, Nadal-Caraballo, & Campbell, 2021; Parker, Ruggiero, Serafin, & Hill, 2019).

The marine science community may be reluctant to adopt ML techniques due the opaque nature of ML algorithms and a resulting lack of trust in ML algorithm predictions. Marine scientists often use theoretical models based on first principles (e.g., Roelvink, McCall, Mehvar, Nederhoff, and Dastgheib (2018); Tolman and Chalikov (1996)), where cause and effect are readily seen. In contrast, ML models are often "black box" techniques. Therefore, in order to adopt an ML technique, it is desirable to develop trust in the proposed ML algorithm by ascertaining that the ML predictions are made in accordance with the established theory or at least with sound reason. Techniques to improve the transparency of artificial intelligence are generally called "explainable AI" (Doran, Schulz, & Besold, 2017; Goebel et al., 2018; Holzinger, 2018). Herein, the ML techniques are queried to determine how the ML predictions were made to increase the transparency of the ML technique.

An additional benefit of increasing the transparency of ML decision making

is leveraging the information about ML decision making for Knowledge Discovery (KD) about marine systems. In a KD process, knowledge about the system is learned by interpreting the results of data mining techniques (Cios, Pedrycz, & Swiniarski, 1998). Blind interpretation of ML results, however, could lead to incorrect conclusions since the ML technique might find unimportant correlations (Wadoux, Samuel-Rosa, Poggio, & Mulder, 2020). Therefore, a robust ML KD process is a hybrid approach which combines established domain knowledge with the pattern recognition ability of an ML technique to lead to new conclusions about a system. An additional limitation in using ML KD is that, while correlations are revealed, the details about the causation of phenomena might remain unknown (Goldstein, Coco, & Plant, 2019). Beuzen et al. (2018a) illustrates a KD example (called a "descriptive" ML application) within coastal geomorphology using Bayesian Networks, where a specialized Bayesian Network is inspected to determine the combination shoreline factors correlated with coastal flooding. Herein, MLKD is performed using hybrid approaches in two ways: in chapter two, an ML routine is coupled with an established numerical model to determine areas of model bias and corrections to wave forecasts; in chapters three and four, the ML technique expands the identification of nearshore morphology based off a well established conceptual model.

A powerful aspect of scientific discovery through ML is increasing the temporal and spatial scales of data analysis. In order to resolve these scales, ML algorithms must be generalizable enough to transfer between locations or time periods. The challenge in applying a ML technique to marine science arises in that each marine ecosystem and landscape is unique; two locations might exhibit similar patterns, but still be subject to phenomena that differentiate the two sufficiently well such that ML trained at one location cannot be applied to a different location. Temporally, forecasting variables faces the challenge of non-stationarity in the time series. If the trend of a time series changes over time, an ML technique trained on a previous time period might make inaccurate predictions for a different time period. Herein, the ML techniques are applied at different geographical locations

or time periods to determine transferability skill.

Advancements in traditional research ML has led to more complex and powerful algorithms called deep ML (LeCun, Bengio, & Hinton, 2015). In traditional machine learning, the raw data are used as input features to an ML task such as classification. In deep ML, the raw data is transformed into representations by several layers of transformations, which are then used as input features to the ML task. This dissertation encompasses the two types; traditional ML in chapter 2 and deep in chapter 3. In this dissertation, the traditional ML technique has the advantage of being more readily transparent and interpretable than the deep ML technique.

The data types, model types, ML routines and domain of the two chapters are distinct. The second chapter focuses on off-shore waves, the model is numerical, and the data type are observations of continuous variables. Wave models, currently providing operational forecasts globally, are numerical models that solve an energy balance equation to predict a wave spectrum (Tolman & Chalikov, 1996). They are subject to ongoing research and improvement, and are validated by observations of wave spectra or bulk parameters made by remote sensors such as offshore buoys or satellites (Björkqvist, Vähä-Piikkiö, Alari, Kuznetsova, & Tuomi, 2020; A. Ellenson & Özkan-Haller, 2018; Ortiz-Royero & Mercado-Irizarry, 2008; Stopa, Ardhuin, Babanin, & Zieger, 2016b). A transparent ML technique, a decision tree, was applied to improve the accuracy of wave model predictions. A decision tree is distinct from other ML tools, such as neural nets, in that it can be visually inspected to determine the parameter combinations used to optimize its predictions (Etemad-Shahidi & Mahjoobi, 2009). The objectives of chapter 2 are to increase the skill of the numerical model, determine areas where wave model improvements can be made by investigating the ML technique decision making, and also to determine the transferability constraints of the technique with respect to different geospatial locations and time periods.

The second and third chapters focus on coastal geomorphology, the model is conceptual, and the data type are images. The dynamics and theory comprising

nearshore geomorphology, in contrast with wave models, include feedback dynamics between solid and aqueous phases (Roelvink et al., 2018). As a result, the governing equations require resolution of complex interactions between sediment transport and hydrodynamics that might be parameterized empirically. The model in consideration, beach state classification, views nearshore morphology on a temporal and spatial scale wherein the aforementioned small-scale dynamics are not resolved.

Beach state classification is a qualitative conceptual model that establishes a categorization system for frequently occurring morphological patterns and a theory on morphological pattern evolution (Lippmann & Holman, 1990; L. D. Wright & Short, 1984). Beach states encompass a description of the nearshore surf zone bathymetric shape and associated physical dynamics. For example, one category might be associated with rhythmicity and a high prevalence of rip currents. Observational studies have found that the beach state affects hydrodynamics at the shoreline (Gomes da Silva, Coco, Garnier, & Klein, 2020) and therefore changes in beach state can alter the shoreline. Quantification of instantaneous and future beach state has been used to model shoreline change (Davidson, Splinter, & Turner, 2013). While the beach state framework is useful in that it simplifies complex morphological patterns into readily understood categories, the qualitative nature of the framework results in a degree of subjectivity in beach state labelling.

A Convolutional Neural Network (CNN) is applied to detect beach states from imagery at two locations. The imagery used is Argus daytime exposure imagery, which is an average of video frames of wave breaking (Holland, Holman, Lippmann, Stanley, & Plant, 1997). Waves generally break on a sandbar, and so morphology can be inferred from wave breaking patterns. The detection technique is applied at two different locations. The objectives of this chapter are to determine the appropriateness of applying a CNN to detect a beach state and to determine its transferability between locations. Given the subjectivity in labelling a beach state, the CNN skill is determined adequate by comparing its predictions with the labelling agreement between the authors. Finally, to determine appropriate decision

making, the CNN was queried to confirm that it was indeed identifying the right characteristics associated with each class.

The simplification of the nearshore morphology into discrete beach states results in a decrease in descriptive accuracy and a mis-perception of the reality that the nearshore morphology exists in more complicated shapes. Nearshore morphologies can exhibit alongshore variability of beach state, and therefore might not be readily categorized into a discrete shape. The fourth chapter builds off the third in that it uses a CNN to detect morphological characteristics. However, the fourth chapter expands the beach state framework to encompass alongshore variable beach states. There are two goals in this chapter: the first is to increase the generalizability of the beach state framework such that the automated technique can be applied to high volume datasets of Argus imagery and maximize the amount of information extracted; the second is to characterize the temporal variability and distribution of beach simplices from a 28 year dataset of Argus imagery.

Chapter 2: An Application of a Machine Learning Algorithm to Determine and Describe Error Patterns within Wave Model Output

A. Ellenson, Y. Pei, G. Wilson, H. Tuba Özkan-Haller, Xiaoli Fern

Coastal Engineering
https://doi.org/10.1016/j.coastaleng.2019.103595
Volume 157

ABSTRACT: This study uses a machine learning algorithm, the bagged regression tree, to detect error patterns within 24-hour forecasts of significant wave height time series. The input to the machine learning algorithm were bulk parameter outputs of the numerical wave model (WaveWatch III) and wind information from the Global Forecast System at buoy locations along the California-Oregon border in the United States. The output of the algorithm are predictions of hourly deviations between numerical model output and buoy observations of significant wave height. When these deviations were applied as corrections to the forecasts, error metrics root-mean-squared-error, bias, percent error, and scatter index were reduced in several different experiments, confirming that the error pattern was successfully detected by the machine learning algorithm. Furthermore, the detected error pattern was consistent between buoys at different locations, as presented in a geo-spatial application of the machine learning algorithm. As a descriptive tool, the algorithm delineated regions of similar error within the context of model phase space (significant wave height and mean wave period). Specifically, the algorithm detected significant wave height overestimations for significant wave heights greater than 3.4 m, wave period greater than 9.1 s, and waves coming from the W-NW quadrant. Also, for significant wave heights greater than the 95th percentile value (5.4 m), the algorithm detected differences in mean error patterns.

2.1 Introduction

Forecasts or hindcasts of ocean wave conditions provide hazard warnings for residential coastal communities, information about environmental conditions for recreational and commercial coastal populations, and information about historical wave conditions in areas where observations of significant wave height do not exist Appendini, Torres-Freyermuth, Salles, López-González, and Mendoza (2014); García-Medina, Özkan-Haller, and Ruggiero (2014); Guedes Soares, Rusu, Bernardino, and Pilar (2011). Coastal development involving structures or marine renewable energy rely on this information to make design decisions.

Ocean wave predictions are products of physics-based numerical wave models

that are governed by equations which describe the physical processes involved in the generation, propagation, and dissipation of wave energy represented by a wave energy density spectrum. Generally, these models require two main inputs, bathymetry and wind fields, and produce a variety of output products, most often in the form of maps or time series of bulk wave parameters such as significant wave height, mean wave direction, and mean wave period. These models are prone to errors which can be due to inaccurate input information or imperfections in the governing equations and parameterizations that are used in the wave model.

Machine learning techniques use statistics to determine patterns between inputs and outputs. One application of these techniques within wave forecasting has been to generate significant wave height forecasts statistically, i.e., without reference to a physics-based model. Environmental information at a specific location, such as time-lagged wind or significant wave height observations are often the primary input, and significant wave height predictions are the primary output. Studies have used time-lagged wind or wave information as input to the machine learning techniques to predict significant wave height into the future Mudronja, Matić, and Katalinić (2017). Machine learning techniques includeArtifical Neural Networks (ANNs), adapative neuro-fuzzy inference systems (ANFIS), support vector machines (SVM), and Bayesian networks (BN). The ability for different machine learning techniques to predict significant wave height given previously observed wind data was studied in Malekmohamadi, Bazargan-Lari, Kerachian, Nikoo, and Fallahnia (2011). They found that ANNs, ANFIS, and SVMs work well, whereas BNs resulted in less accurate results. In Berbić, Ocvirk, Carević, and Lončar (2017), they used SVMs and ANNs to predict significant wave height up to 6 hours into the future using time-lagged wind information as input. In Zamani, Solomatine, Azimian, and Heemink (2008), they use three different ANNs to predict significant wave heights 1, 3, and 6 hours into the future, and contrast these ANNs with an Instance Based Learning technique wherein they included spatial information for significant wave height predictions. In Nikoo, Kerachian, and Alizadeh (2018), they compare bayesian networks, regression tree, fuzzy k-

nearest neighbor, and support vector regression techniques to forecast significant wave height given wind data, including wind direction as well as wind speed. They recommended finding a method that also can predict the spatial structure of the wave field.

Other studies have used neighboring spatial information (wind or significant wave height error) to predict significant wave height at a location of interest Peres, Iuppa, Cavallaro, Cancelliere, and Foti (2015). Another application has involved the generation of data at locations where information was missing. For example, data from neighboring buoys can be used as input to a statistical technique to generate time series for a buoy with a data gap (Kalra & Deo, 2007; Tsai, Lin, & Shen, 2002).

The studies discussed above primarily concerned the ability for machine learning algorithms to produce accurate wave forecasts, without explicit reference to wave physics. In contrast, this study leverages the strengths of both physics-based and machine learning methods in a hybrid approach wherein a variation of the decision tree is used to correct a physics-based model. This hybrid approach can improve predictions, yielding more accurate results than either methods alone (Berbić et al., 2017; Reikard, Pinson, & Bidlot, 2011; Woodcock & Engel, 2005; Woodcock & Greenslade, 2007). In particular, neural networks have been used as a post-processing routine to correct the original wave model predictions of wave parameters (Deshmukh, Deo, Bhaskaran, Nair, & Sandhya, 2016; Hadadpour, Moshfeghi, Jabbari, & Kamranzad, 2013; Moeini, Etemad-Shahidi, Chegini, & Rahmani, 2012; Moeini, Etemad-Shahidi, Chegini, Rahmani, & Moghaddam, 2014; Zhang, Li, Li, & Qi, 2006).

The machine learning algorithm of interest in this study is the bagged decision tree, an ensemble method of the decision tree, which has been applied in a variety of disciplines. Specific to wave forecasting, Mahjoobi and Etemad-Shahidi (2008) and Jain, Deo, Latha, and Rajendran (2011), regression trees were compared with ANNs to predict significant wave height in Lake Superior in Etemad-Shahidi and Mahjoobi (2009), and their performance was marginally better. In Etemad-Shahidi

and Bonakdar (2009), the decision tree was used to predict wave run up. Its predictions were more accurate than an established engineering empirical formula.

As noted in other studies, the advantage to decision tree over other machine learning techniques such as ANNs is that its logic is transparent and can be readily understood through visual inspection or by extracting the decision rules in a post-processing routine Etemad-Shahidi and Bonakdar (2009); Etemad-Shahidi and Mahjoobi (2009). Because of this, the decision tree offers an opportunity to learn about patterns within the data being studied. The decision tree is both predictive and descriptive, since the architecture of the algorithm reveals relationships within the data. Other machine learning algorithms, such as Bayesian networks or fuzzy-logic based systems, share this quality and can be used to infer information about large data sets. In Beuzen et al. (2018a); Gutierrez, Plant, Thieler, and Turecek (2015), the rules of Bayesian networks were interpreted to determine the relevant physical drivers in shoreline change processes. In Cornejo-Bueno, Rodríguez-Mier, Mucientes, Nieto-Borge, and Salcedo-Sanz (2018), a fuzzy logic system was used to predict significant wave height and energy flux. The algorithm was able to capture relevant predictors for the different physical regimes. Sea state was better predicted by local meteorological values (i.e., air temperature, water temperature and wind speed), and swell was better predicted by surrounding spatial information as well as meteorological values (i.e., significant wave height values from surrounding buoys, mean wave direction and atmospheric pressure).

In this study, decision tree is used in both descriptive and predictive senses. It is trained to predict wave model error, and the construction of the algorithm delineates regions of model phase space in which systematic model biases occur. For example, if physics-based model predictions consistently underestimate large significant wave height for a certain wave direction and significant wave height combination, this approach detects this significant wave height/wave direction combination and determines a mean underestimation value. This can provide information to model developers about areas of model inadequacies and potential areas of wave model improvement. This algorithm also shows potential in deter-

mining geo-spatial error patterns, since it can generate accurate error predictions at neighboring locations where data has not been provided. This could be used for correcting gridded wave model output, where measurements are not available.

The technique is applied to several months long 24-hour significant wave height forecasts generated for buoy locations within the Northeast Pacific region, as shown in Figure 2.4. Results are presented for two different wave climates: high and low energy (winter and summer, respectively). The input features are comprised of 24-hour predicted time series of significant wave height, mean wave direction, mean wave period, wind direction, and wind magnitude. The potential for the technique to be extrapolated to other locations is also demonstrated.

2.2 Methods

2.2.1 Bagged Regression Tree

The goal of a machine learning algorithm is to relate a vector of inputs (or "input features," $\mathbf{x}$) and their associated output values (or "targets," y). Input feature-target pairs $((\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N))$ are called instances. The algorithm learns patterns during a training phase, which are then used as a basis upon which to make predictions during the testing phase.

Machine learning algorithms, or "learners," have different strengths and weaknesses depending on base logic used to learn patterns. In order to generate a stronger learner, several different learners or multiple realizations of the same learner can be combined in an ensemble technique. The ensemble technique dictates how the learners are combined. The method used in this study is an ensemble method where many realizations of the same base learner are used. (Dietterich, 2000).

The base learner used in this study is a regression tree, and belongs to a broader class of methods known as Classification and Regression Tree (CART) (Breiman, Friedman, Stone, & Olshen, 1984). The algorithm used by regression tree is to map input features to a predicted target value by splitting the instances into disjoint

sets, or partitions. In the entire data set, T, each partition, t, encompasses a sub-set of instances ($(\mathbf{x}_i, y_i) \in t$) and is associated with a target value prediction. The elements of the hourly input feature vectors ($\mathbf{x}_i$) are modelled environmental parameters associated with the forecast for that hour, i. These include significant wave height, Hs, mean wave direction, MWD, mean wave period $Tm01$, wind magnitude, $wndmag$, and wind direction $winddir$. See A.2 for definitions of the bulk wave parameters. The output target (y_i) for this study is the difference between the significant wave height observations, $H_{s,obs}$, and modelled significant wave height, $H_{s,WW3}$ for a particular hour, i:

$$y_i = H_{s,obs} - H_{s,WW3} \tag{2.1}$$

A negative (positive) target value indicate a significant wave height overestimation (underestimation). In Figure 2.1, the 2012-2013 target values, y_i, are shown within the feature space of the input parameters H_s, T_{m01} and MWD. The target values are the WW3 Hs prediction errors for the 2012-2013 winter seasons (defined by Equation 2.1). This feature space, T, is the space that the decision tree will divide into partitions.

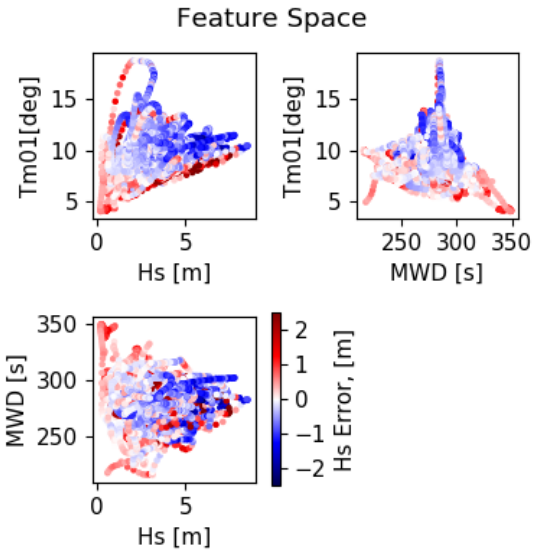


Figure 2.1: The 2012-2013 winter target values (colored points) plotted as a feature space with respect to the input features. The target values are the WaveWatch III Hs prediction error (Equation 2.1), and the three input feature elements are WaveWatch III Hs, MWD and $Tm01$.

The objective is to find the partitions wherein the mean target value of that partition ($\overline{y}(t)$) is most similar to the rest of the target values encompassed within that partition. Specifically, the algorithm seeks the partitions, t, in the entire dataset, T, that minimize the sum of variances in y_i across all the partitions:

$$\frac{1}{N}\sum_{t\in T}\sum_{\mathbf{x}_i\in t}(y_i-\overline{y}(t))^2 \tag{2.2}$$

These partitions are constructed by splitting the target values (y_i) with respect to values of the input features ($\mathbf{x}_i$). The input features can be used more than one time in determining final partitions. During the training phase, the regression tree establishes partitions by determining the maximum variance reduction within the entire data set according to Equation 2.2.1. See Figure 2.2 for an example of how the data is successively split on threshold values of input features, establishing a

tree structure. At each level of the tree, decisions are made to split the data set based on a threshold value of an input feature to minimize the variance of the entire dataset, thus establishing 'branches.' In the final partitions, the mean target value is determined for the data associated with that partition ($\overline{y}(t)$). During testing, each instance is categorized into a partition due to the values of the input features, $\mathbf{x_i}$, that correspond with the thresholds determined during training. Each partition is associated with a mean target value ($\overline{y}(t)$, established during training) which is the prediction for the instances which fall into the associated partition. See Figure 2.3 for an illustration of the final partitions, t, made relative to the entire dataset T.

Through partitioning the data space by minimizing Equation (2), the decision tree finds clusters of points with similar over- or under- estimations and associated with the same environmental context. The environmental context, defined by the tree structure, is readily understood through visual interpretation and extraction of the decisions made to establish the partitions (see Section 2.4.2 for examples). The decision tree structure can therefore be used as a diagnostic tool for model developers to find systematic errors within wave model output. This makes the decision tree method more transparent than other "black box" machine learning algorithms, such as artificial neural networks. The interpretation of the decision making process for artificial neural networks is a subject of ongoing research (Koh & Liang, 2017).

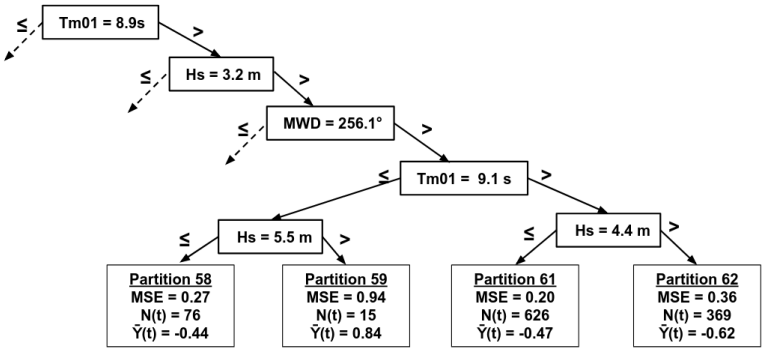


Figure 2.2: Representative tree structure of the splits made between the input features during training on the training data set. The input features which the decision tree splits on are indicated within each node with a threshold value. Branches to the left (right) represent the division of data points less than (greater than) the threshold value. In the final partitions, MSE is the mean squared error between the member target values within the partition and the mean target value associated with that partition. $\overline{y}(t)$ is the mean target value associated with that partition.

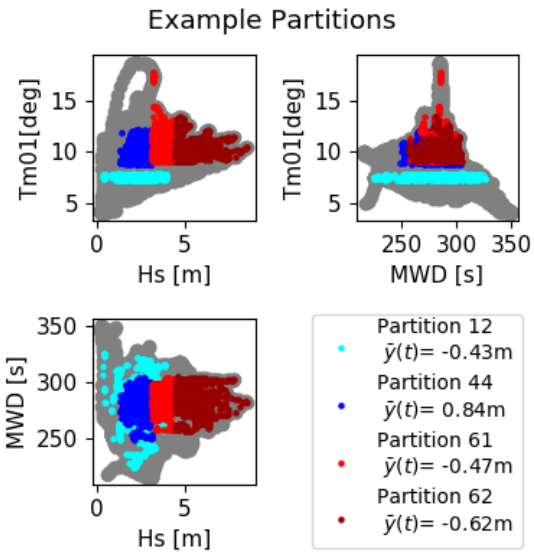


Figure 2.3: An example of three final partitions determined after the decision tree split the entire dataset with respect to values of bulk wave parameters. These partitions correspond with partitions 61 and 62 that are represented in the tree structure in Figure 2.2.

The ensemble technique, wherein many trees are trained, is called bagging (short for bootstrap aggregation). In this technique, each tree is trained on unique subsets of the entire learning dataset. The subsets are comprised of instances that are chosen with replacement. The subset of samples are chosen in such a way that an instance can be found within a subset several times or not at all, according to Breiman (1996). The final prediction is an average of the predictions made by all of the base learners within the bagging scheme. The base learner, the regression tree, combined with the ensemble technique, bagging, is referred to as a "bagged regression tree."

The number of splits made on a learning data set is referred to as the depth of a tree, and the ensemble size used to make the final prediction in the bagged regression tree is referred to as the number of trees. The depth and number of trees of each bagged regression tree is determined during the training phase through

cross validation, as described by Breiman and Spector (1992), and in the following section. Also, the data might be split with respect to one feature more often than another, indicating that the feature is important in its decision making rules.

2.2.2 Parameter Selection

During the cross-validation portion of training optimal tree depths and number of trees were selected by comparing performance on a validation set via root-mean-squared-error. The tree depths tested were 2, 3, 4, 5, 10, 15, 20, 30. The number of trees tested were 10, 30, 50, 100, 150, 200, 250 trees. To determine the optimal forest parameters, five-fold cross validation was implemented during the training phase Breiman and Spector (1992). This consisted of first shuffling the data using a Fisher-Yates shuffling routine, then dividing the learning data into five subsets, or folds. Four folds served as a training set and the fifth fold served as a validation set. Each combination of depth and number of trees was trained on four folds, and tested on the last validation fold. The performances of each bagged regression tree was assessed by the root-mean-squared-error of the predictions of the validation fold. This process was repeated five times, where each fold served as a validation set at least once. The bagged regression tree that resulted in the lowest average RMSE between the five validation runs was the one used for post-processing the wave forecasts.

Each run (where a run consists of a train/test cycle) is associated with one bagged regression tree. The output of a run is stochastic in nature due to the randomness associated with the bagging method as well as shuffling the time series. Therefore, each run was performed thirty times, and the mean value of the thirty runs is reported here as the final value for each experiment.

2.2.3 Experiments

Experiments were designed to determine the performance of the bagged regression tree for different wave climates and different input feature combinations. The data

was divided into summer and winter seasons to test the bagged regression tree on different wave climates. In the Eastern North Pacific, the winter (October-March, inclusively) is characterized by a more energetic wave climate and is associated with larger significant wave heights and longer periods as compared with summer (April-September, inclusively). The winter data set was characterized by an average significant wave height of 3.1 meters and an average mean wave period ($Tm01$) of 9.3 seconds. The summer data set was characterized by an average significant wave height of 1.8 meters and average mean wave period of 7.2 seconds. The seasonal data was further divided into training and testing sets. The training set consisted of the years 2012-2013 and the testing set consisted of the year 2014. This corresponds to training sets comprised of 6,228 and 6,303 instances (input feature/target pairs), and a testing set of 3,864 and 3,855 instances, for winter and summer, respectively.

The characteristics of these experiments are presented in Table 2.1. The first two experiments were single-buoy experiments, and used training data from buoy 46050. In the first experiment, input features included only the forecasted bulk parameters (Hs, $Tm01$ and MWD) as provided by the wave model, and will be called "Waves Only." In the second experiment, input features also included wind information wind direction ($winddir$) and wind magnitude ($wndmag$) in addition to wave information and will be called "With Wind." These first two experiments were performed for summer and winter. The input feature/season combination which resulted in the best performance ("Winter - With Wind") was then selected for the third experiment, whose purpose was to test the ability for the algorithm to be extrapolated to different points in geographical space. The algorithm was trained on the data from four other buoys in the region in different water depths (see Figure 2.4), excluding buoy 46050. The algorithm was then tested on this excluded buoy. Because it was trained on four different buoys, this test will be called "Multiple Buoys."

Table 2.1: The experiment titles, their respective input features, and training and testing data sources.

Experiment Name	Input Features	2012-2013 Training Data Source	2014 Testing Data Source
Waves Only	Hs, MWD, $Tm01$	Buoy 46050	Buoy 46050
With Wind	Hs, MWD, $Tm01$, wind direction, wind magnitude	Buoy 46050	Buoy 46050
Multiple Buoys (Winter Season, "With Wind")	Hs, MWD, $Tm01$, wind direction, wind magnitude	Buoys 46211, 46243, 46029, 46027	Buoy 46050

2.2.4 Data Sources

This study uses wave model output, wind model output, and observations of significant wave height. The wave model output are 24-hour time horizon forecasts of bulk parameters significant wave height, mean wave period ($Tm01$) and mean wave direction, as provided by WaveWatch III (WW3) from the years 2012-2015 (García-Medina, Özkan-Haller, Ruggiero, & Oskamp, 2013). The wave model was developed as an operational forecasting tool for the Pacific Northwest Coast of the US. A validation of the model by García-Medina et al. (2013) using two three-month hindcasts showed a normalized root-mean-squared-error of 0.2 m for significant wave height and 0.15 s for mean wave period ($Tm01$). The wind model input to the wave model are 24-hour time horizon wind forecasts of wind direction and wind magnitude provided by the Global Forecasting System (GFS). These wind data were interpolated into the wave model longitudinal and latitudinal gridpoints and output by the wave model. This forecasting model configuration uses a mosaic of nested grids of increasing resolution, following Tolman (2008). The outer-most grid has a resolution of 1.25 ° longitude by 1.00° latitude and spans from 77°S

to 77°N. The intermediate grid spans the Eastern North Pacific (5° to 60.25°N in latitude and 170°W to 177° in longitude) with a resolution of 15 arc-minute. The third grid, the outer shelf grid, spans a region of 41.45° to 47.50°N and 127° to 123.75°W with a resolution of 3 arc-minute. The model is forced by Tolman and Chalikov ST2 physics Tolman and Chalikov (1996) and 24-hour lead time GFS 3-hour 10 meter winds and air-sea temperature differences at a resolution of 0.5°. Significant wave height observations were taken from National Data Buoy Center (NDBC) buoys along the Oregon-California coast and are shown in Figure 2.4. The single-buoy experiments used data from buoy 46050. Additional buoy data from buoys 46027, 46029, 46211 and 46243 were used in the experiment labelled "Multiple Buoys." These locations were included in training because they were likely to be subject to similar environmental conditions (i.e., wave events and atmospheric patterns) due to their proximity. Table 2.2 lists the buoy label, depth, latitude, longitude, type of buoy and distance from test buoy 46050. For all experiments, the bagged regression tree was tested on data from buoy 46050.

Table 2.2: Depth, latitude, and longitude for each buoy used in the Multiple Buoys experiment

Buoy	Depth [m]	Latitude [°N]	Longitude [°W]	Distance from 46050 [km]	Buoy Type
46050	140	44.677	124.515	0	3m directional discus
46027	46	41.850	124.386	314	3m directional discus
46029	134	46.143	124.485	163	3m directional discus
46211	40	46.858	124.244	243	Directional Waverider
46243	24	46.216	124.128	174	Directional Waverider

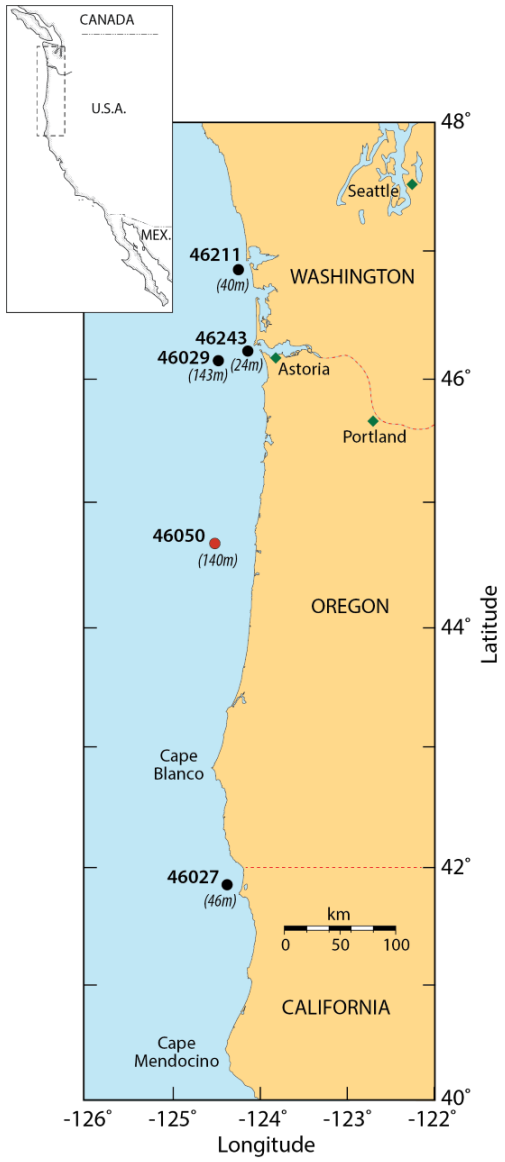


Figure 2.4: The locations and water depths of the buoys implemented in this study. Predictions and error metrics were computed using buoy 46050 (red dot). In the geospatial application, the bagged regression tree was trained with data from the other locations. Inset shows the geographic location of the domain (dashed rectangle) on the U.S. west coast.

2.3 Results

2.3.1 Bagged Regression Tree Structure

The parameters controlling the bagged tree structure (tree depth and number of trees), were determined using cross validation, as explained in section 2.2.2. The tree depths for each respective season were less than five for each of the single-buoy experiments. For summer, the optimal trees tended to have larger depths for the With Wind experiment (average tree depth of 4.0) compared to the Waves Only experiment (average tree depth of 3.6). For winter, the average optimal tree depth was the same (4.9) for both experiments. The number of trees required for the experiments ranged between 109 and 127, where the With Wind experiments required more trees than the Waves Only experiments (127 versus 121 for winter, and 127 versus 109 for summer). For the multiple buoy experiment, the optimal trees had a depth of 10 and the number of trees was 173.

2.3.2 Application to Wave Forecasts

The application of bagged regression tree corrections to the original numerical model output time series resulted in more accurate significant wave height predictions for both summer and winter seasons. See Table 2.3 for error metrics between each modelled and corrected time series and the observed data. Representative portions of each season's time series are presented in Figure 2.5. The improvement in significant wave height predictions due to the application of the bagged regression tree indicates that in the majority of cases the bagged regression tree successfully detected over- and under- estimations in wave model output within the wave model phase space. The experiment wherein the input features included wind information wind magnitude and wind direction (With Wind) resulted in slightly greater accuracy of significant wave height predictions (within one one-hundredth of each error metric, see Table 2) than the experiment that did not include the wind input feature information (the Waves Only experiment). The results of the

With Wind experiment are reported below.

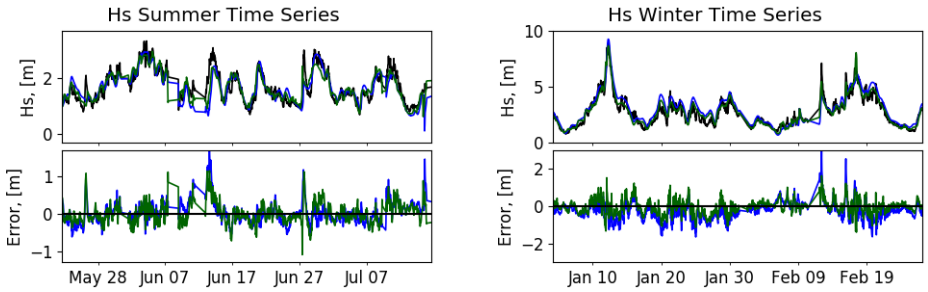


Figure 2.5: Representative time series of significant wave height for summer and winter. Gaps in the data exist due to wave model restarts or gaps in observations. The original model significant wave height output is colored blue, the observations (from NDBC buoy 46050) are colored black, and the corrected significant wave height time series is colored green. Note the y-axis for winter is greater than for summer. The bottom panels show the error as the difference between the observations and the model significant wave height output (blue), and the corrected time series from the With Wind experiment (green). The majority of the time, the application of the corrections resulted in more accurate Hs predictions, but sometimes resulted in less accurate Hs predictions, such as the underestimations in the summer of late June.

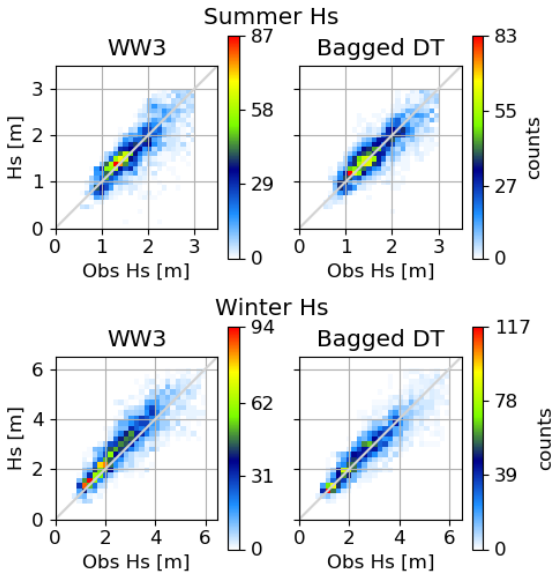


Figure 2.6: Density scatter plots (colors represent histogram counts) of predicted versus observed significant wave height. The data shown comprise all data from the With Wind experiments, separated into summer (top panels) and winter (bottom panels) testing periods. Left-hand panels show results from the original WW3 model, and right-hand panels show results after applying corrections from the bagged decision tree. To create the histograms, data were binned in equally spaced increments of 0.1 m for summer and 0.2 m for winter.

Figure 6 shows density scatter plot comparisons of the significant wave height data before and after applying the bagged regression tree corrections. For both summer and winter, the bagged regression tree reduced the number of overestimated significant wave heights, but increased the number of underestimated significant wave heights. Overall, for both seasons, the application of the significant wave height corrections resulted in an increase in accuracy. The improvement of bias for the corrected time series was statistically significant for winter but not for summer, with a confidence level of 95% (p-values of 0.00 and 0.54 for winter and summer, respectively).

Table 2.3 shows that error metrics root-mean-squared-error (RMSE), bias, scat-

ter index (SI) and percent error (PE) of the significant wave height time series were reduced for both seasons after the application of the bagged regression tree corrections (see A for definitions of error statistics). The application of the bagged regression tree corrections resulted in more accurate predictions of significant wave height error for winter than for summer. For winter, RMSE, bias and SI were reduced by 16%, 92% and 19%, and for summer, RMSE, bias and SI were reduced by 12%, 0% and 12%.

Table 2.3: Error metrics calculated between the original model output and the observations (labelled WW3), and the error metrics calculated between the various bagged regression tree experiments and the observations (labelled Waves Only and With Wind for the Summer and Winter).

Test	RMSE [m]	Bias [m]	PE [%]	SI [–]
Summer WW3	0.33	-0.01	18	0.19
Summer Waves Only	0.29	-0.01	17	0.17
Summer With Wind	0.29	-0.01	17	0.17
Winter WW3	0.58	0.24	23	0.21
Winter Waves Only	0.49	0.02	18	0.18
Winter With Wind	0.48	0.03	17	0.17
Winter Multiple Buoys With Wind	0.48	0.03	17	0.17

For both seasons, the results show that the bagged decision tree had more skill at correcting errors for data associated with wave periods greater than 6 s than the data associated with wave periods less than 6 s. To determine this, the two significant wave height time series (the original and the bagged regression tree) were binned with respect to the observed mean wave period, and error metrics were calculated (see Figure 2.7). The bins consist of very short period (0-6s), short period (6-8s), mid-period (8-12s) and long period (12-18s). For winter, most of the data points (N = 1094 points) were within the middle mean wave period ($Tm01$) bin. For summer, most of the data points (N = 2077) were within the short mean wave period bin (6-8s) and no data points were associated with long

mean wave periods. For summer and winter, Figure 2.7 shows that the application of the bagged regression tree corrections reduced error for the significant wave heights associated with middle mean wave periods, and increased error (i.e., had negative skill) for the significant wave heights associated with short mean wave periods. Similar to the overall bulk parameters, the absolute bias was reduced more than RMSE.

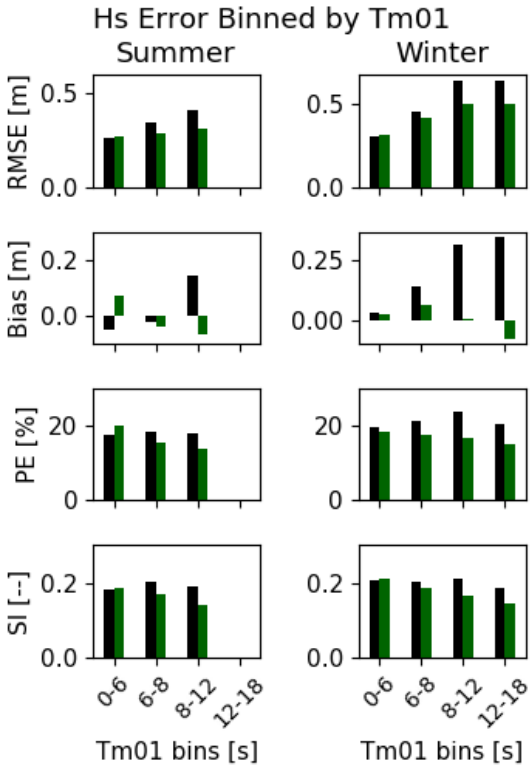


Figure 2.7: Error metrics binned by mean wave period $Tm01$ for summer (left) and winter (right). The WW3 error is colored black and the With Wind experiment is colored green. Note that no data points are associated with the longest period bin for summer. The greatest error reductions are achieved for significant wave heights associated with mean periods greater than 8s.

2.3.3 Results: Geo-spatial Application

The bagged regression tree corrections were also applied in such a way as to explore the potential for a geo-spatial application. In this experiment, the bagged regression tree was applied at a location where no prior training information was provided; instead, the bagged regression tree was trained on four other buoys in the region, and tested at the original location (buoy 46050, see Figure 2.4. Additional geo-spatial information in the form of longitude, latitude and water depth were included as input features. The output target remained the same (significant wave height error). The best performing experiment, the With Wind - Winter season was chosen as the input feature/season set-up. The results presented in Table 2.3 show that inclusion of data from other buoys resulted in similar accuracy of the bagged regression tree corrections as compared with the original Waves Only - Winter experiment. The improvement in significant wave height predictions shows the potential to apply the bagged regression tree to correct wave forecast output at neighboring geographical locations within a modelled region, where the absence of training information at a specific location would be analogous.

2.4 Discussion

2.4.1 Bagged Regression Tree as a Diagnostic Tool of Model Error

The fact that the decision tree is able to detect (and correct) errors in the WW3 model shows that such errors are predictable (i.e., not purely stochastic), and hence may be associated with times when an incomplete understanding of physical processes results in errors within the model parameterizations which represent these physical processes. As a primary example, in the following, we will analyze the instances associated with the most populated partition (partition 31) and where the bagged regression tree skill was highest (data associated with observed mean wave periods ($Tm01$) between 8-12s). Note, the intent of this analysis is to use the decision tree outputs to search for clear patterns in the wave model error, which is

the reason for isolating the data for which the decision tree predictions were most accurate; that is, the algorithm is used here as a "diagnostic" tool rather than a predictive one in order to observe its diagnostic ability (Beuzen et al., 2018a). The 2012-2013 training data from partition 31 are plotted in Figure 2.8. The partition encompasses data associated with GFS winds that are less than 15.6 m/s, modelled significant wave heights that are greater than 3.4 m, modelled mean wave periods ($Tm01$) that are greater than 9.1 s, and modelled mean wave directions are from the W-NW quadrant (greater than 245° and less than 360°). The correction for this partition is -0.57 m, indicating an overestimation of significant wave height. The instances within this partition are also associated with mean wave period overestimations for the lower mean wave periods. The bias at mean wave periods 8-9s was 1.55 +/- 1.00 s, while the bias for mean wave periods greater than 10s was small and negative, on average. The mean wave period overestimation suggests that the distribution of the variance density across frequencies of the wave model was different than those of the observations. One possibility is that the low frequency wave energy was over-estimated by this physics package (Tolman & Chalikov, 1996) which is consistent with the conclusions of Stopa, Ardhuin, Babanin, and Zieger (2016a), although the present analysis using bulk wave parameters cannot directly confirm this. Further analysis of wave spectral data would be required to confirm this hypothesis, and is suggested as future work.

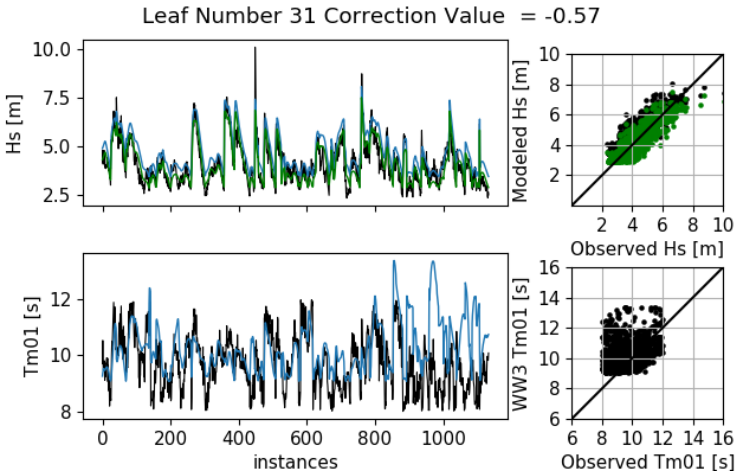


Figure 2.8: Data points of the most populated partition, partition 31. The left panels show the instances of NDBC buoy 46050 observations (black), WW3 (blue), and corrected (green) Hss. The top right panel shows the original WW3 (black) and corrected (green) instances as a scatter plot. The bottom left panel shows the instances of observed (black) and WW3 (blue) $Tm01$. The bottom right panel shows the $Tm01$ instances as a scatter plot.

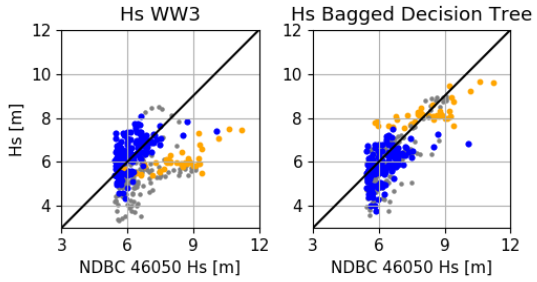
Special attention was given to the bagged regression tree treatment of data for extreme wave heights, due to the interest given to extremes within the wave modeling community (Ruggiero, Komar, & Allan, 2010). To do so, an analysis was performed on the subset of data where the significant wave height exceeded the 95th percentile value, $Hs > 5.39m$. The bagged regression tree detected clusters of different error character (i.e., more or less error) within the data that had these large significant wave heights, as shown in Figure 2.9 and explained in the following. There are a total of 342 points with $Hs > 5.39m$ in the training data, and these were categorized into 23 different partitions by the bagged regression tree. The partitions that have the greatest number of data points associated with these significant wave heights are partitions 31 and 44 (encompassing 173 and 45 points, respectively; see Figure 2.9). Note that the 173 data points are a subset of data within partition 31, the entirety of which is illustrated in Figure 2.8. By definition

the correction value applied to the data in partition 31 reduced the bias from 0.57m to 0.00m for the data in this partition. However, when this bias correction value of -0.57m is applied to the large wave heights only, it results in a decrease of skill for this cluster of points (original bias of 0.24 m to -0.33 m). This decrease in skill is primarily a result of the three data points where NDBC 46050 $Hs > 8m$; when these data points are removed, the application of the bias correction value results in similar skill (albeit bias reverses from 0.27 to -0.29). If the decision tree had applied the same bias to all other points $Hs > 5.39m$, it would have resulted in a decrease in skill. Instead, it differentiated other clusters of points and resulted in an overall increase in skill for these data (bias of -0.64 m to -0.34 m). We now focus on one such partition, partition 44, to determine which environmental context is associated with the cluster of points encompassed within this partition. See Table 2.4 for the respective rules made on the data within each of the partitions.

Table 2.4: Decision rules made to separate the data associated with partitions 31 and 44. The first column displays the specific input feature, and the second and third columns show the rule for the partition specific to that column. '–' indicates that no rule was made on that input feature.

Input Feature	Rule for Partition 31	Rule for Partition 44
Hs	$>$ 3.44 m	$>$ 5.39 m
$Tm01$	$>$ 9.06 s	$>$ 8.32 s
MWD	$> 245°, \leq 360°$	–
$wndmag$	$\leq$ 15 m/s	$>$ 15 m/s
$wnddir$	–	$> 0°, \leq 215°$
Error Correction Value	-0.57 m	2.15 m

The mean wind direction for instances in partition 44 and $Hs > 5.39$ was 187°, with a standard deviation of 7°. Therefore, partition 44 predicts that large wave heights ($Hs > 5.39m$) are strongly under-predicted (bias of -2.15 m) during times when the regional wind is strong ($> 15m/s$) and from the south. Interestingly, this pattern is consistent with a previous study which showed that sig-



H

Figure 2.9: Data points greater than the 95th percentile significant wave height (5.4 m) in the 2012-2013 training data. Data from partitions 31 and 44 are colored blue and orange, respectively. The left panel is Hs scatter plot of the original model output, and the right panel is the Hs after the bagged decision tree corrections.

nificant wave height under-estimations consistently occurred during times when a strong southerly wind was present in the region A. Ellenson and Özkan-Haller (2018). Further, the instances within this partition were also associated with mean wave period underestimations ($Tm01$ bias of -0.5 s, not shown). In A. Ellenson and Özkan-Haller (2018), it was found that the southerly wave energy generated by the low-level southerly wind was under-estimated for this model configuration (ST2-GFS) resulting in both $Tm01$ and Hs under-estimations. Hence the error correction patterns detected by the bagged regression tree method are found to be consistent with known deficiencies in WW3 physics during certain environmental conditions, rather than statistical patterns present in the data alone.

2.4.2 Bagged Regression Tree Structure

The complexity of the bagged regression tree (i.e., the number of trees and tree depth) increased with as the number of input features (see Section 2.3.1). The Multiple Buoys experiment had the greatest complexity, followed by the With Wind and then the Waves Only experiments. A possible explanation for the increase in complexity is that each new feature introduces one more possible split of the feature space. Therefore, the tree complexity increased according to the size of the

input feature space.

For both summer and winter, mean wave period ($Tm01$) provided the most information to the bagged regression tree for the Waves Only and With Wind experiments, and is therefore ranked with the highest feature importance value (see Figure 2.10). Feature importances rank features that provide more information higher than the features that provide less information. This is calculated by summing the total amount of variance reduction by each input feature x_i within one tree, and averaging this value throughout the ensemble, which is further described in Breiman et al. (1984). The feature importance measures are normalized by the greatest feature importance value to sum to 1. A high feature importance value indicates that the pattern of error was best revealed when partitioning the data by that input feature. As shown in Figure 10, for winter, the most informative input features were $Tm01$ and wind magnitude (with feature importances of 0.38 and 0.35, respectively), whereas for summer, the most informative input feature was $Tm01$ (with a feature importance of 0.59).

The bagged regression tree improved the accuracy of the winter dataset more than the summer dataset. Generally, higher significant wave heights are simulated with higher error, and therefore lend themselves more readily to an error correcting technique Hanson, Tracy, Tolman, and Scott (2009); Moeini et al. (2014). Another interpretation for the better winter performance as opposed to summer was that the specific input features used in this study better discriminated between regions of higher or lower target values for winter than for summer. For an example of how an input feature can discriminate a region of high target value, see Figure 2.11, where wind magnitudes $> 15m/s$ delineates a region of greater negative bias in the training data. A possible avenue of future work would be to identify other input variables (beyond those used in this study) that would be more informative for summer data. Specifically, one could follow the approach of Cornejo-Bueno et al. (2018), where corrections were applied differently for regions of wind sea and swell. In this approach, input features are determined for each sea state regime, and reflect physical processes relevant to each sea state regime.

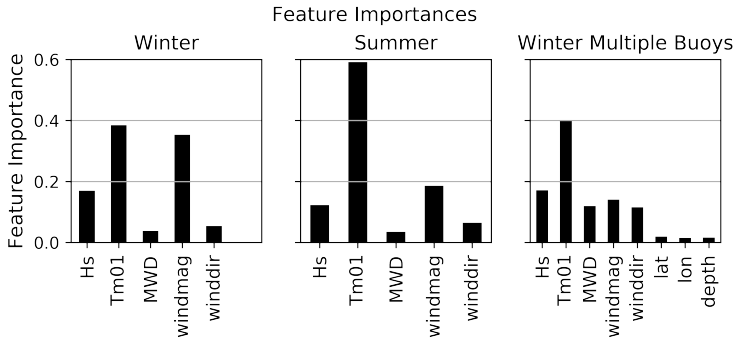


Figure 2.10: Feature importances for the two single buoy experiments for summer, winter, and the multiple buoys experiments at NDBC buoy 46050.

2.4.2.1 Input Features: Multiple-Buoy Experiments

When multiple buoys were considered in the training, mean wave period ($Tm01$) provided the most information, and the geospatial features *lat*, *lon*, and *depth* were also used (feature importances < 0.02) (see Figure 2.10). The nonzero feature importances for these variables indicates that the algorithm detected a spatial dependence in the data set that, when incorporated into the tree structure, resulted in a better fit to the data than without the inclusion of the information from these variables. One bagged regression tree is examined in the following as an example to show how the geo-spatial input features were used. The decision tree split most often on latitudes of 43.998°N and 46.537°N (six times each), which partitioned data associated with model output point located at buoy 46027 (northern California) from the other output points, and model output point located at buoy 46211 (Washington) from the other output points. It split most often on a longitude of 124.186°W (ten times), which partitioned data associated with model output point located at buoy 46243 (closest to Astoria canyon in northern Oregon) from the other output points. It split most often on depths of 33 and 43 m (four times each), which partitioned data associated with model output point located at buoys 46243 and 46211 from model output located at buoys 46027 and 46029.

An analysis of the decision tree structure was performed to determine whether it made a distinction between the error for different buoys, according to their location and depth. The ability to make such distinctions would be a prerequisite to applying the tree in a geospatial application, where predictions are required at locations and depths not included in training. Recall, by definition, data points encompassed within the same final partition had similar input features (i.e., modelled environmental conditions) and mean target values (i.e., significant wave height error). During training, the model output locations that were closest and in the most similar water depths (output locations at buoys 46243 in Washington and 46211 in northern Oregon) had the highest number of data points within the same partitions (4,456 data points out of 12,552 combined data points). The model output locations that were furthest from each other (output points at 46211 in Washington and 46027 in northern California) had the lowest number of data points within the same partitions (2,824 error corrections out of a combined 12,582 data points). During testing, the output location (buoy 46050) had the highest number of data points (2,953 out of 3,864 data points) within the same partitions as model output point at 46029, which was closest (163 km away) . It had the lowest number of data points (2,237 out of 3,864 data points) within the same partitions as model output point 46243, which was second closest (174 km away) but at the shallowest water depth (140 m versus 24 m water depths for buoys 46050 and 46243, respectively). When considering which locations to include in training, locations that experience similar environmental condition and wave transformation processes as the testing location would be the best candidates.

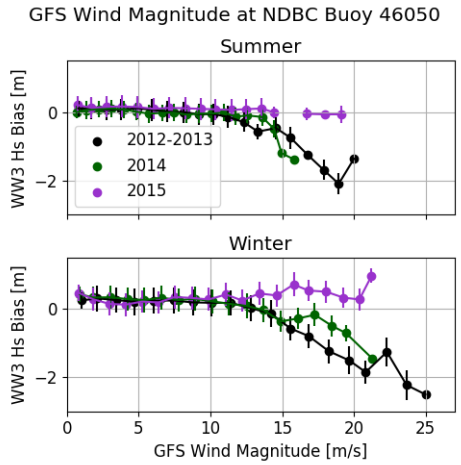


Figure 2.11: Comparisons of the wind magnitude input feature versus the target value (significant wave height error) for summer and winter. Training data from years 2012-2013 are colored black, testing data from 2014 and 2015 are colored green and purple, respectively. The data is binned in 1.25 m/s bins and the average is plotted here. The y-errorbar indicates the standard deviation within each bin.

2.4.3 Generalizability

Recall that the experiments shown in Section 3 used training data from years 2012-2013, and were tested on data from 2014. An additional experiment was performed in which data from 2015 were instead used to assess the degree to which the results were generalizable across multiple years. Interestingly, the 2015 experiment resulted in poor performance, for reasons explained next. When the With Wind bagged regression tree, trained on 2012-2013, was applied to 2015, Hs error increased for both seasons. For winter, SI increased from 0.21 to 0.26, and for summer, SI increased from 0.15 to 0.16 from the original WW3 to the applied bagged regression tree Hs time series (not shown).

To understand the poor performance with 2015 data, we first note that the wind magnitude was the second-most informative feature for the decision tree in

both winter and (to a lesser extent) summer (See Figure 2.10). For winter, the first decision of the majority ($> 90\%$) of trees within its ensemble separated regions of wind magnitude greater than or less than 15 m/s. For summer, when the bagged regression tree split the feature space on wind magnitude, the thresholds ranged between 10-20 m/s. When the wind magnitude is plotted against the target (Hs error), differences between the trends during the training years (2012-2013) and the testing years (2014, 2015) are apparent (see Figure 2.11). The trends of the testing data of 2015 can be seen to be different from that of the 2012-2013 training data, specifically for the wind magnitude values greater than 15 m/s, which is the threshold at which the bagged regression tree made a decision for winter. At these higher wind values, the 2014 data follows the same trend of increasing negative bias for wind speeds $> 15m/s$. In contrast, the 2015 data shows no such trend, and a slight positive bias for wind speeds $> 15m/s$. Therefore, the error corrections of the partitions established during training would overestimate error corrections required for the 2015 testing year.

Climatological differences between the years 2012-2014 and 2015 might explain why the wave model error trends were different for the different years. The years 2012-2014 were correlated with negative Pacific Decadal Oscillation (PDO) indices, and neutral El-Niño (ENSO) indices. During late 2014, PDO and ENSO values shifted to become positive. It is possible that the decrease in wave model error for high wind speeds between the years 2012-14 and 2015 is associated with the changing climate conditions. The results suggest that climatic differences can affect the statistical relationships between environmental conditions and WW3 model error, which in turn can have negative implications for the results. Future methods could account for this by adding a temporal aspect to the feature space (e.g., including time-dependent climate indices as input features), but the present data set is not long enough to investigate such an approach..

Machine learning algorithms cannot be used to extrapolate beyond the information provided to them; if the training input feature/target patterns do not exist in the testing data, the test predictions could be erroneous. The bagged regres-

sion tree is trained to find the patterns between input features and target values for the training data. During testing, the predictions remain consistent with the relationships found within the training data. This becomes a challenge for researchers applying machine learning techniques to environmental contexts where regime shifts occur, as shown in this study. Therefore, if this technique were to be applied in a hindcasting sense, one should carefully consider whether the training and testing time periods are comparable to one another. In particular, results from this study suggest that the training and testing data should be sampled from years that share a similar climatology (in this case PDO/ENSO), although further testing is recommended to confirm this. In a forecasting sense, one would try to ensure that input feature/target relationships within the testing data would be similar to the input feature/target relationships within training data.

2.5 Conclusions

This study demonstrated how a machine learning algorithm, a bagged regression tree, can be used to predict and describe systematic error of model forecasts of significant wave height. The bagged regression tree was used to analyze the errors in 24-hour time horizon significant wave height prediction time series made by a numerical model. In this hybrid approach, numerical model output is used as input features, and the target is the error between the modelled and observed values of significant wave height. The accuracy of the error detection method was confirmed when error metrics of wave model output were reduced after the application of bagged regression tree-based corrections to significant wave height forecasts. The algorithm improved significant wave height predictions more for winter than for summer. For this season, the corrections reduce the winter error metric SI by 19% (from 0.21 to 0.17).

In a descriptive sense, this method can act as a diagnostic tool for finding regions where a wave model improvement is necessary. During training, it isolates hours of similar environmental conditions and consistent over or underestimations. The bagged regression tree successfully differentiated between wave events greater

than the top 95th percentile of significant wave height ($Hs > 5.4m$) that were associated with different background wind conditions (wind mag $\geq 15m/s$ or wind mag $< 15m/s$). These two partitions of large Hs values required different error correction values (-0.57m and 2.15m). The bulk parameter mean wave period ($Tm01$) provided the most information for the bagged regression tree to determine these patterns, meaning that wave model error can be most readily detected when the output is delineated by mean wave period. In this study, the bagged regression tree best detected the error for mid wave periods (6-12s), indicating that the error signal is strongest for waves associated with these periods.

A geospatial application was demonstrated in which the bagged decision tree was trained on data from several other locations and predicted values at the original location. Additional geospatial information (latitude, longitude and water depth) was provided as a basis for extrapolation to other locations. The results showed the same reduction in SI (19%, from 0.21 to 0.17) for winter as the localized application, signifying that the pattern of error within the wave model output is consistent between different locations in a region. In future work, the technique could be applied to correct gridded model output.

The decision tree makes an implicit assumption that the relationships between the input features and target are similar for the training and testing data. When the bagged regression tree was applied to a different testing year (2015), the error associated with the significant wave height predictions increased. This is because the correlations between the input features and target for the training (2012-2013) years and testing year (2015) differed. In applications, a practitioner should try to ensure that the input feature/target relationships for the training and testing data are similar.

Chapter 3: Beach State Recognition Using Argus Imagery and Convolutional Neural Networks

Ashley N. Ellenson, Joshua A. Simmons, Greg W. Wilson, Tyler J. Hesser, Kristen D. Splinter

MDPI Remote Sensing
https://doi.org/10.3390/rs12233953
Volume 12, Issue 23

ABSTRACT: Nearshore morphology is a key driver in wave breaking and the resulting nearshore circulation, recreational safety, and nutrient dispersion. Morphology persists within the nearshore in specific shapes that can be classified into equilibrium states. Equilibrium states convey qualitative information about bathymetry and relevant physical processes. While nearshore bathymetry is a challenge to collect, much information about the underlying bathymetry can be gained from remote sensing of the surfzone. This study presents a new method to automatically classify beach state from Argus *daytimexposure* imagery using a machine learning technique called convolutional neural networks (CNNs). The CNN processed imagery from two locations: Narrabeen, New South Wales, Australia and Duck, North Carolina, USA. Three different CNN models are examined, one trained at Narrabeen, one at Duck, and one trained at both locations. Each model was tested at the location where it was trained in a self-test, and the single-beach models were tested at the location where it was not trained in a transfer-test. For the self-tests, skill (as measured by the F-score) was comparable to expert agreement (CNN F-values at Duck = 0.80 and Narrabeen = 0.59). For the transfer-tests, the CNN model skill was reduced by 24-48%, suggesting the algorithm requires additional local data to improve transferability performance. Transferability tests showed that comparable F-scores (within 10%) to the self-trained cases can be achieved at both locations when at least 25% of the training data is from each site. This suggests that if applied to additional locations, a CNN model trained at one location may be skillful at new sites with limited new imagery data needed. Finally, a CNN visualization technique (Guided-Grad-CAM) confirmed that the CNN determined classifications using image regions (e.g., incised rip channels, terraces) that were consistent with beach state labelling rules.

3.1 Introduction

The temporal evolution of nearshore morphology is a key area of active research within the coastal community. Nearshore morphology dictates wave breaking patterns and nearshore circulation, which is important in understanding nutrient

transport and determining recreational safety and erosion risk (R. A. Holman, Symonds, Thornton, & Ranasinghe, 2006; Turner, Whyte, Ruessink, & Ranasinghe, 2007; Wilson, Özkan-Haller, & Holman, 2010). For example, urban beaches may experience high levels of pollutants entering the surfzone during storms from run-off and pose a health risk to swimmers as well as the local ecosystem. The nearshore morphology affects the generation of bores and nearshore currents that ultimately influence the time and length scales of pollutant mixing, dispersal and advection (Grant et al., 2005; Inman & Brush, 1973). Similarly, rip currents are the leading cause of death at beaches globally and pose a significant risk to swimmer safety (Austin, Scott, Russell, & Masselink, 2013). Bathymetric rip currents are more likely to develop when undulating morphological features are present and are common in the Rhythmic Bar Beach (RBB) and Transverse Bar Rip (TBR) beach states (Austin et al., 2013; Castelle, Scott, Brander, & McCarroll, 2016; L. D. Wright & Short, 1984). Lastly, coastal erosion impacts coastal communities via loss of usable beach amenity and property (Helderop & Grubesic, 2019; Leatherman, 2018). Nearshore morphology preceding a storm has been shown to influence the levels of shoreline and dune erosion (Castelle et al., 2015; Thornton, MacMahan, & Sallenger, 2007).

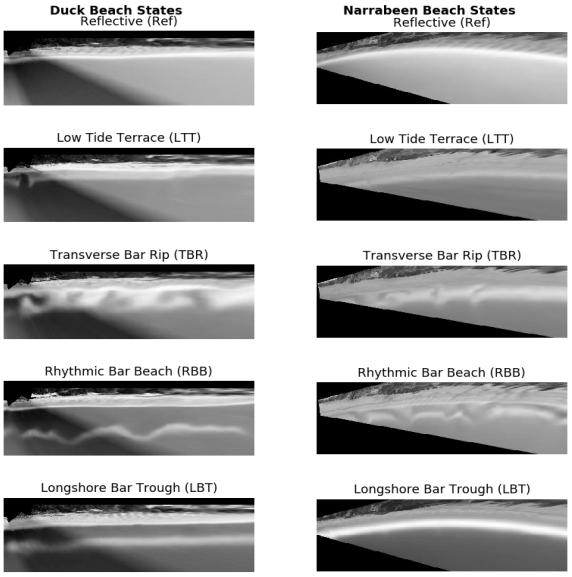


Figure 3.1: Examples of rectified Argus imagery from Duck (left) and Narrabeen (right), illustrating the Wright and Short classification scheme used for labelling. Note that the Duck imagery (left) is merged from multiple cameras.

Nearshore morphology can be detected in the surf zone remotely using video cameras. Time-averaged images of the nearshore surf zone, for example from Argus cameras, can be used to detect the shape of sandbars because of the tendency of waves to preferentially break over these higher topographic features (R. A. Holman & Stanley, 2007). Nearshore morphology is generally considered to exist in consistently-occurring patterns known as beach states (Lippmann & Holman, 1990; L. D. Wright & Short, 1984), and the occurrence of these states has been found to correlate with incident wave conditions and sediment grain size (L. Wright, Short, & Green, 1985). Time series of beach state observations have been used to qualitatively validate modelling studies of sandbar evolution (Dubarbier, Castelle, Ruessink, & Marieu, 2017; Plant, Holland, & Holman, 2006; Ranasinghe, Symonds, Black, & Holman, 2004; Siegle, Huntley, & Davidson, 2007; K. D. Splinter, Holman, & Plant, 2011; Strauss, Tomlinson, & Hughes, 2006), to gain a general un-

derstanding of the behavior of a specific beach system (Armaroli & Ciavola, 2011; Castelle, Bonneton, Dupuis, & Sénéchal, 2007; De Santiago et al., 2013; Lippmann & Holman, 1990; Ojeda, Guillén, & Ribas, 2011; van Enckevort & Ruessink, 2003; **?**), and to determine the level of recreational hazard that exists within the nearshore system (A. Short & Hogan, 1994a). A widely-used beach state classification scheme is that of Wright and Short (L. D. Wright & Short, 1984), who defined five distinct sandbar morphology states illustrated in Figure 3.1. The modal beach state at a given beach depends on the dominant incident wave energy, tidal range, and local sediment size (Loureiro, Ferreira, & Cooper, 2013; Masselink & Short, 1993). The reflective (Ref) state is the lowest energy state, characterized primarily by waves breaking at the shoreline and an absence of offshore sandbar features. The Low Tide Terrace (LTT) state represents intermediate to low wave energy, with bar welding to the shoreline and the potential for weak rips to be present. The Transverse Bar Rip (TBR) state represents intermediate wave energy with rip circulations. The Rhythmic Bar Beach (RBB) state represents intermediate-high energy, and has a contiguous trough separating the bar from the shoreline, and a bar exhibiting rhythmic crescentic patterns due to offshore-directed rip currents. Lastly, the Longshore Bar Trough (LBT) state represents intermediate to high wave energy, with a contiguous trough; however, rip currents are not as strong as in the RBB state, so the sandbar lacks rhythmicity and extends linearly alongshore.

In Argus time exposure (timex) products, sustained breaking over topographic features (sandbars) appears as bright white bands, and the resulting spatial pattern can be categorized into a beach state (Browne, Strauss, Tomlinson, & Blumenstein, 2006; Lippmann & Holman, 1989; Price & Ruessink, 2011a; Ranasinghe et al., 2004). Specific morphological features, such as the position of the sandbar, shoreline or rip locations have been previously derived from Argus timex imagery (Armaroli & Ciavola, 2011; Plant et al., 2006; K. D. Splinter et al., 2011; van Enckevort & Ruessink, 2001). While these previous methods are successful in identifying specific morphological features of interest, they require a number of pre-processing steps to extract features (e.g., quasi-linear features such as sand-

bar crests or shoreline position). Also, the methods are calibrated for specific sites, and may not transfer successfully to other locations (Contardo & Symonds, 2015). Nearshore optical remote sensing data should be exploited in a scalable and generic way, thereby advancing our understanding of coastal processes at different sites (R. Holman & Haller, 2013; Smit et al., 2007; K. D. Splinter, Harley, & Turner, 2018). Our current mathematical formulations cannot extract meaningful physical information (such as depth limited wave breaking) from remotely sensed imagery. Machine learning techniques may offer a potential path forward, however, because their underlying extremely flexible mathematical formulations can be adapted to detect physically relevant patterns in imagery.

Machine learning techniques have been previously used in a variety of coastal applications (A. Ellenson, Pei, Wilson, Özkan-Haller, & Fern, 2020; Molines, Herrera, Gómez-Martín, & Medina, 2019; Peres et al., 2015). In contrast with previous studies, the input to the machine learning alogrithm in this study is imagery. Other machine learning/coastal imaging studies have used machine learning as a measurement technique for hydrodynamic quantities such as significant wave height in laboratory settings (den Bieman, de Ridder, & van Gent, 2020) and in the field (Buscombe, Carini, Harrison, Chickadel, & Warrick, 2020), and for morphological properties such as grain size (Buscombe, 2020) and laboratory bed level (den Bieman, van Gent, & Hoonhout, 2019). Machine learning has also been used for segmenting and classifying coastal images (Hoonhout, Radermacher, Baart, & van der Maaten, 2015), improving shoreline detection (Vos, Splinter, Harley, Simmons, & Turner, 2019), classifying wave breaking type (Buscombe & Carini, 2019), and predicting wave run-up (Beuzen, Goldstein, & Splinter, 2019).

This study uses a convolution neural network (CNN) to automate the classification of beach state from images with limited pre-processing steps. The paper is organized as follows. Section 3.2 describes the dataset in this study that is derived from two different sites: Narrabeen, New South Wales, Australia; and Duck, North Carolina, United States. Section 3.3 describes the methods, including the general CNN model and the specific implementation used for the present study.

The suitability of a CNN to classify beach state at each location is explored in Section 3.4, followed by a discussion of the results in Section 3.5 and conclusions in Section 3.6 .

3.2 Field Sites And Data

3.2.1 Field Sites

Figure 3.2 shows the location of the two study sites. The first site is a sandy barrier beach located at the U.S. Army Engineer Research and Development Center Field Research Facility (FRF) in Duck, North Carolina. The wave climate at Duck is seasonal, with higher incident wave energy in the winter, and lower incident wave energy in the summer (Birkemeier, DeWall, Gorbics, & Miller, 1981). The annual average significant wave height is 1.1 m, with waves tending to come from the south during spring and summer, and from the north during winter. Winter storms consist of both extra-tropical (north-easters) and tropical (hurricanes) cyclones. The mean spring tide range is microtidal at 1.2 m. The beach slope averages 0.108 at the foreshore, and decreases with distance offshore to 0.006 at 8 m depth (Horrillo-Caraballo & Reeve, 2008). The sediment is comprised primarily of medium to fine grain quartz with finer sands further offshore (Stauble, 1992). The median grain size between the bar and the shoreline is approximately 0.5 mm, with 20% carbonate material, while offshore of the bar the median grain size becomes 0.2 mm (Birkemeier et al., 1981). This beach generally is classified as an intermediate beach (Lippmann & Holman, 1990) and can frequently have at least one or two sandbars present (Alexander & Holman, 2004).

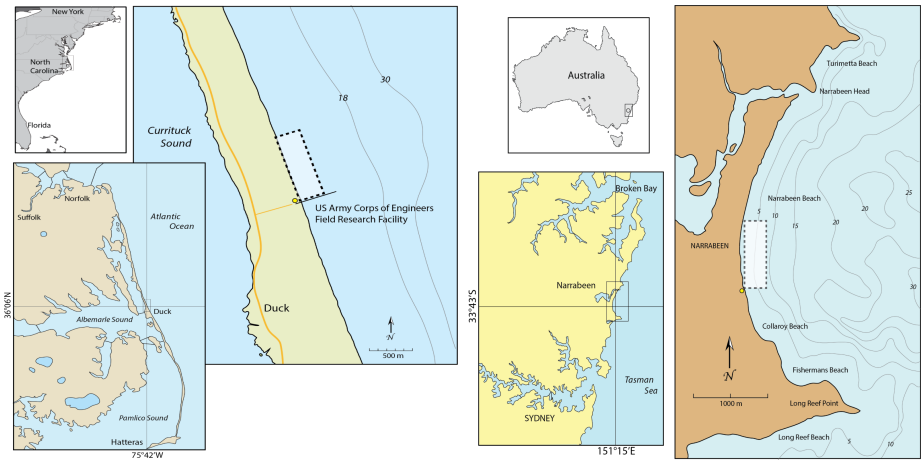


Figure 3.2: Maps illustrating the location of the two study sites, Duck (left panel), and Narrabeen (right panel). Dots show the camera locations and dashed boxes denote imagery location.

The second study site, Narrabeen-Collaroy (herein refered to as Narrabeen) is an embayed beach located in Sydney, Australia (Figure 3.2, right panel). The average annual wave climate is of moderate to high energy (average significant wave height of 1.6 m) (Turner et al., 2016). There is generally a background SSE swell generated by mid-latitude cyclones crossing the Tasman sea, south of Australia. Similar to Duck, Narrabeen sees higher wave energy in winter, and lower wave energy in summer. The summer waves have shorter periods than in winter, and are a combination of both SE swell and local north-easterly sea breeze waves. The winter waves have longer periods, and are generated by storms that consist of mid-latitude cyclones from the south, east-coast lows generated near the NSW coast, and tropical cyclones from the northeast. The mean spring tide range is microtidal at 1.3 m (Turner et al., 2016). The sediment is comprised of fine to medium quartz sand with a median grain size of 0.3 mm, and 30% carbonate fragments. Due to the embayed nature of this beach and the predominant SSE swell, there is an alongshore gradient in wave energy and in beach state caused by wave refraction and headland-induced diffraction effects. As a result the more exposed sections of

the beach in the north of the embayment are commonly dissipative-intermediate beach states, whereas the more sheltered southern end of the embayment (the location of the present study) is typically classified as reflective-intermediate beach states (Harley, Turner, Short, & Ranasinghe, 2011).

3.2.2 Dataset

The dataset in this study is comprised of orthorectified time exposure (timex) grayscale imagery collected hourly from Argus stations (R. A. Holman & Stanley, 2007), over a period 1987–2014 for Duck, and 2004–2018 for Narrabeen. At Duck the images were combined from one to three different camera views (depending on the year and how many cameras were installed at the time) spanning the area of the beach north of the FRF pier, while at Narrabeen the images were from a single camera view looking north towards the stretch of beach in the center of the embayment (see Figure 3.2). Effects of poor lighting (due to sun angle or cloud cover), or a lack of wave-breaking signal (due to low waves or a high tide), were reduced by averaging all hourly timex images taken throughout the day into a single daytimex image for each day. The oblique daytimex images were then orthorectified onto a domain of 900 m alongshore by 300 m cross-shore, with a ground resolution of 2.5 $\times$ 2.5 m grid (same for both sites). In the case of Duck, multiple camera views were merged following (Holland et al., 1997). The orthorectified daytimex images were then bilinearily interpolated to 512 $\times$ 512 pixels, following standard machine learning practices where images are reshaped to be square before being input to convolutional neural networks (Ghosh, Das, & Nasipuri, 2019).

Figure 3.1 shows example imagery from Duck and Narrabeen, which highlight several notable differences between the two sites. At Narrabeen, areas outside of the camera field of view are marked in black, resulting in two black triangular spaces in the left side (south side) of the image at this site. At Duck, artifacts of the camera merging result in a diagonal seam separating the images from the three cameras. Additionally, the three cameras at Duck do not always share the same intensity histogram, which can result in non-uniform shading throughout the

merged image. Finally, the Narabeen beach exhibits a noticeable curvature in its shape, which is in contrast to the straight coastline at Duck.

3.3 Methods

Machine learning is based on learning patterns within data by learning correlations between input and a specified output (also called a 'target'). The branch of machine learning applied in this study is called classification, whereby the targets are discrete classes rather than continuous variables. The Convolutional Neural Network (CNN) technique used here (see Section 3.3.2 for technique details) is a specific deep learning classification technique that uses supervised learning, meaning the outputs are known *a-priori* for a set of pre-defined training cases (see Section 3.3.1 for the explanation of the targets). Once the relationships between input and output are learned from the training data, the CNN is then tested on data from outside the training set to determine its out of sample accuracy. The code for this project is available online at https://github.com/anellenson/DeepBeachState, and the data is available upon request.

3.3.1 Dataset Preparation: Manual Labelling And Augmentation

Supervised learning requires a set of manually-classified images used for training and testing. In this study, the target is one of five beach state classes described in Section 3.1 and the input is an Argus time exposure image. From the 10+ years of daytimex images available from the two sites, 125 images per class (per beach state) per site were selected that were consistent with the description from (L. D. Wright & Short, 1984). Of these, 100 were reserved for training (per class per site, so 1000 total), and 25 for testing (per class per site, so 250 total). The images were labelled by a single person (the first author) in order to minimize labelling inconsistencies that might occur among different labellers. The labelling considered the visible wave breaking patterns, with each image considered independently from the others and in a randomized order. The labeller selected the images wherein only one beach

state was visible (no longshore variability of beach states or shape ambiguity, see Section 3.5.1 for further discussion on beach state labelling challenges). The test dataset was also labelled by the four co-authors independently, for the purposes of benchmarking the CNN model skill in comparison to inter-labeller agreement (see Section 3.4.1 for further explanation and results).

Deep learning requires large (on the order of thousands of images) amounts of input data for training, and performance can usually be improved by the addition of more training data (Goodfellow, Bengio, & Courville, 2016). In this study, the amount of data was limited by the number of times each state occurred over the time period spanned by the dataset, and the number of images which clearly exhibited only one beach state (see Section 3.5.1). CNN performance typically benefits from larger dataset sizes than available for this study. As such, image augmentation (Perez & Wang, 2017) was used to increase the number of images in the dataset. Five total augmentations were applied, increasing the number of images to 3000 images per site, or 6000 images total. The augmentations included: (1) simultaneous horizontal and vertical flip, (2) random rotation up to 15 degrees, (3) random erasing of 2–8% of the image, (4) random horizontal and vertical translations to a maximum of 15% in the horizontal and 20% in the vertical, and (5) image darkening by gamma correction (power law transform) with a gain of 1 and gamma of 1.5.

3.3.2 Convolutional Neural Network

This study uses the CNN architecture Resnet-50 (He et al., 2015); see Appendix B.1 for detailed information about the architecture. The CNN predicts a beach state label (y) based on an input Argus image (x). The algorithm learns this relationship in the form of parameters of a function (f) that maps the input image to the beach state label; $y = f(\mathbf{x})$. Figure B.1 shows the details of this function, which has here been represented as an operator, f. In the case of the CNN, the function f comprises two key steps: (1) feature extraction, in which important spatial structures from the input are extracted (e.g., linear or rhythmic features);

and (2) class prediction, in which a neural network is used to map the extracted features from step 1 into a predicted beach state based on learned relationships between the spatial features found in step 1 and the associated targets (Goodfellow et al., 2016). Historically, when machine learning techniques are applied to imagery, Steps (1) and (2) are performed separately, and so each are separately optimized. In contrast, the deep learning CNN algorithm combines steps (1) and (2) into the same optimization. For each input image, the model outputs a vector of $(k \times 1)$ probabilities, where k is equal to the total number of beach states (5). The entry with highest probability is chosen as the CNN label prediction.

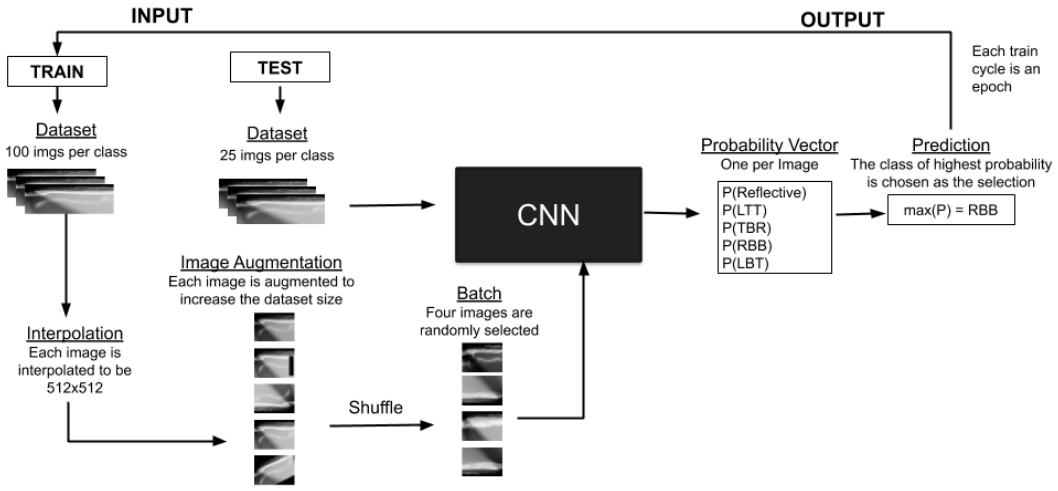


Figure 3.3: The work flow for one training epoch and testing cycle. Each epoch of the training process results in an update of the CNN parameters.

During the supervised training, the performance of the model is optimized by minimization of a cross-entropy cost function with respect to the CNN free parameters. Figure 3.3 shows a schematic of the iterative training procedure. For many CNN studies, a pre-trained CNN is used, meaning that the parameters (the convolutional filter weights) of the CNN have already been optimized to identify objects (e.g., animals, faces, buildings), in a different dataset, such as Im-

ageNet (Simon, Rodner, & Denzler, 2016). Pre-trained models were considered in this study, where the final neural network (classification step) was retrained. However, the pre-trained models failed to focus on the bathymetric regions relevant to beach state classification. Therefore, the entire Resnet-50 CNN was trained 'end-to-end', meaning that all of the parameters of the entire base model were altered, similar to (Buscombe et al., 2020).

During one training epoch, the CNN was fed the entire augmented data set (3000 images) in batches of four randomly-selected images at a time; batch sizes of eight, 12 and 36 images were also tested, but four was found to be the optimal batch size, where training loss was lowest. The training of the CNN is stochastic due to the random seeding of the original parameters, the optimization routine used during training, and the random batches of data selected as input in an epoch. The effects of this randomness on the reported accuracy of each model was assessed by training an ensemble of 10 CNNs for each experiment.

3.3.3 Visualization: Saliency Maps

Model visualization refers to techniques for inspecting a trained CNN model to determine how its predictions are made. Visualization can be used to confirm that a model's classifications are based on appropriate qualitative features of beach states, as opposed to other non-relevant features contained in the training data. The model visualization technique used in this study, Guided Grad-CAM, highlights which pixels are most important in the CNN's final prediction by incorporating information from the two steps of the deep learning process: (1) feature extraction and (2) classification (Selvaraju et al., 2017). Guided Grad-CAM is a combination of two visualization techniques that extract information from images: (1) Guided back propagation (GBP) extracts information from the feature extraction step; and (2) class activation maps (CAMs) extract information from the classification step. GBP identifies specific pixels associated with relevant spatial structures (Springenberg, Dosovitskiy, Brox, & Riedmiller, 2015; Zeiler & Fergus, 2014), which in the present study might include linear or curved sandbar shapes or

shorelines. The pixels identified by GBP are those that provided the strongest signals to the optimization routine during cost function minimization (Springenberg et al., 2015). On the other hand, CAMs identify the pixels which have the largest contribution to classifying an image as a specific class (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). Guided Grad-CAMs combine both GBP and CAM techniques by multiplying their outputs together, resulting in visualizations called "saliency maps" that are both spatially detailed and class-specific. Specific to this study, if the CNN has been trained successfully, the pixels highlighted by the Guided Grad-CAM should correspond to the regions showing patterns of wave breaking associated with each beach state class.

Example saliency maps for Duck and Narrabeen for each state are shown in Figure 3.7 (discussed further in Section 3.5.1). Visual inspection of the saliency maps showed that approximately 70% had highlighted areas of specific relevance to beach states (e.g., incised rip channels for TBR, swash and off-shore bar for LBT), as in the examples shown.

3.3.4 Experiments

Three 10-member ensembles of CNNs were trained. The first two ensembles were single-location ensembles, meaning that the training data came from one location (Duck or Narrabeen). The CNN ensemble trained at Duck is hereafter referred to as Duck-CNN, and the CNN ensemble trained at Narrabeen is hereafter referred to as Nbn-CNN. The third ensemble was a combined-location ensemble, meaning that training data came equally from both Narrabeen and Duck, hereafter referred to as combined-CNN. Each CNN was fed 3000 training images (including augmented images), with the combined-CNN using 1500 images from each location, where the images were chosen randomly.

Each single-location ensemble was tested at both locations in self and transfer tests. As the combined-location ensemble was trained with data from both locations, it did not have a transfer test. Overall performance metrics F1, normalized mutual information, and Matthews correlation coefficient were evaluated

following (Baldi, Brunak, Chauvin, Andersen, & Nielsen, 2000). However, the conclusions drawn from the three metrics were similar, so only the F-score is reported herein. The F-score (See Appendix B.2 for a definition) ranges between 0 and 1 with a higher F-score value indicating better performance. Per-state accuracy is also reported to assess state-specific performance of the CNN, and similarly confusion matrices were calculated to determine if biases were present where two or more states were consistently confused with each other. The F-scores and accuracies presented are the average of the 10 CNNs for each ensemble.

Finally, experiments were performed to assess transfer skill as a function of training data composition, and is presented in Section 3.5.3. Specifically, the goal was to determine the percentage of data required from each site to reach skill comparable to the single-location tests. In these experiments, data was incrementally added from the transfer site as percentages of the total. For example, if the CNN was originally trained at Duck, then data from Narrabeen were added. Eight total experiments were performed, with ratios of Duck to Narabeen training data ranging from 0:100 to 100:0 in 5% increments keeping the total number of training data constant (3000 images). The F-score was assessed at each increment to determine the percentage of data required from each location in order to reach skillful performance.

3.4 Results

3.4.1 Inter-Labeller Agreement

Human performance is commonly used as a benchmark for machine learning skill quantification (L. Chen, Wang, Fan, Sun, & Naoi, 2015). In this study, the "true" label is defined as the label chosen by the primary labeller (the first author). However, it is acknowledged that the same label may not be chosen by another labeller, even among domain experts, owing to the inherently "fuzzy" nature of beach state classification (discussed further in Section 3.5.1). A human-performance skill benchmark was therefore defined by comparing the true labels to ones chosen by

alternative labellers (the co-authors). Figure 3.4 shows the results of the confusion table and per-state accuracy for the labeller agreement on the validation set.

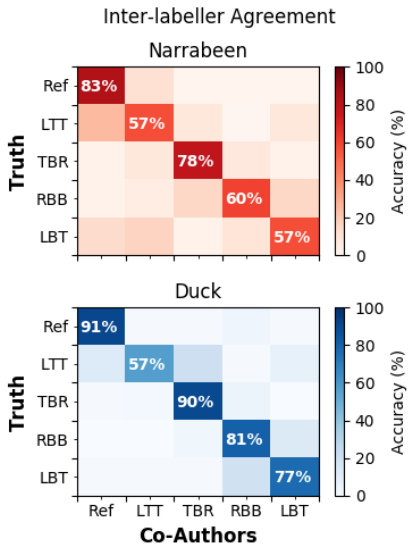


Figure 3.4: Confusion table plotted as truth vs. other labellers, where "Truth" is defined as the labels chosen by the primary labeller.

Specific recurring errors were noted in the inter-labeller comparisons when considering individual beach states. For both locations tested here, the lowest per-state agreement was the LTT state (57%); for Narrabeen, 24% of LTT images were mislabelled as the lower-energy adjacent Reflective state, while for Duck, 21% of LTT images were mislabelled as the adjacent, but higher energy, TBR state. These two cases illustrate the most common reason for misclassification: confusion between adjacent states that belong to a similar energy regime and therefore have similar morphology. Confusion also existed between non-adjacent states, however, that had similar morphological characteristics. For example, at Narrabeen, LBT was confused with either LTT (17% of LBT images) or Reflective (13% of LBT images). LBT and LTT states are similar in that both are linear sandbars, however LBT is found further offshore and is always associated with a trough. LTT is a

lower energy sandbar configuration with bar welding to the shoreline, and can have a trough during high tide when the terrace is flooded. The confusion at Narrabeen may be due to the relatively narrow surfzone width, and so the distance between the shoreline and the sandbar was small in many images. This confusion was also reported in another study (Ranasinghe et al., 2004) at nearby Palm Beach, NSW, a site that displays similar nearshore morphology to Narrabeen.

3.4.2 CNN Skill

Figure 3.5 summarizes the F-scores obtained by the CNN. In general, the skill was highest when the training and testing data were from the same location, in which case, CNN F-scores were comparable to inter-labeller agreement as reported in Section 3.4.1. Duck-CNN F-score was 0.80 compared with 0.79 for manual agreement, and Nbn-CNN F-score was 0.59 compared with 0.57 for manual agreement. The CNN trained on the combined training dataset (combined-CNN) predicted beach state more accurately on the combined test dataset (test data from both locations, black boxes Figure 3.5) than either of the single-site CNNs. Specifically on the combined dataset, the combined-CNN reached an F-score of 0.68, compared with Duck-CNN F-score of 0.61 and Nbn-CNN F-score of 0.53. Interestingly, the combined-CNN also slightly outperformed the Nbn-CNN at Narrabeen (F-Score = 0.61), although the two were equally skillful to within the variability of their respective CNN ensembles.

The transfer skill of the CNN, defined as the skill when trained at one location and then tested at another location, varied depending on the training data set. The Nbn-CNN skill was reduced by 24% when transferred to Duck, which was less than the 48% reduction in skill for the Duck-CNN when transferred to Narrabeen. This suggests that the correlations between sandbar characteristics and beach state learned at Narrabeen were more informative when predicting beach state at Duck than vice versa. It is speculated this could be due to the relatively smaller length scales at Narrabeen, and therefore the requirement that the CNN learn relatively finer sandbar features at Narrabeen compared to Duck. That is, the finer features

learned by Nbn-CNN may have remained applicable when predicting beach states at Duck, while larger features learned by Duck-CNN were less applicable when predicting beach states at Narrabeen. See Section 3.5.2 for further discussion and quantification of length scales at the two sites.

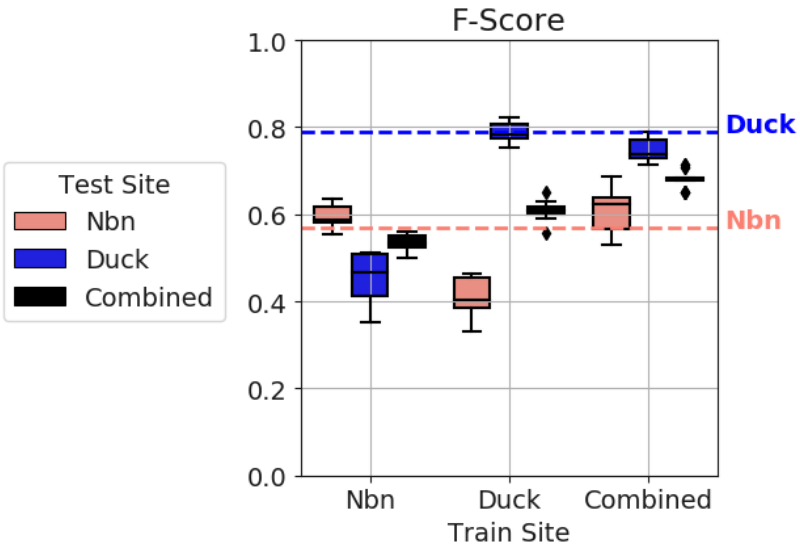


Figure 3.5: F-score performance values from tests at individual and combined datasets. The x-axis shows the location of training data. The boxes show the quartiles of the F-scores from the ensembles and the whiskers the rest of the distribution within 1.5 × the interquartile range. The horizontal dashed lines correspond to the inter-labeller agreement F-scores.

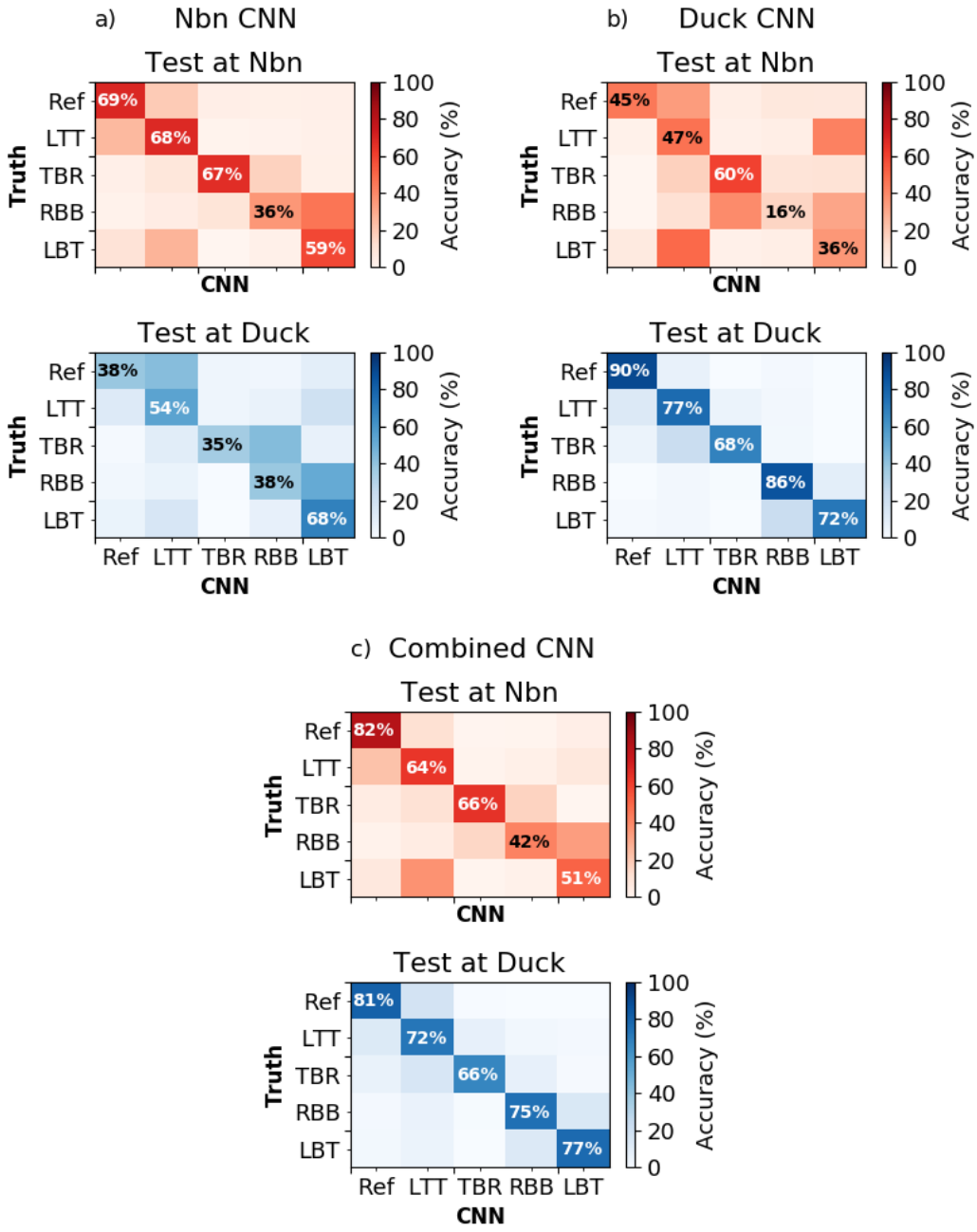


Figure 3.6: Confusion table results from the Nbn-CNN, Duck-CNN and combined-CNN (panels **a**, **b** and **c**, respectively). Top panel (red) shows results for tests at Nbn, and bottom panel (blue) shows results from tests at Duck. For each matrix, the label provided by the CNN is counted in the columns, and the true label is counted in the rows. Per-state accuracies are within the diagonal.

Figure 3.6 shows the confusion tables from each of the six tests. Overall, for all self-tests, the CNN outperformed random choice (accuracy >20%) and the accuracy was comparable to inter-labeller agreement (Figure 3.4). The self tests (i.e., trained and tested at the same locations) resulted in slightly higher per-state accuracy at Duck (68–90%) than at Narrabeen (36–69%). For both sites, the highest accuracy was in the classification of the low-energy Ref state, while the lowest accuracy of the CNN was in classifying the rhythmic states of RBB (Nbn) and TBR (Duck). The Nbn-CNN confused the RBB state most often with LBT, with 47% of the RBB images misclassified as LBT. Note that the RBB and LBT states both correspond to an offshore sand bar with a distinct trough, with the differentiating factor being the degree of bar curvature. Also at Narrabeen, 27% of LBT images were confused for LTT, an error that also occurred in the manual classification experiment (17% of LBT images confused for LTT). The Duck-CNN confused the TBR state most often with LTT, with 28% of the TBR images classified as LTT. The TBR and LTT states both correspond with bar welding, and both may include rip currents, with the differentiating factor being a larger number and size of rip currents present in TBR.

For the transfer tests, the per-state accuracy decreased for the majority of states compared to the self tests. Per-state accuracy ranged between 35–68% (Nbn-CNN at Duck) and 16–60% (Duck-CNN at Nbn). Similar to the overall skill trends (see Figure 3.5), Nbn-CNN transferred better than the Duck-CNN when assessed with per-state accuracy. The per-state transfer accuracy for Nbn-CNN was highest for LBT (68%), and lowest for the Ref, TBR and RBB states (accuracies of 38%, 35% and 38%, respectively) when tested at Duck. The confusion for these states was primarily adjacent and up-state (Ref confused as LTT, TBR confused as RBB, and RBB confused as LBT). Additionally, the LTT images were confused as LBT 48% of the time. For the Duck-CNN transfer test, the bar states characterized by rhythmic features (TBR and RBB) were mainly labelled as TBR (44% percent of RBB images were classified as TBR), and the linear states (Ref, LTT, LBT) were labelled as LTT. Overall, the Nbn-CNN had higher per-state accuracy at Duck than

the Duck-CNN had at Narrabeen. The main confusion for Nbn-CNN at Duck was up-state, adjacent-state confusion, whereas the Duck-CNN at Nbn had an LTT bias. Possible explanations for state confusions are detailed in Section 3.5.2.

The combined-CNN that was trained equally with data from Narrabeen and Duck showed overall good skill and per-state accuracy at both beaches (42–82% at Nbn and 66–81% at Duck). Compared to the Nbn-CNN at Narrabeen, it exhibited less confusion for the Ref class, and similar confusion for RBB (34% for combined-CNN versus 47% for Nbn-CNN of RBB images were confused for LBT) and LBT (38% for combined-CNN versus 27% for Nbn-CNN of LBT images were confused for LTT). Compared to the Duck-CNN at Duck, it exhibited similar confusion for TBR (16% for combined-CNN versus 28% for Duck-CNN of images were confused for LTT). The combined-CNN also resulted in slightly lower per-state accuracy for Ref, LTT and RBB states (accuracy reductions of 9%, 5%, and 11% for each state, respectively) at Duck.

3.5 Discussion

3.5.1 Beach State Classification

It is notable that many of the misclassifications made by the CNN (Figure 3.6) and the disagreement between labellers (Figure 3.4) can mostly be attributed to states that are adjacent to one another in the ordered list defined by Wright & Short (L. D. Wright & Short, 1984). This implies that, as expected, adjacent states have similar morphology and may be easily mistaken for one another. The use of the classification system of Wright and Short (L. D. Wright & Short, 1984) in this study implicitly assumes that instantaneous beach morphologies, as observed by Argus imagery, can be categorized into discrete states. This is an approximation, however, as sandbars exhibit a continuum of shapes as they evolve between configurations (Armaroli & Ciavola, 2011; Price & Ruessink, 2011a; L. Wright et al., 1985), and may never reach a true equilibrium state. This is especially true during the slower, down-state evolution of sandbar configuration (i.e., RBB to TBR to

LTT) as described by (Lippmann & Holman, 1990). During the labelling process, the most dominant state was used as the classification, but in practice 'pure' beach states are not present in all images. Therefore, achieving 100% accuracy for an ambiguous beach state is impossible. Three predominant issues were identified by the authors as challenges in determining beach state: (1) labeller perception, (2) alongshore variability of bar state and (3) sandbar state ambiguity due to the nearshore evolving between states.

Since there are no rigorous, quantifiable rules which delineate each beach state, a state identification for a specific image could vary due to labeller perception. Differences in labeller perception can occur either because the labeller is different (a different person), or the labeller might have a different perception on a different day. While the labels used to train and test the CNN in this study were made solely by the first author to limit labeller perception bias within the model training, the co-author labelling experiments (see Figure 3.4) highlight the challenges with labeller perception and the overall complexity in classifying unique beach states. Specifically, the co-authors did not achieve 100% agreement on any one state. Most notably, there was more confusion among the different labellers for Narrabeen than Duck, suggesting this beach may exhibit more complex or ambiguous beach states (discussed further in Section 3.5.2).

Alongshore gradients and irregularities in hydrodynamic forcing might impose alongshore variability of sediment transport and resultant sandbar patterns in one image. Therefore, one image might clearly exhibit more than one beach state in the alongshore (for examples, see Figure 3.7). For the two test sites presented here, the pier at Duck can affect sediment transport at a distance of up to 500 m (Pianca, Holman, & Siegle, 2015) and at Narrabeen strong alongshore gradient of breaking wave height due to wave sheltering from the adjacent headland results in alongshore variability in the dominant beach states (Harley et al., 2011). Additionally, as the morphology evolves between states, the sandbar shape can exhibit characteristics of adjacent classes in one image and therefore have an ambiguous classification. In either of these cases (i.e., alongshore variability or shape ambiguity), a more

accurate classification would be a mixture of classes. However, since the CNN in this study is a single-label classification tool, only a single discrete target class can be given for training, potentially causing model and labeller disagreement.

Saliency maps can be used to troubleshoot model/labeller disagreement (see Section 3.3.3 for a discussion on saliency maps). The representative saliency maps in Figure 3.7 illuminate which regions of the image were chosen by the CNN to be important for classification. Specifically, they show that the CNN can differentiate between states due to physical characteristics such as sandbar welding or rhythmicity. For images labelled as Ref, salient (warm color) areas are focused on the breakers at the shoreline or dark areas just offshore that indicate a conspicuous absence of breaking. Salient areas for images labelled as LTT are also focused on shorebreak, which can be associated with few rips and bar welding that characterize the LTT state. The curved wave breaking patterns connected to the shoreline that can be associated with rip currents and intermittent welded sandbars which characterize the TBR state are the salient areas for the TBR state. For images labelled RBB salient areas are focused on rhythmic features. Salient areas are focused on an offshore bar and the shoreline, which indicates the existence of a trough for images labelled LBT.

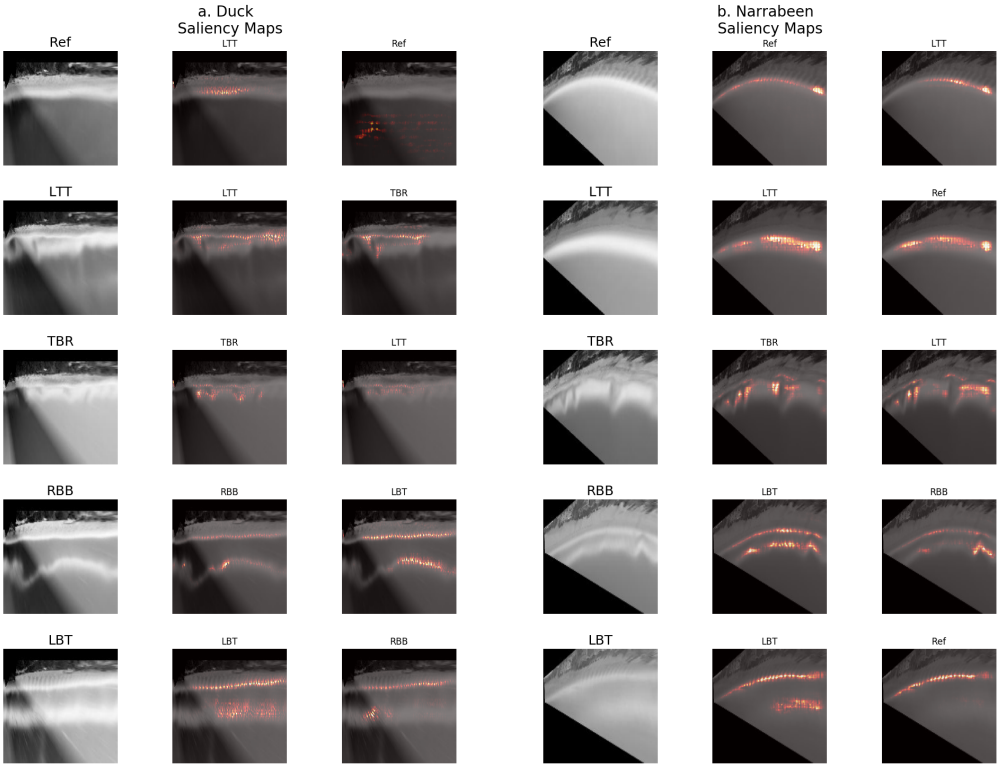


Figure 3.7: Saliency maps showing the pixels most relevant for classification decisions for Duck in subfigure (a), and Narrabeen in subfigure (b). The original image fed into the CNN, the first classification choice, and the second classification choice are in the first, second and third columns, respectively. The saliency maps are generated by the single-location CNNs.

Despite the CNN being a discrete classification tool, the saliency maps suggest it can also detect the presence of multiple beach states within one image, and/or beach state ambiguity. For example, the LTT image at Duck (Figure 3.7a, second row) could plausibly be labelled as either LTT or TBR, depending on which side of the image the labeller focuses. Incised rip channels, characteristic of the TBR state, exist in the left side of the image, while a terrace, characteristic of the LTT state, exists on the right side of the image. The resulting first and second choice

saliency maps for this image highlight the LTT and TBR features, respectively. Both of these choices have validity, but as currently implemented the CNN only reports the first choice as its output. The use of saliency maps to develop a multi-output classification or a non-discrete labelling system (Armaroli & Ciavola, 2011) are possible improvements to the present model. For example, object localization, a deep learning technique developed by (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2015), is a technique wherein the CNN identifies the location of objects within a picture using class activation maps. Object localization could potentially be adapted as a way to identify and quantify alongshore variable bar states.

3.5.2 Site Imagery Differences Affecting State Identification

The lower skill at beach state identification at Narrabeen by both the CNN and by the co-authors suggests that classifying beach state at Narrabeen is a more ambiguous problem than at Duck. Similarly, the probabilistic output of the CNN (the step before the final maximum-likelihood selection step) showed that the average probability assigned to its classification was 85% and 76% at Duck and Narrabeen, respectively. The different choice probabilities imply that the CNN had slightly less confidence in its predictions at Narrabeen, which is consistent with the relatively lower skill that was obtained. The ambiguity of beach state at Narrabeen was possibly due to two reasons: (1) difficulty in consistently identifying the shoreline; and (2) smaller length scales at Narrabeen.

As shown in the saliency maps (Figure 3.7), the CNN typically identifies the shoreline and offshore features when classifying the states. However, the optical signature of the shoreline at these sites can be quite different. The shoreline at Narrabeen can be identified in time exposure imagery in two ways: by the higher image intensities associated with swash motions or, more frequently, by lower intensities associated with wet, dark sand. In contrast, the shoreline at Duck is consistently identifiable by higher image intensities due to swash motions (Madsen & Plant, 2001; Pianca et al., 2015; Plant & Holman, 1997). The lack of consistency of shoreline intensity at Narrabeen results in greater difficulty in identifying the

shoreline, thus making beach state labelling more difficult. Specifically, the key difference between LTT/Ref classes and LBT is the distance between the shoreline and the sandbar. At Narrabeen, however, the separation between shoreline and sandbar is less obvious than at Duck, due to the former's lack of a consistent optical signature of the shoreline.

As described in Section 3.2.1, the modal bar states at Narrabeen and Duck differ, with Duck existing in a slightly higher (intermediate) energy state more consistently than Narrabeen (intermediate-reflective). Narabeen generally has a morphology that is contained closer to the shoreline, consistent with it being a generally more reflective beach than Duck due to a larger grain size and lower-energy wave climate. A variogram analysis (Balaguer, Ruiz, Hermosilla, & Recio, 2010; Bohling, 2005; Wu, Peng, Shan, & Cui, 2015) was performed in the cross-shore and the mean length scales were quantified as the average range of the variograms for the test dataset. The mean length scales of sandbar features at Narrabeen were smaller than at Duck by 5%, 22%, 24%, 13%, and 16% for states Ref, LTT, TBR, RBB and LBT, respectively. This suggests that the cross-shore position of the sandbar is generally further offshore at Duck than at Narrabeen, and so the physical sandbar features, such as rhythmicity or welding, can be more exaggerated at Duck than at Narrabeen. Furthermore, individual beach states at Narrabeen have length scales (as defined by the average range of the variograms) that are more similar to one another (range between 607–627 m), compared to at Duck (range between 664–805 m). Overall, the differences in inter-state length scale variability and overall length scales magnitudes could contribute to the relative clarity of beach state at Duck compared with Narrabeen, leading to higher CNN performance and inter-labeller agreement.

3.5.3 Data Requirements for Skillful Transfer of the CNN to New Sites

Section 3.4 presents results from experiments where the composition of training data from each beach was even or completely from one beach. A third set of experiments were performed wherein the percentage of data from each beach in

the training set differed. The intention of the experiments was to determine how much data from a different (or new) location was necessary to obtain adequate test skill. This is important when considering the use of such a CNN on new sites where limited training data may be available. Figure 3.8 shows the F-scores for different ratios of data added to the training set and then tested on each of the three single-location test data sets (Nbn, Duck, Combined).

For the training set ratios which consisted of at least 5% of data from either location, when the CNN was tested on the combined data set the F-score remained within 15% of the max skill (F-score = 0.69) suggesting the model was relatively insensitive to training data composition when presented with a range of diverse images from both sites. In contrast, when the CNN was tested on the individual sites (Duck or Nbn) F-scores decreased if the percentage of images from one location dropped under a certain threshold. At Duck and Narrabeen (blue line in Figure 3.8), skill became lower than 10% of its maximum value (maximum values of $\sim$0.8 and $\sim$0.64 at Narrabeen, respectively) if fewer than 25% of than training data was from Duck. Overall, this suggests for reasonable transferability of the model, a minimum of 25% of the data should come from any new sites proposed.

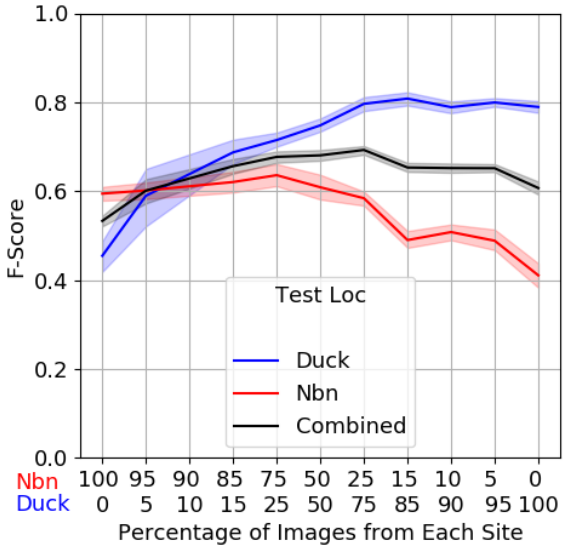


Figure 3.8: F-scores for CNN ensembles trained with varying ratios of training data and tested at Narrabeen (red), Duck (blue) and the combined data set (black). The shading represents the 95% confidence interval for the ensembles for each test. The x-axis shows the number of training images per class per location.

3.6 Conclusions

This study applied a convolutional neural network (CNN) to recognize beach states from daytime exposure (daytimex) Argus imagery at two contrasting beaches, Duck, NC, USA and Narrabeen, NSW, Australia. Three CNNs were considered: two trained with data from each of the sites individually, and one trained with data from both sites. The trained models were then tested on images from each site that were not included in the training data. The model results were compared with each other and to the agreement between four domain experts who manually labelled the dataset. The results showed that CNN ensembles that were trained and tested at the same site had skill that was comparable to inter-labeller agreement of the same test data set, and that the overall skill was higher at Duck

(F-score = 0.80) than at Narrabeen (F-score = 0.61). CNN per-state accuracy was comparable to inter-labeller per-state agreement for Duck and slightly lower for Narrabeen. For both sites, the highest accuracy was in the classification of the low-energy Ref and LTT states, while the lowest skill of the CNN was in classifying the rhythmic states of RBB (Nbn) and TBR (Duck). The combined-CNN had the highest skill on the combined test dataset (F-score = 0.68, versus F-score = 0.59 and 0.53 for Duck-CNN and Nbn-CNN, respectively). Compared with the self-trained CNNs, it similarly confused RBB with LBT at Narrabeen, and TBR with LTT at Duck, with slightly lower per-state accuracy for Ref, LTT and RBB states at Duck.

When the single-site trained CNN was transferred to the other site, the F-scores dropped by 20% (Nbn-CNN tested at Duck), and 58% (Duck-CNN tested at Narrabeen). In additional transferability tests, the composition of the training dataset was altered to contain different proportions of training images from each of the two locations. Overall, comparable skill (10% of the maximum skill) to the self-trained CNN tests was achieved when at least 25% of the data came from the transfer site. Further,

Saliency maps were used to identify the specific image regions that were used for CNN decision making. They showed that relevant regions were highlighted for determining beach state classification (e.g., the swash region for the Reflective state, or the offshore bar and shoreline for Longshore Bar Trough), suggesting that the CNN had accurately identified key features that distinguish beach states at these two contrasting sites.

Additionally, the alongshore variability of beach state should be considered in beach state detection in order to extract the most information available from the image. Ultimately, a globally applicable CNN that would be able to detect beach state for all locations could be developed by using labelled data from sites around the world.

Chapter 4: From Beach State to Breach Probability Simplex with a Convolutional Neural Network and Argus Imagery

Ashley N. Ellenson, Greg W. Wilson, Joshua A. Simmons, Tyler J. Hesser, Kristen D. Splinter

In preparation for JGR Earth Surface

ABSTRACT: Nearshore hydrodynamics govern the dynamics of nearshore ecology, surf zone pollutant transport, recreational safety, and coastal erosion. Beach morphology and incident wave conditions drive nearshore dynamics and can be qualitatively described using a categorization system called beach states. Beach states encompass discrete shapes frequently observed nearshore patterns. However, beach morphologies can exhibit alongshore variability of dominant state, and are not always accurately described by a single beach state. This study expands beach state framework of (L. D. Wright & Short, 1984) to acknowledge morphologic complexity by quantifying the alongshore variability of beach state within a probabilistic framework called a beach probability simplex. A beach probability simplex is a five dimensional vector where each entry corresponds to the planform area of a unique state from (L. D. Wright & Short, 1984) relative to the total area of a given section of beach. A convolutional neural network (CNN) is used to detect the beach probability simplex from a 28-year time series of remotely sensed morphology from Duck, North Carolina, USA. Properties of the beach probability simplex include entropy (E) and mean simplex value ($\overline{S}$) and a simplex label. Entropy quantifies the spread of states within a probability beach simplex, and $\overline{S}$ is the weighted mean value of the entries of the probability simplex. A majority (67%) of the 28-year dataset of remotely sensed morphologies exhibited alongshore variability of beach state, either with two states present as bi-modal simplices, or three states present as tri-modal simplices. The simplices that encompassed the most off-shore beach states (e.g., Longshore Bar Trough) corresponded to higher significant wave heights and northerly wave angles (average $Hs > 1.20$, average $MWD > 0°$), and the simplices that encompassed the most on-shore beach states (e.g., Low Tide Terrace) corresponded to lower significant wave heights and southerly wave angles (on average, $Hs < 1.00$, $-3.47° > MWD > -12.61°$). A spectral analysis of $\overline{S}$ showed that the variability was dominated by an annual peak, and a slightly lower four year peak that did not rise above the 95% confidence level. Mean simplex values and entropy were partitioned by unique day of year, and averaged over the 28 years. Twenty-eight year daily averaged mean simplex value correlated with

significant wave height and mean wave direction (R=0.77 and 0.85, respectively). In contrast, 28 year daily averaged entropy correlated slightly with significant wave height (R=0.44), and did not correlate with mean wave direction (R=0.24). The most complex morphologies were found in fall and spring (average entropy was 0.67 and 0.61 for fall and spring, respectively, versus 0.51 and 0.56 for summer and winter).

4.1 Introduction

Knowledge of nearshore morphology and associated hydrodynamics is crucial for coastal environmental health and safety. Recrational hazards can be determined by morphological configurations; accreted versus erosive morphological configurations determine hazard to beach users (Benedet et al., 2006; A. Short & Hogan, 1994b). Nearshore ecology is affected by surf zone width (Morgan et al., 2017; Shanks, Morgan, MacMahan, & Reniers, 2017; Talbot & Bate, 1987) and rip currents (Shanks et al., 2018). Pollutant fate and transport is partially governed by the relative degree of alongshore versus rip current circulation patterns (Boehm, 2003; Boehm, Keymer, & Shellenbarger, 2005; Grant et al., 2005; Winckler et al., 2013). Beach erosion, a hazard for coastal communities, is affected by sub-aqueous morphology affects total water levels (Gomes da Silva et al., 2020; Hughes et al., 2014; Ruggiero, Komar, McDougal, Marra, & Beach, 2001).

Surf zone morphologies have been observed to generally exist in similar patterns called beach states (Chappell & Eliot, 1979; Greenwood & Davidson-Arnott, 1979; Lippmann & Holman, 1990; Price & Ruessink, 2011b; Ranasinghe et al., 2004; L. D. Wright & Short, 1984). Beach states rely on spatial patterns and associated hydrodynamics of surf zone morphology to delineate beach states, grouping bars into categories based on specific combinations of morphological characteristics (e.g., rhythmicity or trough continuity). Beach states have been used within the coastal community to qualitatively describe the beach and associated hazards (Benedet et al., 2006; Scott, Russell, Masselink, Wooler, & Short, 2007; A. Short & Hogan, 1994b) validate models (Dubarbier et al., 2017) or as a driving force in shoreline

change modeling (Davidson et al., 2013).

A four year observational study by (L. D. Wright & Short, 1984) in southeastern Australia, hereafter referred to as WS84, established the most commonly used beach state framework. WS84 observed six beach states, and ordered them with respect to incident dimensionless fall velocities:

$$\Omega = \frac{H_b}{w_s T_p} \tag{4.1}$$

H_b is wave height at breaking, w_s is sediment fall velocity and T_p is peak wave period. On the two ends of the beach state continuum are dissipative (where the bar is most offshore, $\Omega \geq 6$) and reflective states (where the bar is most onshore, $\Omega \sim 1$). Dissipative configurations mainly occur in the winter, and reflective configurations in summer. The nearshore morphology evolves between dissipative and reflective states through intermediate states that exhibit various degrees of rip currents and welding ($1< \Omega <6$). Herein, higher energy beach states will be referred to as "offshore" and lower energy beach states will be referred to as "onshore."

Lippmann and Holman (1990), hereafter LH90, expanded the WS84 classification system to consider the degree of longshore rhythmicity in categorizing the nearshore morphology shapes. LH90 analyzed a two year time series of beach states from remotely sensed nearshore morphology at Duck NC, also the study site in this work. The LH90 classification rules consider the following attributes of sandbar spatial structure: 1) existence or absence of a bar; 2) location of the sandbar relative to the shoreline, specifically if the cross-shore location had was close ($\sim O(10m)$) or relatively further away ($\sim O(10^2 m, 10^3 m)$); 3) the longshore structure of the sandbar, specifically the degree of linearity or rhythmicity; and 4) the existence of trough, specifically whether the trough was continuous or discontinuous. Like WS84, LH90 also arranged the beach states in a specific order, but arranged the beach states with respect to average incident significant wave height (Hs) instead of Ω, finding that both Ω and Hs parameterized the continuum

equivalently.

In reality, surfzone bathymetry can exhibit complexity that are not described with the beach state framework (Armaroli & Ciavola, 2011; Price & Ruessink, 2011b), because the beach state framework uses a prescribed assembly of morphological attributes for each individual beach state. However, the attributes from the beach state framework can be applied to complex shapes in an alongshore variable way, thus framing the morphology as a non-uniform beach state. Morphologies that have alongshore non-uniform beach state can emerge due to alongshore variability of incident forcing conditions (e.g., due to physical structures or shoreline curvature) (Pianca et al., 2015; Thornton et al., 2007), and can cause a high degree of alongshore variability in wave breaking patterns and associated nearshore hydrodynamics (Gomes, Mulligan, Brodie, & McNinch, 2016; Mulligan, Gomes, Miselis, & McNinch, 2019; Quartel, Kroon, & Ruessink, 2008). Additionally, the over-simplification of the morphology into a discrete beach state results in labelling subjectivity; two people might notice different characteristics of the morphology and assign a beach state differently (Lippmann & Holman, 1990; Price & Ruessink, 2011b; Ranasinghe et al., 2004)

Herein, we extend the qualitative beach state framework to a quantitative beach probability simplex in order to represent "mixed-state" morphology types that are otherwise not included in the LH90 and WS84 systems. Generally, a probability simplex is an n-dimensional vector, where each entry corresponds to a probability, and the vector sums to one (Aitchison, 1982). For example, in geologic applications, a sediment sample might be described by different percentages of silt, sand, or clay. In the present study, the overall morphology is described by different percentages of individual beach states. A beach probability simplex is therefore flexible enough to represent an alongshore variable beach-state, and can encompass discrete beach states as an endmember case. The level of complexity, or the relative degree of disparate states, of the beach probability simplex is quantified with Shannon's entropy (Shannon, 1948). Shannon's entropy originally quantified the level of information within a telecommunication, where a diversity of signals

will provide the most information and have the highest entropy. Herein, a diversity of characteristics from different states results in higher entropy. For example, morphology with a uniform length scale of rhythmicity (exhibiting features of one state) is less complex than morphology that exhibits linearity, a trough, and rip currents in the alongshore (exhibiting features of three states).

This study builds off (A. N. Ellenson, Simmons, Wilson, Hesser, & Splinter, 2020), hereafter referred to as E20, wherein a machine learning algorithm, a Convolutional Neural Network (CNN) was applied to identify single beach states from remotely sensed wave breaking patterns. Given the high volume of data available to the coastal community (R. Holman & Haller, 2013; K. Splinter, Harley, & Turner, 2018), it is important to develop methods that can extract useful information from these data. More information can be extracted from a single image by using a probability simplex representation of the beach configuration, as it leverages the CNN's ability to detect a variety of morphological characteristics within an image. Additionally, the flexible and automated nature of this method reduces the subjective nature of beach state labelling due to state oversimplification and inter-labeller biases.

This contribution first describes the field site and dataset, the windowing routine used to generate the beach probability simplex, and then applies the detection technique to a 28-year dataset of remotely sensed morphology from Duck, NC. Each probability simplex is associated with a categorical representation, a mean simplex value, and a measure of entropy. The simplex distribution and dynamics are presented with associated wave conditions, followed by a seasonal analysis of the mean simplex value.

4.2 Data and Field Site

The field site is a sandy barrier beach located at the U.S. Army Engineer Research and Development Center Field Research Facility (FRF) in Duck, NC. The mean spring tide range is microtidal at 1.2 m. The beach slope ranges between 0.108 (at the foreshore) and 0.006 (8 m depth) (Stauble, 1992). The medium to fine grain

quartz sediment has a median grain size of 0.5 mm between the bar and the shoreline (20% carbonate material) and an offshore grain size of 0.2 mm (Birkemeier et al., 1981).

Wave data (significant wave height, H_s, mean wave direction MWD, peak wave period, T_p) were taken from the 8-m array at the Field Research Facility in Duck NC. Any gaps within the 8-m array time series were filled with data from the former linear array (before it was upgraded in 1990), or by a 17-m or 24-m waverider buoy (29 months out of 336), where these wave data were shoaled to 8m depth using linear wave theory and assuming an alongshore-uniform coastline. Mean wave direction was calculated following Kuik, Van Vledder, and Holthuijsen (1988), and converted to the angle from shore-normal (north positive and south negative) accounting for the 72° shoreline orientation relative to north. The dimensionless fall velocity was calculated as in equation 4.1. In this formulation, H_b was approximated following Komar (1998) as $H_b = 0.39 g^{\frac{1}{5}} (T_p H_0)^{\frac{2}{5}}$, g is the acceleration due to gravity of $9.8 m/s^2$ and H_0 is the deep water significant wave height, where the observed H_s was back-shoaled to deep water. The sediment fall velocity, w_s, dependent on d_{50} (taken to be 0.3mm), was $0.04 m/s$.

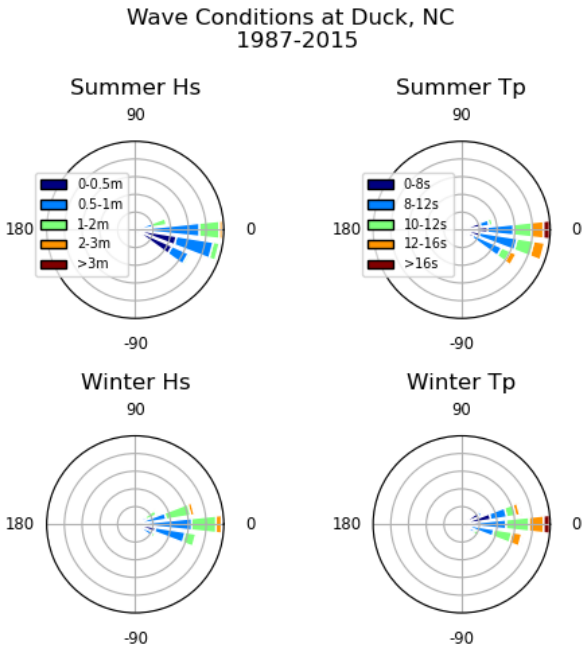


Figure 4.1: Daily averaged wave conditions at Duck, NC for 1987-2015. Total points in summer is 4943 and winter is 4848. Each bin on the radial axis represents 314 data points for summer plots (top panel), and 333 data points for winter (bottom panel). Directional bin widths are 15°. East is 0°, north is 90°, and south is −90°.

Wave data were taken between October 1, 1987 to December 31, 2015, resampled to daily averages, and cleaned to remove measurement errors. Mean period, T_m, was used in the place of erroneous T_p values. The cleaned data are presented as a wave rose in Figure 4.1, totalling 4848 points for winter and 4943 points for summer. The 28-year average Hs is 0.84 m, T_p is 9.04s, and MWD is -4.84°. Winter (months September-February, inclusive) waves have an average Hs of 0.95m, Tp of 9.5s, and generally come from the east. Summer (March-August, inclusive) waves have a lower than winter average Hs of 0.74m, similar to winter Tp of 9.4s, and come from the E/ESE. In winter, higher wave heights coming from the NE are due to extra-tropical cyclones (north-easters). Distant tropical cyclones (hur-

ricanes) can generate higher wave heights associated with long periods ($Tp > 16s$) during late summer and fall.

4.2.1 Imagery Dataset and Cleaning: Argus Daily Time Exposure Images

The data is comprised of ortho-rectified, grayscale, daytime exposure images from the Argus station at the Army Corps Engineering Field Research Facility in Duck, North Carolina. The data set spans from January 1, 1987 to December 30, 2015, totalling 9531 images. Argus images capture wave breaking patterns in the surf zone, from which morphology can be inferred (Holland et al., 1997). The images are evenly distributed across the 28-year time span, with approximately 700-840 images per month, and a median number of 810 images per month.

The data were quality controlled to remove images where the bar shape was not visible due to a variety of reasons: image distortion due to drops on the camera lens; images where the data was never received from the camera; and high Hs conditions, where wave dissipation saturated the surf zone and bar shape was obscured. These occurrences were automatically detected using a purpose-made convolutional neural network (called a data quality CNN), trained following E20. The data quality CNN was trained with seven classes, where the training dataset numbered 100 images per class; the classes comprised the five WS84 beach states, a no visibility class, and a fully dissipative (storm) class. The data quality CNN identified a total of 7% (658) of images as low-quality and removed from the entire dataset, although it was possible some poor quality images remained due to imperfections in the classifier. The number of poor quality images remaining are thought to be small enough in number to not affect the final results of this study; a sample of 100 non-removed images were inspected and of those only six were found to be false negatives. The number of images removed were not distributed evenly throughout the seasons. Storms that rendered images low quality occurred more frequently during winter months. Specifically, the the highest proportion of images

removed was in the months February, March and April. On average, $\sim 9\%$ of data were removed from these months. The lowest proportion of images removed was in August ($\sim 4\%$ of data were removed from this month).

Steps were also taken to remove images that were: 1) misclassified by the CNN; and 2) exhibited only partial visual signature of the nearshore morphology due to waves breaking preferentially on shallow bar areas due to low Hs. Simplices (see Section 4.3.4) that encompassed states from either end of the WS84 beach state order (e.g., RBB and LTT), called "non-adjacent simplices," were usually due to preferential wave breaking or CNN mis-classification and were removed from analysis. For an example, see Figure 4.6, purple box. A manual inspection of 268 non-adjacent simplex LBT-LTT images showed that 30% were correctly labelled, 20% were due to low incident wave energy and 50% were labelled incorrectly. Also, weak wave breaking by definition was also the main feature of simplices encompassing the Reflective state. Therefore, simplices encompassing the reflective state as well as non-adjacent simplices (N=2592) were removed, resulting in a total of 6370 data points for analysis.

4.3 Methods

4.3.1 Convolutional Neural Network

Machine learning algorithms generate a prediction given an input after learning relationships between the input and desired predictions in a process called training. The study herein uses a subset of machine learning called deep learning. In E20, a deep learning CNN called ResNet-50 (He et al., 2015) was trained to predict beach states given an Argus daytimex image of nearshore wave breaking patterns.

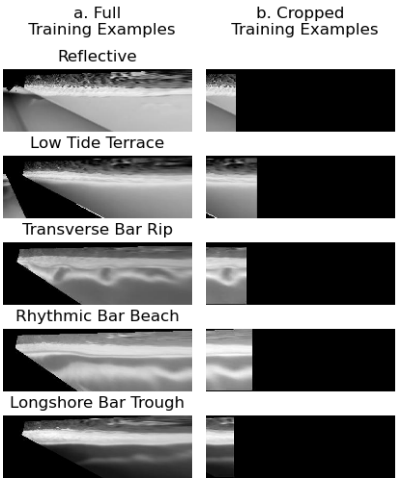


Figure 4.2: Training examples for the "full" dataset and the cropped images that were fed into the windowing routine.

The CNN architecture, described in more detail in E20, Appendeix A, consists of the following: 1) a convolutional layer; 2) a max-pooling layer; 3) four convolutional blocks/ReLU combination; 4) global average pooling layer; 5) a fully connected layer; and 6) a softmax layer. The convolutional blocks are connected through residual connections and consist of: 1) 1x1 convolution operation/batch normalization; 2) 3x3 convolution operation/batch normalization; and 3) 1x1 convolution operation/batch normalization. The final output of the CNN is a 5x1 probability vector produced by the softmax layer, where each entry corresponds to the likelihood that the CNN detects the beach state in an image. For most machine learning applications, the entry with the highest probability is taken as the prediction of the image. However, this study used the full softmax output (the probability vector) to allow for the potential of more than one state within the CNN's field of view (see section 4.3.3).

This model was trained 'end-to-end', meaning all of the parameters of the algorithm were trained with supervised learning. The process of supervised learning requires two datasets (training and validation) and assessing model skill requires a third (testing). The datasets used herein are described in Section 4.3.2. The

training dataset provides the information for the CNN to learn relationships between the inputs and desired output. The skill on the validation dataset is used to determine if the CNN parameters need to be altered; as long as the validation skill increases, the CNN is learning, and the learning continues. If the validation skill plateaus, the CNN has finished learning, and training ends in a process called "early stopping." In E20, the CNN was optimized on the training dataset (not the validation dataset). Herein, the CNN was optimized over the validation dataset. The hyperparameters (batch size, learning rate, early stopping) for the supervised learning routine are the same as in E20.

4.3.2 Training and Validation Datasets

In E20, the machine learning routine was provided full images of the beach and predicted a single discrete label, such as the examples provided in Figure 4.2a. In contrast, the present study seeks to capture the alongshore variability of a beach state. To do so, the full image is cropped in the alongshore into sub-images as in Figure 4.2d, and provided as input to the CNN in a process called a "windowing routine" as described in Section 4.3.3. The CNN was trained to recognize charactreistics of beach states as detailed in Table 4.1.

In order to keep the training and test datasets consistent (i.e., cropped images), the training dataset from E20 (100 images per class) were cropped for training in the present study, such as in Figure 4.2, right panel. In E20, each full image was considered a "pure" (i.e., alongshore uniform) beach state. The crops were made manually such that the characteristics of each beach state, as detailed in Table 4.1, were apparent in the cropped image. Since the identifying characteristics for each class can have different length scales (e.g., TBR rhythmic cross-sectional lengths can vary in the alongshore) or more than one characteristic (e.g., LTT contains small rhythmic features and linearity), the training dataset encompassed variety of morphological length scales and and identifying attributes for each state that are detailed in Table 4.1, "Variation" column. The training dataset also encompassed a variety of window sizes in the sub-images. Each training image from E20

was cropped at least once and maybe twice, depending on if the characteristics described in Table 4.1 could be seen in one the full image more than once. Additionally, cropping increased the number of images available for use, so this study included a validation set in addition to the training and test sets. The CNN used in this study had the same architecture but different internal parameters than the one in E20 due to the different training images and the optimization of the CNN on the validation set.

Each cropped image was overlaid onto a black background in a composite image such as in Figure 4.2d. The training dataset totalled 100 composite images per beach state, the validation set totalled 20 composite images per beach state. The training and validation images were augmented with the four augmentations that were used in E20, specifically: 1) random horizontal or vertical flips; 2) horizontal or vertical translations; 3) random rotation up 45 degrees; 4) random erasing and 5) gamma darkening. With the augmentations, there were 500 training images per class and 100 validation images per class. In total, there were 2500 training images and 500 validation images.

4.3.3 Simplex Detection: Image Windowing Routine

In the windowing routine, alongshore sections of the image are fed to the CNN serially, and the CNN therefore makes alongshore variable predictions. The process is illustrated in Figure 4.3.3. The images were first cropped by alongshore windows that spanned the entire cross-shore, with an alongshore window length of 320m (128 pixels) and at a series of alongshore lags spaced 56m (32 pixels) apart (cyan square, Figure 4.3.3a). The window length was chosen such that it would be larger than the smallest alongshore length scales in a typical image. Observations have found that the smallest alongshore distance of a a rip current is $\sim$50 m (MacMahan, n.d.). The image crops were then overlaid onto a black background to make a "composite image," (Figure 4.3.3d). The composite images were necessary to make the total input size of the image consistent with the image size in the training dataset (512x512 pixels); since the CNN was trained on images that are 512x512 pixels,

the predictions should also be made on 512x512 images.

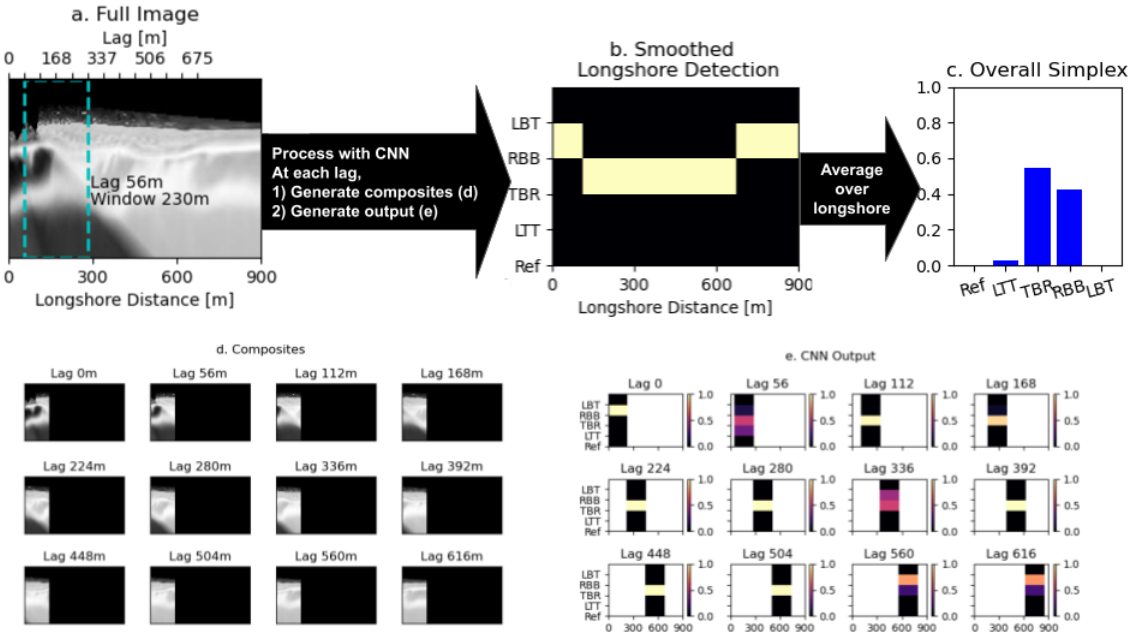


Figure 4.3: Example of CNN detecting the beach state simplex; panel (a) is an example full image overlaid with a cropping window size 320m at a lag of 56m; (b) is the longshore simplex after smoothing; (c) the overall simplex; (d) shows the composite images fed into the CNN (i.e., the cropped sub-images from (a)); and (e) is the CNN output for each composite image shown in (d).

The composite images of Figure 4.3.3d were fed to the CNN to produce probabilistic output as in Figure 4.3.3e. The final output from the CNN totalled thirteen 5x1 probability vectors (twelve of which are shown in Figure 4.3.3e). As a result of the lags being shorter than the windows, the windows would overlap in the alongshore, and so the CNN would produce approximately five outputs for each location alongshore. For example, in Figure 4.3.3e, the welded region of the shoreline at the right side of lag 56m composite image can also be seen on the right side of lag 0m, right center of lag 112m, left center of lag 168m and slightly on the left of lag 280m. In order to aggregate the overlapping predictions, the CNN output for each

window and lag combination were averaged to produce an alongshore detection as in Figure 4.3.3b, where each alongshore position was associated with a beach probability simplex.

The original alongshore detection at times had noisy results, therefore it was smoothed with respect to alongshore position in such a way that the beach probability simplex in each position summed to one. Specifically, only the highest 95th percentile estimations were chosen for each alongshore position. If one of the cameras was missing when the data was collected, a triangular black region would result from the rectification routine, cutting off informative sections of the bar signal in the cross-shore (not shown); such estimations were removed from further analysis. The final, smoothed version of the alongshore detection is shown in Figure 4.3.3b. The smoothed alongshore detection was then averaged over the alongshore to produce a beach probability simplex for the entire image, shown in Figure 4.3.3c.

In summary, the steps of the image windowing routine were the following:

1. Crop images with window 320m at lags of 56 m (Figure 4.3.3a) to make composite images (Figure 4.3.3d).

2. Generate CNN probabilistic output for composite images (Figure 4.3.3e).

3. Average CNN probabilistic output with respect to lags to produce alongshore detection, and smooth alongshore detection with respect to each alongshore position (Figure 4.3.3b).

4. Average alongshore detection to produce beach probability simplex (Figure 4.3.3c).

4.3.4 Definitions and Terms: Probability Simplex, Simplex Label, Entropy and Mean Simplex Value

A beach state probability simplex, $\overrightarrow{S}$, is a probability mass function that represents the overall beach state as the relative contribution each state. Each entry

corresponds to a probability, and the simplex entries sum to one:

$$\overrightarrow{S} = \{(S_0, .., S_4) : S_i > 0 (i = 0, ..., 4), S_0 + ... + S_4 < 1\} \tag{4.2}$$

The probability simplex is built off the characteristics of five of the original six states in WS84. Duck is primarily characterized by one bar, and therefore dissipative conditions often coincide with storm conditions and saturated surf zones, obscuring the bar shape. Each entry within the simplex represents the relative alongshore length of a certain WS84 state within an entire given span of beach alongshore length. The identification of a beach state was defined as the identification of the characteristics listed in Table 4.1. The WS84 states were used, as opposed to the LH90 states, due to their being more commonly used in coastal studies in comparison to LH90 (Garnier, Calvete, Falqués, & Dodd, 2008; Masselink & Short, 1993; Price & Ruessink, 2011b; Ranasinghe et al., 2004). The alongshore length of all detections of a given beach state in the image, as produced by the windowing routine (Section 4.3.3) are summed to give a total length, L_i. The simplex entry was calculated by dividing L_i by the total length of the beach in the image:

$$S_i = \frac{L_i}{L_{total}} \tag{4.3}$$

Several useful products can be derived from the beach probability simplex: the simplex label, the mean simplex value $\overline{S}$, and entropy E, which are described in Table 4.2.

The simplex can be represented with a categorical representation, which is useful in the following ways: conveying which attributes are present in the image colloquially; for categorizing the frequently emerging shapes; and directly comparing the results here with previous state studies. The categorical representation, referred to as the label, is the combination of the nonzero entries associated with the states that make up the simplex. For example, if an off-shore barline is both rhythmic (RBB) and linear (LBT) in the alongshore, its probability simplex may read [0, 0, 0, 0.3, 0.7], and in that case its label would be LBT-RBB. Labels that

Class #	WS84 Beach State	LH90 Beach State	Characteristic	Variability
0	Reflective (Ref)	A	Lack of wave breaking, no offshore bar	Alongshore uniform
1	Low Tide Terrace (LTT)	B	Terracing, small rhythmic features within the terrace	Cross-shore distance of terrace, number and size of rhythmic features
2	Transverse Bar Rip (TBR)	C, D	Attachment to the shoreline and rhythmic features that exit the surf zone	Rhythmic length scales, number of rhythmic features
3	Rhythmic Bar Beach (RBB)	E	Trough, rhythmic bar line	Cross shore distance between shoreline and barline, length scales of rhythmic features, extent of curvature
4	Longshore Bar Trough (LBT)	F, G	Trough, linear bar line	Cross-shore distance between shoreline and bar line, strength of breaking signal on bar

Table 4.1: Class numbers, WS84 Beach state, LH90 beach state, characteristics and the variability thereof associated with each state. Note that LH90 further subdivided the WS84 categories, and so more than one state might correspond with an WS84 state.

encompass one state are called "uni-modal", two states called "bi-modal" and three states called "tri-modal."

The mean simplex value is a simplified representation of the five dimensional vector that represents the entries within the simplex. The mean simplex value is useful in that it is more readily plotted and used in models than a five dimensional vector ($\overrightarrow{S}$). The mean simplex value is defined as a weighted sum of the entries

Term	Definition	Use
Probability simplex ($\overrightarrow{S}$)	Probability vector where each entry corresponds to WS84 beach state, eqn. 4.2, eqn. 4.3	Flexible framework that represents nuances in morphological characteristics
Label (e.g., LBT-RBB)	Combination of state names of non-zero entries in probability simplex	Categorical representation of the probability simplex; colloquially useful since simplex characteristics are readily identifiable by name; directly comparable to previous studies
Mean simplex value ($\overline{S}$)	Weighted mean of simplex, eqn. 4.4	Continuous representation of simplex, readily modelled
Entropy (E)	Shannon's entropy of probability simplex, eqn. 4.5	Quantification of probability simplex complexity

Table 4.2: Glossary of terms specific to the simplex, its derived quantities, and how each are used

within the simplex.

$$\overline{S} = \sum_{i=0}^{4} iS_i \tag{4.4}$$

Where i is the state value number, defined in Table 4.1, and S_i is the simplex entry associated with that state (equation 4.3). The mean simplex value is only calculated for the probability simplices that have entries adjacent to each other; if the entries are not adjacent to each other, the mean simplex value would represent a simplex that encompasses two end member beach states as an intermediate beach simplex. The diversity of states within the simplex can be measured with Shannon's entropy:

$$E = \sum_{i=0}^{4} S_i log_2 \frac{1}{S_i} \tag{4.5}$$

Entropy is different from other definitions that relate the physical alongshore and cross-shore variability of morphologies in that entropy represents the level of state mixing.

The mean simplex value is a more stable measurement than the entropy measurement. While the mean simplex value represents the average state represented in an image, entropy describes the spread of states within a simplex. Since the entropy is calculated by taking the logarithm of the state probability ($S_i < 1$), slight changes in the values within the probability simplex are amplified in the calculation of entropy. Therefore the entropy measurement is also sensitive to the CNN detection of each state which can potentially be a noisy measurement.

4.4 Results and Discussion

4.4.1 Simplex Distribution

Uni-modal states, the original WS84 states, comprised 32% of the adjacent states. Simplices that encompassed two states, bi-model simplices, were most prevalent, comprising 51% of the data, and tri-modal simplices comprised 16% of the data. Overall, the majority (79%) of the simplices were comprised of states that were adjacent to each other on the WS84 beach state continuum (e.g., RBB combined with TBR to form RBB-TBR). The remaining 21% of the simplices were made up of non-adjacent states (e.g., RBB combining with low energy state LTT to form LBT-LTT) and were removed from the analysis as described in Section 4.2. Figure 4.4 shows the simplex "family tree," illustrating how the states combined to form simplices and the associated proportion of the dataset. Two of the four bi-modal simplices were also identified in LH90 (indicated in Figure 4.4). The states original to this study include the two tri-modal simplices and one bi-modal simplex.

The simplex labels were grouped into sub-groups called "families" for descriptive purposes, indicated by the font color in Figure 4.4. The physical attributes from the LH90 framework, outlined in Section 4.1, provided an initial basis for the families, except for the designation of the RBB simplex. Based on LH90's

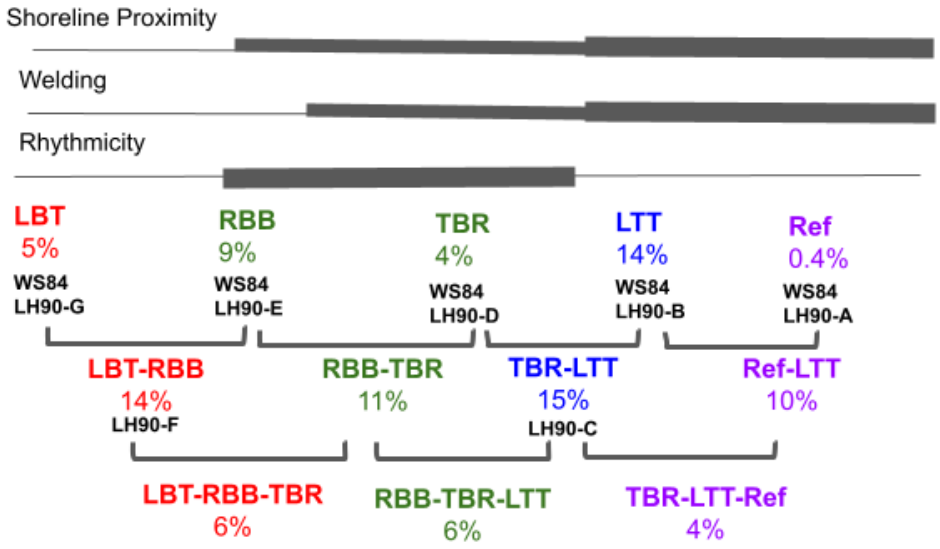


Figure 4.4: Simplex "family tree" illustrating how the WS84 states were most likely to combine to form simplices. WS84 and LH90 states are indicated in black boldface below the simplex name. The font color corresponds to state family; LBT is red, TBR is green, LTT blue, and Ref purple. The horizontal lines indicate the morphological characteristics (e.g., degree of linearity) differentiating the families, where the line weight corresponds with degree within the family.

classification, the RBB simplex would be grouped with the LBT family since the simplices in the LBT family were furthest away from the shoreline with a continuous trough. However, in this study the RBB simplex was designated within the TBR family. The RBB simplex was more similar to TBR simplices than to the LBT simplices in that the RBB simplex, similar to the rest of the simplices within the TBR family, coincided with, on average, a decrease in significant wave height from background conditions and average shore-normal wave height. In contrast, the simplices in the LBT family were associated with an increase in dimensionless fall velocity, and slightly northern wave directions. See Figure C.1c in Appendix C.1 for further detail).

The physical attributes of the first family, LBT (20% of the entire dataset), had varying degrees of linearity, was furthest away from the shoreline and had a pronounced trough. The second family, TBR (24% of the dataset) had a high degree of rhythmicity, intermediate in distance to the shoreline, and varying degrees of welding. The third family, LTT (22%), had a high degree of linearity, the closest proximity to the shoreline, and a high degree of shoreline attachment. The final family, Ref ($\sim$ 11% of the dataset), was characterized by a lack of wave breaking.

LH90 found that states within the TBR family occurred most often, followed by states in the LBT family, and that LTT states occurred least often. This study found that the simplices of the TBR and LTT families occurred with similar frequency (30% and 29% of the dataset were TBR and LTT simplices, respectively), followed by LBT simplices (25%). The high occurrence of the uni-modal LTT simplex could be due to the fact that low wave breaking images remained in the dataset despite data cleaning (see Section 4.2), resulting in the detection of the LTT state.

4.4.2 Simplex Shapes, Entropy and Mean Simplex Values

Each simplex is associated with a mean simplex value, $\overline{S}$, and entropy quantity, E, as defined in equations 4.4 and 4.5. Mean simplex value and entropy were not calculated for simplices encompassing the Reflective state due to their removal

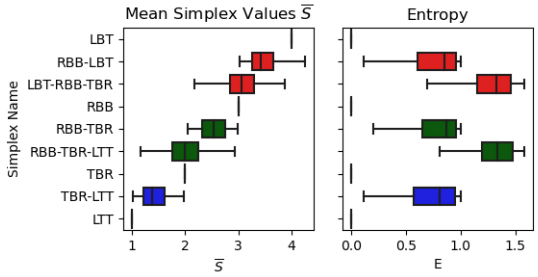


Figure 4.5: Simplices ordered by average mean simplex values and their associated entropy ranges. The vertical lines in the boxes represent the median, and the whiskers the 5th and 95th percentile values.

from further analysis (see Section 4.2). As shown in Figure 4.5, as the number of states in the probability simplices increased from two (bi-modal) to three (tri-modal), the range of their mean values also increased, due to the varying portion of each image that may be classified as a specific state. Entropy, the measurement of probability simplex diversity, by definition increases with the number of states in a simplex. The median entropy for tri-modal simplices is 1.3, bi-modal is 0.86 and uni-modal, by definition, is zero. The mean simplex value by definition scaled between the families. LBT, TBR and LTT families had a mean $\overline{S}$ values of 3.44, 2.44 and 1.06. Mean entropy was higher for LBT and TBR families (0.86, 0.81, respectively) than for the LTT family (0.33). Both the LBT and TBR families encompassed tri-modal simplices that contributed to their high entropy values, while the LTT family only encompassed bi- and uni-modal simplices.

The LBT family (red box, Figure 4.6) was characterized mainly by the existence of a trough and was the furthest off-shore from the shoreline amongst all the families. The bi-modal simplex occurred most frequently within this family. The three simplices with the LBT state and their proportion of the entire dataset were: 1)LBT-RBB (14%) pronounced trough, rhythmic and linear bar; 2) LBT-RBB-TBR (6%) some welding, rhythmic and linear offshore bar; and 3) LBT (5%) pronounced trough, straight and linear offshore bar. Figure 4.6, panel A (red outline) shows examples of the LBT simplices. Generally, the wave breaking

patterns of the LBT-RBB-TBR had long O($\sim$ 900m) length scales of rhythmicity (such as RBB-LBT-TBR in Figure 4.6), and the bar would be attached only in one region in the alongshore.

The TBR family (green box, Figure 4.6) was characterized mainly by rhythmicity and with varying degrees of welding, and was of intermediate distance to the shoreline (relative to the other families). It encompassed the highest number (four) of simplices among the families, two bi-modal, one uni-modal, and one tri-modal: 1) RBB-TBR (11% of the dataset) rhythmic, alternating sections of trough or welding; 2) RBB (9%) rhythmic, pronounced trough; 3)RBB-TBR-LTT (6%) rhythmic sections had alternating sections of trough and welding, linear sections attached to the shoreline; and 4) TBR (4%) rhythmic and welding. Often, oblique transverse bars were included in RBB-TBR-LTT, as the bar would detach from the shoreline in the alongshore, thus forming a trough and identified as the RBB state by the CNN.

The LTT family (blue box, Figure 4.6) was characterized mainly by varying degrees of shoreline attachment and was closest to the shoreline. The LTT family encompassed the least number (two) of simplices among the families. The two simplices and proportion of the dataset were TBR-LTT (15%) alternating rhythmic and linear sections entirely welded to the shoreline and LTT (14%) linear welded to the shoreline.

Any simplex that had the Ref state was categorized in the Ref family, so it encompassed many non-adjacent simplices, two adjacent-member simplices (TBR-LTT-Ref, LTT-Ref), one uni-modal simplex (Ref) The existence of the Reflective state within a simplex usually indicated that there were low wave heights that resulted in preferential wave breaking in the alongshore. For example, in Figure 4.6d, offshore breaking occurred within several spots throughout the image, but the wave signal is not visual along the entire barline. The section of visible barline was labelled RBB, and the other sections were labelled LTT and Ref.

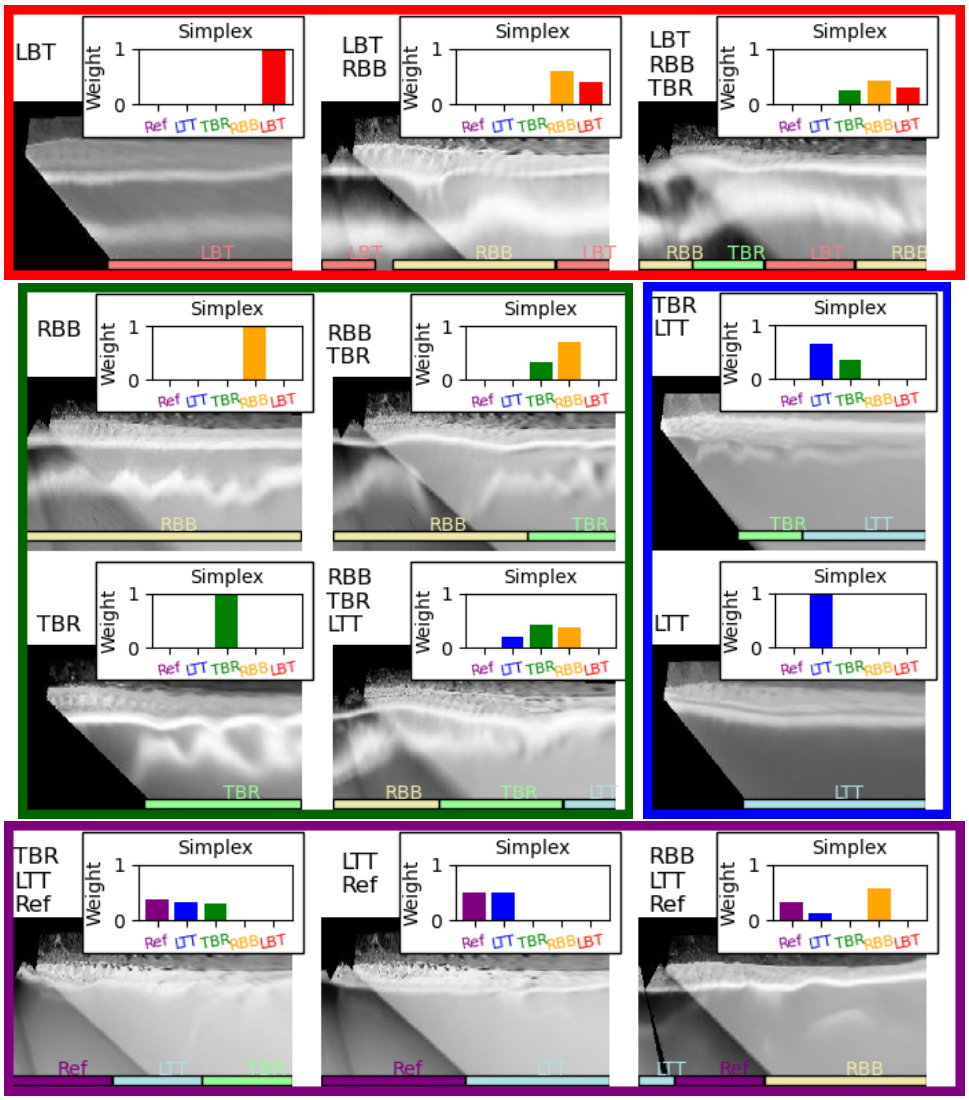


Figure 4.6: Examples of each adjacent simplex used in the analysis. Within each image, horizontal colored boxes indicate which state was detected in the alongshore. The simplex distribution is plotted in the top right corner. Each box outline indicates a family: red corresponds to the LBT family; green to TBR; blue to LTT; and purple to Ref.

4.4.3 Simplex Dynamics

The simplices were ordered with respect to mean $\overline{S}$ that corresponded with each simplex (centerline of boxes, Figure 4.5), following how WS84 ordered states. Simplices evolved between each other along the mean $\overline{S}$ order. The transition table in Figure 4.7 shows the probability of the simplex evolving from one to another, and the number of times the evolution occurred is shown in parantheses. The probability of moving to a simplex from a specific one can be calculated by dividing the number of movements to a simplex by the total number of movements from the first simplex:

$$P_{i,j} = \frac{C_{i,j}}{\sum_{i=1}^{N} C_j} \tag{4.6}$$

In this equation, the simplex transitioned to would be the i, j entry in the table, C is the number of occurrences that corresponds to that entry, and N is the total number of evolutions in the j^{th} row. A relative stability of each simplex is calculated as the probability of the simplex evolving to itself (i.e., the diagonal of Figure 4.7). The entries in the off diagonal reflect the probability of a simplex evolving to a different one. Transitions that occurred less than 1% (45 times) of the total number of transitions were removed from the analysis.

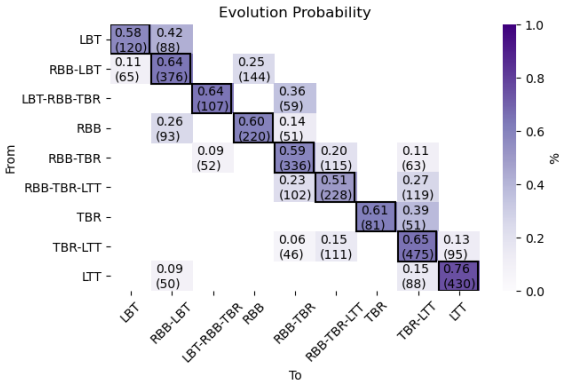


Figure 4.7: Simplex transition table. Top number in cells is the transition probability, bottom number in parantheses in cell is the number of times the transition occurred.

Previous studies, including LH90 and (Price & Ruessink, 2011b; Ranasinghe et al., 2004) used mean residence time as a measure of stability. Mean residence time was measured the average number of consecutive days that a morphology remained in a state. In this study, the probability of self-transitioning, not mean residence time, is a more appropriate measure of stability due to the approach; slight changes in the alongshore, which happens on shorter time scales than full state transitions, might lead to an observation of a different simplex. Therefore, relative stabilities between simplex shapes, as opposed to direct comparisons of mean residence time, are compared to previous literature in the following.

The simplices were more likely to remain stable than evolve to another, as 66% of the total evolutions were to the same simplex (self-evolutions). Overall, bi-modal simplices were equally stable as their uni-modal counterparts, and the most ephemeral simplex was the tri-modal simplex RBB-TBR-LTT (50% likelihood for self evolution). Price and Ruessink (2011b); Ranasinghe et al. (2004) and LH90 found that simplex stability generally decreased from LTT to LBT. The results herein are in agreement with respect to the two end-members; the uni-modal LTT simplex had the highest stability (76%) and LBT had lower stability (58%).

However, the other simplices in the LBT family had relatively higher stabilities (average of 62%) than the simplices in the TBR family (average of 58%). The low stability of the TBR family was mainly due to the RBB-TBR-LTT, which had the lowest stability at 51%.

Both down and up simplex evolution occurred mainly through adjacent or "one over" simplices. Down simplex evolution mostly occurred through bi-modal simplices with the exception of RBB-TBR-LTT, and generally followed the accepted accretive behavior introduced by WS84. Starting from the highest energy configuration, the configuration LBT was likely to increase in rhythmicity to evolve into a bi-modal simplex RBB-LBT before evolving to RBB. Thereafter, down simplex movement was most likely to occur through bi- and tri-modal states (RBB-TBR to RBB-TBR-LTT to TBR-LTT) to reach the lowest energy configuration, LTT. Up simplex evolution mainly (89%) occurred between adjacent simplices or skipped one simplex during "slow" up-simplex evolution, in contrast with the generally accepted "fast" reset events as in LH90 and in (Ranasinghe et al., 2004). However, fast (skipping 7 simplices) up-simplex evolutions (LTT to LBT-RBB) did occur 50 times. Observational studies (Price & Ruessink, 2011b) and laboratory studies (Michallet, Castelle, Barthélemy, Berni, & Bonneton, 2013) have also observed erosional sequences coinciding with an increase in rhythmicity and trough continuity, and the physical dynamics of the process has been described in detail by A. D. Short (1979). During slow up-simplex transitions, the lowest energy configuration, LTT, increased in rhythmicity to TBR-LTT. The TBR-LTT simplex was most likely to evolve to the tri-modal to RBB-TBR-LTT, further increase in rhythmicity to RBB-TBR, and then move off-shore to LBT-RBB-TBR. The LBT-RBB-TBR state did transition to a fully offshore position, LBT-RBB 38 times, but it is not reported in Figure 4.7 since it occurred slightly less than 1% of the total transitions (45 occurrences).

4.4.4 Temporal Variability of Mean Simplex Value

The 28-year record of daily mean simplex values was dominated by an annual signal. Figure 4.8 shows the power spectrum of the 28-year record of daily mean simplex values $\overline{S}$. The time series was first interpolated to a daily time series (N=9531) and detrended with a linear fit. The spectrum was calculated using Welch's method, using 2555 point segments with 638 overlapping points using a Hann window. The most energetic peak that rose above the 95% confidence interval was the annual peak, and the second harmonic of the annual peak at two cycles per year was also found. Consistent with Pianca et al. (2015), spectra of the wave forcing parameters Hs, Tp and MWD also had annual peaks (not shown), suggesting a forced annual oscillation of beach state.

One interannual at peak at 4.2 years was also found. While not significant at the 95% confidence level in this data set, it is notable that a 4.2 year oscillation was also observed in shoreline observations at this field site by Pianca et al. (2015); hence it is plausible that a related 4.2-year process is being weakly reflected in beach states. Inter-annual peaks were not evident in the wave forcing (Hs, Tp and MWD).

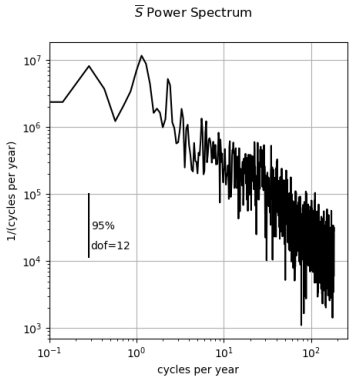


Figure 4.8: Power spectrum of mean simplex value time series.

4.4.5 Annual Variability of Simplex Shapes

The time series exhibited an annual seasonal cycle between the families as the simplices move onshore in the summer to the LTT family through the TBR family and offshore in the winter to the LBT family also through the TBR family. Figure 4.9, top panel, shows the full simplex, mean simplex value and entropy time series averaged over each unique day of year for the 28-year dataset. Simplex entropy was highest for the fall months during the offshore transition.

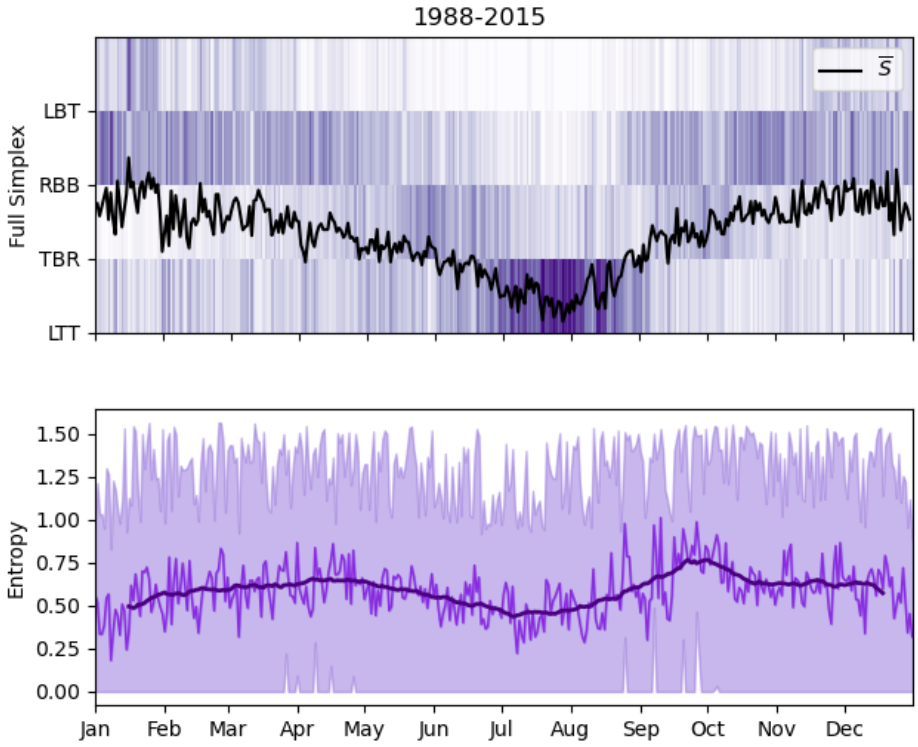


Figure 4.9: (Upper panel) Mean simplex values (black line), full simplex, averaged by day for the 28 year dataset. (Lower panel) daily-averaged entropy, where bold line is 30-day rolling average and the shaded areas represent a 90% confidence interval of data from the 28 years.

During the winter months (December, January, February), the simplices were

mostly offshore in the LBT family (44% of winter simplices were from the LBT family, 31% within the TBR family and 25% in the LTT family), and resulted in highest seasonal mean simplex values ($\overline{S} = 2.7$ for winter), albeit somewhat lower than the average $\overline{S}$ value for the LBT family (3.4). Specifically, LBT-RBB and LBT simplices comprised 30% and 10% of the winter dataset, respectively. Low wave days lead the CNN to detect the LTT simplex. As a result, the winter months, the mean simplex jumps between 4 or 5 (RBB or LBT) and 0 (LTT), resulting in a monthly average mean simplex value for the time series of $\sim$3. During the summer months (June, July, August), the simplices were the most onshore of the entire year in the LTT and TBR families (summer simplices were comprised of 4% LBT family, 36% TBR family and 62% LTT family), with the lowest average mean simplex value of the year ($\overline{S} = 1.7$). Specifically, TBR-LTT and LTT simplices comprised 33% and 29% of the summer dataset, respectively.

During the transition seasons (spring and fall), the simplices evolved mainly through the TBR family, and the distribution of simplices was skewed towards LTT in spring (spring simplices 29% LTT and fall simplices 14% LTT) and slightly towards LBT in fall (spring simplices 24% LBT and fall simplices 28% LBT). As a result, the spring months had a lower average mean simplex value than in fall ($\overline{S} = 2.5$ for fall versus $\overline{S} = 2.3$ in spring). For spring, bi-modal simplices from each family made up the highest proportion of the dataset: LBT-RBB (14%); RBB-TBR (19%); and TBR-LTT (17%). For fall, a bi-modal simplex from the LBT, one bi-modal and one tri-modal from the TBR family made up the highest proportion of the dataset: LBT-RBB (15%); RBB-TBR (19%); RBB-TBR-LTT (14%).

The entropy measurement had a slight annual signal, although exhibited high variability through the 28 years, as shown in the bottom panel of Figure 4.9. During the winter months, when $\overline{S}$ was at its highest values, entropy was lower than average (28-year winter average $E = 0.56$ versus an overall average of 0.58). During spring, entropy rose slowly as $\overline{S}$ fell, reaching the highest value of spring (28-year monthly average $E = 0.66$) in April. As $\overline{S}$ continued to decrease after

April and into the summer months, E decreased as well, reaching the lowest values in the summer months (28-year summer average $E = 0.51$ for summer). As $\overline{S}$ increased in the fall months, E increased the most quickly to the highest values of the year in September (28-year monthly average $E = 0.72$).

Overall, entropy reached higher values in spring and fall months and lower values in winter and summer (28-year winter average $E = 0.67, 0.61$ for fall and spring, respectively, and $E = 0.51, 0.56$ for summer and winter). The elevated entropy values for fall and spring were due to the occurrence of tri-modal simplices (35% and 28% of all tri-modal configurations occurred in fall and spring, respectively, compared with 20% and 16% in winter and summer, respectively). During the fall, the further off-shore tri-modal simplex, LBT-RBB-TBR, was more likely to occur than in spring (38% for fall versus 30% in spring). The forces driving entropy are not clear (no high correlations were found with wave forcing parameters), but it is hypothesized that a combination of factors interact to contribute to higher entropy. Specifically, the development of high entropy simplices in fall months is likely due to an alongshore variable migration of the bar off-shore (off-shore migration through up-simplex movements as described in Section 4.4.3) due to higher background wave heights, longer incident wave periods, and a greater likelihood of high wave events.

4.4.6 Environmental Conditions associated with Simplices

The incident environmental conditions (Ω, MWD, Hs, Tp) corresponding with the simplex families overlapped but were distinct in their mean values (p=0.00 between all populations of parameters Ω, MWD, Hs associated with families). Figure 4.10 shows the histograms of environmental parameters for each family. The mean values of Hs, MWD and Ω were more distinct between the families (average p value between the families was 0.00) than the mean Tp values (average p value between the populations was 0.04). Differences in Ω populations between the families were largely due to differences in Hs populations between the families rather than differences in Tp populations between the families, except for a region

of the highest Ω values for LBT family simplices (driven by smaller Tp values).

Between the three families, the simplices with mostly offshore states (LBT family) corresponded with the highest wave heights (average $Hs = 1.15 + / - 0.15$m) and dimensionless fall velocities (average $\Omega = 5.21 + / - 1.86$), and the most northern to shore-normal wave directions (average $MWD = 4.88 + / - 16.62°$). The simplices with mostly intermediate states (TBR family) corresponded with intermediate wave heights (average $Hs = 0.84 + / - 0.38m$) and dimensionless fall velocities (average $\Omega = 3.93 + / - 1.58$), and shore-normal to southern wave directions (average $MWD = -2.87 + / - 16.40°$). The simplices with mostly onshore states (LTT family) corresponded with the lowest wave heights (average $Hs = 0.62 + / - 0.37$m) and dimensionless fall velocities (average $\Omega = 3.17 + / - 1.65$), and southern wave directions (average $MWD = -10.95 + / - 17°$).

The 28-year daily averaged $\overline{S}$ was correlated with the 28-year daily averaged forcing parameters Hs, MWD and Ω, while entropy was not. Figure 4.11 shows scatterplots between the 28 year daily averaged $\overline{S}, E$ and 28 year daily averaged environmental parameters. Mean simplex value correlated highest with mean wave direction (R=0.85), equally with Hs and Ω (R=0.77) and lowest with Tp (R=0.04). The relationship between $\overline{S}$ and Ω is generally in agreement with WS84, who showed that Ω ranged between 2-6 for intermediate states. Values for Ω overlapped for each state, with the mean Ω values progressively increasing (LTT to TBR to RBB). However, in contrast with WS84, the highest Ω values are for slightly lower $\overline{S}$ values (i.e., not the most offshore simplices). The LBT-RBB-TBR simplex (average $\overline{S} = 3$) was associated with highest Hs and Tp values, resulting in higher Ω values overall. Entropy correlated slightly with Hs and Ω (R=0.44, 0.38, respectively), and had low correlation ($R < 0.4$) with MWD, Tp and Ω.

The correspondence between Ω and beach configuration is well established in the literature (Lippmann & Holman, 1990; Ranasinghe et al., 2004; L. D. Wright & Short, 1984), however some studies have found Ω does not alone explain beach state (Jiménez, Guillén, & Falqués, 2008; Masselink & Pattiaratchi, 2001). Mean wave direction has been observed to cause changes in sandbar rhythmicity, with oblique

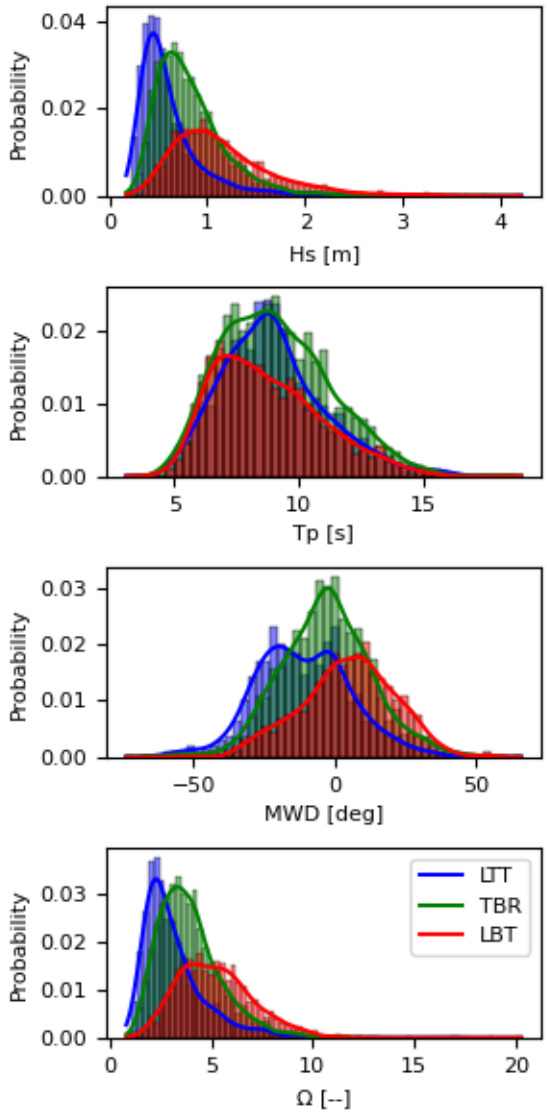


Figure 4.10: Histogram and kernel density estimates of Hs, Tp, MWD and Ω with respect to each simplex family.

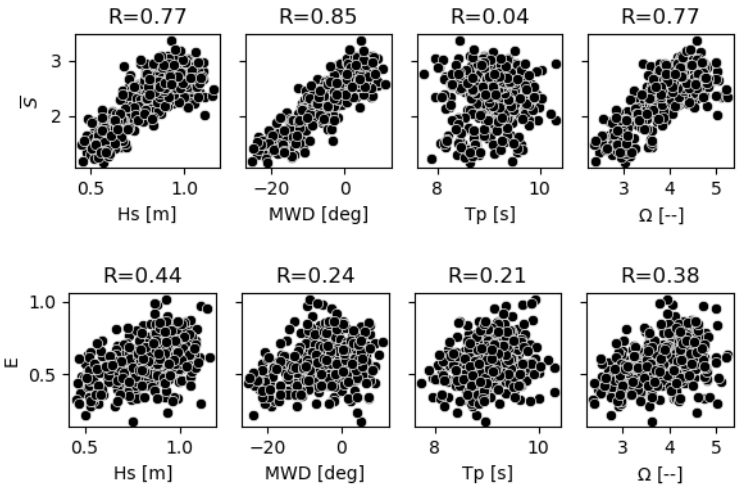


Figure 4.11: Twenty-eight year daily averaged $\overline{S}$ and E values with respect to forcing parameters Hs, MWD, Tp and Ω. Number of data points is 365.

waves causing sandbar straightening Contardo and Symonds (2015) and normally incident waves causing rip currents, or rhythmicity in the bar line Thornton et al. (2007). Within the LBT and LTT families, the simplices with the most linear states (LBT, LTT) were on average associated with the most oblique average mean wave directions of all the simplices ($MWD = 4.18°$ for LBT and $MWD = -12.61°$ for LTT). Within the TBR family, the simplices with the most rhythmic states (RBB-TBR and TBR) were on average associated with the most shore normal mean wave direction ($MWD \sim 2°$) of the simplex family. Price and Ruessink (2011b) also found that incident wave angle corresponded to beach state, however, found that oblique wave angles did not necessarily correspond with more linear bar configurations, but rather more upstate bar configurations.

4.5 Conclusions

This study derives a beach probability simplex to represent surf zone morphologies that exhibit alongshore variability of dominant beach state. The beach probability

simplex is a probability mass function where each entry corresponds to relative contribution (quantified as total length within the alongshore) of one of the beach states defined by L. D. Wright and Short (1984). Additionally, the beach probability simplex was detected with a low level of subjectivity and in an automated way, where a convolutional neural network (CNN) quantified beach probability simplices from morphology inferred from remotely sensed wave breaking patterns (Argus daytimexposure images). Useful properties of the simplex included: a categorical representation of the simplex (the state names of the non-zero values of the simplex); a mean simplex value ($\overline{S}$, a scalar representation of the simplex); and simplex entropy (E) (Shannon's entropy that quantified the diversity of states within a beach probability simplex).

A 28-year dataset of daily beach simplices from Duck, NC was presented. The majority (79%) of the simplices were comprised of states mutually adjacent to each other in the ordering of states defined by L. D. Wright and Short (1984). Of the adjacent states, the majority (67%) encompassed two or three states. The simplices were grouped into families (Longshore Bar Trough, LBT, family, Transverse Bar Rip, TBR, family and Low Tide Terrace, LTT family), mainly following Lippmann and Holman (1990), where the families were differentiated from each other based on morphological factors: degree of shoreline proximity; shoreline attachment; and linearity versus rhythmicity of the barline.

A spectral analysis of the 28-years of daily mean beach state ($\overline{S}$) showed a statistically significant annual signal and a weaker 4.3 year signal which, while not significant at the 95% confidence level, was consistent with past observations of shoreline motion at the field site (Pianca et al., 2015). An annual oscillation between the simplex configurations occurred the simplices transitioned between onshore, summer LTT configurations and offshore, winter LBT configurations, and travelled through TBR in both spring and fall. The simplices in fall and spring showed higher levels of entropy (greater amounts of complexity), due to a greater occurrence of tri-modal simplices. The tri-modal simplices in fall were more offshore than in spring.

On average, the daily significant wave height was highest for simplices within the LBT family, lowest for simplices within the LTT family, and intermediate for simplices within the TBR family. These findings are consistent with LH90, WS84, Ranasinghe et al. (2004) that found beach state is largely driven by wave height, often parameterized by the non-dimensional fall velocity $\Omega = H_s/wT_p$. A strong relationship between beach state and wave angle ($R = 0.85$) was also observed. The average associated mean wave direction was most oblique for simplices dominated by linear states and shore-normal for simplices dominated by rhythmic states. A different long term (9 year) observational study found that incident wave angle corresponded with beach state (Price & Ruessink, 2011b), however, that oblique wave angle caused upstate bar migration. Further research is recommended to investigate the mechanisms by which oblique waves drive beaches to higher-energy states (e.g., (Ranasinghe et al., 2004)), and to confirm the observation using long-term data sets from additional beaches.

Mean simplex value correlated with Hs and MWD ($R = 0.77, 0.85$, respectively), while entropy showed slight correlation with Hs ($R = 0.44$), but no clear relationship with forcing parameters. Mean simplex value conveys information about the proximity of the bar to the shoreline, which both LH90 and WS84 have shown to be related to incident wave conditions. In contrast, entropy conveys information about alongshore variability in beach state, which did not seem to have a direct relationship with any of the measured wave parameters (Hs, MWD, or Tp). The emergence of morphologies with high entropy is more likely due to a combination of factors (e.g., alongshore variability of wave breaking patterns, the emergence and development of rip currents, storm sequencing) which should be explored in future work.

Chapter 5: Conclusions

The dissertation has demonstrated how pattern recognition technology and large volumes of data can assist in knowledge discovery within marine science. Machine learning algorithms were applied in three applications. In the applications, the ML algorithm was applied with transparency, the limits of its transferability were determined, and knowledge about the marine system or model was gained.

5.1 Chapter Conclusions

In the second chapter, a decision tree was applied to correct the output of hourly 24-hour time horizon predictions of significant wave height made by a numerical wave model by detecting areas of wave model bias. The application of the decision tree corrections improved numerical model skill more for winter than for summer. Upon inspection of the decision tree decision making, consistent model over-estimations for data points associated with mid wave periods (6-10s) and model under-estimations for high wave heights ($Hs > 5.4m$) associated with high wind speeds (wind speed $> 15m/s$) were found. Geospatially, the decision tree algorithm improved model skill at a location that was witheld from training, after being trained on neighboring buoys. Temporally, the decision tree algorithm did not improve model skill when applied to make predictions for a different year.

In the third chapter, an image processing ML algorithm, a Convolutional Neural Network (CNN) was applied to detect beach state from remotely sensed wave breaking patterns in Argus imagery for Narrabeen, Sydney, Australia and Duck, North Carolina, USA. Three CNNs were trained, two with data from a single location, and one with data from both locations. The CNNs accurately predicted beach state at the location where it was trained. Transferability skill varied, depending on where the CNN was trained; the Narrabeen-trained CNN made more accu-

rate predictions of beach state when applied to Duck than when the Duck-trained CNN was applied to Narrabeen. Additionally, the study used Gradient-Class Activation Maps to confirm that the CNN was detecting morphological characteristics appropriate to each beach state.

The fourth chapter builds off the third by applying the CNN to a long (28 year) record of Argus imagery, but first expands the beach state framework from discrete to continuous by developing a beach probability simplex. The beach probability simplex is a probability mass function that quantifies the relative degree of alongshore variability of beach state. The majority (67%) of the resulting beach probability simplices encompassed at least two or more beach states, suggesting that a beach probability simplex is more appropriate in describing nearshore morphology than discrete beach states. A spectral analysis of the time series showed that the beach morphology had the highest peak at an annual cycle, and a less energetic at a four year cycle that did not rise above the 95% confidence interval. The simplices with the greatest alongshore variability occurred in the fall time.

Transferability skill depended on the relationship between the training data used in ML development and the data to which the ML model was transferred. In the wave model application, the decision tree's geospatial transferability skill was higher than the temporal transferability skill. In the geospatial transfer experiment, relationships between the input and desired output in the training data from neighboring buoys held when the decision tree made predictions at a buoy omitted from training. However, when transferred temporally, the relationships learned during training did not hold. In the beach state application, the degree to which a CNN could detect beach state at a different location depended on where the CNN was trained. The reasons for the different transferability skill were unclear. Transferability skill increased as images from the alternative site were included in the training dataset, demonstrating that the CNN became more generalizable as a variety of training data is presented, which is generally accepted within the ML community. As a guideline, transferability skill of an ML technique must be tested for each application, and a variety of training data will increase the generalizability

of an ML technique.

5.2 General Conclusions and Future Work

The ML applications simultaneously improved the marine models and increased the understanding of each system in a knowledge discovery process. In the first application (chapter two) the ML algorithm improved the numerical wave model skill. Additionally, the ML algorithm identified consistent wave model bias with respect to model phase space, pointing to areas of potential model improvement with respect to modeled environmental conditions. In the second application (chapters three and four) the ML algorithm increased the accuracy of the beach state conceptual model by detecting alongshore variability of beach state, progressing the beach state representation from discrete to continuous, therefore creating a more flexible framework that captures the continuous nature of morphological evolution and appearance. Application of the flexible framework resulted in higher resolution of morphological evolution, in that the most likely transitions between complex morphological shapes were quantified.

The ML algorithms were applied with transparency, which is important for adoption of ML algorithms, and also a method to discover knowledge about the system. In the first application, the transparency of the decision tree lies in the architecture itself. The decision making of the decision tree can be understood with manual inspection of its architecture, which was performed to find the areas of wave model bias. In the second application, an outside technique, Guided Grad-CAM, was required to determine areas of input images that were relevant for CNN decision making. The Guided Grad-CAMs confirmed that the CNN identified the morphological characteristics appropriate to its associated label prediction.

Knowledge Discovery (KD) can be a powerful outcome when ML is applied to large datasets. However, a limitation arises; while ML techniques can provide information about potential correlations in the data, mathematical or physical models are more useful in providing information about causation since the underlying forcing can be directly observed and controlled. Additionally, domain

knowledge is required to ensuring that the interpretation of ML results is in accordance with reason. Therefore, hybrid approaches are most robust in discovering knowledge about a system. In chapter two, the ML routine post-processes the output of a well established numerical model in a hybrid method. In chapters three and four, the ML routine advances a well established conceptual model, but does not replace it. Finally, as a limitation in chapter four, the causation of system dynamics was not determined. This was because the method was unable to isolate and directly observe the simultaneous contributions of physical forcings and feedbacks within the system, therefore making it difficult to ascertain which variable or combination of variables were responsible for alongshore variable bar evolution. Future research could explore the causes of the dynamics of system evolution by simulating the observed phenomena and querying the underlying processes.

A powerful way that data science techniques can aid in marine science research is to increase the spatial and temporal scales at which observations can be analyzed. Future work could increase the transferability skill of the algorithms, and then apply the algorithms to larger volumes of data for inter-site or temporal comparisons. In the first application, the decision tree could be trained at additional locations using satellites and other buoys, and applied to predict errors on gridded model output. This would increase the accuracy of significant wave height predictions with respect to spatial resolution, and also could reveal if any spatial parameters correspond to model bias (e.g., specific locations see consistent model over or under estimations for certain environmental conditions). In the second application, the CNN could be trained with and applied to imagery from other Argus sites around the world that have different beach characteristics. This would increase both the temporal and spatial variability of beach morphology records, allowing for the comparison of nearshore bathymetry dynamics on a potentially global scale.

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139–160. Retrieved 2021-07-05, from https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1982.tb0119 (_eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1982.tb01195.x) doi: 10.1111/j.2517-6161.1982.tb01195.x

Alexander, P. S., & Holman, R. A. (2004, July). Quantification of nearshore morphology based on video imaging. *Marine Geology*, *208*(1), 101–111. Retrieved 2020-06-12, from https://linkinghub.elsevier.com/retrieve/pii/S0025322704001185 doi: 10.1016/j.margeo.2004.04.017

Appendini, C. M., Torres-Freyermuth, A., Salles, P., López-González, J., & Mendoza, E. T. (2014). Wave climate and trends for the Gulf of Mexico: A 30-yr wave hindcast. *Journal of Climate*, *27*(4), 1619–1632.

Armaroli, C., & Ciavola, P. (2011, March). Dynamics of a nearshore bar system in the northern Adriatic: A video-based morphological classification. *Geomorphology*, *126*(1-2), 201–216. Retrieved 2020-06-18, from https://linkinghub.elsevier.com/retrieve/pii/S0169555X10005076 doi: 10.1016/j.geomorph.2010.11.004

Austin, M. J., Scott, T. M., Russell, P. E., & Masselink, G. (2013, March). Rip Current Prediction: Development, Validation, and Evaluation of an Operational Tool. *Journal of Coastal Research*, *29*(2), 283–300. Retrieved 2020-09-23, from https://meridian.allenpress.com/jcr/article/29/2/283/144696/Rip-Current-Pre (Publisher: Allen Press) doi: 10.2112/JCOASTRES-D-12-00093.1

Balaguer, A., Ruiz, L., Hermosilla, T., & Recio, J. (2010, February). Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. *Computers & Geosciences*, *36*(2), 231–240. Retrieved 2020-05-26, from https://linkinghub.elsevier.com/retrieve/pii/S0098300409002374 doi: 10.1016/j.cageo.2009.05.003

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, *16*(5), 412–424.

Benedet, L., Finkl, C., & Klein, A. (2006). Morphodynamic Classification of Beaches on the Atlantic Coast of Florida: Geographical Variability of Beach Types, Beach Safety and Coastal Hazards. *Journal of Coastal Research*, 360–365. Retrieved 2021-07-04, from https://www.jstor.org/stable/25741596 (Publisher: Coastal Education & Research Foundation, Inc.)

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

Bergen, K. J., Johnson, P. A., Hoop, M. V. d., & Beroza, G. C. (2019, March). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433). Retrieved 2021-07-21, from https://science.sciencemag.org/content/363/6433/eaau0323 (Publisher: American Association for the Advancement of Science Section: Review) doi: 10.1126/science.aau0323

Beuzen, T., Goldstein, E. B., & Splinter, K. D. (2019). Ensemble models from machine learning: an example of wave runup and coastal dune erosion. *Natural Hazards and Earth System Sciences*, *19*(10), 2295–2309.

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

Beuzen, T., Splinter, K. D., Marshall, L. A., Turner, I. L., Harley, M. D., & Palmsten, M. L. (2018b, May). Bayesian Networks in coastal engineering: Distinguishing descriptive and predictive applications. *Coastal Engineering*, *135*, 16–30. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S0378383917303678 doi: 10.1016/j.coastaleng.2018.01.005

Biel, R. G., Hacker, S. D., & Ruggiero, P. (2019). Elucidating Coastal Foredune Ecomorphodynamics in the U.S. Pacific Northwest via Bayesian Networks. *Journal of Geophysical Research: Earth Surface*, *124*(7), 1919–1938. Retrieved 2021-07-21, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JF004758 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JF004758) doi: 10.1029/2018JF004758

Birkemeier, W. A., DeWall, A. E., Gorbics, C. S., & Miller, H. C. (1981, 10). *A*

user's guide to cerc's field research facility. (Tech. Rep.). Coastal Engineering Research Center, For Belvoir VA.

Björkqvist, J.-V., Vähä-Piikkiö, O., Alari, V., Kuznetsova, A., & Tuomi, L. (2020, January). WAM, SWAN and WAVEWATCH III in the Finnish archipelago – the effect of spectral performance on bulk wave parameters. *Journal of Operational Oceanography*, *13*(1), 55–70. Retrieved 2021-07-21, from https://doi.org/10.1080/1755876X.2019.1633236 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/1755876X.2019.1633236) doi: 10.1080/1755876X.2019.1633236

Boehm, A. B. (2003, December). Model of Microbial Transport and Inactivation in the Surf Zone and Application to Field Measurements of Total Coliform in Northern Orange County, California. *Environmental Science & Technology*, *37*(24), 5511–5517. Retrieved 2021-07-04, from https://doi.org/10.1021/es034321x (Publisher: American Chemical Society) doi: 10.1021/es034321x

Boehm, A. B., Keymer, D. P., & Shellenbarger, G. G. (2005, September). An analytical model of enterococci inactivation, grazing, and transport in the surf zone of a marine beach. *Water Research*, *39*(15), 3565–3578. Retrieved 2021-07-04, from https://www.sciencedirect.com/science/article/pii/S0043135405003611 doi: 10.1016/j.watres.2005.06.026

Bohling, G. (2005). Introduction to geostatistics and variogram analysis. *Kansas geological survey*, *1*, 1–20.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression trees*. CRC press.

Breiman, L., & Spector, P. (1992). Submodel selection and evaluation in regression. The X-random case. *International Statistical Review*, 291–319.

Browne, M., Strauss, D., Tomlinson, R., & Blumenstein, M. (2006, November). Objective Beach-State Classification From Optical Sensing of Cross-Shore Dissipation Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(11), 3418–3426. (Conference Name: IEEE Transactions on Geoscience and Remote Sensing) doi: 10.1109/TGRS.2006.877758

Bulteau, T., Baills, A., Petitjean, L., Garcin, M., Palanisamy, H., & Le Cozannet, G. (2015, January). Gaining insight into regional coastal changes on La Réunion island through a Bayesian data mining approach. *Geomorphology*, *228*, 134–146. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S0169555X14004607

doi: 10.1016/j.geomorph.2014.09.002

Buscombe, D. (2020). SediNet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry. *Earth Surface Processes and Landforms*, *45*(3), 638–651. Retrieved 2020-06-29, from https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.4760 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/esp.4760) doi: 10.1002/esp.4760

Buscombe, D., & Carini, R. J. (2019, January). A Data-Driven Approach to Classifying Wave Breaking in Infrared Imagery. *Remote Sensing*, *11*(7), 859. Retrieved 2020-03-09, from https://www.mdpi.com/2072-4292/11/7/859 (Number: 7 Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/rs11070859

Buscombe, D., Carini, R. J., Harrison, S. R., Chickadel, C. C., & Warrick, J. A. (2020, January). Optical wave gauging using deep neural networks. *Coastal Engineering*, *155*, 103593. Retrieved 2020-03-09, from http://www.sciencedirect.com/science/article/pii/S0378383919301243 doi: 10.1016/j.coastaleng.2019.103593

Castelle, B., Bonneton, P., Dupuis, H., & Sénéchal, N. (2007, November). Double bar beach dynamics on the high-energy meso-macrotidal French Aquitanian Coast: A review. *Marine Geology*, *245*(1), 141–159. Retrieved 2020-06-25, from http://www.sciencedirect.com/science/article/pii/S002532270700148X doi: 10.1016/j.margeo.2007.06.001

Castelle, B., Marieu, V., Bujan, S., Splinter, K. D., Robinet, A., Sénéchal, N., & Ferreira, S. (2015, June). Impact of the winter 2013–2014 series of severe Western Europe storms on a double-barred sandy coast: Beach and dune erosion and megacusp embayments. *Geomorphology*, *238*, 135–148. Retrieved 2020-09-23, from http://www.sciencedirect.com/science/article/pii/S0169555X15001385 doi: 10.1016/j.geomorph.2015.03.006

Castelle, B., Scott, T., Brander, R. W., & McCarroll, R. J. (2016, December). Rip current types, circulation and hazard. *Earth-Science Reviews*, *163*, 1–21. Retrieved 2020-07-01, from http://www.sciencedirect.com/science/article/pii/S0012825216303117 doi: 10.1016/j.earscirev.2016.09.008

Chappell, J., & Eliot, I. G. (1979, July). Surf-beach dynamics in time and space — An Australian case study, and elements of a predictive model. *Marine Geology*, *32*(3), 231–250. Retrieved 2021-07-05, from

https://www.sciencedirect.com/science/article/pii/0025322779900665 doi: 10.1016/0025-3227(79)90066-5

Chen, L., Wang, S., Fan, W., Sun, J., & Naoi, S. (2015). Beyond human recognition: A cnn-based framework for handwritten character recognition. In *2015 3rd iapr asian conference on pattern recognition (acpr)* (pp. 695–699).

Chen, S.-T. (2019, February). Probabilistic forecasting of coastal wave height during typhoon warning period using machine learning methods. *Journal of Hydroinformatics*, *21*(2), 343–358. Retrieved 2021-07-21, from https://doi.org/10.2166/hydro.2019.115 doi: 10.2166/hydro.2019.115

Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (1998). Data mining and knowledge discovery. In *Data mining methods for knowledge discovery* (pp. 1–26). Springer.

Contardo, S., & Symonds, G. (2015, October). Sandbar straightening under wind-sea and swell forcing. *Marine Geology*, *368*, 25–41. Retrieved 2020-06-19, from http://www.sciencedirect.com/science/article/pii/S0025322715300025 doi: 10.1016/j.margeo.2015.06.010

Cornejo-Bueno, L., Rodríguez-Mier, P., Mucientes, M., Nieto-Borge, J., & Salcedo-Sanz, S. (2018). Significant wave height and energy flux estimation with a genetic fuzzy system for regression. *Ocean Engineering*, *160*, 33–44.

Davidson, M. A., Splinter, K. D., & Turner, I. L. (2013, March). A simple equilibrium model for predicting shoreline change. *Coastal Engineering*, *73*, 191–202. Retrieved 2020-12-07, from http://www.sciencedirect.com/science/article/pii/S0378383912001676 doi: 10.1016/j.coastaleng.2012.11.002

den Bieman, J. P., de Ridder, M. P., & van Gent, M. R. A. (2020, June). Deep learning video analysis as measurement technique in physical models. *Coastal Engineering*, *158*, 103689. Retrieved 2020-06-25, from http://www.sciencedirect.com/science/article/pii/S037838391930612X doi: 10.1016/j.coastaleng.2020.103689

den Bieman, J. P., van Gent, M. R. A., & Hoonhout, B. M. (2019, December). Physical model of scour at the toe of rock armoured structures. *Coastal Engineering*, *154*, 103572. Retrieved 2020-06-29, from http://www.sciencedirect.com/science/article/pii/S0378383919300651 doi: 10.1016/j.coastaleng.2019.103572

De Santiago, I., Morichon, D., Abadie, S., Castelle, B., Liria, P., & Epelde, I. (2013, January). Video monitoring nearshore sand-

bar morphodynamics on a partially engineered embayed beach. *Journal of Coastal Research*, *65*, 458–463. Retrieved 2020-06-25, from http://www.bioone.org/doi/10.2112/SI65-078.1 doi: 10.2112/SI65-078.1

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

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, *40*(2), 139–157.

Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable ai really mean? a new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*.

Dubarbier, B., Castelle, B., Ruessink, G., & Marieu, V. (2017). Mechanisms controlling the complete accretionary beach state sequence. *Geophysical Research Letters*, *44*(11), 5645–5654. Retrieved 2020-06-25, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073094 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL073094) doi: 10.1002/2017GL073094

Eadi Stringari, C. (2020). *Data driven Investigations of Broken Wave Behaviour in the Surf and Swash Zones* (Unpublished doctoral dissertation).

Ellenson, A., & Özkan-Haller, H. T. (2018). Predicting large ocean wave events characterized by bimodal energy spectra in the presence of a low-level southerly wind feature. *Weather and Forecasting*, *33*(2), 479–499.

Ellenson, A., Pei, Y., Wilson, G., Özkan-Haller, H. T., & Fern, X. (2020). An application of a machine learning algorithm to determine and describe error patterns within wave model output. *Coastal Engineering*, *157*, 103595.

Ellenson, A. N., Simmons, J. A., Wilson, G. W., Hesser, T. J., & Splinter, K. D. (2020, January). Beach State Recognition Using Argus Imagery and Convolutional Neural Networks. *Remote Sensing*, *12*(23), 3953. Retrieved 2021-01-03, from https://www.mdpi.com/2072-4292/12/23/3953 (Number: 23 Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/rs12233953

Etemad-Shahidi, A., & Bonakdar, L. (2009). Design of rubble-mound breakwaters using M5 machine learning method. *Applied Ocean Research*, *31*(3), 197–201.

Etemad-Shahidi, A., & Mahjoobi, J. (2009). Comparison between M5 model tree and neural networks for prediction of significant wave height in Lake

Superior. *Ocean Engineering*, *36*(15-16), 1175–1181.

García-Medina, G., Özkan-Haller, H. T., & Ruggiero, P. (2014). Wave resource assessment in Oregon and southwest Washington, USA. *Renewable Energy*, *64*, 203–214.

García-Medina, G., Özkan-Haller, H. T., Ruggiero, P., & Oskamp, J. (2013). An inner-shelf wave forecasting system for the US Pacific Northwest. *Weather and Forecasting*, *28*(3), 681–703.

Garnier, R., Calvete, D., Falqués, A., & Dodd, N. (2008). Modelling the formation and the long-term behavior of rip channel systems from the deformation of a longshore bar. *Journal of Geophysical Research: Oceans*, *113*(C7). Retrieved 2020-06-25, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JC004632 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JC004632) doi: 10.1029/2007JC004632

Ghosh, S., Das, N., & Nasipuri, M. (2019, September). Reshaping inputs for convolutional neural network: Some common and uncommon methods. *Pattern Recognition*, *93*, 79–94. Retrieved 2020-04-17, from http://www.sciencedirect.com/science/article/pii/S0031320319301505 doi: 10.1016/j.patcog.2019.04.009

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., ... Holzinger, A. (2018). Explainable ai: the new 42? In *International cross-domain conference for machine learning and knowledge extraction* (pp. 295–303).

Goldstein, E. B., Coco, G., & Plant, N. G. (2019, July). A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Reviews*, *194*, 97–108. Retrieved 2021-07-07, from https://www.sciencedirect.com/science/article/pii/S001282521830391X doi: 10.1016/j.earscirev.2019.04.022

Gomes, E. R., Mulligan, R. P., Brodie, K. L., & McNinch, J. E. (2016, October). Bathymetric control on the spatial distribution of wave breaking in the surf zone of a natural beach. *Coastal Engineering*, *116*, 180–194. Retrieved 2021-07-05, from https://www.sciencedirect.com/science/article/pii/S0378383916301235 doi: 10.1016/j.coastaleng.2016.06.012

Gomes da Silva, P., Coco, G., Garnier, R., & Klein, A. H. F. (2020, May). On the prediction of runup, setup and swash on beaches. *Earth-Science Reviews*, *204*, 103148. Retrieved 2021-07-04, from https://www.sciencedirect.com/science/article/pii/S0012825219306828

doi: 10.1016/j.earscirev.2020.103148

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (http://www.deeplearningbook.org)

Grant, S. B., Kim, J. H., Jones, B. H., Jenkins, S. A., Wasyl, J., & Cudaback, C. (2005). Surf zone entrainment, along-shore transport, and human health implications of pollution from tidal outlets. *Journal of Geophysical Research: Oceans*, *110*(C10). Retrieved 2020-09-23, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JC002401 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004JC002401) doi: 10.1029/2004JC002401

Greenwood, B., & Davidson-Arnott, R. G. D. (1979, February). Sedimentation and equilibrium in wave-formed bars: a review and case study. *Canadian Journal of Earth Sciences*, *16*(2), 312–332. Retrieved 2021-07-05, from https://cdnsciencepub.com/doi/abs/10.1139/e79-030 (Publisher: NRC Research Press) doi: 10.1139/e79-030

Guedes Soares, C., Rusu, L., Bernardino, M., & Pilar, P. (2011). An operational wave forecasting system for the Portuguese continental coastal area. *Journal of Operational Oceanography*, *4*(2), 17–27.

Gutierrez, B. T., Plant, N. G., Thieler, E. R., & Turecek, A. (2015). Using a Bayesian network to predict barrier island geomorphologic characteristics. *Journal of Geophysical Research: Earth Surface*, *120*(12), 2452–2475.

Hadadpour, S., Moshfeghi, H., Jabbari, E., & Kamranzad, B. (2013). Wave hindcasting in Anzali, Caspian Sea: A hybrid approach. *Journal of Coastal Research*, *65*(sp1), 237–243.

Hanson, J. L., Tracy, B. A., Tolman, H. L., & Scott, R. D. (2009). Pacific hindcast performance of three numerical wave models. *Journal of Atmospheric and Oceanic Technology*, *26*(8), 1614–1633.

Harley, M. D., Turner, I. L., Short, A. D., & Ranasinghe, R. (2011). A reevaluation of coastal embayment rotation: The dominance of cross-shore versus alongshore sediment transport processes, Collaroy-Narrabeen Beach, southeast Australia. *Journal of Geophysical Research: Earth Surface*, *116*(F4). Retrieved 2020-06-12, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JF001989 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011JF001989) doi: 10.1029/2011JF001989

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*. Retrieved 2020-04-17, from http://arxiv.org/abs/1512.03385 (arXiv: 1512.03385)

Helderop, E., & Grubesic, T. H. (2019). Social, geomorphic, and climatic factors driving us coastal city vulnerability to storm surge flooding. *Ocean & Coastal Management*, *181*, 104902.

Hey, A. J., Tansley, S., Tolle, K. M., et al. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Microsoft research Redmond, WA.

Holland, K., Holman, R., Lippmann, T., Stanley, J., & Plant, N. (1997, January). Practical use of video imagery in nearshore oceanographic field studies. *IEEE Journal of Oceanic Engineering*, *22*(1), 81–92. (Conference Name: IEEE Journal of Oceanic Engineering) doi: 10.1109/48.557542

Holman, R., & Haller, M. C. (2013, January). Remote Sensing of the Nearshore. *Annual Review of Marine Science*, *5*(1), 95–113. Retrieved 2020-06-25, from https://www.annualreviews.org/doi/10.1146/annurev-marine-121211-172408 (Publisher: Annual Reviews) doi: 10.1146/annurev-marine-121211-172408

Holman, R. A., & Stanley, J. (2007). The history and technical capabilities of argus. *Coastal engineering*, *54*(6-7), 477–491.

Holman, R. A., Symonds, G., Thornton, E. B., & Ranasinghe, R. (2006). Rip spacing and persistence on an embayed beach. *Journal of Geophysical Research: Oceans*, *111*(C1). Retrieved 2020-09-23, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JC002965 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JC002965) doi: 10.1029/2005JC002965

Holzinger, A. (2018). From machine learning to explainable ai. In *2018 world symposium on digital intelligence for systems and machines (disa)* (pp. 55–66).

Hoonhout, B. M., Radermacher, M., Baart, F., & van der Maaten, L. J. P. (2015, November). An automated method for semantic classification of regions in coastal images. *Coastal Engineering*, *105*, 1–12. Retrieved 2020-06-24, from http://www.sciencedirect.com/science/article/pii/S0378383915001313 doi: 10.1016/j.coastaleng.2015.07.010

Horrillo-Caraballo, J., & Reeve, D. (2008). An investigation of the link between beach morphology and wave climate at duck, nc, usa. *Journal of Flood Risk Management*, *1*(2), 110–122.

Hughes, M. G., Aagaard, T., Baldock, T. E., & Power, H. E. (2014, September). Spectral signatures for swash on reflective, intermediate and dissipative beaches. *Marine Geology*, *355*, 88–97. Retrieved 2021-07-05, from https://www.sciencedirect.com/science/article/pii/S0025322714001492 doi: 10.1016/j.margeo.2014.05.015

Inman, D. L., & Brush, B. M. (1973). The coastal challenge. *Science*, *181*(4094),

20–32.

Ioffe, S., & Szegedy, C. (2015, March). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]*. Retrieved 2020-10-01, from http://arxiv.org/abs/1502.03167 (arXiv: 1502.03167)

Jain, P., Deo, M., Latha, G., & Rajendran, V. (2011). Real time wave forecasting using wind time history and numerical model. *Ocean Modelling*, *36*(1-2), 26–39.

James, S. C., Zhang, Y., & O'Donncha, F. (2018, July). A machine learning framework to forecast wave conditions. *Coastal Engineering*, *137*, 1–10. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S0378383917304969 doi: 10.1016/j.coastaleng.2018.03.004

Jiménez, J. A., Guillén, J., & Falqués, A. (2008, September). Comment on the article "Morphodynamic classification of sandy beaches in low energetic marine environment" by Gómez-Pujol, L., Orfila, A., Cañellas, B., Alvarez-Ellacuria, A., Méndez, F.J., Medina, R. and Tintoré, J. Marine Geology, 242, pp. 235–246, 2007. *Marine Geology*, *255*(1), 96–101. Retrieved 2021-06-30, from https://www.sciencedirect.com/science/article/pii/S0025322708001151 doi: 10.1016/j.margeo.2008.04.002

Kalra, R., & Deo, M. (2007). Genetic programming for retrieving missing information in wave records along the west coast of India. *Applied Ocean Research*, *29*(3), 99–111.

Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. In *Proceedings of the 34th international conference on machine learning-volume 70* (pp. 1885–1894).

Komar, P. D. (1998). *Beach processes and sedimentation*.

Kuik, A., Van Vledder, G. P., & Holthuijsen, L. (1988). A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography*, *18*(7), 1020–1034.

Kyprioti, A. P., Taflanidis, A. A., Nadal-Caraballo, N. C., & Campbell, M. O. (2021, January). Incorporation of sea level rise in storm surge surrogate modeling. *Natural Hazards*, *105*(1), 531–563. Retrieved 2021-07-21, from https://doi.org/10.1007/s11069-020-04322-z doi: 10.1007/s11069-020-04322-z

Leatherman, S. P. (2018). Coastal erosion and the united states national flood insurance program. *Ocean & Coastal Management*, *156*, 35–42.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553),

436–444.

Lippmann, T. C., & Holman, R. A. (1989). Quantification of sand bar morphology: A video technique based on wave dissipation. *Journal of Geophysical Research: Oceans*, *94*(C1), 995–1011. Retrieved 2020-06-26, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC094iC01p00995 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JC094iC01p00995) doi: 10.1029/JC094iC01p00995

Lippmann, T. C., & Holman, R. A. (1990). The spatial and temporal variability of sand bar morphology. *Journal of Geophysical Research: Oceans*, *95*(C7), 11575–11590. Retrieved 2020-01-24, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC095iC07p11575 doi: 10.1029/JC095iC07p11575

Loureiro, C., Ferreira, , & Cooper, J. A. G. (2013, December). Applicability of parametric beach morphodynamic state classification on embayed beaches. *Marine Geology*, *346*, 153–164. Retrieved 2020-06-22, from http://www.sciencedirect.com/science/article/pii/S0025322713001989 doi: 10.1016/j.margeo.2013.09.005

MacMahan, J. H. (n.d.). Field observations of rip currents. , 236.

Madsen, A. J., & Plant, N. G. (2001, March). Intertidal beach slope predictions compared to field data. *Marine Geology*, *173*(1), 121–139. Retrieved 2020-09-02, from http://www.sciencedirect.com/science/article/pii/S0025322700001687 doi: 10.1016/S0025-3227(00)00168-7

Mahjoobi, J., & Etemad-Shahidi, A. (2008). An alternative approach for the prediction of significant wave heights based on classification and regression trees. *Applied Ocean Research*, *30*(3), 172–177.

Malde, K., Handegard, N. O., Eikvil, L., & Salberg, A.-B. (2020). Machine intelligence and the data-driven future of marine science. *ICES Journal of Marine Science*, *77*(4), 1274–1285.

Malekmohamadi, I., Bazargan-Lari, M. R., Kerachian, R., Nikoo, M. R., & Fallahnia, M. (2011). Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. *Ocean Engineering*, *38*(2-3), 487–497.

Masselink, G., & Pattiaratchi, C. B. (2001, February). Seasonal changes in beach morphology along the sheltered coastline of Perth, Western Australia. *Marine Geology*, *172*(3), 243–263. Retrieved 2021-07-01, from https://www.sciencedirect.com/science/article/pii/S0025322700001286 doi: 10.1016/S0025-3227(00)00128-6

Masselink, G., & Short, A. D. (1993). The Effect of Tide Range on Beach

Morphodynamics and Morphology: A Conceptual Beach Model. *Journal of Coastal Research*, *9*(3), 785–800. Retrieved 2020-06-22, from https://www.jstor.org/stable/4298129 (Publisher: Coastal Education & Research Foundation, Inc.)

Michallet, H., Castelle, B., Barthélemy, E., Berni, C., & Bonneton, P. (2013). Physical modeling of three-dimensional intermediate beach morphodynamics. *Journal of Geophysical Research: Earth Surface*, *118*(2), 1045–1059. Retrieved 2021-07-03, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrf.20078 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/jgrf.20078) doi: 10.1002/jgrf.20078

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

Moeini, M. H., Etemad-Shahidi, A., Chegini, V., Rahmani, I., & Moghaddam, M. (2014). Error distribution and correction of the predicted wave characteristics over the Persian Gulf. *Ocean Engineering*, *75*, 81–89.

Molines, J., Herrera, M. P., Gómez-Martín, M. E., & Medina, J. R. (2019). Distribution of individual wave overtopping volumes on mound breakwaters. *Coastal Engineering*, *149*, 15–27.

Morgan, S. G., Shanks, A. L., MacMahan, J., Reniers, A. J. H. M., Griesemer, C. D., Jarvis, M., & Fujimura, A. G. (2017). Surf zones regulate larval supply and zooplankton subsidies to nearshore communities. *Limnology and Oceanography*, *62*(6), 2811–2828. Retrieved 2021-07-04, from https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lno.10609 (_eprint: https://aslopubs.onlinelibrary.wiley.com/doi/pdf/10.1002/lno.10609) doi: 10.1002/lno.10609

Mudronja, L., Matić, P., & Katalinić, M. (2017). Data-based modelling of significant wave height in the Adriatic Sea. *Transactions on Maritime Science*, *6*(01), 5–13.

Mulligan, R. P., Gomes, E. R., Miselis, J. L., & McNinch, J. E. (2019, April). Non-hydrostatic numerical modelling of nearshore wave transformation over shore-oblique sandbars. *Estuarine, Coastal and Shelf Science*, *219*, 151–160. Retrieved 2021-07-05, from https://www.sciencedirect.com/science/article/pii/S0272771417312271 doi: 10.1016/j.ecss.2019.01.027

Nieves, V., Radin, C., & Camps-Valls, G. (2021, April). Predicting regional coastal sea level changes with machine learn-

ing. *Scientific Reports*, *11*(1), 7650. Retrieved 2021-07-21, from https://www.nature.com/articles/s41598-021-87460-z (Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computer science;Physical oceanography;Projection and prediction Subject_term_id: computer-science;physical-oceanography;projection-and-prediction) doi: 10.1038/s41598-021-87460-z

Nikoo, M. R., Kerachian, R., & Alizadeh, M. R. (2018). A fuzzy knn-based model for significant wave height prediction in large lakes. *Oceanologia*, *60*(2), 153–168.

Ojeda, E., Guillén, J., & Ribas, F. (2011, February). Dynamics of single-barred embayed beaches. *Marine Geology*, *280*(1), 76–90. Retrieved 2020-06-18, from http://www.sciencedirect.com/science/article/pii/S0025322710003130 doi: 10.1016/j.margeo.2010.12.002

Ortiz-Royero, J. C., & Mercado-Irizarry, A. (2008, March). An intercomparison of swan and wavewatch iii models with data from ndbc-noaa buoys at oceanic scales. *Coastal Engineering Journal*, *50*(01), 47–73. Retrieved 2021-07-21, from https://www.worldscientific.com/doi/abs/10.1142/S0578563408001739 (Publisher: World Scientific Publishing Co.) doi: 10.1142/S0578563408001739

Pape, L., Plant, N. G., & Ruessink, B. G. (2010). On cross-shore migration and equilibrium states of nearshore sandbars. *Journal of Geophysical Research: Earth Surface*, *115*(F3). Retrieved 2020-06-25, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JF001501 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JF001501) doi: 10.1029/2009JF001501

Parker, K., Ruggiero, P., Serafin, K. A., & Hill, D. F. (2019, August). Emulation as an approach for rapid estuarine modeling. *Coastal Engineering*, *150*, 79–93. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S0378383918305532 doi: 10.1016/j.coastaleng.2019.03.004

Peres, D., Iuppa, C., Cavallaro, L., Cancelliere, A., & Foti, E. (2015). Significant wave height record extension by neural networks and reanalysis wind data. *Ocean Modelling*, *94*, 128–140.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Pianca, C., Holman, R., & Siegle, E. (2015). Shoreline variability from

days to decades: Results of long-term video imaging. *Journal of Geophysical Research: Oceans*, *120*(3), 2159–2178. Retrieved 2020-06-05, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JC010329 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014JC010329) doi: 10.1002/2014JC010329

Plant, N. G., Holland, K. T., & Holman, R. A. (2006). A dynamical attractor governs beach response to storms. *Geophysical Research Letters*, *33*(17). Retrieved 2020-01-24, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL027105 doi: 10.1029/2006GL027105

Plant, N. G., & Holman, R. A. (1997, July). Intertidal beach profile estimation using video images. *Marine Geology*, *140*(1), 1–24. Retrieved 2020-09-02, from http://www.sciencedirect.com/science/article/pii/S0025322797000194 doi: 10.1016/S0025-3227(97)00019-4

Price, T. D., & Ruessink, B. G. (2011a, April). State dynamics of a double sandbar system. *Continental Shelf Research*, *31*(6), 659–674. Retrieved 2020-06-25, from http://www.sciencedirect.com/science/article/pii/S0278434310003924 doi: 10.1016/j.csr.2010.12.018

Price, T. D., & Ruessink, B. G. (2011b, April). State dynamics of a double sandbar system. *Continental Shelf Research*, *31*(6), 659–674. Retrieved 2020-06-18, from http://www.sciencedirect.com/science/article/pii/S0278434310003924 doi: 10.1016/j.csr.2010.12.018

Quartel, S., Kroon, A., & Ruessink, B. G. (2008, April). Seasonal accretion and erosion patterns of a microtidal sandy beach. *Marine Geology*, *250*(1), 19–33. Retrieved 2021-07-05, from https://www.sciencedirect.com/science/article/pii/S0025322707002812 doi: 10.1016/j.margeo.2007.11.003

Ranasinghe, R., Symonds, G., Black, K., & Holman, R. (2004, September). Morphodynamics of intermediate beaches: a video imaging and numerical modelling study. *Coastal Engineering*, *51*(7), 629–655. Retrieved 2020-06-19, from http://www.sciencedirect.com/science/article/pii/S0378383904000900 doi: 10.1016/j.coastaleng.2004.07.018

Reikard, G., Pinson, P., & Bidlot, J.-R. (2011). Forecasting ocean wave energy: The ECMWF wave model and time series methods. *Ocean Engineering*, *38*(10), 1089–1099.

Roelvink, D., McCall, R., Mehvar, S., Nederhoff, K., & Dastgheib, A. (2018, April). Improving predictions of swash dynamics in XBeach: The role of groupiness and incident-band runup. *Coastal Engineering*, *134*, 103–123. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S0378383917301321 doi: 10.1016/j.coastaleng.2017.07.004

Ruggiero, P., Komar, P. D., & Allan, J. C. (2010). Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest. *Coastal Engineering*, *57*(5), 539–552.

Ruggiero, P., Komar, P. D., McDougal, W. G., Marra, J. J., & Beach, R. A. (2001). Wave Runup, Extreme Water Levels and the Erosion of Properties Backing Beaches. *Journal of Coastal Research*, *17*(2), 407–419. Retrieved 2021-07-05, from https://www.jstor.org/stable/4300192 (Publisher: Coastal Education & Research Foundation, Inc.)

Scott, T., Russell, P., Masselink, G., Wooler, A., & Short, A. (2007). Beach Rescue Statistics and their Relation to Nearshore Morphology and Hazards: A Case Study for Southwest England. *Journal of Coastal Research*, 1–6. Retrieved 2021-07-04, from https://www.jstor.org/stable/26481545 (Publisher: Coastal Education & Research Foundation, Inc.)

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the ieee international conference on computer vision* (pp. 618–626).

Shanks, A. L., Morgan, S. G., MacMahan, J., Reniers, A. J., Jarvis, M., Brown, J., ... Griesemer, C. (2018, January). Persistent Differences in Horizontal Gradients in Phytoplankton Concentration Maintained by Surf Zone Hydrodynamics. *Estuaries and Coasts*, *41*(1), 158–176. Retrieved 2021-07-04, from https://doi.org/10.1007/s12237-017-0278-2 doi: 10.1007/s12237-017-0278-2

Shanks, A. L., Morgan, S. G., MacMahan, J., & Reniers, A. J. H. M. (2017). Alongshore variation in barnacle populations is determined by surf zone hydrodynamics. *Ecological Monographs*, *87*(3), 508–532. Retrieved 2021-07-04, from https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1265 (_eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecm.1265) doi: 10.1002/ecm.1265

Shannon, C. E. (1948, July). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. (Conference Name: The Bell System Technical Journal) doi: 10.1002/j.1538-7305.1948.tb01338.x

Short, A., & Hogan, C. (1994a). Rip currents and beach hazards: their impact on public safety and implications for coastal management. *Journal of Coastal Research*, 197–209.

Short, A., & Hogan, C. (1994b). Rip Currents and Beach Hazards: Their Impact on Public Safety and Implications for Coastal Management. *Journal of Coastal Research*, 197–209. Retrieved 2020-09-23, from https://www.jstor.org/stable/25735599 (Publisher: Coastal Education & Research Foundation, Inc.)

Short, A. D. (1979). Three Dimensional Beach-Stage Model. *The Journal of Geology*, *87*(5), 553–571. Retrieved 2021-07-03, from https://www.jstor.org/stable/30070742 (Publisher: The University of Chicago Press)

Siegle, E., Huntley, D. A., & Davidson, M. A. (2007, February). Coupling video imaging and numerical modelling for the study of inlet morphodynamics. *Marine Geology*, *236*(3), 143–163. Retrieved 2020-06-25, from http://www.sciencedirect.com/science/article/pii/S002532270600274X doi: 10.1016/j.margeo.2006.10.022

Simon, M., Rodner, E., & Denzler, J. (2016, December). ImageNet pre-trained models with batch normalization. *arXiv:1612.01452 [cs]*. Retrieved 2020-09-23, from http://arxiv.org/abs/1612.01452 (arXiv: 1612.01452)

Smit, M. W. J., Aarninkhof, S. G. J., Wijnberg, K. M., González, M., Kingston, K. S., Southgate, H. N., ... Medina, R. (2007, June). The role of video imagery in predicting daily to monthly coastal evolution. *Coastal Engineering*, *54*(6), 539–553. Retrieved 2020-06-26, from http://www.sciencedirect.com/science/article/pii/S0378383907000221 doi: 10.1016/j.coastaleng.2007.01.009

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427–437.

Splinter, K., Harley, M., & Turner, I. (2018, November). Remote Sensing Is Changing Our View of the Coast: Insights from 40 Years of Monitoring at Narrabeen-Collaroy, Australia. *Remote Sensing*, *10*(11), 1744. Retrieved 2020-06-29, from http://www.mdpi.com/2072-4292/10/11/1744 doi: 10.3390/rs10111744

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

Splinter, K. D., Holman, R. A., & Plant, N. G. (2011). A behavior-oriented

dynamic model for sandbar migration and 2DH evolution. *Journal of Geophysical Research: Oceans*, *116*(C1). Retrieved 2020-01-10, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006382 doi: 10.1029/2010JC006382

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2015, April). Striving for Simplicity: The All Convolutional Net. *arXiv:1412.6806 [cs]*. Retrieved 2020-02-05, from http://arxiv.org/abs/1412.6806 (arXiv: 1412.6806)

Stauble, D. K. (1992, 7). *Long-term profile and sediment morphodynamics: Field research facility case history* (Tech. Rep. No. 92-7). Coastal Engineering Research Center Vicksburg MS.

Stokes, C., Russell, P., & Davidson, M. (2016, March). Subtidal and Intertidal Three-Dimensionality at a High Energy Macrotidal Beach. *Journal of Coastal Research*(75 (10075)), 472–476. Retrieved 2020-06-30, from https://meridian.allenpress.com/jcr/article/doi/10.2112/SI75-095.1/29807/Su (Publisher: Allen Press) doi: 10.2112/SI75-095.1

Stopa, J. E., Ardhuin, F., Babanin, A., & Zieger, S. (2016a). Comparison and validation of physical wave parameterizations in spectral wave models. *Ocean Modelling*, *103*, 2–17.

Stopa, J. E., Ardhuin, F., Babanin, A., & Zieger, S. (2016b, July). Comparison and validation of physical wave parameterizations in spectral wave models. *Ocean Modelling*, *103*, 2–17. Retrieved 2021-07-21, from https://www.sciencedirect.com/science/article/pii/S1463500315001614 doi: 10.1016/j.ocemod.2015.09.003

Strauss, D., Tomlinson, R., & Hughes, L. (2006). Numerical modelling and video analysis of intermediate beach state transitions. In *The 7th international conference on hydroscience and engineering.*

Talbot, M. M. B., & Bate, G. C. (1987, December). Rip current characersitics and their role in the exchange of water and surf diatoms between the surf zone and nearshore. *Estuarine, Coastal and Shelf Science*, *25*(6), 707–720. Retrieved 2021-07-04, from https://www.sciencedirect.com/science/article/pii/0272771487900175 doi: 10.1016/0272-7714(87)90017-5

Thornton, E. B., MacMahan, J., & Sallenger, A. H. (2007, June). Rip currents, mega-cusps, and eroding dunes. *Marine Geology*, *240*(1), 151–167. Retrieved 2020-07-01, from http://www.sciencedirect.com/science/article/pii/S0025322707000539 doi: 10.1016/j.margeo.2007.02.018

Tolman, H. L. (2008). A mosaic approach to wind wave modeling. *Ocean Modelling*, *25*(1-2), 35–47.

Tolman, H. L., & Chalikov, D. (1996). Source terms in a third-generation wind wave model. *Journal of Physical Oceanography*, *26*(11), 2497–2518.

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

Turner, I. L., Harley, M. D., Short, A. D., Simmons, J. A., Bracs, M. A., Phillips, M. S., & Splinter, K. D. (2016, December). A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen, Australia. *Scientific Data*, *3*(1), 160024. Retrieved 2020-03-12, from http://www.nature.com/articles/sdata201624 doi: 10.1038/sdata.2016.24

Turner, I. L., Whyte, D., Ruessink, B. G., & Ranasinghe, R. (2007, February). Observations of rip spacing, persistence and mobility at a long, straight coastline. *Marine Geology*, *236*(3), 209–221. Retrieved 2020-09-23, from http://www.sciencedirect.com/science/article/pii/S0025322706002957 doi: 10.1016/j.margeo.2006.10.029

van Enckevort, I. M. J., & Ruessink, B. G. (2001, August). Effect of hydrodynamics and bathymetry on video estimates of nearshore sandbar position. *Journal of Geophysical Research: Oceans*, *106*(C8), 16969–16979. Retrieved 2020-06-25, from http://doi.wiley.com/10.1029/1999JC000167 doi: 10.1029/1999JC000167

van Enckevort, I. M. J., & Ruessink, B. G. (2003, March). Video observations of nearshore bar behaviour. Part 2: alongshore non-uniform variability. *Continental Shelf Research*, *23*(5), 513–532. Retrieved 2020-06-25, from http://www.sciencedirect.com/science/article/pii/S0278434302002352 doi: 10.1016/S0278-4343(02)00235-2

Vos, K., Splinter, K. D., Harley, M. D., Simmons, J. A., & Turner, I. L. (2019). Coastsat: A google earth engine-enabled python toolkit to extract shorelines from publicly available satellite imagery. *Environmental Modelling & Software*, *122*, 104528.

Wadoux, A. M.-C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science*, *71*(2), 133–136.

Wilson, G. W., Özkan-Haller, H. T., & Holman, R. A. (2010). Data assimilation and bathymetric inversion in a two-dimensional horizontal surf zone model. *Journal of Geophysical Research: Oceans*, *115*(C12). Retrieved 2020-09-23, from

https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006286 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JC006286) doi: 10.1029/2010JC006286

Winckler, P., Liu, P. L.-F., & Mei, C. C. (2013, November). Advective Diffusion of Contaminants in the Surf Zone. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *139*(6), 437–454. Retrieved 2021-07-04, from

Woodcock, F., & Engel, C. (2005). Operational consensus forecasts. *Weather and Forecasting*, *20*(1), 101–111.

Woodcock, F., & Greenslade, D. J. (2007). Consensus of numerical model forecasts of significant wave Heights. *Weather and Forecasting*, *22*(4), 792–803.

Wright, L., Short, A., & Green, M. (1985, January). Short-term changes in the morphodynamic states of beaches and surf zones: An empirical predictive model. *Marine Geology*, *62*, 339–364. doi: 10.1016/0025-3227(85)90123-9

Wright, L. D., & Short, A. D. (1984, April). Morphodynamic variability of surf zones and beaches: A synthesis. *Marine Geology*, *56*(1), 93–118. Retrieved 2020-07-01, from http://www.sciencedirect.com/science/article/pii/0025322784900082 doi: 10.1016/0025-3227(84)90008-2

Wu, X., Peng, J., Shan, J., & Cui, W. (2015). Evaluation of semivariogram features for object-based image classification. *Geo-spatial Information Science*, *18*(4), 159–170.

Zamani, A., Solomatine, D., Azimian, A., & Heemink, A. (2008). Learning from data for wind–wave forecasting. *Ocean Engineering*, *35*(10), 953–962.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision – eccv 2014* (pp. 818–833). Cham: Springer International Publishing.

Zhang, Z., Li, C.-W., Li, Y.-S., & Qi, Y. (2006). Incorporation of artificial neural networks and data assimilation techniques into a third-generation wind–wave model for wave forecasting. *Journal of Hydroinformatics*, *8*(1), 65–76.

Zheng, J., Zhang, C., Demirbilek, Z., & Lin, L. (2014). Numerical study of sandbar migration under wave-undertow interaction. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *140*(2), 146–159.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015, April). Object Detectors Emerge in Deep Scene CNNs. *arXiv:1412.6856 [cs]*. Retrieved 2020-07-01, from http://arxiv.org/abs/1412.6856 (arXiv: 1412.6856)

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the ieee*

conference on computer vision and pattern recognition (pp. 2921–2929).

APPENDICES

Appendix A: Wave Model Decision Tree

A.1 Errors

In the following definitions, N is the number of samples, $\hat{y}$ is the estimated value, and y is the true value. Root mean squared error (RMSE) is defined as

$$RMSE = \sqrt{\frac{\sum_i^N (y_i - \hat{y}_i)^2}{N}}; \tag{A.1}$$

the percent error (PE) as

$$PE = 100\sqrt{\frac{1}{N}\sum_i^N \left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \tag{A.2}$$

the scatter index (SI) as

$$SI = \frac{RMSE}{\overline{y}} \tag{A.3}$$

and bias as

$$Bias = \frac{1}{N}\sum_i^N \hat{y}_i - y_i \tag{A.4}$$

.

Lower error values indicate greater model skill.

A.2 Wave Parameter Definitions

The bulk parameter MWD is defined as the vectorial mean of the directional spectrum as in Kuik et al. (1988). Splitting on wave direction poses a logical problem in that the variable is periodic, such that partitions can only be logically defined after an initial split into sectors. Mean wave direction is defined using a nautical convention in the range 0-360 degrees. This could have a minor (unavoidable) effect on the overall partition structure; partitions abutting the zero degree line cannot be joined together, hence the data in these partitions will be considered as separate branches of the tree. In our case, MWD values ranged from 220-360 for the entire dataset, so this problem did not arise. The bulk parameters H_s and

T_{m01} are defined as follows.

$$H_s = 4\sqrt{(m_0)} \tag{A.5}$$

$$T_{m01} = \frac{m_0}{m_1} \tag{A.6}$$

where the nth moment, m_n, is defined as

$$m_n = \int_\theta \int_f f^n F(f,\theta)\,\mathrm{d}f\mathrm{d}\theta \;\; for \;\; n = 0, 1, 2... \tag{A.7}$$

In this formulation, f is frequency and θ is direction. The integration limits for frequency are 0.03-0.4Hz for the observations, and 0.041-0.411Hz for the wave model.

A.3 Computational Library

The methods described here were implemented using python's scikit-learn library. This library includes the base decision tree, the bagged regression tree, a bagging regressor ensemble method, and a grid-search cross validation routine. Cross-validation was executed first, in order to find the parameters that resulted in the most accurate bagged regression tree. The optimal parameters were then used in training and testing. This cross validation, train and test phases were executed 30 times, as explained in section 2.2.2. The algorithm can be run on one core, and the run time is approximately 6000 seconds for one validation, train and test phase. The majority of the computations occurred during cross-validation.

Appendix B: Chapter 3 Appendix

B.1 Convolutional Neural Network Theory

This study uses a pre-defined CNN called Resnet-50 He et al. (2015), shown schematically in Figure B.1. The CNN begins with a feature extraction step, consisting of a convolutional layer followed by convolutional blocks. The convolutional layer includes kernel of weights (treated as free parameters) that are convolved over the image intensity values, resulting in a representation of the image into a new feature space called a feature map. For example, a kernel might act to find edges within the image, resulting in a feature map that contains only the edges of high contrast from the original image. The first convolutional layer in Resnet-50 has 64 kernels that span 7x7 pixels, and are applied sequentially to the image with a stride (or lag) of two pixels. The output from the convolutional layer is then fed into a batch normalization layer. The batch normalization step acts to redistribute the values of the feature maps to a Gaussian distribution, which aids in optimization Ioffe and Szegedy (2015). Next, an activation function (here a $ReLU$ function, $ReLU(x) = max\{0, x\}$) is applied to the batch-normalized values to filter only the positive values output from the convolution step. This allows for the convolutional neural network to calculate non-linear relationships between the input and output. The $ReLU$ output is then fed into an max pooling layer, which reduces the dimensions of the feature map by providing summary statistics of the salient portions of the feature maps. Specifically, the max pooling layer consists of convolving a filter over the image. The filter outputs the maximum value of a feature map within a neighborhood of 3x3 pixels and with a stride of two pixels.

The output of the first convolutional layer is then passed to a series of "bottleneck" convolutional blocks, Figure B.1. As the information passes through the CNN, the original 3x512x512 block of data (the image) increases in the first dimension due to the number of kernels used, and decreases in the two spatial dimensions due to the use of the 3x3 kernels with a stride of two in the convolutional steps. In addition to the sequential connections between each block, the blocks are also connected via skip, or "residual" connections (hence the name ResNet). These residual connections occur through an identity mapping $(y = F(\mathbf{x}) + \mathbf{x}), 0)$; meaning that the information from an earlier block ($\mathbf{x}$) is directly passed to a later block, skipping the transformations made within the blocks along the way ($F(\mathbf{x})$).

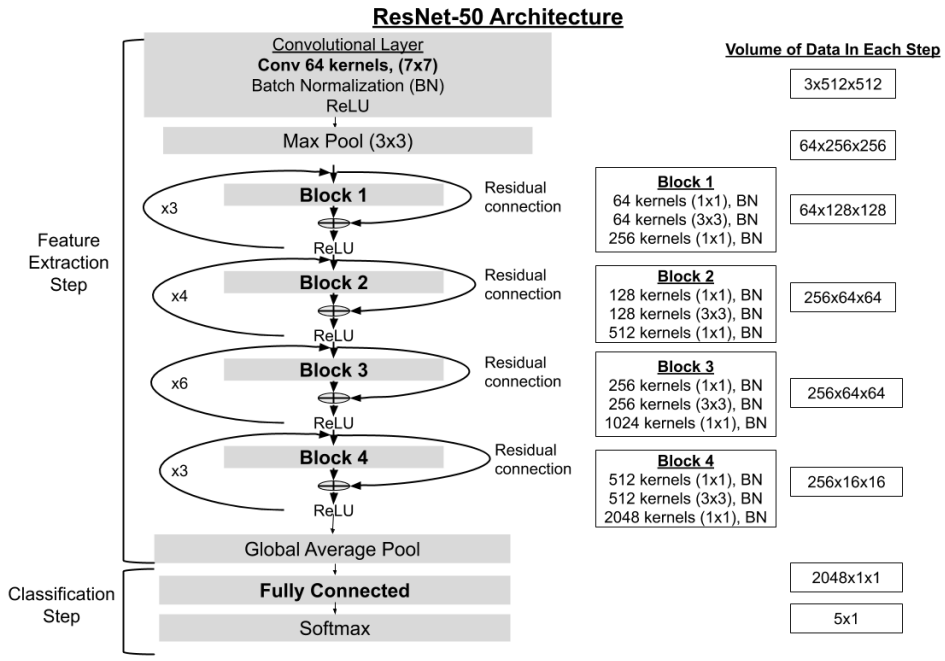


Figure B.1: ResNet architecture (adapted from Table 1 and Figure 3 of He et al. (2015)). The modules with learnable parameters are boldface.

The residual connections help the algorithm to reduce training error quicker, since they enable information to flow more efficiently and directly through the network to allow better adjustment of the kernel weights in the first layers. The output from the final convolutional block is fed to a global average pooling layer. Similar to the max pooling layer, it serves to reduce the dimensions of the feature maps by providing summary statistics about the feature maps. It outputs the average of each feature map, resulting in a $(k \times 1 \times 1)$ vector, where k is the number of feature maps (2048 in ResNet50).

After the feature extraction step comes the classification step. This is performed with a traditional machine learning (i.e., not deep learning) technique: a neural network. The output of the global average pooling layer, the flattened vector, is fed into a fully connected neural network that has one layer of neurons. The output of the neural network is in turn passed to a softmax function ($softmax = \frac{exp(\mathbf{z}_c)}{\sum_c(\mathbf{z_c})}$), which outputs a probability mass vector corresponding to the predicted probability for each class, $\hat{\mathbf{y}}$. The entry with the highest probability is taken as the class prediction.

Maximum likelihood estimation of the CNN parameters is performed during training by minimization of a cost function calculated over the training data set. For one training example (image, n) with a total of five classes, the cost function is the cross entropy function:

$$J_n(\mathbf{\Theta}) = -\sum_{c=1}^{5} y_c log(\hat{y_c}) \tag{B.1}$$

In this function, the CNN prediction (defined as the entry with the maximum value of the softmax output) is $\hat{y_c}$, and the true value is y_c. The softmax output can be thought of as a modelled probability distribution, where the model is defined by the free parameters of the CNN. The target can be thought of as a Dirac delta function with a '1' entered in the position of the true class. Maximum likelihood estimation is used to determine the free parameters of the CNN that are most likely to predict the true distribution Goodfellow et al. (2016). The maximum likelihood estimation is made by minimizing the cost function with an iterative scheme called stochastic gradient descent (SGD) with momentum. At each iteration step of SGD, the parameters are adjusted according to the value of the gradient of the cost function with respect to those parameters. In this study, mini-batch gradient descent is used, so the parameters are updated with a gradient value that is averaged from four training examples: $\frac{1}{4}\sum_{n=1}^{4} \nabla_{\mathbf{\Theta}} J_n(\mathbf{\Theta})$. Momentum is used to achieve minimization more efficiently, by adding a weighted estimate of the gradient from the previous batch to the adjustment. The specific SGD parameters used in this study were a learning rate of 0.01, and a momentum weight of 0.9. The learning rate was reduced by a factor of 10 if the training loss did not improve (decrease) after 8 epochs. The training was run for 120 epochs, and generally converged after approximately 40 epochs.

B.2 Skill Metrics

For a multi-class classification problem, a confusion matrix provides information about which classes were categorized correctly and incorrectly. A confusion matrix $\mathbf{Z}$, is a KxK matrix where K is the number of classes considered in the multi-class problem. The entry in the confusion table z_{ij} represents how many images of class i were classified as class j. The total number of images belonging to a certain class i is defined as $x_i = \sum_i z_{ij}$, while the total number of images classified (by the CNN) as a certain class i is defined as $y_i = \sum_j z_{ji}$. The specificity, or precision (P), is defined as the number of images that were labelled correctly divided by the

number of images labelled as that category:

$$P = \frac{z_{ii}}{y_i} \tag{B.2}$$

The per-class accuracy, or recall (R), is defined as the number of images that were labelled correctly divided by the total number of images with that label in the dataset:

$$R = \frac{z_{ii}}{x_i} \tag{B.3}$$

The F-score is defined as the harmonic mean between precision and recall:

$$F = 2\frac{P * R}{P + R} \tag{B.4}$$

See references Sokolova and Lapalme (2009) and Baldi et al. (2000) for further information on accuracy measures for multi-class classification problems.

Appendix C: Chapter 4 Appendix

C.1 Wave Conditions Corresponding with Simplex Dynamics

L. D. Wright and Short (1984) (WS84) observed that the current morphology is dependent upon the historical environmental conditions, called "equilibrium conditions." Specifically, equilibrium dimensionless fall velocity ($\overline{\Omega}$) corresponds to which state the beach was currently in. This study calculates equilibrium conditions following L. Wright et al. (1985):

$$\overline{\Omega} = [\sum_{i=1}^{D} 10^{\frac{-i}{\phi}}]^{-1} \sum_{i=1}^{D} [\Omega_i 10^{\frac{-i}{\phi}}] \tag{C.1}$$

In this formulation, D=30 and ϕ=5. Equilibrium conditions are exponentially weighted in such a way that the values ϕ days ago had a weighting factor of 10%, and the conditions from the last thirty days are taken into account.

Studies have observed that changes in the hydrodynamics induce a lagged response of the morphology (Stokes, Russell, & Davidson, 2016; L. Wright et al., 1985). Therefore, prediction of beach state can be undertaken using an equilibrium approach, whereby the next beach state is a function of the equilibrium conditions and the difference between instantaneous conditions and equilibrium conditions (Davidson et al., 2013), called "disequilibrium."

$$\Delta\Omega = \overline{\Omega} - \Omega \tag{C.2}$$

The simplices were arranged in order of average $\overline{S}$ (centerline of boxes, Figure 4.5) and the corresponding equilibrium and disequilibrium wave conditions were calculated and presented in Figure C.1.

The simplices corresponded with $\overline{\Omega}$ and $\overline{MWD}$, as defined in equation C.1. Figure 4.7b and c show $\overline{\Omega}$ and $\overline{MWD}$ when the evolution occurred. Generally, the time averaged wave conditions corresponded to the simplex from the previous day (i.e., the rows in Figure 4.7). WS84 observed that intermediate beach states generally correspond with Ω values between 1-6, and the values herein correspond with that range ($\overline{\Omega}$ values between 2.97-5.03), with LTT simplices corresponding to lower $\overline{\Omega}$ and LBT simplices corresponding higher $\overline{\Omega}$. However, in contrast with

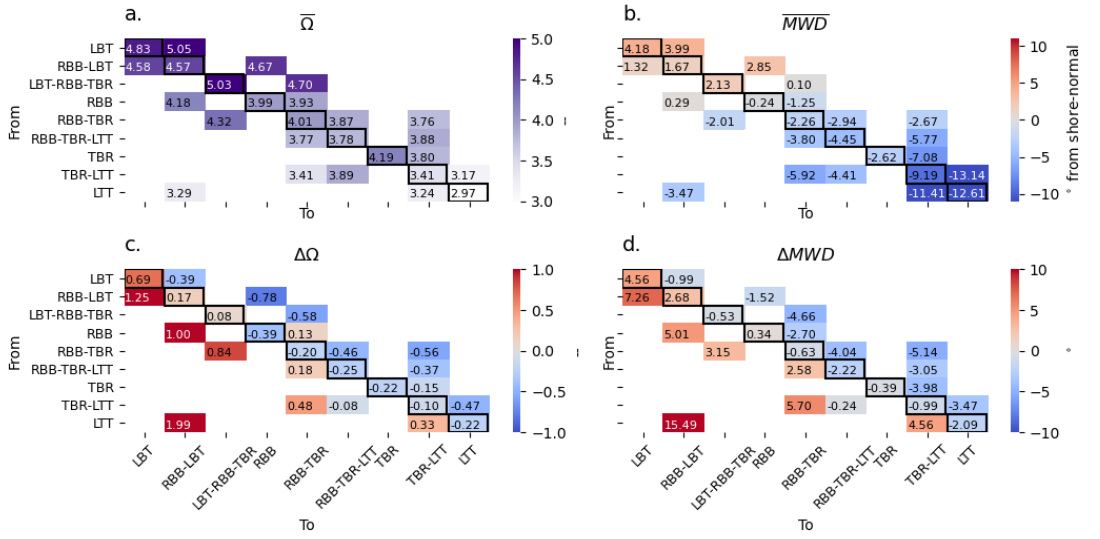


Figure C.1: Wave conditions corresponding to the simplex transitions. (Top panels) Time averaged dimensionless fall velocity in (a) and mean wave direction in (b). (Bottom panels) The difference between the time-averaged and instantaneous fall velocity in (c) and mean wave direction in (d).

WS84, the highest dimensionless fall velocity also corresponded with slightly more onshore, rhythmic and even terraced shapes, specifically LBT-RBB and LBT-RBB-TBR (average $\Omega = 5.05$ and 5.03, respectively). The LBT-RBB-TBR simplex was found to occur on average for the highest wave heights and longest wave periods (average $Hs = 1.15m, Tp = 9.12s$) within the LBT family, suggesting that low steepness, high wave events permit rhythmic barlines. This observation is consistent with Price and Ruessink (2011b), who observed an erosive TBR state (eTBR), a state encompassing rip currents that coincided with offshore bar movement.

The simplex ordering also corresponded with $\overline{MWD}$. Specifically, slightly northern wave directions ($4.18 > \overline{MWD} > 2.13°$) correspond with the LBT simplices, slightly southern ($-4.45° < MWD < -0.24°$) with TBR simplices, and southern ($-12.61 < \overline{MWD} < -9.19°$) with LTT simplices.

Simplex stability (i.e., the diagonal of the tables in Figure C.1) corresponded to the smallest disequilibrium values (equation C.2). Higher levels of disequilibria have been hypothesized to induce a greater response in the morphology Davidson et al. (2013); L. Wright et al. (1985), which is generally confirmed herein; on average, $\overline{\Delta\Omega} = 0.26, \overline{\Delta MWD} = 1.60°$ when the simplex remained in the same configuration

versus $\overline{\Delta\Omega} = 0.58, \overline{\Delta MWD} = 4.33°$ when the simplex evolved to a different one.

Down-simplex evolutions occurred for reductions in dimensionless fall velocity and increasingly southern incident wave angles. The greatest reduction in dimensionless fall velocity among all the transitions ($\Delta\Omega = -0.78$) occurred as RBB-LBT evolved to RBB. The up simplex evolutions corresponded to increasing dimensioness fall velocities and wave angles becoming more northerly. The highest increase in dimensionless fall velocity and mean wave direction ($\Delta\Omega = 1.99, \Delta MWD = 15.49°$) coincided in the evolution between LTT to LBT-RBB, which occurred 50 times. This transition was likely due to storm events arriving after an extended period of lower wave heights associated with the LTT configuration.

The transition between TBR-LTT to RBB-TBR-LTT occurred for slight reductions in dimensionless fall velocity and wave angle ($\Delta\Omega = -0.08, \Delta MWD = -0.24°$). This transition occurred most often in the month of August (20% of all TBR-LTT to RBB-TBR-LTT evolutions occurred in this month, with the rest distributed equally through the months), and contributed to annual entropy growth. During this month, the barline was evolving offshore from its most onshore position in summer. At times, the bar moved offshore at a a small distance such that the CNN labelled the configuration LTT, but acted dynamically as an LBT in that onshore migration occurred. Therefore, the evolution between TBR-LTT and RBB-TBR-LTT was often a slight onshore movement as the barline became more rhythmic and accreted. Note that the differences between LTT and LBT are shoreline proximity, and based scaling bar distance to shoreline with incident infragravity waves. However, undertow is now considered to be the main drive of offshore bar migration Zheng, Zhang, Demirbilek, and Lin (2014), and so the differentiation between LTT and LBT being shoreline proximity, following LH90, is likely outdated. An improvement to beach state theory could potentially be the inclusion of a linear state with intermediate proximity to the shoreline.