

# INVESTIGATING MACHINE LEARNING FOR VIRTUAL WAVE MONITORING

Peach L<sup>1,2</sup>, Cartwright N<sup>1,2</sup>, Strauss, D<sup>2</sup>

Wave monitoring is a time consuming and costly endeavour which, despite best efforts, can be subject to occasional periods of missing data. This paper investigates the application of machine learning to create "virtual" wave height (Hs), period (Tz) and direction (Dp) parameters. Two supervised machine learning algorithms were applied using long term wave parameter datasets sourced from four wave monitoring stations in relatively close geographic proximity. The machine learning algorithms demonstrated reasonable performance for some parameters through testing, with Hs performing best overall followed closely by Tz; Dp was the most challenging to predict and performed relatively the poorest. The creation of such "virtual" wave monitoring stations could be used to hindcast wave conditions, fill observation gaps or extend data beyond that collected by the physical instrument.

*Keywords: Machine Learning; Wave Monitoring; Neural Network*

## INTRODUCTION

Recording and understanding wave conditions is important for many ocean and coastal activities, from recreational fishing and boating, to shipping, disaster management and renewable energy. Long term wave monitoring is also important for understanding the wave climate which can help inform coastal engineering design and beach management. Wave monitoring is a costly and time-consuming exercise and despite best efforts can suffer from data loss due to instrument malfunction or damage. In some cases, near real-time wave condition data loss can be of immediate concern to the data user, depending on the application, for example safe ship transit (Barnes et al. (2015)). Often, wave monitoring sites provide in-situ data used for high-resolution wave model verification or data assimilation (e.g. WAVEWATCH Tolman and Group (2014), SWAN Booij et al. (1999)). Such models can supplement wave monitoring networks, forecast conditions, and provide estimates of conditions for a large spatial area, but they require observations for calibration and validation. Short-term wave monitoring deployments are also common, they can range in time from a few days to years depending on the requirements of the user. Some applications may include coastal engineering works, scientific investigations, dredging or ship-to-ship transfer. Short-term deployments can be necessary if there are no long-term monitoring sites nearby, or in nearshore areas where islands and reefs affect wave conditions.

The aim of this work is to investigate whether machine learning (ML) approaches can be used to help fill gaps, extend datasets, and create virtual monitoring sites to continue estimating wave conditions after physical monitoring has ceased. A merit of ML approaches is, unlike physics-based approaches, they do not necessarily need other datasets such as bathymetry to develop them. The effective application of such ML approaches could also help inform future monitoring site positioning as discussed in Londhe and Panchang (2007). The use of ML for estimating wave conditions has been the subject of some research, particularly in recent years, as it has grown in popularity. Several papers have focused on reconstruction or correlation using ML algorithms, with reasonable results (Abhigna et al. (2018), Berbić et al. (2017)). This work aims to build on that research and establish the suitability of such techniques in the context of nearshore applications and for developing virtual wave monitoring sites. In general, the application of ML techniques in this field has been focused on providing forecasts, particularly for the growing renewable energy sector (Hatalis et al. (2014), Cornejo-Bueno et al. (2016)). Increasingly more complex ML approaches are being applied as different techniques are leveraged to improve performance (Salah et al. (2016)) or facilitate more widescale applicability (Pirhooshyaran and Snyder (2020)).

The specific objective of this work is to establish whether offshore long-term monitoring sites can be used as inputs into ML models to provide accurate estimates of wave conditions at nearshore monitoring sites (virtual wave monitoring). An application of these techniques is also presented by creating a virtual wave monitoring site for a short-term nearshore dataset, extending it beyond the deployment of the physical wave monitoring device.

## STUDY AREA & DATA

The study area is located on the South East Queensland Coast, Australia (Figure 1). The region has several wave monitoring sites both long and short-term, operated by state governments. All the sites are

<sup>1</sup>School of Engineering and Built Environment, Griffith University, Australia

<sup>2</sup>Griffith Centre for Coastal Management, Griffith University, Australia

situated on the continental shelf, several in shallow waters near the coast (< 25m depth) and two in deeper water (> 50m), the approximate depth of long-term monitoring sites is outlined in Table 1.

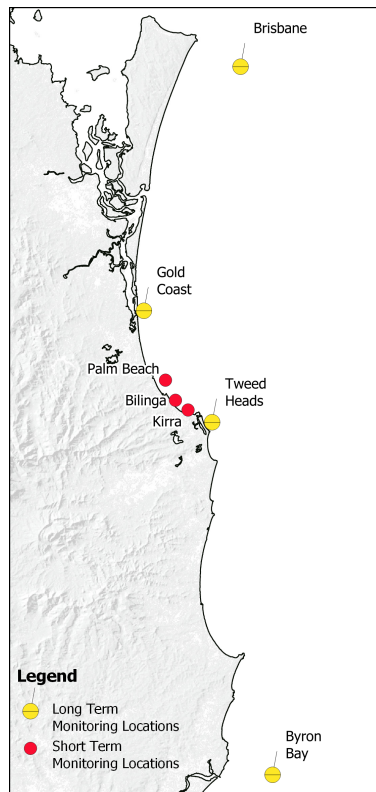


Figure 1: Map depicting wave monitoring locations in the study area with available wave data.

Table 1: Long-term monitoring sites approximate depth (short-term monitoring site depth is unknown).

Monitoring Site	Depth (m)
Brisbane	80
Byron Bay	62
Tweed Heads	25
Gold Coast	16

The wave monitoring data includes the following wave parameters: zero up-crossing significant wave height ( $H_s$ ), mean wave period ( $T_z$ ), maximum wave height ( $H_{max}$ ), peak wave period ( $T_p$ ) and spectrally derived peak wave direction ( $D_p$ ). Each of the wave monitoring sites collects data using a directional wave rider buoy developed by Datawell in the Netherlands (DES (2018)). Although some sites may use a slightly different measurement technology (i.e. GPS or accelerometer), differences outside of gross errors are typically minimal (Andrews and Peach (2019)). The data used for analysis was between 2000 and 2019. Quality control was undertaken to remove erroneous data and consisted of three key elements: automated identification of spikes, inter-site comparison, and visual inspection. Spikes were identified using a threshold, above which values are likely to be erroneous. This threshold was established by multiplying the standard deviation for a given month of  $H_s$  parameter data by five, as used by the Queensland Government wave monitoring program (Andrews and Peach (2019)). Comparisons between monitoring sites were undertaken both for quality control and to analyse the wave climate (discussed in the next section). After quality control, the final dataset used for ML was a coincident time series of data at 1-hour frequency interval between all the monitoring sites. Approximately 64% of the possible data (assuming no instrument outages) over a period of 17 years was usable for training the ML models and is displayed in Figure 2. This period was

selected since all wave monitoring data at the chosen sites recorded wave direction, and it is a desirable parameter. In addition, wave parameters for the time period selected are all recorded at approximately a 30 minute or 1-hour frequency, whereas some of the older historical data has a variable frequency thus making comparison more challenging. For consistency a 1-hour frequency interval was used to develop the ML models.

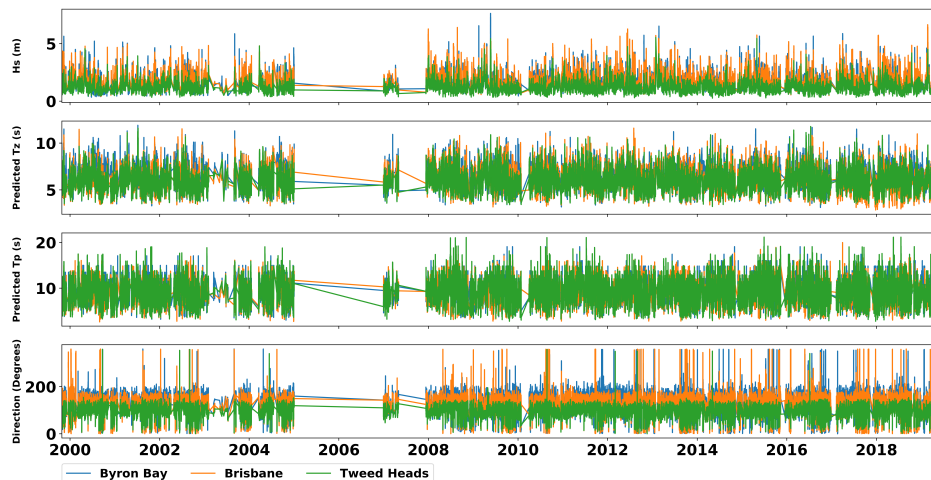


Figure 2: Timeseries of wave parameters used for training, after quality control.

## METHODOLOGY

The methodology outlined below is broken down into several components in the order they were conducted. Put simply, the objective is to 'map' wave parameters from various source locations to a parameter at a single target location, an example configuration is depicted in Figure 3. Notice that parameters from the two source locations (in this case the monitoring locations of Byron Bay and Brisbane) are the input and a single parameter (Hs) as Tweed Heads is the output. An initial investigation demonstrated that predicting a single parameter performs better, which is corroborated in the literature (Berbić et al. (2017), Rao et al. (2013)).

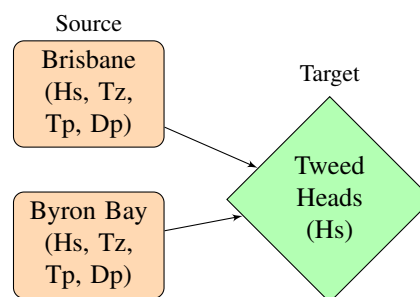


Figure 3: Diagram and data flow from Source locations to Target locations, an example for Tweeds Heads, using the Brisbane and Byron Bay monitoring sites.

## Wave Climate Analysis

Wave climate analysis was undertaken for Hs, Tz and Dp wave parameters, to establish the appropriate machine learning techniques and help with data quality control. Specifically, this analysis helps understand some of the key differences between the various parameters at different locations, informing which ML techniques might be suitable. Figure 4 shows the normalised distributions for three wave sites; two offshore

(Brisbane and Byron Bay) and one nearshore (Tweed Heads). As we might expect the distributions for the offshore monitoring sites are more similar than the nearshore due to the effect of the changing bathymetry on waves as they approach the coast. The Tweed Heads monitoring site is in shallower water and therefore waves measured at that monitoring location are subject to transformations due to varying depth (e.g. shoaling and refraction). This location also has an island and reef system to the south, which provides some sheltering from waves arriving from that direction. The  $T_z$  parameter is more similar across all the sites, this is expected as wave period is theoretically constant from offshore to nearshore, however some differences occur due to each site's position (as each could feasibly be impacted by slightly different, localised forcing conditions). Wave direction is perhaps one of the most important parameters, especially when considering nearshore locations in shallow water.

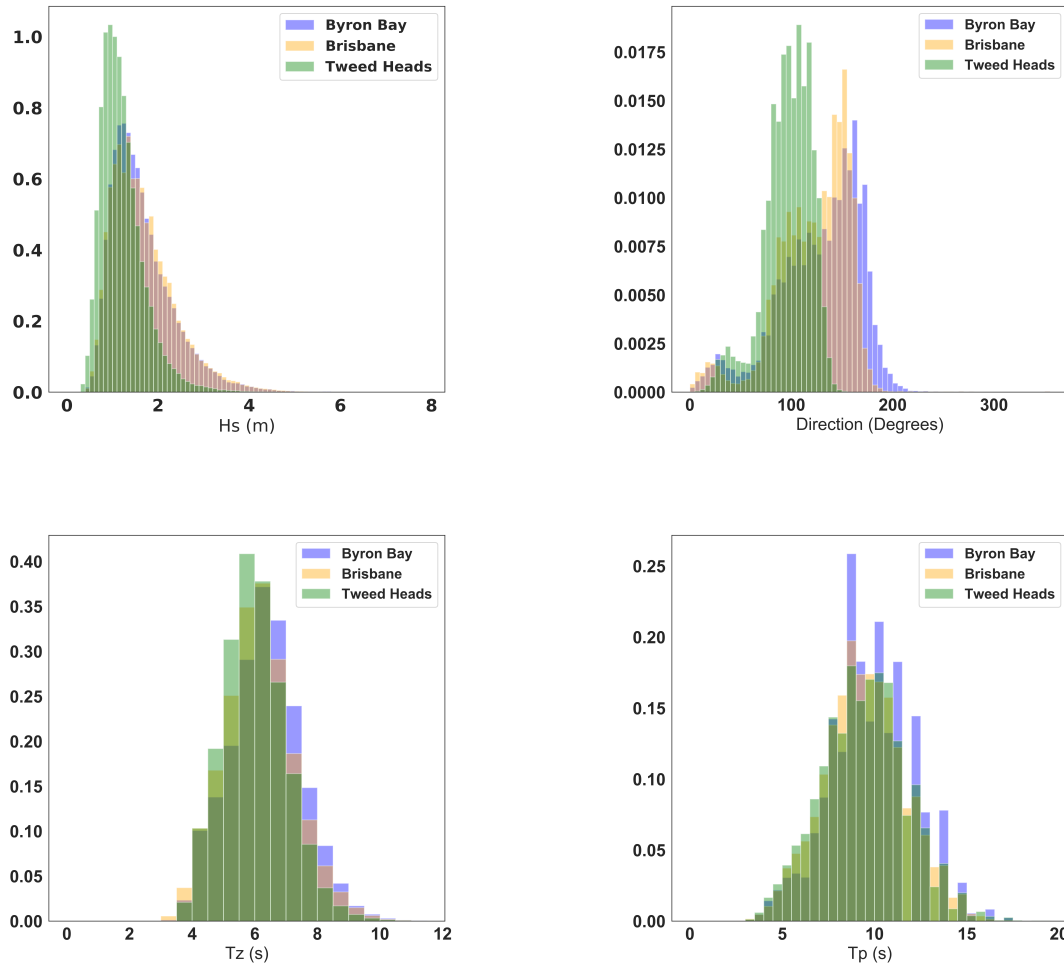


Figure 4: Wave parameter normalized distributions

### Feature Engineering

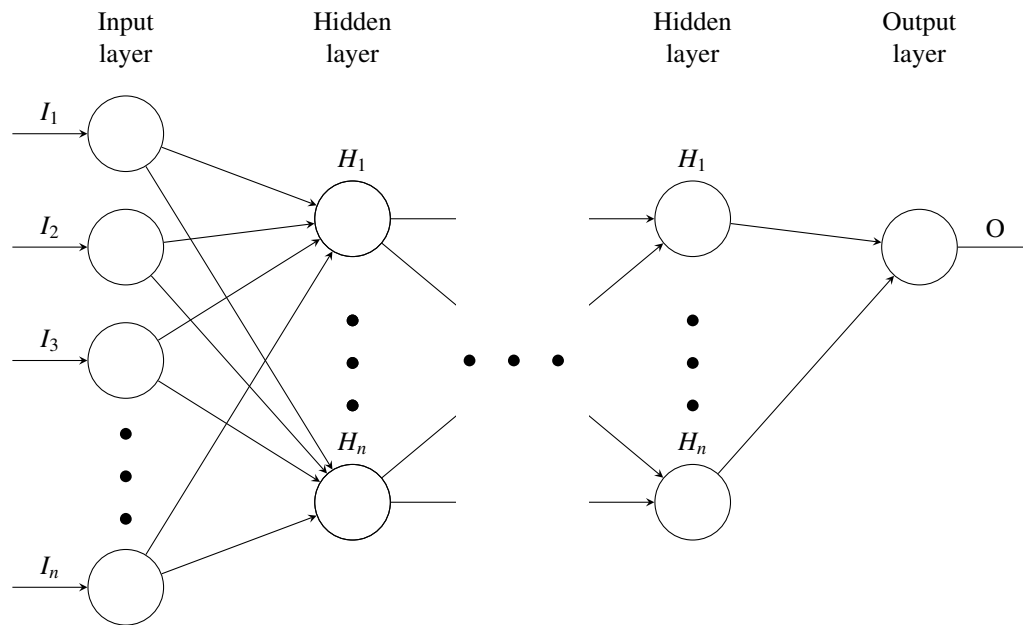
One of the first steps in choosing the type of ML approach to consider is the type of data used when training a ML model. Often data will need to be transformed or transposed to a format suitable for ML and to promote a good result, this process is referred to as Feature Engineering. In this case our objective is to create numerical predictions, therefore regression ML algorithms were selected. Wave parameter values for training typically need to be standardised or normalised, to allow comparisons of the differing parameters which have different units (e.g. metres to seconds), and doing so can improve gradient descent performance



and therefore speed of convergence (Pedregosa et al. (2011)). There are also some specific considerations for directional parameters, for example wave direction can often be an issue for data driven techniques due to its circular nature (north is both 0 and 360 degrees). It was therefore necessary to convert the angles into vectors, making wave direction two parameters (cosine(Dp) and sine(Dp)).

#### Machine Learning Model Selection and Training

Two ML models were selected based on a review of the available literature, with consideration of the available data and wave climate; a neural network (Deo et al. (2001), Londhe and Panchang (2007), Berbić et al. (2017)) and decision trees (James et al. (2017), Pirhooshyaran and Snyder (2020)). Both approaches have both been applied with some success, in this case using a regression. The Multilayer Perceptron (MP) neural network was applied using the scikit-learn python library (Pedregosa et al. (2011)), in the configuration in Figure 5. It was necessary to use more than one hidden layer as this is required for non-linear problems, a final configuration was selected through iterative testing of a 3-year subset of the data. The best performing configuration was 5 hidden layers each containing 50 weights ( $H_n$  in Figure 5), had the best overall skill metrics as discussed in detail in the results section.



**Figure 5: Neural Network Configuration.** Where  $I_n$  represent the input parameters from the source monitoring sites,  $H_n$  the hidden weights and  $O$  the single output parameter at the target site.

The Random Forest (RF) regression decision tree approach was also applied using the scikit-learn Python library (Pedregosa et al. (2011)). Model parameters were chosen iteratively by testing a 3-year subset of the data, this helped to quickly identify the best model configurations before undertaking training on the full dataset. For the RF approach the number of estimators was established at 200 per validation iteration. Typically, more estimators produces a better result, but there is often a point of diminishing returns where the model size is too large making training times prohibitively long.

#### Cross-Validation

The entire available timeseries (where all three monitoring sites had overlapping data) was split, and an independent timeseries (excluded from testing and training) of approximately 6 months was kept aside for assessing model performance. Time series split cross-validation was used for model training, this approach does not treat the data as identically distributed as other cross-validation like Kfolds approaches (Pedregosa et al. (2011)). Time series split cross-validation trains the model for a discrete period of data (e.g. 1 month at a time) and then iterates through the data by a defined period until the dataset is exhausted. Through testing, 3 months was chosen as the cross-validation time window; this was selected as a balance between the minimum number of data points required by the model to train, and consideration of seasonal variability.

## RESULTS AND DISCUSSION

When assessing model performance, an important consideration is model overfitting, this can often lead to ML models that predict closer to the mean of the training data than intended, affecting performance. To help assess this, storm conditions were identified in the independent data used for assessment. They provide a good test of performance as they often have values that deviate from the mean. Storm conditions were identified using a threshold defined by the monitoring site operator (Queensland Government) for the target wave monitoring site (e.g. records where  $H_s$  is greater than 2.5 meters at Tweed Heads, DES (2017)). This threshold is defined by an assessment of historical wave conditions at the monitoring location. Assessing performance during storms also helps assess the suitability of such ML techniques to predict those conditions; some data users are also particularly interested in storm conditions.

Several performance metrics were selected to assess model performance, these are based on those commonly used in the literature for assessing numerical model performance (Alexandre et al. (2015), Saulter (2012), James et al. (2017)).

$$BIAS = \frac{1}{N} \sum_{i=1}^N (O_i - S_i) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - S_i)^2} \quad (2)$$

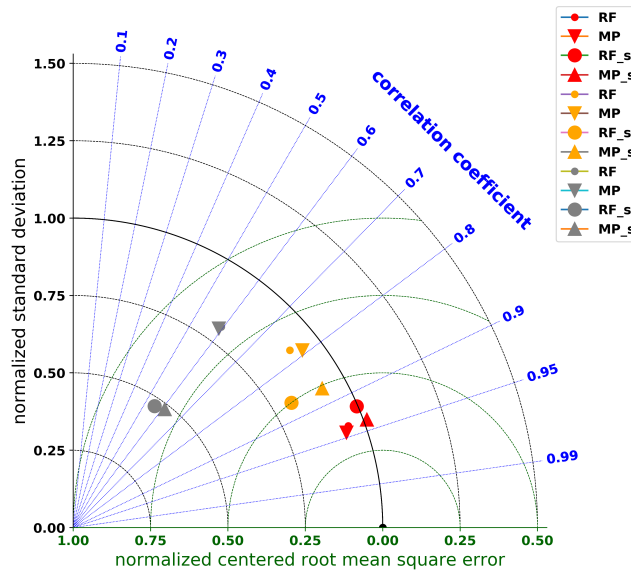
**Where:**

$O$  = Observed

$S$  = Simulated

### Overall Performance Long Term Wave Site

Overall performance is important in the context of developing a virtual wave monitoring site for multiple parameters, as this can provide an indication of confidence in a particular wave parameter in reference to another. In other words, you might have more confidence in the prediction of one parameter over another. The normalized Taylor diagram in Figure 6 displays a comparison of overall performance. Both ML modelling algorithms performed comparably, with  $H_s$  predictions outperforming the others. Peak wave direction is noticeably the worst performing parameter.



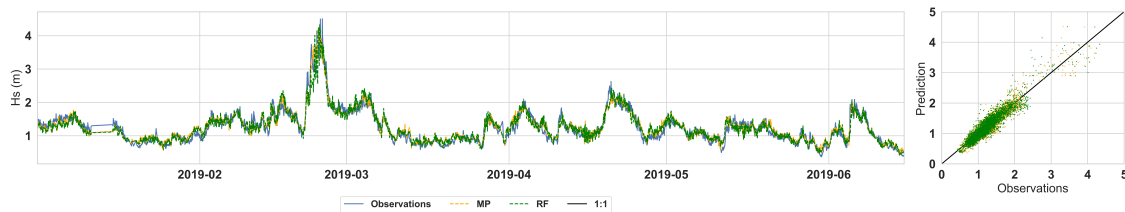
**Figure 6: Normalized Taylor diagram comparing the performance of different parameters, storm conditions are denoted with an s. Red markers denote  $H_s$ , Orange  $T_z$ , and Grey  $D_p$ . The RF models are denoted by circles and MP with triangles.**

**Table 2: Machine Learning Model Performance for Tweed Heads.**

Model	Hs (m)	Hs (m), Storm	Tz (s)	Tz (s), Storm	Dp (°)	Dp (°), storm
MP (BIAS)	0.02	-0.05	-0.06	0.0	-0.21	-1.38
RF (BIAS)	0.01	-0.08	-0.06	-0.03	-0.13	-0.49
MP (RMSE)	0.16	0.26	0.49	0.46	14.77	12.42
RF (RMSE)	0.17	0.3	0.51	0.47	14.75	12.93

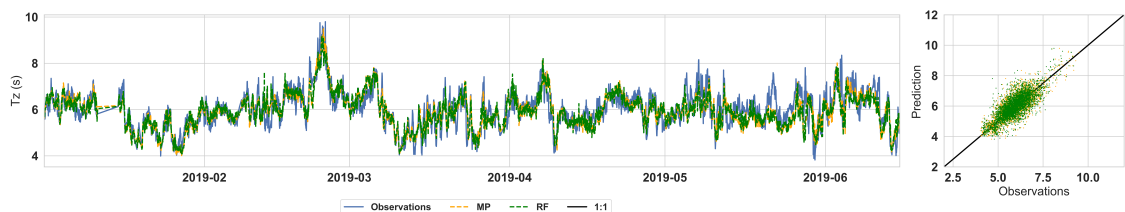
Hs is often one of the better performing parameters from similar research (Londhe and Panchang (2007)), which was also present in this study. The overall model performance is good, with the MP approach slightly outperforming the RF as shown in Table 2. The overall performance statistics (Table 2) also include storm conditions observed in the month of February associated with the passage of tropical cyclone (TC) Oma.

Model performance for storm conditions during TC Oma was slightly reduced, both MP and RF demonstrated similar performance. One consideration worth further exploration is the instability of observations for wave height, in other words the noise in the wave statistic calculated from the instrument and can be observed in late February in Figure 7. Performance may yet improve further if a 3 hourly average was used instead of 1 hourly observations.



**Figure 7: (left) A timeseries plot of the observed Hs, MP and RF for Tweed Heads. (right) A scatter plot depicting a comparison between each ML model and the observed Hs for Tweed Heads.**

Performance of the two ML approaches is again similar for mean wave period with comparable RF and MP. Consideration should be given to the negative bias present in overall results, which may suggest some slight overfitting, but this theory is counteracted somewhat by improved performance during storm conditions. Notably there is a very slight improvement in performance for storm conditions as outlined in Table 2 and can be observed at the end of February in Figure 8. This slight improvement may be due to waves caused by the storm conditions dominating the measured wave periods that make up the hourly average, but it is difficult to infer too much as the difference to the overall performance is small.

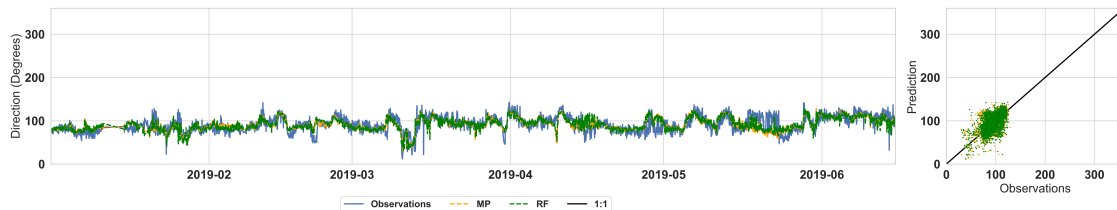


**Figure 8: (left) A timeseries plot of the observed Tz, MP and RF time series for Tweed Heads. (right) A scatter plot depicting a comparison between each ML model and the observed at Tweed Heads.**

Wave direction is a challenging parameter to predict, mean wave direction would have been a preferred parameter as it is notably more 'stable' (less prone to sudden shifts by its nature). Peak wave period is based on the direction of the peak wave energy, which can fluctuate, especially in multimodal sea conditions. However, mean wave direction was not available for all wave monitoring sites as a parameter and therefore

could not be used. In addition, as previously mentioned, it was necessary to encode the direction as two parameters, and predicting two output labels in ML can be more challenging than a single parameter.

Model performance for wave direction was perhaps not as good as for some of the other parameters, but is still reasonably close to the observations, with an RMSE of less than 15 degrees. For reference the quoted resolution of directional wave monitoring by the manufacturer is between 0.1 and 1.5 degrees depending on the instrument used (Datawell (2012a), Datawell (2012b), Datawell (2012c)). Performance suffered most during periods when the peak direction fluctuated over a short period of time, in particular this can be observed during May 2019 in Figure 9.

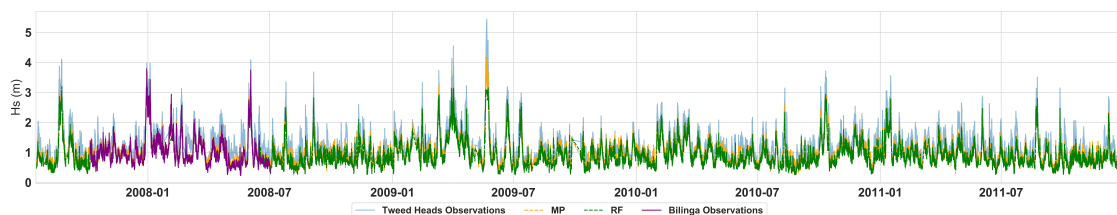


**Figure 9:** (left) A timeseries plot of the observed Dp, MP and RF time series for Tweed Heads. (right) A scatter plot depicting a comparison between each ML model and the observed at Tweed Heads.

Performance during storm conditions is markedly improved, this is expected as wave energy from storms typically dominate the directional wave spectrum. However, of particular interest in the assessment of storm conditions, is an apparent leading of the ML model results to the observations (a change in direction is seen in the prediction prior to being seen in the observations). This is potentially a result of time not being explicitly included in either of the ML approaches used here and where it has been included in cross validation, it was only during a 3-month time window. Therefore, it is likely not granular enough to accommodate sudden shifts in peak wave direction.

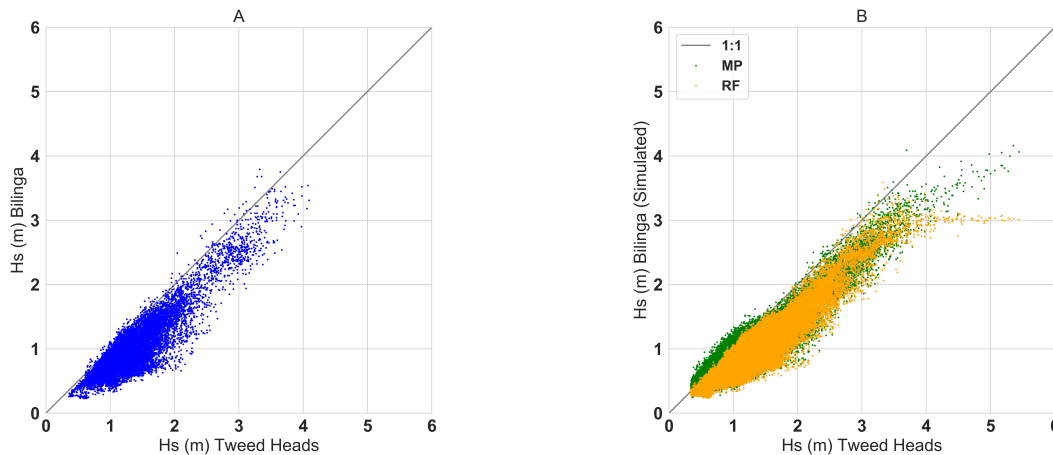
#### An application of ML to extend a short term deployment

For comparison ML models were also developed for a short-term monitoring site (Bilinga in Figure 1). The methodology is similar to that previously outlined, except for the time cross-validation period, which was reduced to 3 days due to the limited availability of data. Also due to the limited data, the entire timeseries was used for the training dataset with the nearby Tweed Heads monitoring site used for comparison and validation. The ML models were used to extend the Bilinga wave site data for 3 years, wave heights for the Tweed Heads monitoring site are generally slightly larger than those for Bilinga (Figure 10), despite this it offers a reasonable proxy for wave conditions at Bilinga, due to its close geographic proximity.



**Figure 10:** A time series of the virtual wave monitoring results for Bilinga, extending the Hs parameter and compared with Tweed Heads Hs for the same period.

The performance of the ML models in Figure 11 below show that generally the MP model outperformed the RF model, with a wider margin than demonstrated in the previous comparison with the long-term monitoring site at Tweed Heads. Clear differences can be observed where wave heights were greater than 3 metres between the Tweed Heads observations and RF model. This is a symptom of one of the limitations of the RF regression method, it is unable to predict values outside of the training data range, in this case it is limited to wave heights just over 3 metres. This phenomenon was not observed in the model constructed from the long-term wave monitoring site, likely because a much larger range of possible values was available, but it is a significant consideration when constructing models for short term sites.



**Figure 11: (A) A scatter plot comparing overlapping observations of Hs between Bilinga and Tweed Heads. (B) A comparison of simulated Hs at Bilinga using the MP and RF approach compared with observations at Tweed Heads (Note: The data used in A is a different time to B).**

## CONCLUSIONS

Although ML models typically require a reasonable size dataset to develop, they don't necessarily require other datasets, such as bathymetry or wind and current information required by the physics-based models, like SWAN or WAVEWATCH. They are also considerably more computationally efficient, taking seconds to produce results once trained as opposed to the equivalent high-resolution physics-based model which could take minutes or hours with the same computational power. However, there are specific requirements to this type of approach such as; several wave monitoring sites in reasonably close proximity (in a similar wave climate), with coincident dataset lengths of several months as a minimum, and ideally with a reasonable estimate of the possible range of conditions.

The application of ML techniques to develop virtual wave monitoring sites shows promise for this region and for the translation of waves from offshore monitoring sites to nearshore ones. However, whether the performance of these ML approaches meets a user's need will depend on the application of the user. These types of ML approaches do show promise in being able to considerably extend the capability of existing wave monitoring networks; by developing virtual wave monitoring sites through careful application of short-term deployments. This could in turn prove useful as input for high-resolution physics-based models that may be developed at a later stage. Future work could include using each of the wave monitoring sites in Figure 1 to develop a network of virtual monitoring sites. Due to the computational efficiency of the ML approaches applied, these virtual wave monitoring sites could produce results in near real-time.

## ACKNOWLEDGEMENTS

The authors express their gratitude to: the Queensland Government Hydraulics Laboratory, Department of Environment and Science Wave Monitoring Program and the Manley Hydraulics Laboratory, NSW Coastal Data Network Program managed by the NSW Department of Planning, Industry and Environment (CCSD). They provided wave monitoring data without which this work would not have been possible.

## References

P. Abhigna, S. Jerritta, R. Srinivasan, and V. Rajendran. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the Moored Buoys in Bay Of Bengal. In *Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017*, volume 2018-Janua, pages 1856–1860. Institute of Electrical and Electronics Engineers Inc., feb 2018. ISBN 9781509038008. doi: 10.1109/ICCSP.2017.8286717.

E. Alexandre, L. Cuadra, J. C. Nieto-Borge, G. Candil-García, M. Del Pino, and S. Salcedo-Sanz. A

- hybrid genetic algorithm-extreme learning machine approach for accurate significant wave height reconstruction. *Ocean Modelling*, 92:115–123, 2015. doi: 10.1016/j.oceomod.2015.06.010. URL <http://dx.doi.org/10.1016/j.oceomod.2015.06.010>.
- E. Andrews and L. Peach. Wave Monitoring Equipment Comparison: A comparison between in-situ wave measurement equipment. Technical report, Queensland Government, Department of Environment and Science, Brisbane, Queensland, 2019. URL [https://www.publications.qld.gov.au/dataset/a719a64a-bbaf-4c02-9f33-39d2ab737c90/resource/4b9a79fd-7245-42cb-b60e-c83f84d82011/fs\\_{\\_}download/wave-monitoring-equipment-comparison.pdf](https://www.publications.qld.gov.au/dataset/a719a64a-bbaf-4c02-9f33-39d2ab737c90/resource/4b9a79fd-7245-42cb-b60e-c83f84d82011/fs_{_}download/wave-monitoring-equipment-comparison.pdf).
- M. Barnes, I. Teakle, P. Wood, and C. Voisey. Assessment of capital works options to mitigate shoaling at Mooloolaba Harbour Entrance. Technical report, nov 2015. URL <http://eisdocs.dsdip.qld.gov.au/SunshineCoastAirportExpansion/EIS/VolumeBchapters/ChapterB4-Coastalprocesses18Sep14.pdf>.
- J. Berbić, E. Ocvirk, D. Carević, and G. Lončar. Application of neural networks and support vector machine for significant wave height prediction. *Oceanologia*, 59(3):331–349, 2017. ISSN 23007370. doi: 10.1016/j.oceano.2017.03.007.
- N. Booij, R. C. Ris, and L. H. Holthuijsen. A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research*, 104(C4):7649–7666, 1999. ISSN 0148-0227. doi: 10.1029/98JC02622. URL <http://doi.wiley.com/10.1029/98JC02622>.
- L. Cornejo-Bueno, J. C. Nieto-Borge, P. García-Díaz, G. Rodríguez, and S. Salcedo-Sanz. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm - Extreme Learning Machine approach. *Renewable Energy*, 97:380–389, nov 2016. ISSN 18790682. doi: 10.1016/j.renene.2016.05.094.
- Datawell. Directional Waverider MkIII Datawell-Oceanographic Instruments The Directional Waverider DWR-MkIII: Over three years of continuous operation. Technical report, Datawell, Haarlem, Netherlands, 2012a. URL [www.datawell.nl](http://www.datawell.nl).
- Datawell. Directional Waverider GPS Datawell-Oceanographic Instruments Measuring waves with GPS. Technical report, Datawell, Haarlem, Netherlands, 2012b. URL [www.datawell.nl](http://www.datawell.nl).
- Datawell. DWR4 with ACM Datawell-Oceanographic Instruments. Technical report, Datawell, Haarlem, Netherlands, 2012c. URL [www.datawell.nl](http://www.datawell.nl).
- M. C. Deo, A. Jha, A. S. Chaphekar, and K. Ravikant. Neural networks for wave forecasting. *Ocean Engineering*, 28(7):889–898, 2001. ISSN 00298018. doi: 10.1016/S0029-8018(00)00027-5.
- DES. Wave Monitoring Annual Summary 2016 - 2017. Technical Report November 2016, Department of Environment and Science, Queensland Government, Brisbane, Queensland, 2017. URL <https://www.publications.qld.gov.au/dataset/a719a64a-bbaf-4c02-9f33-39d2ab737c90/resource/ee424256-7a47-482c-9cd5-795663db65f4/download/annual-wave-report-2016-17-final.pdf>.
- DES. Queensland wave climate, wave monitoring annual summary, November 2016-October 2017. Technical report, Queensland Government, Department of Environment and Science, 2018.
- K. Hatalis, P. Pradhan, S. Kishore, R. S. Blum, and A. J. Lamadrid. Multi-step forecasting of wave power using a nonlinear recurrent neural network. In *IEEE Power and Energy Society General Meeting*, volume 2014-Octob. IEEE Computer Society, oct 2014. doi: 10.1109/PESGM.2014.6939370.
- S. C. James, Y. Zhang, and F. O’Donncha. A Machine Learning Framework to Forecast Wave Conditions. *Coastal Engineering*, 2017.
- S. N. Londhe and V. Panchang. Correlation of wave data from buoy networks. 2007. doi: 10.1016/j.ecss.2007.05.003. URL [www.elsevier.com/locate/ecss](http://www.elsevier.com/locate/ecss).

- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(1):1–5, 2011. ISSN 0717-6163. doi: 10.1007/s13398-014-0173-7.2. URL <https://scikit-learn.org/>.
- M. Pirhooshyaran and L. V. Snyder. Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks. *Ocean Engineering*, 207, 2020. ISSN 00298018. doi: 10.1016/j.oceaneng.2020.107424. URL <https://doi.org/10.1016/j.oceaneng.2020.107424>.
- A. D. Rao, M. Sinha, and S. Basu. Bay of Bengal wave forecast based on genetic algorithm: A comparison of univariate and multivariate approaches. *Applied Mathematical Modelling*, 37(6):4232–4244, mar 2013. ISSN 0307904X. doi: 10.1016/j.apm.2012.09.001.
- P. Salah, A. Reisi-Dehkordi, and B. Kamranzad. A hybrid approach to estimate the nearshore wave characteristics in the Persian Gulf. *Applied Ocean Research*, 57:1–7, 2016. doi: 10.1016/j.apor.2016.02.005. URL <http://dx.doi.org/10.1016/j.apor.2016.02.005>.
- A. Saulter. Current and future verification of operational wave models. (June):25–27, 2012. URL [www.globwave.org](http://www.globwave.org).
- H. L. Tolman and W. D. Group. User manual and system documentation of WAVEWATCH III version 4.18. *NOAA/NWS/NCEP/MMAB Technical Note*, (333):311, 2014.