Coastal Engineering 157 (2020) 103595

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: http://www.elsevier.com/locate/coastaleng

An application of a machine learning algorithm to determine and describe error patterns within wave model output

Ashley Ellenson [a,*], Yuanli Pei [b], Gregory Wilson [c], H. Tuba Özkan-Haller [d], Xiaoli Fern [e]

[a] College of Civil and Construction Engineering. Oregon State University, 104 COAS Administration Building, Corvallis, OR, 97330, USA
[b] Kensho Technologies, LLC, Kensho Technologies, LLC, Cambridge, MA, USA
[c] College of Earth, Ocean and Atmospheric Sciences, Oregon State University, USA
[d] College of Earth, Ocean and Atmospheric Sciences, College of Civil and Construction Engineering, Oregon State University, USA
[e] College of Electrical Engineering and Computer Science. Oregon State University, 104 COAS Administration Building, Corvallis, OR, 97330, USA

ARTICLE INFO

Keywords:
Wave forecasting
Machine learning
Decision tree
Regression tree

ABSTRACT

This study uses a machine learning algorithm, the bagged regression tree, to detect error patterns within 24-h 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 were 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 (*Tm*01)). 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 model phase space associated with mean error patterns.

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 et al., 2014; García-Medina et al., 2014; Guedes Soares et al., 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

* Corresponding author.
E-mail address: ellensoa@oregonstate.edu (A. Ellenson).

https://doi.org/10.1016/j.coastaleng.2019.103595
Received 25 January 2019; Received in revised form 5 November 2019; Accepted 7 November 2019
Available online 27 January 2020

(Mudronja et al., 2017). Machine learning techniques used include Artifical 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 et al. (2011). They found that ANNs, ANFIS, and SVMs work well, whereas BNs resulted in less accurate results. In (Berbić et al., 2017), they used SVMs and ANNs to predict significant wave height up to 6 h into the future using time-lagged wind information as input. In (Zamani et al., 2008), they use three different ANNs to predict significant wave heights 1, 3, and 6 h 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 et al., 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 et al., 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 (Tsai et al., 2002; Kalra and Deo, 2007).

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 et al., 2011; Woodcock and Greenslade, 2007; Woodcock and Engel, 2005). In particular, neural networks have been used as a post-processing routine to correct the original wave model predictions of wave parameters (Moeini et al., 2012, 2014; Deshmukh et al., 2016; Hadadpour et al., 2013; Zhang et al., 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 et al., 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 Mahjoobi, 2009; Etemad-Shahidi and Bonakdar, 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., 2018; Gutierrez et al., 2015), the rules of Bayesian networks were interpreted to determine the relevant physical drivers in shoreline change processes. In (Cornejo-Bueno et al., 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 determining 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-h significant wave height forecasts generated for buoy locations within the Northeast Pacific region, as shown in Fig. 4. Results are presented for two different wave climates: high and low energy (winter and summer, respectively). The input features are comprised of 24-h 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. Methods

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 $((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N))$ 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 et al., 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 Appendix B 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:

$$y_i = H_{s,obs} - H_{s,WW3} \tag{1}$$

A negative (positive) target value indicate a significant wave height overestimation (underestimation). In Fig. 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 (1)). This feature space, T, is the space that the decision tree will divide into partitions.

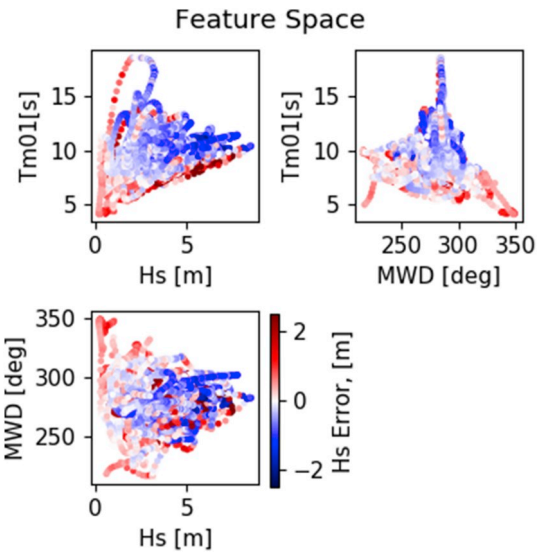

Fig. 1. The 2012–2013 winter target values (colored points) plotted as a feature space with respect to the input features. The target values are the WaveWatch III *Hs* prediction error (Equation (1)), and the three input feature elements are WaveWatch III *Hs*, *MWD* and *Tm*01.

The objective is to find the partitions wherein the mean target value of that partition ($\bar{y}(t)$) 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:

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

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). See Fig. 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 ($\bar{y}(t)$). During testing, each instance is categorized into a partition with respect 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 ($\bar{y}(t)$, established during training) which is the prediction for the instances which fall into the associated partition. See Fig. 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 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 and Liang, 2017).

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 (Breiman, 1996). The final prediction is

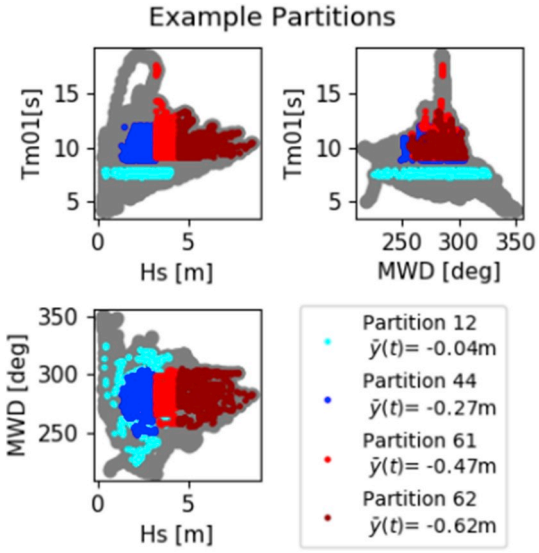

Fig. 3. An example of three final partitions determined after the decision tree split the entire dataset with respect to values of bulk wave parameters. These partitions correspond with partitions 61 and 62 that are represented in the tree structure in Fig. 2.

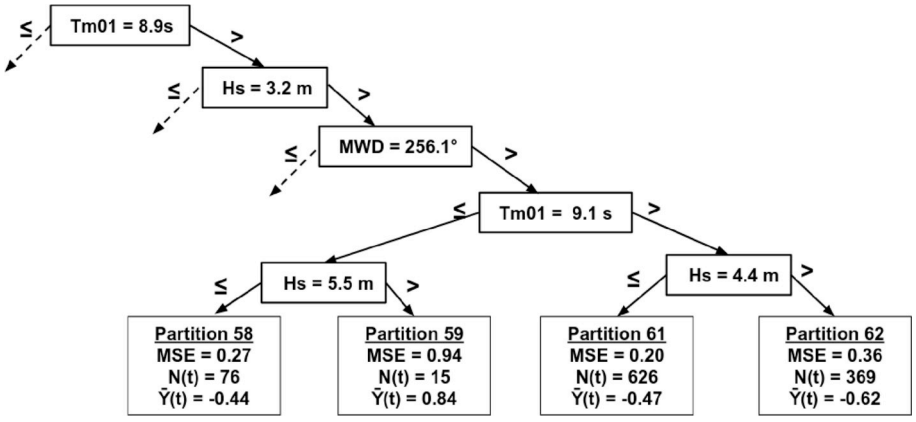

Fig. 2. Representative tree structure of the splits made between the input features during training on the training data set. The input features which the decision tree splits on are indicated within each node with a threshold value. Branches to the left (right) represent the division of data points 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. $\bar{y}(t)$ is the mean target value associated with that partition.

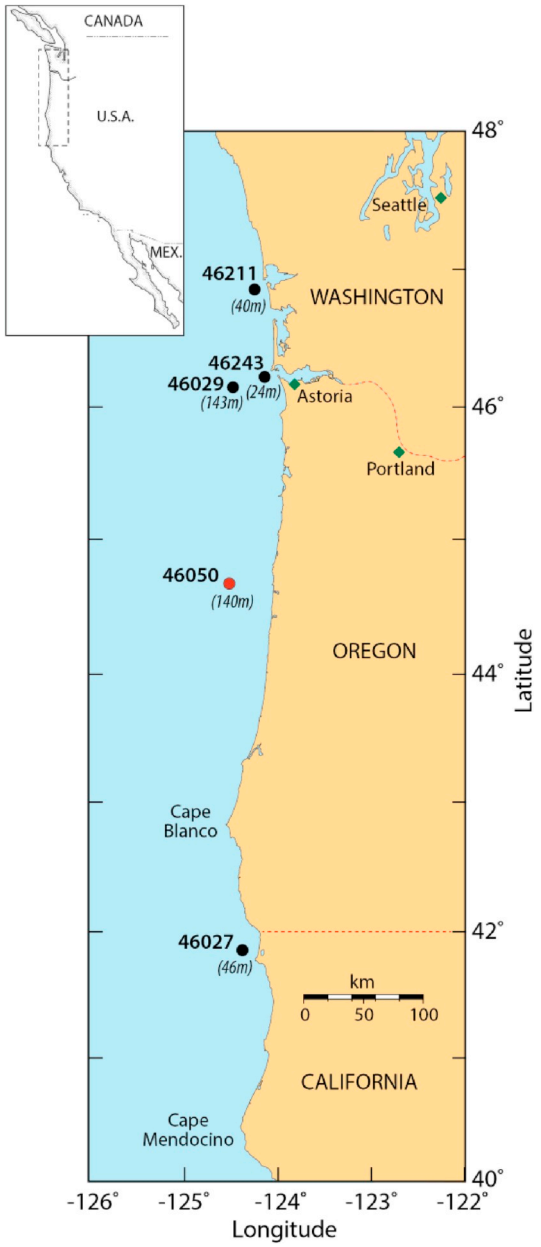

Fig. 4. The locations and water depths of the buoys implemented in this study. Predictions and error metrics were computed using buoy 46050 (red dot). In the geospatial application, the bagged regression tree was trained with data from the other locations. Inset shows the geographic location of the domain (dashed rectangle) on the U.S. west coast. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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. 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.

2.3. Experiments

Experiments were designed to determine the performance of the bagged regression tree for different wave climates and different input feature combinations. The data were 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 m and an average mean wave period ($Tm01$) of 9.3 s. The summer data set was characterized by an average significant wave height of 1.8 m and average mean wave period of 7.2 s. The seasonal data were 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 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

Table 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

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 Fig. 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."

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-h time horizon forecasts of bulk parameters significant wave height, mean wave period (*Tm*01) and mean wave direction, as provided by WaveWatch III (WW3) from the years 2012–2015 (García-Medina et al., 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 (*Tm*01). The wind model input to the wave model are 24-h 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-h lead time GFS 3-h 10 m 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 Fig. 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 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
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

3. Results

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. 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. Further discussion on bagged regression tree architecture can be found in Section 4.2.

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 3 for error metrics for the original modelled (WW3) and corrected time series as compared with the observed data. Representative portions of each season's time series are presented in Fig. 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 majority of the time, the application of the corrections resulted in more accurate significant wave height predictions, but sometimes resulted in less accurate significant wave height predictions, such as the underestimations in the June month of the summer time series. 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.

Fig. 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 3 shows that error metrics root-mean-squared-error (RMSE),

Table 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

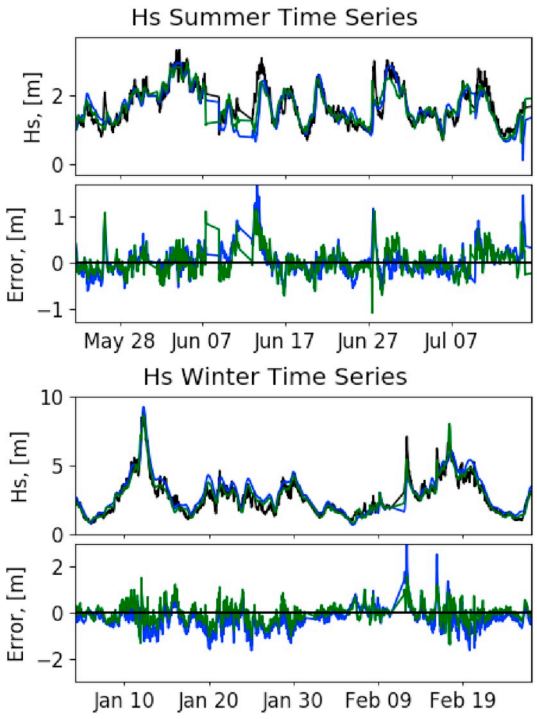

Fig. 5. Representative time series of significant wave height for summer and winter. Gaps in the data exist due to wave model restarts or gaps in observations. The original model significant wave height output is colored blue, the observations (from NDBC buoy 46050) are colored black, and the corrected 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). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

bias, scatter 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 Appendix 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%.

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 (Tm01), and error metrics were calculated (see Fig. 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 (*Tm*01) 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, Fig. 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.

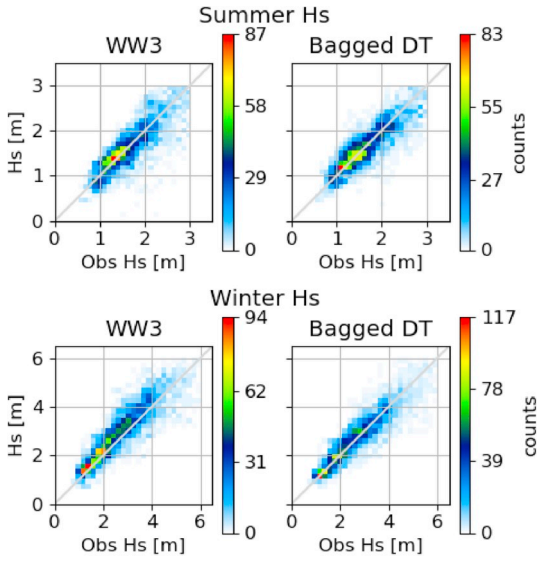

Fig. 6. Density scatter plots (colors represent histogram counts) of predicted versus observed significant wave height. The data shown comprise all data from the With Wind experiments, separated into summer (top panels) and winter (bottom panels) testing periods. Left-hand panels show results from the 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. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

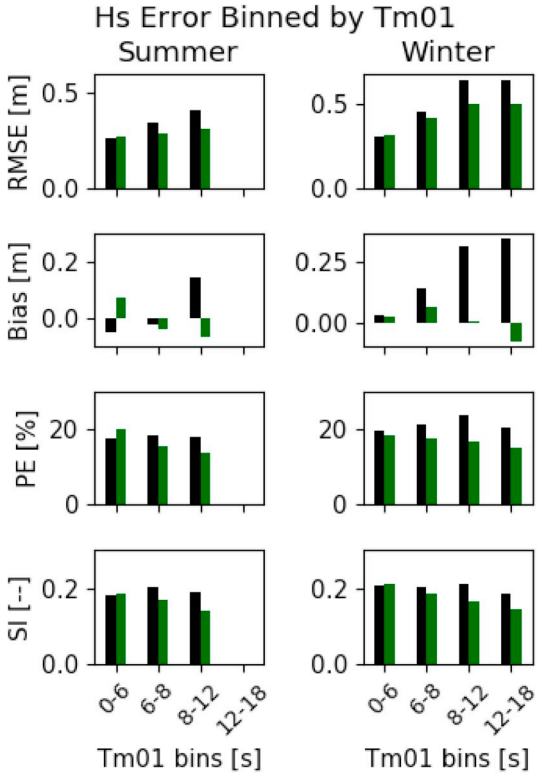

Fig. 7. Error metrics binned by mean wave period *Tm*01 for summer (left) and winter (right). The WW3 error is colored black and the With Wind experiment is colored green. Note that no data points are associated with the longest period bin for summer. The greatest error reductions are achieved for significant wave heights associated with mean periods greater than 8s. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3. Results: geospatial 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 Fig. 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 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.

4. Discussion

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 (*Tm*01) between 8 and 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., 2018). The 2012–2013 training data from partition 31 are plotted in Fig. 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 (*Tm*01) 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 and Chalikov, 1996) which is consistent with the conclusions of (Stopa et al., 2016), 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.

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 et al., 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 Fig. 9 and explained in the following. There are a

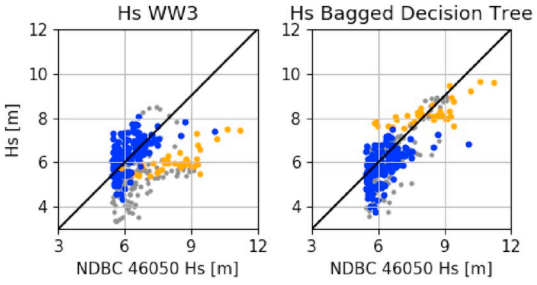

Fig. 9. Data points greater than the 95th percentile significant wave height (5.4 m) in the 2012–2013 training data. Data from partitions 31 and 44 are colored blue and orange, respectively. The left panel is *Hs* scatter plot of the original model output, and the right panel is the *Hs* after the bagged decision tree corrections. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

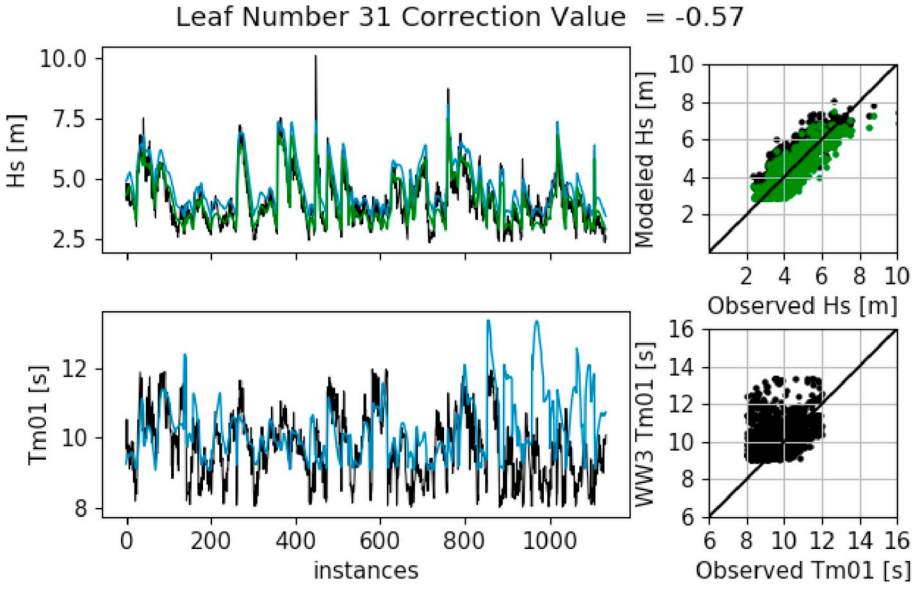

Fig. 8. Data points of the most populated partition, partition 31. The left panels show the instances of NDBC buoy 46050 observations (black), WW3 (blue), and corrected (green) *Hss*. The top right panel shows the original WW3 (black) and corrected (green) instances as a scatter plot. The bottom left panel shows the instances of observed (black) and WW3 (blue) *Tm*01. The bottom right panel shows the *Tm*01 instances as a scatter plot. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

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 Fig. 9). Note that the 173 data points are a subset of data within partition 31, the entirety of which is illustrated in Fig. 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 error 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 4 for the respective rules made on the data within each of the partitions.

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 local wind is strong ($> 15m/s$) and from the south. Interestingly, this pattern is consistent with a previous study which showed that significant wave height underestimations consistently occurred during times when a strong southerly wind was present in the region (Ellenson and Özkan-Haller, 2018). Further, the instances within this partition were also associated with mean wave period underestimations (*Tm*01 bias of −0.5 s, not shown). In (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 *Tm*01 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.

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 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.

Table 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
*Tm*01	> 9.06 s	> 8.32 s
MWD	$> 245°, \leq 360°$	–
wndmag	≤ 15 m/s	> 15 m/s
wnddir	–	$> 0°, \leq 215°$
Error Correction Value	−0.57 m	2.15 m

For both summer and winter, mean wave period (*Tm*01) 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 Fig. 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 Fig. 10, for winter, the most informative input features were *Tm*01 and wind magnitude (with feature importances of 0.38 and 0.35, respectively), whereas for summer, the most informative input feature was *Tm*01 (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 et al., 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 Fig. 11, where wind magnitudes $> 15m/s$ delineate 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.

4.2.1. Input features: Multiple-buoy experiments

When multiple buoys were considered in the training, mean wave period (*Tm*01) provided the most information, and the geospatial features *lat*, *lon*, and *depth* were also used (feature importances < 0.02) (see Fig. 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

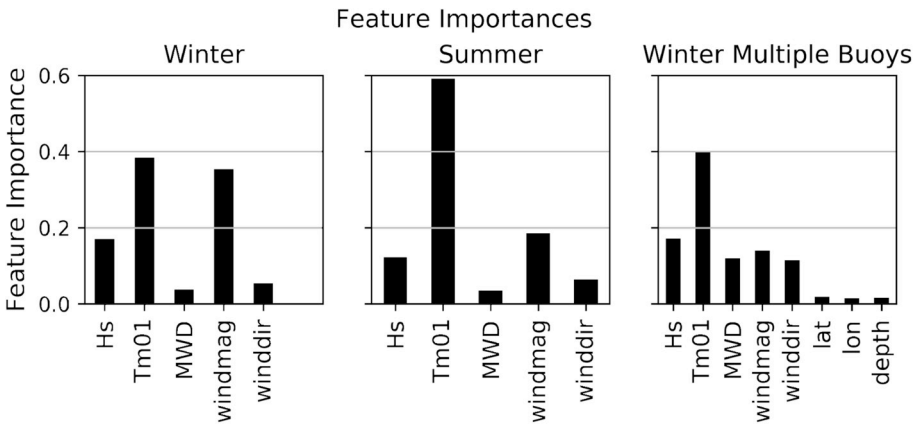

Fig. 10. Feature importances for the two single buoy experiments for summer, winter, and the multiple buoys experiments at NDBC buoy 46050.

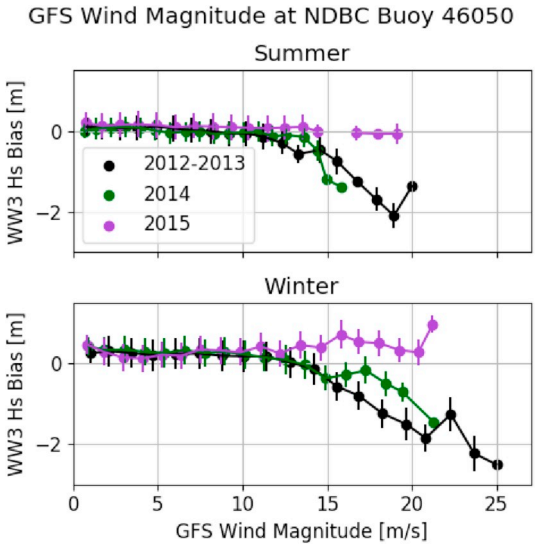

Fig. 11. Comparisons of the wind magnitude input feature versus the target value (significant wave height error) for summer and winter. Training data from years 2012–2013 are colored black, testing data from 2014 to 2015 are colored green and purple, respectively. The data is binned in 1.25 m/s bins and the average is plotted here. The y-errorbar indicates the standard deviation within each bin. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

number of data points within the same partitions (4456 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 (2824 error corrections out of a combined 12,582 data points). During testing, the output location (buoy 46050) had the highest number of data points (2953 out of 3864 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 (2237 out of 3864 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.

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 Fig. 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 and 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 Fig. 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 regression 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.

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-h time horizon significant wave height prediction time series made by a numerical model. In this hybrid approach, numerical model output was used as input features, and the target was 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 scatter index by 19% (from 0.21 to 0.17).

In a descriptive sense, this method can act as a diagnostic tool for finding regions of model output phase space where a wave model improvement is necessary. During training, it isolated 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 scatter index (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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge David Reinert for his help in making the map in the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (1545188).

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.coastaleng.2019.103595 and at the following github repository: https://github.com/anellenson/DecisionTree_WaveForecasts.

Appendix A. Error Metrics

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

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

and bias as

$$Bias = \frac{1}{N}\sum_{i}^{N} \widehat{y}_i - y_i \tag{A.4}$$

Lower error values indicate greater model skill.

Appendix B. 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°. 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 to 360 for the entire dataset, so this problem did not arise. The bulk parameters H_s and T_{m01} are defined as follows.

$$H_s = 4\sqrt{(m_0)} \tag{B.1}$$

$$T_{m01} = \frac{m_0}{m_1} \tag{B.2}$$

where the nth moment, m_n, is defined as

$$m_n = \int_{\theta}\int_{f} f^n F(f,\theta)\mathrm{d}f\,\mathrm{d}\theta \quad for \quad n = 0, 1, 2... \tag{B.3}$$

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.

Appendix C. 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. The algorithm can be run on one core, and the run time is approximately 6000 s for one validation, train and test phase. The majority of the computations occurred during cross-validation.

References

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. J. Clim. 27 (4), 1619–1632.

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.

Beuzen, T., Splinter, K., Marshall, L., Turner, I., Harley, M., Palmsten, M., 2018. Bayesian networks in coastal engineering: distinguishing descriptive and predictive applications. Coast Eng. 135, 16–30.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.

Breiman, L., Spector, P., 1992. Submodel selection and evaluation in regression. The X-random case. Int. Stat. Rev. 291–319.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

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. Eng. 160, 33–44.

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 J. Ocean. Eng. 41 (4), 944–953.

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

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 Forecast. 33 (2), 479–499.

Etemad-Shahidi, A., Bonakdar, L., 2009. Design of rubble-mound breakwaters using M5 machine learning method. Appl. Ocean Res. 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. Eng. 36 (15–16), 1175–1181.

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

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

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. J. Geophys. Res.: Earth Surface 120 (12), 2452–2475.

Hadadpour, S., Moshfeghi, H., Jabbari, E., Kamranzad, B., 2013. Wave hindcasting in anzali, caspian sea: a hybrid approach. J. Coast Res. 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. J. Atmos. Ocean. Technol. 26 (8), 1614–1633.

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

Kalra, R., Deo, M., 2007. Genetic programming for retrieving missing information in wave records along the west coast of India. Appl. Ocean Res. 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. JMLR. org, pp. 1885–1894.

Kuik, A., Van Vledder, G.P., Holthuijsen, L., 1988. A method for the routine analysis of pitch-and-roll buoy wave data. J. Phys. Oceanogr. 18 (7), 1020–1034.

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

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. Eng. 38 (2–3), 487–497.

Moeini, M.H., Etemad-Shahidi, A., Chegini, V., Rahmani, I., 2012. Wave data assimilation using a hybrid approach in the Persian Gulf. Ocean Dynam. 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. Eng. 75, 81–89.

Mudronja, L., Matić, P., Katalinić, M., 2017. Data-based modelling of significant wave height in the Adriatic Sea. Trans. Marit. Sci. 6, 5–13, 01.

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.

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

Reikard, G., Pinson, P., Bidlot, J.-R., 2011. Forecasting ocean wave energy: the ECMWF wave model and time series methods. Ocean. Eng. 38 (10), 1089–1099.
Ruggiero, P., Komar, P.D., Allan, J.C., 2010. Increasing wave heights and extreme value projections: the wave climate of the US Pacific Northwest. Coast Eng. 57 (5), 539–552.
Stopa, J.E., Ardhuin, F., Babanin, A., Zieger, S., 2016. Comparison and validation of physical wave parameterizations in spectral wave models. Ocean Model. 103, 2–17.
Tolman, H.L., 2008. A mosaic approach to wind wave modeling. Ocean Model. 25 (1–2), 35–47.
Tolman, H.L., Chalikov, D., 1996. Source terms in a third-generation wind wave model. J. Phys. Oceanogr. 26 (11), 2497–2518.
Tsai, C.-P., Lin, C., Shen, J.-N., 2002. Neural network for wave forecasting among multi-stations. Ocean. Eng. 29 (13), 1683–1695.
Woodcock, F., Engel, C., 2005. Operational consensus forecasts. Weather Forecast. 20 (1), 101–111.
Woodcock, F., Greenslade, D.J., 2007. Consensus of numerical model forecasts of significant wave Heights. Weather Forecast. 22 (4), 792–803.
Zamani, A., Solomatine, D., Azimian, A., Heemink, A., 2008. Learning from data for wind–wave forecasting. Ocean. Eng. 35 (10), 953–962.
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. J. Hydroinf. 8 (1), 65–76.