Data-driven models for short-term ocean wave power forecasting

Chenhua Ni1,2

1 National Ocean Technology Center, Tianjin, China
2 Engineering Department, Lancaster University, Bailrigg, Lancaster, UK

Correspondence
Chenhua Ni, National Ocean Technology Center, 213 Jieyuanxi Rd, Nankai, Tianjin 300112, China.
Email: jeff_ni@126.com

Funding information
SFMRE (Special Funds for Marine Renewable Energy) from MNR, Grant/Award Number: GHME2020ZC01

Abstract
In order to integrate wave farms into the grid, the power from wave energy converters (WEC) must be forecasted. This study presents a novel data-driven modelling (DDM) method to predict very short-term (15 min–4 h) and short-term (0–72 h) power generation from a WEC. The model is tested using data from an oscillating body converter. Several other methods are tested as well. These include support vector machines (SVM), neural networks (NN), and recurrent neural networks (RNN). Of these, the best is the long-short-term memory (LSTM) network, which is trained and updated on observed values. The experiments demonstrate both the SVM and NN forecast well. However, the proposed deep learning models predict them more accurately. The models work well over short horizons. At horizons longer than three days, accuracy deteriorates, and the models cannot fit the data well.

1 | INTRODUCTION

While wind and solar have been the leading sources of renewable energy up to now, waves are increasingly being recognised as a viable source of power for coastal regions [1]. The global potential of wave energy has been estimated at 3.7 terawatts, almost double the world electrical consumption in 2010. Ocean wave energy is more predictable than wind and solar [2]. Further, the energy density of water is greater than wind [3]. Although there is no single preferred design for wave energy converters (hereafter WECs), the technologies fall into three main categories: oscillating water columns (OWC), oscillating body systems (OBS), and overtopping converters [4]. One key aspect of integrating wave farms into the grid is power forecasting, which is used for planning and calculating reserves [5–7]. Wave energy has some of the same problems as wind and solar. Waves can show non-linear variability and intermittency. In particular, waves are more volatile at night when wind speed is often higher. The horizons involved in the management of power grids range from as little as a few hours to as long as several days. The models used here are therefore tested for very short-term (15 min–4 h) and short-term (0–72 h) horizons.

The methods for forecasting wave energy comprise two major approaches, physics-based models and data-driven techniques. Physics models have been in operation since the 1960s, with the theoretical basis established in the late 1950s [8]. As the models have become more sophisticated, they have emerged as one of the most powerful tools for the study of surface water waves [9]. The major physics models include WAVEWATCH III in the United States, and the ECMWF (European Centre for Medium-Range Weather Forecasts) wave model [10, 11]. In addition, some small-scale models such as SWAN (Simulating Waves Near Shore) have been used to study coastal conditions [12, 13]. The physics models predict the standard wave parameters, the significant wave height and the wave period. The wave energy flux, and the power output from WECs can be calculated from these values based on the power matrix of an operating device [14, 15].

Among data-driven models, a variety of techniques have been used. These include regression-based methods, such as autoregressive moving average (ARMA), Kalman filters, and Markov chains [15–21]. More recently, artificial intelligence (AI) and machine learning methods have been tested. These include neural networks and support vector machines, and more recently the new deep learning methods such as recurrent neural network and LSTM [22–30]. Prior studies have found that physics models predict more accurately beyond the first few hours, while DDMs are more accurate at short horizons [31–33].

This study extends the analysis of data-driven methods, focusing on machine learning and artificial intelligence, to...
2 METHODOLOGY

In theory, the DDM related approaches should be part of machine learning (ML) and data mining methods. As Figure 1 shows, the DDM algorithm is used to determine the relationship between a system’s inputs and outputs using a training data set. Hence, they consider influential approaches to do classification and regression based on historical data [34]. Recent practices show that the DDM yields good forecast results, in particular for very short-term prediction up to several hours [35]. SVM, NN as well as LSTM, have been considered as the main representatives to be forecast.

2.1 Support vector machine

SVM is a linear classifier proposed by Vapnik. To date, SVMs have been widely applied to pattern classification problems as well as non-linear regressions because they have the potential to handle very large feature spaces and dimensional vectors [36].

For a given set of samples \( D = \{ (x_i, y_i) \}_{i=1}^{N} \), the ideal method of SVM for regression is mapping the \( x \) into a high dimensional feature space.

\[
    f(x) = \sum_{i=1}^{D} w_i \phi_i(x) + b
\]

where \( \{ \phi_i(x) \}_{i=1}^{N} \) denote features, \( b \) is a threshold and \( \{ w_i(x) \}_{i=1}^{N} \) denote coefficients that have to be predicted from the samples. Since \( \phi \) is fixed, \( \{ w_i(x) \}_{i=1}^{N} \) is determined from the data by minimising the sum of the empirical risk \( R[w] \) and a complexity term \( \|w\|^2 \), which enforces flatness in feature space

\[
    [w] = \frac{1}{D} \sum_{i=1}^{D} |f(x_i) - y_i|^2 + \lambda \|w\|^2
\]

where \( \lambda \) is a regularisation constant and \( |f(x_i) - y_i|^2 \) is cost function. When the “\( \varepsilon \) – insensitive loss function is used the form shows the minimising function.

\[
    f(x, \alpha, \alpha^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \sum_{j=1}^{D} \phi_j(x_i) \phi_j(x) + b
\]

with \( \alpha, \alpha^* = 0, \alpha_i, \alpha_i^* > 0, i, j = 1, ..., N, \) \( D \) means the \( D \)-dimension feature space. Finally, the coefficients \( \alpha, \alpha^* \) are obtained by maximising the following form

\[
    R(\alpha^*, \alpha) = -\varepsilon \sum_{i=1}^{N} (\alpha_i^* + \alpha_i) + \sum_{j=1}^{N} y_j (\alpha_i^* - \alpha_i)
    - \frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i^* + \alpha_i) \times (\alpha_j^* - \alpha_j) k(x_i, x_j)
\]

where, \( \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha, \alpha^* \leq C \). The data points associated with them are called support vectors. Parameters \( C \) and \( \varepsilon \) are free and have to be decided by the user. More details of SVM for regression can be found in [37].

2.2 Neural network

Artificial neural network is a type of DDM whose structure and calibration procedure have some analogies with brain neural networks. As shown in Figure 2, every neuron represents a single unit of NN that has a non-linear element with multiple inputs but single output. The coefficients of the linear combination are named synaptic weights, and together with the network biases they represent the free parameters of the model. The input layers contain input samples and the output layer gives the result. When the NN is created and trained with input and output data, the predicting result is made.

Fundamentally, the NN mathematical model shows as follow:

\[
    y_i = f \left( \sum_{i=1}^{n} w_{i, j} x_i - \theta_j + b \right)
\]

where \( x_i, y_i (i = 1, 2, ..., n) \) are the input and output signals of the network respectively. \( w_{i, j} \) denotes the weight between neurons and \( \theta_j \) denotes the threshold. \( f(x) \) is effect function. The
network is trained based on error back-propagation (BP) and weight or threshold adjustment. When a sample is given and input into the model, then the output is computed. All samples are inputted once the total error is estimated [39].

2.3 Long-short-term memory

The LSTM is a family member of RNN that uses their feedback connection store representations of recent input events. The specification makes it potentially useful for time series forecasting. As seen, there are many advantages of this particular RNN approach. The result of traditional RNN may leads to oscillating. As seen, there are many advantages of this particular RNN specification makes it potentially useful for time series forecasting.

Each LSTM includes input layers, hidden layers and an output layer. Figure 3a illustrates the additional ‘cell status’ of LSTM. In the left subplot, an RNN cell $A_t$ receives two inputs $A_{t-1}$ and $X_t$, then output $A_{t+1}$ and $H_t$. The LSTM cell in the right subplot additionally adds parameter $C$ to RNN cell, and saves cell status from previous cell. Besides, Figure 3b depicts a single LSTM cell at time step $t$. The LSTM cell obtains the ability to add or remove information from the cell status by structures called gates. There are three gates designed in LSTM architecture: a forget gate ($f_t$) means how many memories will be reserved from $c_{t-1}$ to $c_t$; an output gate ($o_t$) decides how many memories will be output to $H_t$; an input gate ($i_t$) means how many memories will be reserved from $c_t'$. As shown in the right subplot of Figure 3b, the long-term memory is calculated with $c_t'$ multiply by $f_t$. The short-term memory is calculated with $c_t'$ multiplied by $i_t$, which means the input gate. As a result, the $c_t$ combines long-term with short-term information. In terms of output $H_t$, it is decided by cell status $c_t$ and output gate $o_t$. With the Equation (6)–(8), the LSTM architecture follows the steps to pass the information to the next one based on the three functional gates [41, 42].

The mathematical representation of LSTM can be obtained as follows:

$$f_t = \sigma(x_t \cdot \theta_{xf} + H_{t-1} \cdot \theta_{hf} + b_f)$$ \hspace{1cm} (6)$$

$$i_t = \sigma(x_t \cdot \theta_{xi} + H_{t-1} \cdot \theta_{hi} + b_i)$$ \hspace{1cm} (7)$$

$$o_t = \sigma(x_t \cdot \theta_{xo} + H_{t-1} \cdot \theta_{ho} + b_o)$$ \hspace{1cm} (8)$$

Note that $\theta_{xf}$ and $\theta_{xi}$ represent weights of input and output from previous cells, and $b_f$ denotes bias with $n \in \{i, f, o\}$.

Cell memory updates itself recursively by interaction of its previous value with forget and other values.

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t'$$ \hspace{1cm} (9)$$

$$c_t' = \sigma_t(W_t' \odot x_t + U_t \odot H_{t-1} + b_t)$$ \hspace{1cm} (10)$$

where $W_t', U_t, b_t$ denote weights and biases, and $\sigma_t$ denotes tangent function. Finally, the cell output can be computed by the equation below:

$$H_t = o_t \odot \phi(c_t)$$ \hspace{1cm} (11)$$

3 Wave Power Forecasting Framework

Three models are tested, the neural net, the SVM, and the LSTM. Figure 4 illustrates the forecasting strategy of this paper. The three DDM models are explained and employed; the different results are given separately. According to the forecasting purposes, the prediction strategy includes different horizon ranges from a few hours to up to 3 days.

3.1 Feature selection

The performance of DDM highly depends on its inputs. There are several observations collected from the WEC such as wave height, rotor speed, pressure etc. Network training with redundant inputs not only reduces the predicting accuracy but also increases the computation. Hence, each observation has several lags with different features that need to be determined. The principal component analysis (PCA) method is adopted to select the main predictors. PCA is widely used to filter redundant factors, details can be found in ref. [43]. Then the Granger causality test (GCT) is conducted to determine correlations between the remaining variables and the wave parameters. It is used to keep the testing horizon stationary; details can be founded in the literature [44]. As a result, the data collected from WEC are
TABLE 1 The overview of variables collected from the WEC

| Inputs                  | Time interval | Max/Min values | Units  |
|-------------------------|---------------|----------------|--------|
| Hydraulic pressure     | 1 min         | 7.14/0.04      | bar    |
| Rotor speed             | 1 min         | 152.51/0.01    | r/min  |
| Hydraulic flow          | 1 min         | 1.99/0.00      | m³/s   |
| Motor torque            | 1 min         | 141.01/0.00    | N·m    |

TABLE 2 The main hyper-parameters of purposed LSTM

| Hyper-parameters                  | Values |
|-----------------------------------|--------|
| Input sequence                    | 4      |
| Hidden LSTM units                 | 200    |
| Fully connected layer             | 1      |
| Output layers                     | 1      |
| Batch size                        | 200    |
| Number of epochs                  | 100    |

optimised for four predictors: pressure, flow, rotor speed and motor torque, Table 1.

3.2 Model architecture

In order to assess the performance of each model, the three purposed network are adopted separately. SVM uses wave power time-series data as training data. The polynomial kernel and the Gaussian radial basis function are used as the kernel function of this SVM. The standard three-layer BP networks are also selected as benchmarks.

For the NN modelling, the inputs are the pressure, flow, rotor speed and torque (see Table 1). The lagged input data are used to predict the power. For example, for the 1-hour ahead prediction, the previous 60 minutes of the input series are used. The inputs are classified into train value and predicting value. The three-layer feed-forward networks with a non-linear sigmoid transfer function are deployed.

The LSTM structure is a two-dimensional network. The four time-series inputs represent the vertical dimension, mentioned in Table 1. The lateral dimension indicates the changes in the time domain. The network structure is composed of input layers, LSTM layers with 200 hidden units, one fully connected layer and an output layer. The main hyper-parameters of the network are denoted in Table 2.

The basic LSTM block uses logistic sigmoid and hyperbolic tangent as an activation function. Here, the ReLU activation function is deployed because it is easy to optimise and is not saturated. The traditional optimisation method for a deep network is stochastic gradient descent (SGD), which is the batch version of gradient descent. The batch method can speed up the convergence when training network. Here, the Adagrad optimisation method [45] is adopted, which can adapt the learning rate to the parameters and perform larger updates. To train the LSTM units, initialise weight matrices and bias vectors that include \( \theta_{x,y} \), \( \theta_{h,y} \) and \( b_y \) randomly. Then train the parameters by using BP method with the gradient-based optimisation, to minimise the cost function. For different observation points, the corresponding LSTM units are trained.
3.3 Evaluation for forecast result

The criteria used to evaluate the models include the mean absolute error (MAE), the mean absolute percent error (MAPE), and the root mean squared error (RMSE) [46].

4 FORECASTING EXPERIMENTS

4.1 Data acquisition and preparation

The data for modelling is collected from an operational WEC with 10 kW total capacity. Figure 5 shows the schematic diagram of the WEC. It is an oscillating buoy device (similar as Wavestar) and extracts wave energy from the ocean by using an up-and-down motion absorber. In theory, the WEC contains an energy absorbing system, energy conversion system and energy transmission system. The energy-absorbing system takes the responsibility to acquire energy from wave resource. Meanwhile, the energy conversion system converts kinetic energy to electrical power. Then the power is transmitted from device to grid using energy transmission system.

The data lasted for approximately three months from February to April 2017 in sequence. The quality control procedures were applied to the measurements to fill the gaps and to replace the unrealistically large values. Some of the rectified observations are shown in Figure 6. It is clear that the type of signals shown in Table 1 are recorded at a one-minute interval.

The power signals represent the predictors of the proposed model. Time series of data in total are divided into two non-overlapping data sets. The first 90% samples are selected for training the model and the remaining samples served for testing model.

4.2 Prediction results

Figure 7a illustrates the very short-term multi-step predicting results from 15 min up to approximately 5 h ahead. Figure 7b illustrates the short-term multi-step predicting results from three hours up to 90 hours ahead. Obviously, both the very short-term and short-term predictions (red lines) cannot fit the real values well in almost horizons. And the errors increase conspicuously as the forecast horizon increases. This tendency can be found notably in Figure 7b. The observing curve and forecasting curve have similar tendency between time point 370 and 380. But the errors between the two curves begin to increase sharply after time point 380, showing that the SVM model performs unsatisfactory. Furthermore, for very short-term prediction, less than 4-hour ahead values seem believable by using SVM method. For short-term prediction, less than 3-day ahead values seem believable by using SVM method.

Figure 8 shows the very short-term and short-term power prediction results from NN. It seems much better than SVM results while using the same training and testing datasets. Obviously, the prediction lines fit the observed data properly both for very short-term and short-term prediction. Compared with
Figure 7, the curves from NN model are much closer to the real curves than from SVM model throughout the prediction horizon. Yet the red curve in Figure 8a means forecasting values fluctuates much heavier in order to adapt to the observing values. The forecasts, shown graphically in Figure 8b, demonstrate that the model is extremely accurate.

Normally, each LSTM block trains at each time step based on the previous period’s input signals. In order to achieve a more accurate forecast, the models used here are all updated on actual observations. Figure 9 explains the very short-term prediction and short-term prediction wave power prediction from LSTM network. Similar as NN, each time sequence in Figure 9a represents 15 minutes and the vertical axis means the power output from WEC. The blue line represents the observing data and the red line denotes the predicting data from the end of the observing data. The time point 600 means the end of observing data and from 601 to 620 time-point means the forecasting results of very short-term up to 5 hours (15 min×20). Similarly in Figure 9b, time-point 370 means the end of observing data and from 371 to 400 time-point means the results of short-term up to 90 hours (30×3 h), more than three days. It can be concluded that the proposed deep LSTM performs more accurately than NN and SVM models. In particular, the forecasting performance remains stable during the entire forecasting period. For comparison, the traditional LSTM model were given results which uses predicting data as update strategy. The results are shown in Table 3.

The evaluation index in Table 3 also confirm the results. The 1st, 2nd and 3rd comments represent the model trained with different hyper-parameters. For NN method, the MAPE values (mean the model accuracy) from very short-term demonstrate slightly better than short-term prediction. Similarly, the LSTM evaluation values from very short-term significantly surpass the short-term values. The RMSE values of very short-term obtained 15.04, 16.24, and 16.26, compared with the short-term 23.08, 29.06 and 24.12. Furthermore, all the metrics show that the proposed LSTM model predicts more accurately than NN.

4.3 Analysis and discussion

In case of mid-term prediction, Figure 10 gives the comparative results up to 16.5 hours in advance in forms of curves and bar charts. Each curve means different multi-step forecasting results and each bar chart means the errors between real values and forecasting values. The SVM model has the smoothest tendency while NN and LSTM models fluctuate dramatically. For very short-term prediction, the LSTM methods can fit better than other methods in this first 10 time steps (2.5 hours). For the entire view of the raw data, the very short-term prediction model cannot fit the raw data well more than two days ahead.

Figure 11 also gives the compatible short-term results up to 24 days (each time stamp means 3 hours). The prediction for first 36 time-points fit well from purposed models. For the large values power output, all purposed models cannot fit well.

The evaluation index in Table 3 also confirm the results. The 1st, 2nd and 3rd comments represent the model trained with different hyper-parameters. For NN method, the MAPE values (mean the model accuracy) from very short-term demonstrate slightly better than short-term prediction. Similarly, the LSTM evaluation values from very short-term significantly surpass the short-term values. The RMSE values of very short-term obtained 15.04, 16.24, and 16.26, compared with the short-term 23.08, 29.06 and 24.12. Furthermore, all the metrics show that the proposed LSTM model predicts more accurately than NN.
TABLE 3  Error estimated results of the predictions of the power output

| Indexes          | NN (Very short-term) | NN (Short-term) | LSTM (Very short-term) | LSTM (Short-term) | Traditional LSTM (Very short-term) | Traditional LSTM (Short-term) |
|------------------|----------------------|-----------------|------------------------|-------------------|-----------------------------------|-------------------------------|
|                  | 1st                  | 2nd             | 3rd                    | 1st               | 2nd                              | 3rd                          |
| MAPE(%)          | 41.52                | 44.62           | 40.02                  | 52.27             | 63.03                            | 68.79                        |
| MAE(W)           | 18.22                | 20.27           | 19.15                  | 19.99             | 18.83                            | 20.77                        |
| RMSE(W)          | 23.71                | 26.03           | 25.52                  | 30.20             | 27.09                            | 27.16                        |
|                  | LSTM (Very short-term) |                |                        | LSTM (Short-term) |                                  |                               |
| MAPE(%)          | 28.02                | 36.91           | 29.57                  | 43.65             | 55.69                            | 55.78                        |
| MAE(W)           | 11.75                | 16.42           | 12.91                  | 17.09             | 19.25                            | 15.57                        |
| RMSE(W)          | 15.04                | 16.24           | 16.26                  | 23.08             | 29.06                            | 24.12                        |
|                  | Traditional LSTM (Very short-term) |            |                        | Traditional LSTM (Short-term) |                                  |                               |
| MAPE(%)          | 39.05                | 31.68           | 28.49                  | 45.53             | 68.39                            | 44.26                        |
| MAE(W)           | 18.22                | 13.83           | 12.14                  | 17.09             | 19.25                            | 15.57                        |
| RMSE(W)          | 22.49                | 16.24           | 15.13                  | 23.08             | 29.06                            | 24.12                        |

FIGURE 10  Model comparison of very short-term prediction: (a) curves comparison; (b) errors comparison

other words, the short-term prediction has reliable confidence interval within the first 72 hours and the further predicting results seems suspicious. Furthermore, there are few abnormal values such as minus predicted by models and future work will fix these problems.

The calculations were implemented in Matlab environment in an Inter® Core™ i7-8550U CPU at 1.80 GHz mobile workstation with 16 GB RAM and a Radeon Pro X3100 GPU. For the computing time, the LSTM seems much quicker than SVM and NN. It costs 82.47 s to running purposed NN network and is typically time-consuming and expensive. The SVM method also costs average 40 s in total. Besides, the LSTM only costs approximately 25 s working in parallel computing mode.
5 | CONCLUSION

The analysis has produced several findings. First, the new artificial intelligence methods tested here predict very accurately at extremely short horizons, which are needed for grid integration, but are generally shorter than the horizons at which physics models can forecast. Second, the models work well over slightly longer horizons of up to two days. However, forecast accuracy deteriorates rapidly at horizons of three days or more. Of the methods tested, the neural net and SVM do well. However, the LSTM is consistently more accurate. With the advent of new open source software which makes the LSTM readily available to users, this method can easily be adopted for wave energy forecasting.

ACKNOWLEDGEMENTS

This study was supported by SFMRE (Special Funds for Marine Renewable Energy) from MNR (Grant number GHME2020ZC01).

ORCID

Chenhua Ni https://orcid.org/0000-0002-7474-8973

REFERENCES

1. Annagia, R.A., Cheung, K.F.: Atlas of global wave energy from 10 years of reanalysis and hindcast data. Renew. Energy 39, 49–64 (2012)
2. Reikard, G., et al.: Combining wave energy with wind and solar: Short-term forecasting. Renew. Energy 81, 442–456 (2015)
3. Marcus, L., et al.: Ocean wave energy in the United States: Current status and future perspectives. Renewable Sustainable Energy Rev. 74, 1300–1315 (2017)
4. Sproul, A., Weise, N.: Analysis of a wave front parallel WEC prototype. Sustain. Energy IEEE Trans. 6(4), 1183–1189 (2015)
5. Magagna, A.: 2014 JRC Ocean Energy Status Report – Technology, market and economic aspects of ocean energy in Europe. European Commission, EUR 26983, (2015)
6. Ocean Energy Systems (OES): Ocean Energy Annual Report 2018 (2018)
7. Reikard, G., et al.: Wave energy worldwide: Simulating wave farms, forecasting, and calculating reserves. Int. J. Mar. Energy 17, 156–185 (2017)
8. Phillips, O.M.: On the generation of waves by a turbulent wind. J. Fluid Mech. 2, 417–445 (1957)
9. Wolf, J.: Coupled wave and surge modelling and implications for coastal flooding. Adv. Geosci. 17, 19–22 (2008)
10. Tolman, H.L.: User manual and system documentation of WAVEWATCH-III version 2.22. NOAA /NWS / NCEP / MMAB Technical Note, pp. 222, 133 (2002)
11. Komen, G.J., et al.: Dynamics and Modelling of Ocean Waves, Cambridge, pp. 534 (1994)
12. Surf’s Up: Professor Using Models To Predict Huge Waves, ScienceDaily, Texas A&M University, 23 Feb 2005
13. Thomas, J.T., Dwarkanath, G.S.: Numerical wave modeling: A review. Aquat. Procedia. 4, 443–448 (2015)
14. Pinson, P., et al.: Probabilistic forecasting of the wave energy flux. Appl. Energy 93, 364–370 (2012)
15. Hatalis, K.: Multi-step forecasting of wave power using a nonlinear recurrent neural network. PES General Meeting | Conference & Exposition, IEEE (2014)
16. Hu, Q., et al.: Pattern-based wind speed prediction based on generalized principal component analysis. IEEE Trans. Sustain Energy 5(3), 866–874 (2014)
17. Erdem, E., Shi, J.: ARMA based approaches for forecasting the tuple of wind speed and direction. Appl. Energy 88(4), 1405–1414 (2011)
18. Liu, H., et al.: Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. Appl. Energy 88(3), 724–732 (2011)
19. Silaghi, H., Costea, C.: Wind speed prediction using Box-Jenkins method. J. Comput. Sci. Control Syst. (1), 208 (2008)
20. Poncele, M., et al.: Automatic tuning of Kalman filters by maximum likelihood methods for wind energy forecasting. Appl. Energy 108, 349–362 (2013)
21. Carpinon, A., et al.: Very short-term probabilistic wind power forecasting based on Markov chain models. IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, pp. 107–112 (2010)
22. Song, Z., et al.: Short-term wind speed forecasting with markov switching model. Appl. Energy 130, 103–112 (2014)
23. Scott, C., et al.: A machine learning framework to forecast wave conditions. Coastal Eng. 157, 1–10 (2018)
24. Peres, D.J., et al.: Significant wave height record extension by neural networks and reanalysis wind data. Ocean Modell. 94, 128–140 (2015)
25. Fischer, B., et al.: Online-algorithm using adaptive filters for short-term wave prediction and its implementation. 4th International Conference on Ocean Energy, Dublin, 17 October 2012
26. Zhou, J., et al.: Fine tuning support vector machines for short-term wind speed forecasting. Energy Convers. Manage. 52(4), 1990–1998 (2011)
27. Krishna, K.N., et al.: Regional ocean wave height prediction using sequential learning neural networks. Ocean Eng. 129, 605–612 (2017)
28. Shikhar, S., Stefan, L.A.: Comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol. Energy 162, 232–247 (2018)
29. Wang, H., et al.: Deep learning based ensemble approach for probabilistic wind power forecasting. Appl. Energy 188, 56–70 (2017)
30. Wu, T., et al.: Deep learning for inversion of significant wave height based on actual sea surface backscattering coefficient model. Multimed. Tools. Appl. 79, 34173–34193 (2019)
31. Zaytar, M., Amrani, C.: Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int. J. Comput. Appl. 143(11), (2016)
32. Reikard, G., et al.: Forecasting ocean wave energy: The ECMWF wave model and time series methods. Ocean Eng. 38, 1089–1099 (2017)
33. Reikard, G., Robertson, B., et al.: Simulating and forecasting ocean wave energy in western Canada. Ocean Eng. 103, 223–236 (2015)
34. Solomatine, D., et al.: Data-driven modelling: Concepts, approaches and experiences. In: Abrahart, R.J., See, L.M., Solomatine, D.P. (eds.) Practical Hydroinformatics. Water Science and Technology Library, vol. 68, Springer, Heidelberg (2009)
35. Ahsan, M., et al.: Developing computational intelligence for smart qualification testing of electronic products. IEEE Access 8, 16922–16933 (2020)
36. Kumar, R., et al.: Symbolization assisted SVM classifier for noisy data. Pattern Recognit. Letters 25, 495–504 (2004)
37. Mohandes, M.A., et al.: Support vector machines for wind speed prediction. Renew. Energy 29, 939–947 (2004)
38. Castros, A., et al.: Performance of artificial neural networks in nearshore wave power prediction. Appl. Soft Comput. 23, 194–201 (2014)
39. Cudera, L., et al.: Computational intelligence in wave energy: Comprehensive review and case study. Renew. Sustain. Energy Rev. 58, 1223–1246 (2016)
40. ElSaad, A., et al.: Using LSTM recurrent neural networks to predict excess vibration events in aircraft engines. IEEE 11th International Conference on e-Science, pp. 260–269, Baltimore (2016)
41. Zhao, Z., et al.: LSTM network: A deep learning approach for short-term wind power forecasting. Appl. Energy 188, 56–70 (2017)
42. Kiperwasser, E., Goldberg, Y.: Simple and accurate dependency parsing using bidirectional LSTM feature representations. Trans. Assoc. Comput. Linguistics. 4, 313–327 (2016)
43. Ni, C., Ma, X.: An integrated long-short-term memory algorithm for predicting polar westerlies wave height. Ocean Eng. 215, 107715 (2020)
44. Liu, D., et al.: Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew. Energy 62, 592–597 (2014).
45. Duchi, J., et al.: Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12(7), 257–269 (2010).

46. Ni, C., et al.: Integrated deep learning model for predicting electrical power generation from wave energy converter. The 25th International Conference on Automation & Computing, Lancaster University, Lancaster, UK, 5–7 September (2019).

How to cite this article: Ni C. Data-driven models for short-term ocean wave power forecasting. IET Renew. Power Gener. 2021;15:2228–2236. 
https://doi.org/10.1049/rpg2.12157