Proceedings of the ASME 2024 43rd International
Conference on Ocean, Offshore and Arctic Engineering
OMAE2024
June 9-14, 2024, Singapore, Singapore

OMAE2024-127479

ENHANCING OFFSHORE WIND FARM MET-OCEAN DATA ACCESSIBILITY: A MACHINE LEARNING APPROACH WITH SATELLITE-DERIVED WAVE MEASUREMENTS IN THE CELTIC SEA

Sophie Whistler[1,*], Ian G.C. Ashton[1], Ajit C. Pillai[1] and Graham D. Quartly [2],

[1]Renewable Energy Group, Faculty of Environment, Science, and Economy, University of Exeter, United Kingdom
[2]Plymouth Marine Laboratory, United Kingdom

ABSTRACT

Marine operations are a significant expense for offshore wind farms, representing up to one third of total project costs. An improved understanding of the variation of met-ocean conditions across a wind farm site offers the potential to reduce weather downtime and associated costs. This work employs a machine learning approach utilising a surrogate wave model trained on the relationship between the wave conditions at discrete measurement locations to wave conditions across the entire model domain. The surrogate model can then be run with real-time data inputs from the discrete measurement locations to provide a spatial dataset for waves, without the high computational power needed to run the physics-based wave model itself. This new method enhances the accessibility of met-ocean data to allow more informed decision making for the installation, operation, and maintenance of offshore wind farms.

The approach has already proven successful with fixed measurement buoys, and work is ongoing to adapt the modelling framework to use satellite-derived wave data as an input. With freely available global coverage, satellite data is a useful complementary data source to wave buoy data. Several Earth Observation satellite missions host radar altimeters that report significant wave height along the satellite's ground track. The first step towards utilising radar altimeter data with the machine learning framework is assessing the impact of using only significant wave height data as measurement inputs. This paper compares the model outputs from running the model with wave height, period, and direction data, and with wave height data only. The results show that running the model with wave height data only produces a small reduction in the accuracy of output wave predictions in coastal areas.

Keywords: wave measurement, satellite data, machine learning

*Corresponding author: sw694@exeter.ac.uk

1. INTRODUCTION

Expanding the use of renewable energy technologies is a crucial component in our race to net-zero carbon emissions and in the global agreement to limit the rising temperature of the planet to combat catastrophic climate change. Electricity produced by offshore wind turbines already contributes to this, meeting energy demands from renewable sources of power, and the industry is gearing up for global growth.

Characterisation of the met-ocean conditions in which the wind turbines operate is crucial information required throughout the life cycle of an offshore wind energy project. As projects are developed in deeper waters, further from shore, and occupying larger areas, understanding the spatial variation of the wave conditions across a site becomes more important.

At present, met-ocean conditions are derived from numerical wave models that represent the physics of ocean wave energy transfer. These models can be on global, regional, or local scales, and use various methods of data assimilation to incorporate in-situ and remotely sensed measurements to produce hindcasts (estimates of past conditions), nowcasts (estimates of current conditions), and forecasts (estimates of future conditions) for waves. The high computational power required to run these models limits the frequency with which the resulting forecasts are available for making decisions in the marine environment.

Chen et al. [1] developed a new machine learning framework for nowcasting and forecasting waves for marine renewable energy applications called MaLCOM: **Ma**chine **L**earning for Low-**C**ost **O**ffshore **M**odelling (hereby referred to as MaLCOM). This modelling framework uses a random forest algorithm to learn the spatial relationships between wave conditions at defined locations within a wave hindcast to wave conditions across the entire domain of the model. In implementation, it can then predict wave conditions across the area using only concurrent data at the discrete input locations. The output predictions were found to be of similar accuracy to traditional nowcasts and forecasts, and when compared to a hindcast model offered significant improvements in accuracy; enabling a low-computational cost nowcast methodology for wave modelling. As the surrogate wave model does not have the high computational demands of the numerical wave model, it can be run on any computer, more frequently, updating based on real-time measured conditions within the domain.

Whilst wave buoys provide continuous data, the single point location limits the geographical spread of measurements available as input to the MaLCOM framework. Mounet et al. [2] explored the advantages of incorporating wave data measured by ships to expand the spatial coverage of input data. Satellite-derived wave data presents another potentially useful and complementary source of measurements to input into the framework and would enable immediate application of the method globally. Satellite data provides consistent coverage and is freely available to the public in many cases. In a review of satellite data for the offshore renewable energy sector by Medina-Lopez et al. [3], the current limited use of satellite-derived measurements in the offshore wind energy industry was identified, along with numerous possibilities to employ this data for much-needed cost reductions in the sector.

The primary method of measuring wave heights from satellites is radar altimetry, a widely applied and verified remote sensing technique celebrating 30 years of research and development this year. Though initially designed to measure sea surface height, the return signal from the radar pulse can be processed to yield wave height, providing a data point approximately every 7 km along the ground track of the satellite. This track is repeated in accordance with the revisit cycle of the satellite, for example, approximately every 10 days for Jason-3 ([4], and every 27 days for Sentinel-3 ([5]). The distance between successive ground tracks of each satellite varies in accordance with the latitude of the ground area in question. At the equator, the ground track separation of these two examples is 315 km for Jason-3 and 104 km for Sentinel-3 ([4], [5]), which translates to ground track separation distances in the case study area of the Southwest UK of around 180 km for Jason-3 and 120 km for Sentinel-3. Evidently, satellite-derived wave measurements offer consistent and regular spatial coverage, but coverage that introduces significant spatial and temporal gaps when investigating wave conditions for a specific site, in particular for marine operations where regular updates at a specific location are desirable.

Furthermore, radar altimeter-derived wave data is principally a measure of significant wave height (H_s). Algorithms to derive a wave period from satellite data have been published (e.g [6–8]), and some newer remote sensing instruments, such as CFOSAT's radar scatterometer SWIM (Surface Waves Investigation and Monitoring) have the ability to capture additional information. However, it is H_s from radar altimeters that is currently the most consistent and widely available satellite-derived wave data set.

It follows that key initial considerations of utilising satellite-derived wave data in a MaLCOM style setup are: how to optimise the benefits from the varying locations and times at which the data reports and; how much impact only receiving H_s will have on the machine learning model training and operation. This paper focuses on the second of those challenges. Using a limited implementation of a MaLCOM framework, it evaluates the impact of using only H_s to train the surrogate by directly comparing an 'all parameter' model, using 4 wave parameters (significant wave height, peak period, mean period, and mean direction) to train the surrogate model (referred to as Model_all) and a single parameter, 'H_s only', model (referred to as Model_single). The outputs from these two models are compared to assess the impact of running the method with only wave height data, towards the long term goal of implementing the MaLCOM framework with radar-altimeter H_s data.

2. METHOD

2.1 Case Study Area

The waters around the county of Cornwall, the westernmost part of the Southwest Peninsula of the UK, are used as the case study area in this paper (1). The total area of the model domain is bounded by 49°N to 51°N and 4°W to 7°W, extending up to around 100 km offshore from the Cornish coastline. The previous work conducted to develop the MaLCOM framework [1] also used this area as a case study. The Northern and Western area of the domain is within the Celtic Sea, a region currently preparing for the commercial development of up to 4.5 GW of floating offshore wind energy capacity as established in The Crown Estate's Offshore Wind Leasing Round 5 [9]. Therefore, the area is attracting significant attention to substantially contribute towards national targets to deliver 5 GW of floating offshore wind energy capacity by 2030 [10].

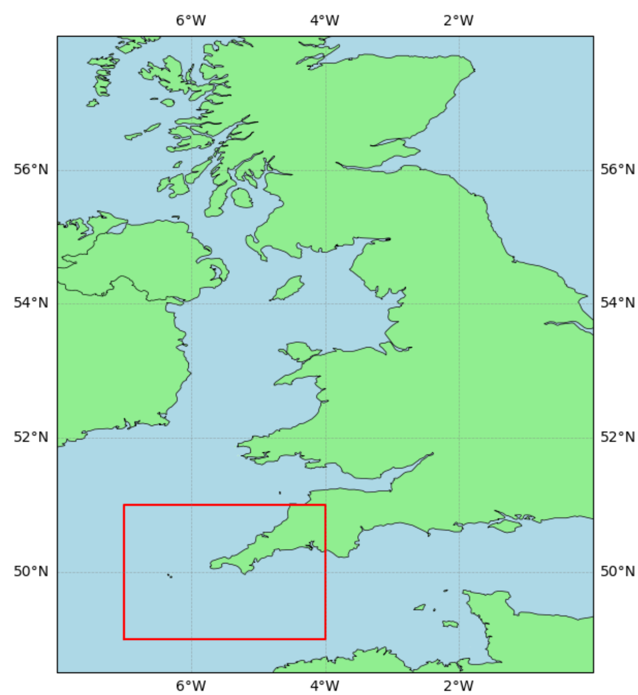

FIGURE 1: MAP OF THE UK WITH MODEL DOMAIN AREA SHOWN IN RED.

2.2 Model Set Up

As outlined in the introduction, this study uses the MaLCOM modelling framework which is fully described by Chen et al. [1]. The approach employs a random forest machine learning algorithm to train the surrogate model. Firstly, the surrogate is trained on a physics-based wave hindcast to learn the spatial relationships between wave conditions at three set locations to wave conditions across the model domain. Secondly, the trained surrogate predicts a spatial nowcast for the domain using time-series data from wave buoys located at the three set locations as input. This model execution process is referred to as 'inference' in machine learning parlance.

In the original MaLCOM method, 4 wave parameters from the hindcast were used to train 4 spatial surrogate models, all of which were then combined to produce the estimates in inference, resulting in a spatial nowcast for all 4 parameters. These parameters were significant wave height (H_s), mean wave period (T_{m02}), peak wave period (T_p) and wave direction (M_{dir}). For the present study, this is termed the 'all parameter model' and referred to as 'Model_all'. To understand the impact of only using the H_s measurements as input into the MaLCOM method, the second case was run that uses only H_s data from the hindcast to train a single H_s only surrogate model. In inference, only the H_s data from the buoys is then used to make the predictions, resulting in a spatial nowcast for H_s only. This arrangement is represented by the flow chart in Figure 3.

2.3 Input Data

The two MaLCOM-approach models created for this paper were trained using the UK Met Office European North West Shelf Wave Physics Reanalysis product, hosted by the Copernicus Marine Service [11]. This offers a long term hindcast, of which 1 year of hindcast data (2018) has been used in this work. Three Channel Coastal Observatory (CCO) wave buoys [12] that are operational within the domain were used to run and test the surrogate models. All three are Datawell Directional Wave Rider Mk III buoys which are located in Penzance, Perranporth, and Looe Bay (Figure 2). Each of these buoys is configured to report integrated wave parameters every 30 minutes. Data from 01 January 2019 to 31 December 2020 were used in inference to run the models, producing a spatial nowcast for the same period.

The MaLCOM methodology for spatial modelling with the inputs and outputs for the two model cases in this study are outlined in Figure 3.

3. RESULTS AND DISCUSSION

3.1 Analysis Conducted

The analysis conducted first compares the outputs from the four parameter model (Model_all) and the single parameter model (Model_single) against one another to assess the spatial performance of the H_s only implementation. Secondly, data from the CCO wave buoy at Porthleven for the concurrent time period was used as an evaluation site to assess the model outputs with data distinct from that used to run the model. Comparing the nowcasts produced from these two cases contributes towards evaluating the limitations of using radar altimeter data, by determining the impact of running a MaLCOM-style model with H_s data only.

3.2 Spatial Comparison

The annual average H_s as predicted by both models present a similar spatial pattern (Figure 4). Both maps indicate the expected

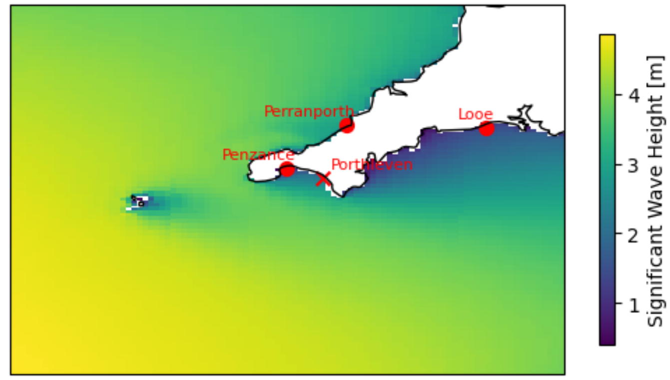

FIGURE 2: MAP OF THE CASE STUDY AREA SHOWING WAVE BUOY LOCATIONS AND AVERAGE SIGNIFICANT WAVE HEIGHT FOR 2018 ACCORDING TO THE WAVE HINDCAST. RED CIRCLES MARK WAVE BUOYS USED TO RUN THE MODELS. THE RED CROSS MARKS THE WAVE BUOY USED TO ASSESS THE MODEL OUTPUTS.

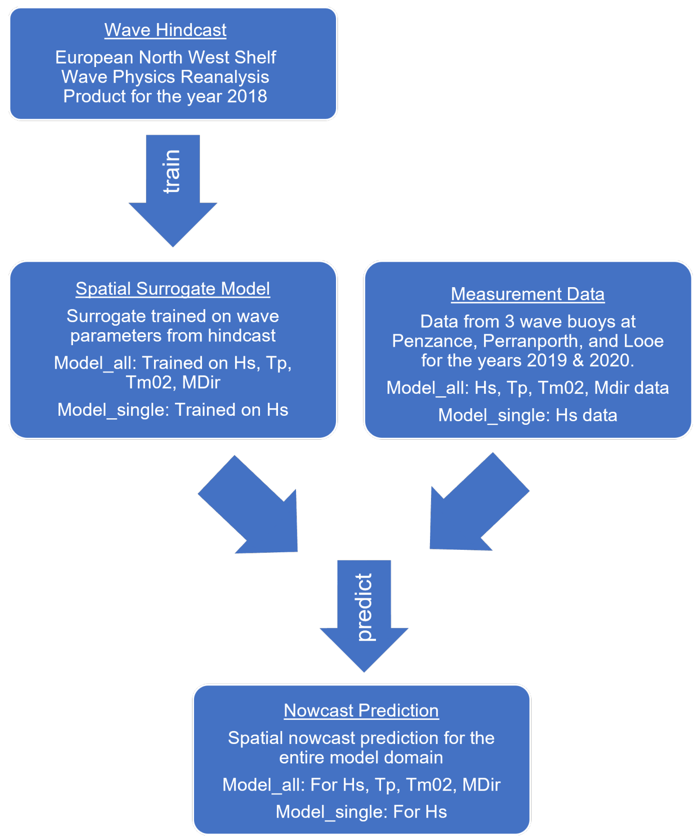

FIGURE 3: FLOWCHART TO SHOW THE MALCOM FRAMEWORK IMPLEMENTATION FOR THE TWO MODEL CASES IN THIS STUDY.

larger waves in the exposed area in the southwest of the domain, with more sheltered areas experiencing smaller average wave height, particularly in the coastal areas on the south coast of the Cornwall peninsula.

Direct comparison of the results in the form of the average value of $\Delta H_s = Model\_all\_H_s - Model\_single\_H_s$ when calculated on a record-by-record basis shows that the differences are

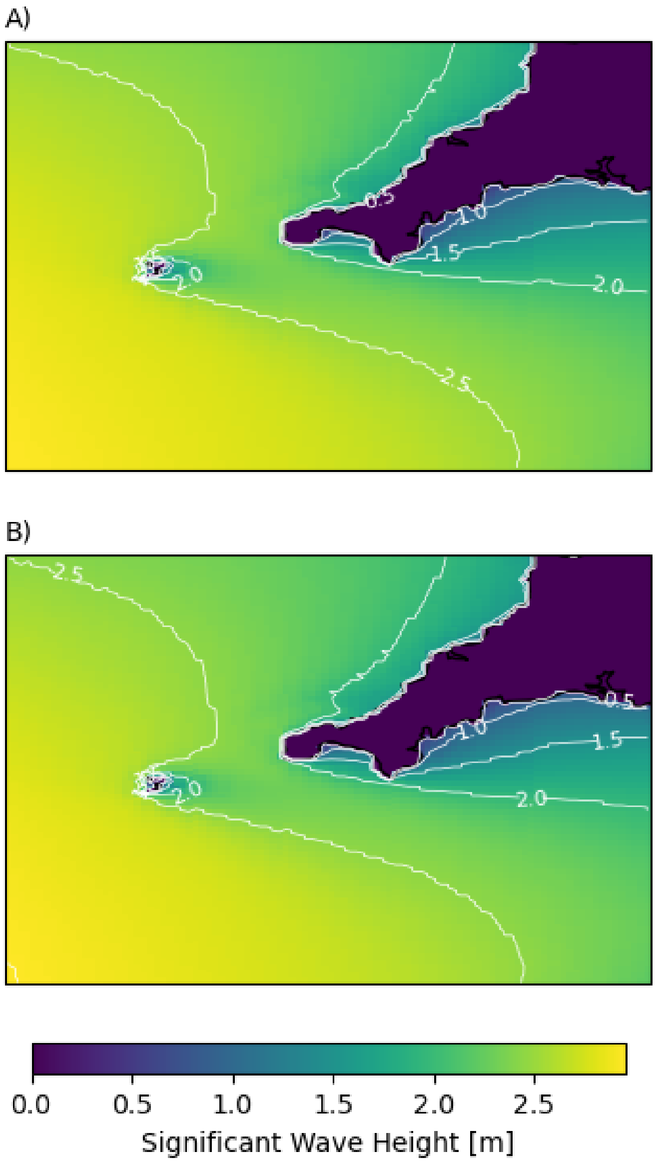

FIGURE 4: TIME-AVERAGED H_S DURING 2019-2020, AS PREDICTED BY THE ALL PARAMETER MODEL (A) AND THE SINGLE PARAMETER (H_S ONLY) MODEL (B)

greatest towards the northeast of the domain and in the slightly sheltered coastal area on the far southwest coast of Cornwall. These differences indicate that Model_single is slightly under-predicting wave heights, with a maximum difference less than 6 cm. The smallest differences were in the more exposed, open areas of sea to the west of the domain, where Model_single slightly over-predicted wave heights compared to Model_all. These differences are shown in Figure 5.

For a relative comparison not affected by the size of the waves, Figure 6 shows the average percentage difference in the results from the two models. This is calculated as the mean of $(Model\_all\_H_s - Model\_single\_H_s)/Model\_all\_H_s$ at each time step. This result retains a broadly similar pattern to Figure 5, although it highlights the largest under-prediction by Model_single occurs in areas close to the southern coast.

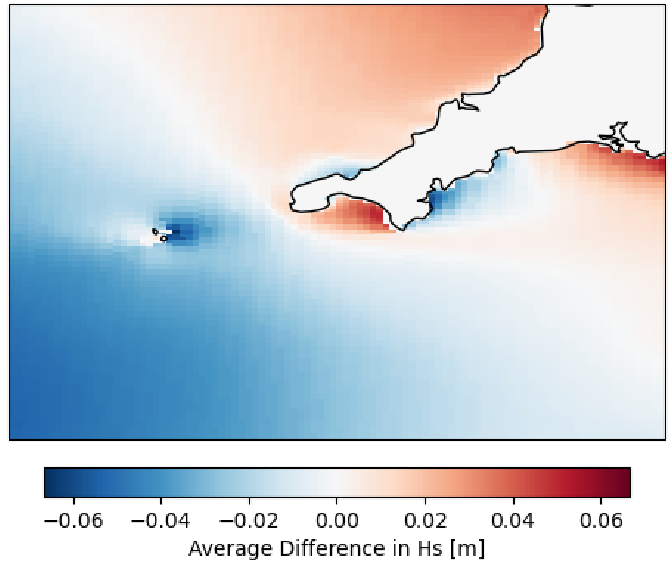

FIGURE 5: MEAN H_S DIFFERENCE FOR THE TWO MODEL CASES, MEAN OF $MODEL\_ALL\_H_S - MODEL\_SINGLE\_H_S$

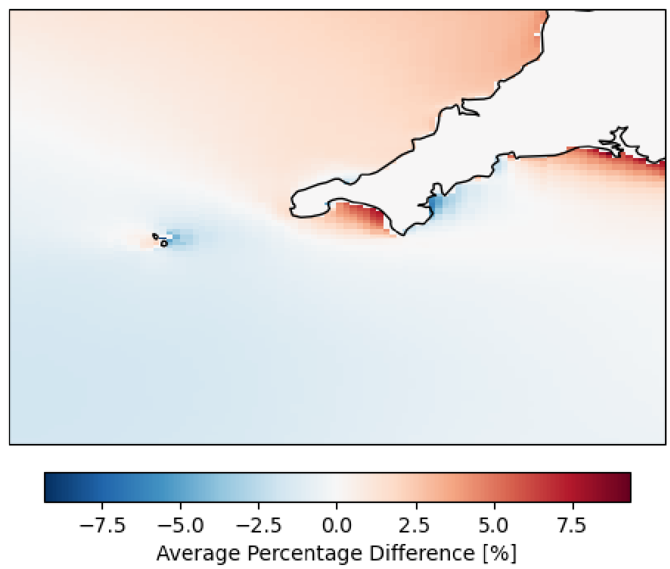

FIGURE 6: AVERAGE PERCENTAGE DIFFERENCE FOR H_S FOR THE TWO MODELS, MEAN OF $(MODEL\_ALL\_H_S - MODEL\_SINGLE\_H_S)/MODEL\_ALL\_H_S$

3.3 Temporal Comparison

The results were compared against in-situ buoy data recorded at Porthleven (Figure 2), a location that was was not included in the training and is not used to run the model. As such, model output at this location is used as an independent dataset to analyse the accuracy of the model, whilst buoy data is treated as ground truth for the purpose of this analysis.

While both models predict the general temporal patterns of H_s values, they both under-predict larger values of H_s when compared to the buoy data (Figure 7). This is evident in the scatter diagram (Figure 8), although it appears that Model_single has a slightly lower under-prediction at higher H_s values than Model_all. Some of the under-prediction is a product of the under-prediction in the hindcast (that was used in model training) at this particular location when compared to the buoy measurements, as seen in Figure 9. Additionally, Figure 10 shows the model outputs over-predicting at small wave heights and under-predicting at large wave heights for this location when evaluated against the hindcast. The combined impact of both under-predictions at larger values of H_s (Figure 9, Figure 10) leads to the under-prediction of the models evident in both Figure 7 and Figure 8.

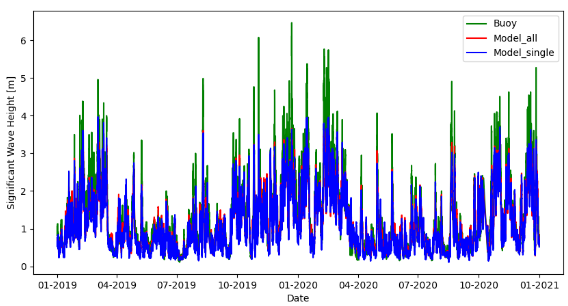

FIGURE 7: TIME SERIES DATA FOR PORTHLEVEN FROM A WAVE BUOY AND THE TWO MODELS

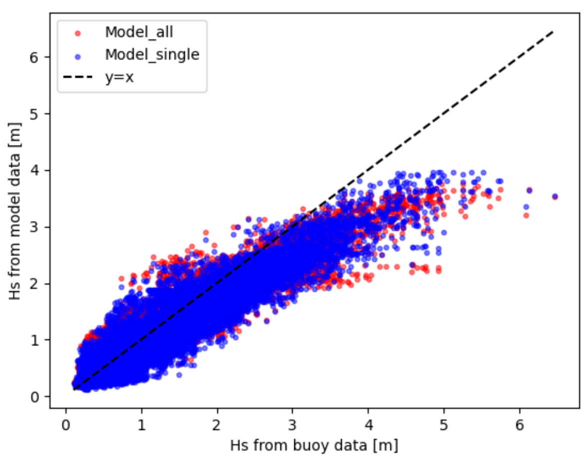

FIGURE 8: H_S SCATTER PLOT TO SHOW MODEL PREDICTIONS AT PORTHLEVEN AGAINST WAVE BUOY OBSERVATIONS

To assess the error between the buoy data and the model results, three statistics were calculated: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). RMSE and MAE are scale-dependent, whilst R^2 is scale-independent. The three error metrics were calculated by Equation (1), Equation (2), and Equation (3), where n is the number of predicted values, y_i the buoy observation representing the 'true' value, $\hat{y}$ the model prediction, and $\bar{y}$ the mean of the

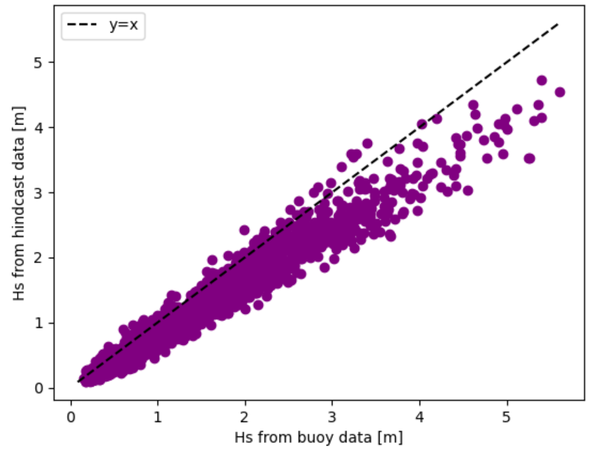

FIGURE 9: H_S SCATTER PLOT TO SHOW HINDCAST VALUES AT PORTHLEVEN AGAINST WAVE BUOY OBSERVATIONS FOR H_S

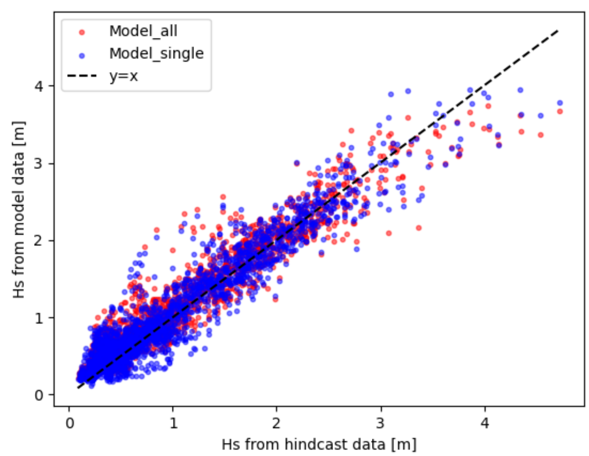

FIGURE 10: H_S SCATTER PLOT TO SHOW H_S VALUES FOR PORTHLEVEN FROM MODEL PREDICTIONS AGAINST HINDCAST VALUES

buoy observation ([13]. Table 1 shows the comparison of these metrics.

Both RMSE and MAE are increased in the H_s only model and the R^2 is also decreased. This indicates greater variability and a less accurate prediction when using H_s only. The differences observed are relatively small compared to the overall accuracy of the surrogate model predictions, which is in agreement with the results shown in Figure 7.

	Porthleven to Model_all	Porthleven to Model_single
RMSE	0.3184	0.3434
MAE	0.2164	0.2488
R^2	0.8616	0.8427

TABLE 1: ERROR STATISTICS COMPARING THE H_s PREDICTIONS FROM MODEL_ALL AND MODEL_SINGLE AGAINST PORTHLEVEN WAVE BUOY OBSERVATIONS.

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

$$\text{MAE} = MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

4. CONCLUSION

This work has used a case study in the Celtic Sea to demonstrate that running a MaLCOM style surrogate model, as initially described in Chen et al. [1] that uses only significant wave height (H_s) produces a working model that can spatially predict significant wave height values across a model domain. Running with H_s only (Model_single) compared to running with 4 parameters (Model_all) reduces accuracy of the model predictions, but that reduction is small relative to the overall accuracy of operational wave models.

The differences introduced by omitting wave period and direction parameters were most pronounced in coastal areas on the south coast of the Cornwall peninsula. In these zones, the H_s only model (Model_single) under-predicts wave heights, potentially due to the reduced learning power of wave refraction and shoaling that are dependent on wavelength and direction.

Both surrogate models matched in-situ measurements but under-predicted higher wave heights. This zone is an area where both surrogate models have relatively low accuracy and although the H_s only model (Model_single) had a slightly lower under-prediction, it is not possible to conclude that this is more accurately representing conditions, rather that there is a general issue with prediction of larger wave states in this area from the surrogate model procedure. The hindcast used to train the surrogate model also under-predicts for large wave heights at the evaluation site, contributing to the models lower accuracy here.

This work demonstrates the feasibility of running surrogate models using H_s only, while allowing analysis of the expected uncertainties both in terms of spatial changes and the absolute uncertainty values. It provides a key step towards deploying a MaLCOM style model using radar-altimeter derived significant wave height (H_s) data. Further work must focus on the potential for improvements on the results displayed, by using the full spatial coverage available from these data sets.

ACKNOWLEDGEMENTS

This work has received funding from the Natural Environment Research Council (NERC) through United Kingdom Research and Innovation (UKRI) for the Centre of Doctoral Training in Sustainable Management of UK Marine Resources (CDT SuMMeR) under grant agreement NE/W007215/1.

A.C. Pillai acknowledges support from the Royal Academy of Engineering under the Research Fellowship scheme (award number: RF\202021\20\175).

REFERENCES

[1] Chen, Jiaxin, Pillai, Ajit C., Johanning, Lars and Ashton, Ian. "Using machine learning to derive spatial wave data: A case study for a marine energy site." *Environmental Modelling & Software* Vol. 142 (2021): p. 105066. DOI 10.1016/j.envsoft.2021.105066. Accessed 2022-10-11, URL https://www.sciencedirect.com/science/article/pii/S1364815221001092.

[2] Mounet, Raphaël E. G., Chen, Jiaxin, Nielsen, Ulrik D., Brodtkorb, Astrid H., Pillai, Ajit C., Ashton, Ian G. C. and Steele, Edward C. C. "Deriving spatial wave data from a network of buoys and ships." *Ocean Engineering* Vol. 281 (2023): p. 114892. DOI 10.1016/j.oceaneng.2023.114892. Accessed 2023-12-15, URL https://www.sciencedirect.com/science/article/pii/S0029801823012763.

[3] Medina-Lopez, E., McMillan, D., Lazic, J., Hart, E., Zen, S., Angeloudis, A., Bannon, E., Browell, J., Dorling, S., Dorrell, R.M., Forster, R., Old, C., Payne, G.S., Porter, G., Rabaneda, A.S., Sellar, B., Tapoglou, E., Trifonova, N., Woodhouse, I.H. and Zampollo, A. "Satellite data for the offshore renewable energy sector: Synergies and innovation opportunities." *Remote Sensing of Environment* Vol. 264 (2021): p. 112588. DOI 10.1016/j.rse.2021.112588. Accessed 2022-10-18, URL https://linkinghub.elsevier.com/retrieve/pii/S0034425721003084.

[4] PO.DAAC. "Jason 3 | PO.DAAC / JPL / NASA." (2020). Accessed 2024-02-27, URL https://podaac.jpl.nasa.gov/JASON3.

[5] European Space Agency. "User Guides - Sentinel-3 Altimetry - Revisit Time - Sentinel Online." (2024). Accessed 2024-02-27, URL https://copernicus.eu/user-guides/sentinel-3-altimetry/coverage/revisit-time.

[6] Gommenginger, C. P., Srokosz, M. A., Challenor, P. G. and Cotton, P. D. "Measuring ocean wave period with satellite altimeters: A simple empirical model." *Geophysical Research Letters* Vol. 30 No. 22 (2003). DOI 10.1029/2003GL017743. Accessed 2023-01-06, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2003GL017743. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003GL017743.

[7] Mackay, E. B. L., Retzler, C. H., Challenor, P. G. and Gommenginger, C. P. "A parametric model for ocean wave period from Ku band altimeter data." *Journal of Geophysical Research: Oceans* Vol. 113 No. C3 (2008). DOI 10.1029/2007JC004438. Accessed 2022-12-13, URL https://onlinelibrary.

wiley.com/doi/abs/10.1029/2007JC004438. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007JC004438.

[8] Badulin, S. I. "A physical model of sea wave period from altimeter data." *Journal of Geophysical Research: Oceans* Vol. 119 No. 2 (2014): pp. 856–869. DOI 10.1002/2013JC009336. Accessed 2024-02-29, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009336. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JC009336.

[9] The Crown Estate. "Floating Offshore Wind Leasing Round 5." (2023). Accessed 2024-01-08, URL https://www.thecrownestate.co.uk/our-business/marine/round-5.

[10] HM Government. "British Energy Security Strategy." (2022).

[11] UK Met Office. "Atlantic- European North West Shelf- Wave Physics Reanalysis." (2021). DOI 10.48670/moi-00060. Accessed 2024-02-26, URL https://data.marine.copernicus.eu/product/NWSHELF_REANALYSIS_WAV_004_015/description.

[12] Channel Coastal Observatory. "National Coastal Monitoring - Map Viewer & Data Catalogue." (2023). Accessed 2024-01-09, URL https://coastalmonitoring.org/cco/.

[13] Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed. Vol. 100. Elsevier (2011).