

Integrated deep learning model for predicting electrical power generation from wave energy converter

Chenhua Ni¹, Xiandong Ma², Ji Wang¹

¹National Ocean Technology Center, Tianjin 300112, China
nichiva@126.com

²Engineering Department, Lancaster University, Lancaster LA1 4YW, UK
xiandong.ma@lancaster.ac.uk

Abstract—The continuous improvement of wave and tidal energy technologies has widely boosted the development of marine energy plants. An accurate predication of the electrical power generation of marine energy not only saves costs for operation and maintenance but also improves manage the electricity consumption and reduce the uncertainty due to the intermittency of the ocean wave and tidal resources. This paper presents an integrated deep learning (DL) network comprising the long short term memory (LSTM) algorithm and the principal component analysis (PCA) to predict the electrical power generation from a wave energy converter (WEC). The results from this integrated data-driven model show the remarkable performance compared with the LSTM alone and other machine learning models. Furthermore, the experiments have shown that the high-frequency oscillating waves and the long term features in the wave have a significant impact on the model's accuracy. This finding demonstrates the superiority of the proposed model in its ability to deal with time sequence data and the effect of the high-frequency oscillating signals on the results over other machine learning methods.

Keywords—Principal component analysis, long short term memory, deep learning, wave energy converter, marine energy, power prediction

I. INTRODUCTION

Marine energy, also known as ocean energy (OE), has been considered as a renewable energy resource to generate clear electricity but remains widely untapped [1]. It is estimated that the potential of OE can generate electrical power almost 80,000 TWh per year by converting ocean temperatures, salt contents, movements of tides, currents, waves and swells [2]. Wave energy, as one of the OE resources, represents the vast majority of the resource potential and is a relatively mature technology among these OE technologies. Electrical power prediction from wave energy would play a crucial role in controlling the balance between gird and consumption and managing the energy usage and shortage. Besides, an accurate prediction of power from WECs is able to help organize the operation and reduce the maintenance costs because of the unpredictable resource and harsh sea conditions [3].

The traditional way to predict wave energy power usually consists of two steps. Firstly, wave parameters, such as wave height, period and direction, can be forecasted by either statistical algorithms or physical models. Then specific regression methods are developed to predict the

electrical power generation through historical data from these wave parameters. The wave forecasting models have become stable and highly credible after decades of considerable improvement. However, the models require large amount of data to train, which usually takes longer time even for a small-scale forecast. For that reason, the machine learning and deep learning algorithms were emerged and bloomed to improve accuracy and promptness of the prediction.

Machine learning (ML), a kind of ANN, is a specific study of computer systems to simulate the behaviors and spirits of human being based on historical data. It performs the forecasting by characterizing a nonlinear and complex relationship between input data and output targets [4]. Until now, artificial intelligence (AI) methods including ANN, extreme learning machines (ELM), support vector machines (SVM) and DL have been used to learn features of the wave power. The DL architectures cover deep neural networks, deep belief networks and recurrent neural networks (RNN), which construct more hidden layers and neurons than those shallow network. DL models have been widely used in computer vision, speech processing, natural language recognition and machine translation primarily because DL structures are able to extract basic features and integrate virtual features from historical data [5]. LSTM method was considered as one of the most appropriate methods to address prediction problems because the method has shown potential to solve a range of problems involving sequential learning in recent years. The LSTM, derived from the RNN family, shares parameters across time series between hidden units. The LSTM algorithm has been applied to forecast day-ahead global horizontal irradiance using satellite data and proved it is a promising technique in solar power prediction. A Gaussian process regression model incorporating with the LSTM has been implemented in a wind farm located in Zhangjiakou, China in order to forecast the wind speed more accurately and to outperform the traditional models [6][7].

For this purpose, the remainder of the paper is organized as follows. Section 2 depicts the methodology integrated with the LSTM algorithm and PCA analysis. Section 3 describes the model architecture and hyper-parameters used in this study. Section 4 shows the performance of the LSTM network using practical sea testing data collected from a WEC deployed in the South China Sea. Finally, Section 5 summarizes the conclusions resulting from the study.

II. METHODOLOGY

A. LSTM network

A RNN network can theoretically use its feedback connections to store representations of recent input events in the form of activations, mostly by changing weights for short-term memory and long-term memory. However, it still may lead to oscillating weights in some cases, resulting in unacceptable amount of time or even not working at all [8]. For these reasons, Sepp Hochreiter and Jürgen Schmidhuber proposed a novel recurrent network architecture called LSTM to tackle these error back-flow problems in 1997. It can learn to make the connections between time indices beyond 1000 steps and keep the specific architecture constant against exploding and vanishing gradients [9]. In 1999, Felix A. Gers, et al, adapted a “forget gate” that enabled an LSTM cell to learn to reset itself at appropriate times and thus releasing internal resources. All AI algorithms alone including LSTM were not able to solve continual versions of these problems; however the LSTM combined with forget gates could easily solve them in a simple and convenient way [10].

As a family member of RNN, LSTM network is formed by an input layer, hidden layers, and an output layer. Each hidden layer of a traditional RNN contains one short-term memory vector h . As shown in Figure 1, the left side illustrates the structure of a RNN cell, which means A_t not only receives information from A_{t-1} but also sends information to A_{t+1} [11]. As shown in the right side in Figure 1, the key element of the LSTM is the cell state c , which can remember long-term information by means of cell state and inputs from previous time series or sequences. There are three inputs at time step t , the output h_{t-1} from the previous time series, the cell state c_{t-1} of the previous time series and the present input x_t . Clearly, the output of this layer includes output value h_t and cell state c_t , which are, in turn, inputs for the next hidden layer. The LSTM does have the ability to remove or add information to the cell state, as carefully regulated by structures called gates. The gates are a forget gate (f) implying how many memories will be reserved from c_{t-1} to c_t , an output gate (o) deciding how many memories will be outputted to h_t , and an input gate (i) implying how many memories will be reserved from c' . The interaction (element-wise multiplication) of input and update gate perform “write functions” into memory. The input gate, using the hard sigmoid, decides which values to write and the update gate, using hyperbolic tangent (\tanh) activation, creates a vector of new cell values [12].

The input gate (i_t), forget gate (f_t) and output gate (o_t) of the LSTM can be obtained as follows.

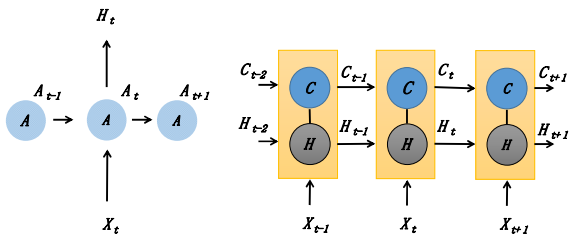


Figure 1. The basic architecture of LSTM algorithm

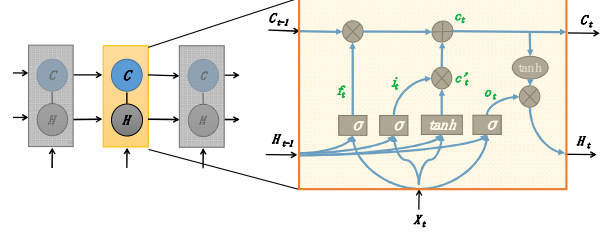


Figure 2. The gates in an LSTM network

$$i_t = \sigma(\omega_i X_t + U_i S_{t-1} + V_i \cdot c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma_g(\omega_f X_t + U_f S_{t-1} + V_f \cdot c_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma_g(\omega_o X_t + U_o S_{t-1} + V_o \cdot c_{t-1} + b_o) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \bar{c}_t \quad (4)$$

$$c'_t = \sigma_c(W_c X_t + U_c S_{t-1} + b_c) \quad (5)$$

$$s_t = o_t \cdot \tanh(c_t) \quad (6)$$

From the above equations, x_t represents the input vector, W , U , V and b denote the hyper-parameters for weights and biases. The symbol \cdot represents the scalar product of the two vectors. The variable σ_g is the sigmoid function, and σ_h and σ_c are the hyperbolic tangent functions, respectively [13][14].

Figure 2 illustrates diagrammatically how these equations work, where each rectangle represents a single LSTM cell and three σ denote three gates, respectively. As shown in the middle rectangle, the long-term memory is calculated with c_{t-1} multiplied by f_t , which means the forget gate. The short-term memory is calculated with c'_t multiplied by i_t , which means the input gate. As a result, the c_t representing the new cell state is able to combine both long-term and short-term information. In terms of output h_t , it is decided by cell state c_t and output gate o_t . With the equations 1-6, the LSTM network can follow the steps to pass the information to the next one based on the three functional gates.

B. PCA method

PCA is a statistical method that uses an orthogonal transformation to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs). It was originally invented as an analogue of the principal axis theorem in mechanics by Karl Pearson in 1901 and improved by many times until now [15]. If a multivariate dataset with a high dimension is given, the PCA is able to result in a low-dimension dataset while still preserving the features of the original dataset with first few PCs. More specifically, given a set of observations x with dimension m dataset, it is transformed to a set of observations x with

dimension n ($n < m$), which called n PCs. The PCA method can reconstruct dataset with the n -dimension PCs (new dataset) rather than simply remove the $(m-n)$ dimension from the original m -dimension dataset. Consequently, PCA method can be used to reduce the complexity of inputs and thus decrease the processing time notably.

C. Performance evaluation index

We choose three mainstream performance evaluation indices to measure the accuracy of prediction, including the coefficient of determination (R-squared), root mean square error (RMSE) and the mean absolute error (MAE). R-squared value provides a measure of how well the observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model [16]. RMSE is more sensitive to a large deviation between the predicted values and the actual outcomes. The MAE measures the absolute difference value between the forecasts and the actual values. These measures are calculated as follows.

$$R_T^2 = 1 - \frac{\sigma_e^2}{\sigma_y^2} \quad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |I_{(pred,i)} - I_{meas,i}| \quad (8)$$

$$RMSE = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (I_{(pred,i)} - I_{meas,i})^2} \quad (9)$$

Here σ_e^2 represents the variance of the residuals between predicted and actual values and σ_y^2 is known as the variance of the actual output. $I_{(pred,i)}$ is the predicted value while $I_{meas,i}$ is the corresponding measurement data.

III. PCA-LSTM PREDICTION ARCHITECTURE

The PCA based LSTM method is shown in Figure 3. As shown, the model is classified into two sections, namely, the feature extraction by PCA and the deep learning algorithm by LSTM. The feature extraction takes the responsibility to extract the PCs from the original dataset while the DP framework makes the prediction using PCs as the input signals.

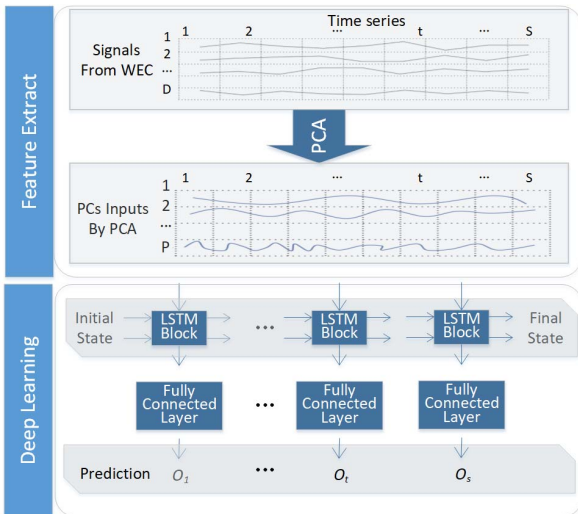


Figure 3. The PCA based deep-learning forecasting method

TABLE I. LIST OF THE VALUES OF LAYERS IN THE NETWORK

Layers	Values
Sequence Input layers	1
LSTM Layers	200
Fully Connected Layers	1
Regression Layers	1

The proposed LSTM network starts with an input layer with multi-variant time-series followed by an LSTM layer. An LSTM layer learns long-term dependencies between time steps in time series and sequence data. Then, a fully connected layer connects every neuron in the LSTM layer to every neuron in the next layer. For a regression prediction, this network ends with a regression output layer.

In the figure, the stream of time series data S with D features of length S are the signals from a WEC. The PCA extracts the useful information from D PCs to P PCs, which means P features are finally treated as the flow of inputs of the LSTM layer. The variable O represents the output (also known as the hidden state) of the network. In the LSTM layer, the first LSTM block takes the initial state of the network and the first time series data from feature extraction section. It also calculates the result h_t and cell state c_t . If it is at time step t , the t block receives the current state of c_{t-1} and h_{t-1} , as well as the input signals at time step t . After processing, the state c_t and result h_t are updated and the calculation is completed. At each step, the output state (h) and the cell state (c) are included in the state of layer. The output state h contains the output information while the cell state c contains information learned from the previous time steps. The gates that represent the specific features of the LSTM network control addition or removal of the information from the cell state. The number of the layers used for this study is given in Table I.

IV. CASE STUDY: POWER PREDICTION FOR A WEC

A. Data acquisition

WECs are typically to convert kinetic energy into electricity through three procedures: energy capture, power take-off (PTO) and electrical energy generating system. A data collection and management system such as supervisory control and data acquisition system (SCADA)

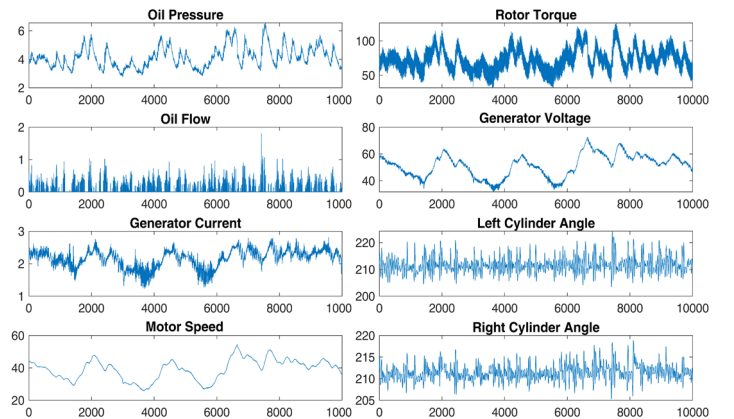


Figure 4. The testing output signals from WEC after pre-processing

TABLE II. THE COEFFICIENTS AND CONTRIBUTION VALUES OF PCA PROCESS

Inputs		PC1	PC2	PC3	PC4	Variance Contribution Rate (%)	Cumulative Contribution Rate (%)
Y1	Oil Pressure	0.2993	0.3311	0.4859	-0.1632	42.1385	42.1385
Y2	Rotor Torque	0.2958	0.3154	0.4522	-0.3236	20.9077	63.0462
Y3	Oil Flow	-0.0041	0.1938	0.3620	0.8987	15.9887	79.0349
Y4	Generator Voltage	0.4790	-0.1587	-0.0083	0.0147	10.1207	89.1556
Y5	Generator Current	0.4050	-0.2674	-0.1241	0.0706	3.5813	92.7369
Y6	Left Cylinder Angle	0.1089	0.5472	-0.4377	0.0170	2.9625	95.6994
Y7	Motor Speed	0.4840	-0.1803	-0.0251	0.0224	2.4528	98.1522
Y8	Left Cylinder Angle	0.1605	0.5443	-0.3963	0.0658	1.8203	99.9725
Y9	Transmission Voltage	0.3970	-0.1687	-0.2522	0.2252	0.0275	100.0000

has commercially been used for wind farms during operation. This study includes a WEC deployed for open sea testing in South China Sea for roughly three months. The capacity of the WEC is 10 kW and the numerical model associated with real observation data shows that the wave condition changes vary during seasons over there.

Nine crucial parameters during testing and operation were collected and recorded in the recording system with a sampling frequency of 10 Hz, including oil pressure and flow, rotor speed and torque, current and voltage, cylinder angles, mechanical power, and inverse power. It is found some signals are correlated and hence the PCA method is adopted to extract the PCs and get rid of the redundant data. Inevitably, abnormal and constant data may exist due to the WEC in inactive or malfunction conditions. Therefore pre-processing of the signals are necessarily used to obtain the best accuracy from the model before the raw data are inputted into the network. Furthermore, normalization is used to regulate the raw data with different scale into range of [0, 1] or [-1, 1]. The results after data pre-processing are shown in Figure 4, indicating the tendency and relationship among the variables recorded from the WEC testing.

B. PCA result

The 9 time-series variables are necessarily normalized to unify dimensions before the PCA method is applied. The coefficients of first four PCs and their corresponding contribution rates are given in Table II. The variance contribution rate, which reflects the correlation between input signals and results, shows the contribution rates of the first 4 PCs are at 42%, 21%, 16% and 10% respectively, dominating nearly 90% of the contributions. Consequently, the 4 PCs contain almost the total information and will be used as the inputs of the proposed LSTM. Let's take PC1 as an example, the higher four scores contributing to the PC1 are Y7 (motor speed), Y4 (generator voltage), Y5 (generator current) and Y9 (transmission voltage), as marked with yellow in Table II. This means the most significant parameter to affect PC1 is motor speed, i.e., the generator speed. The generator voltage, generator current and transmission voltage are regarded as the 2nd to 4th most effective parameters to affect the PC1.

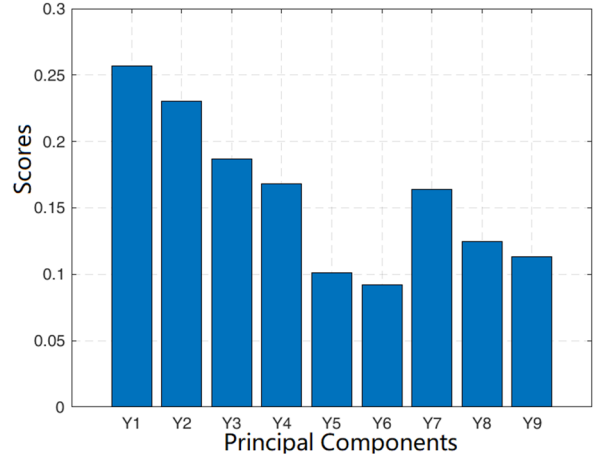


Figure 5. The total PC scores of the 9 variables

As a result, the total scores of Y1 to Y9, i.e., 9 parameters recorded from the operating WEC, are shown in Figure 5. The figure illustrates the rank of the PCs in this case and we can easily find the first 4 signals containing more than 90% information. The 4 signals are Y1, Y2, Y3 and Y4, namely, the oil pressure, rotor torque, generator voltage and accumulated power, respectively. After processing 5 variables are removed due to being considered as redundant signals. Furthermore, the PCA method can not only remove superfluous inputs but also reduce the burden of modelling and processing unnecessary signals.

C. Experimental prediction results

This section describes the electrical power prediction results based on the proposed PCA-LSTM network. Here, the oil pressure, rotor torque, voltage and generator power are the input (as the single time-series data) to the model

TABLE III. PERFORMANCE EVALUATIONS OF SINGLE FEATURE FORECASTING MODEL

Feature	RMSE	MAE	R ²
Oil Pressure	0.415	0.026	0.988
Rotor Torque	23.829	1.861	0.944
Voltage	11.175	1.053	0.243
Accumulated Power	0.004	0.003	0.902

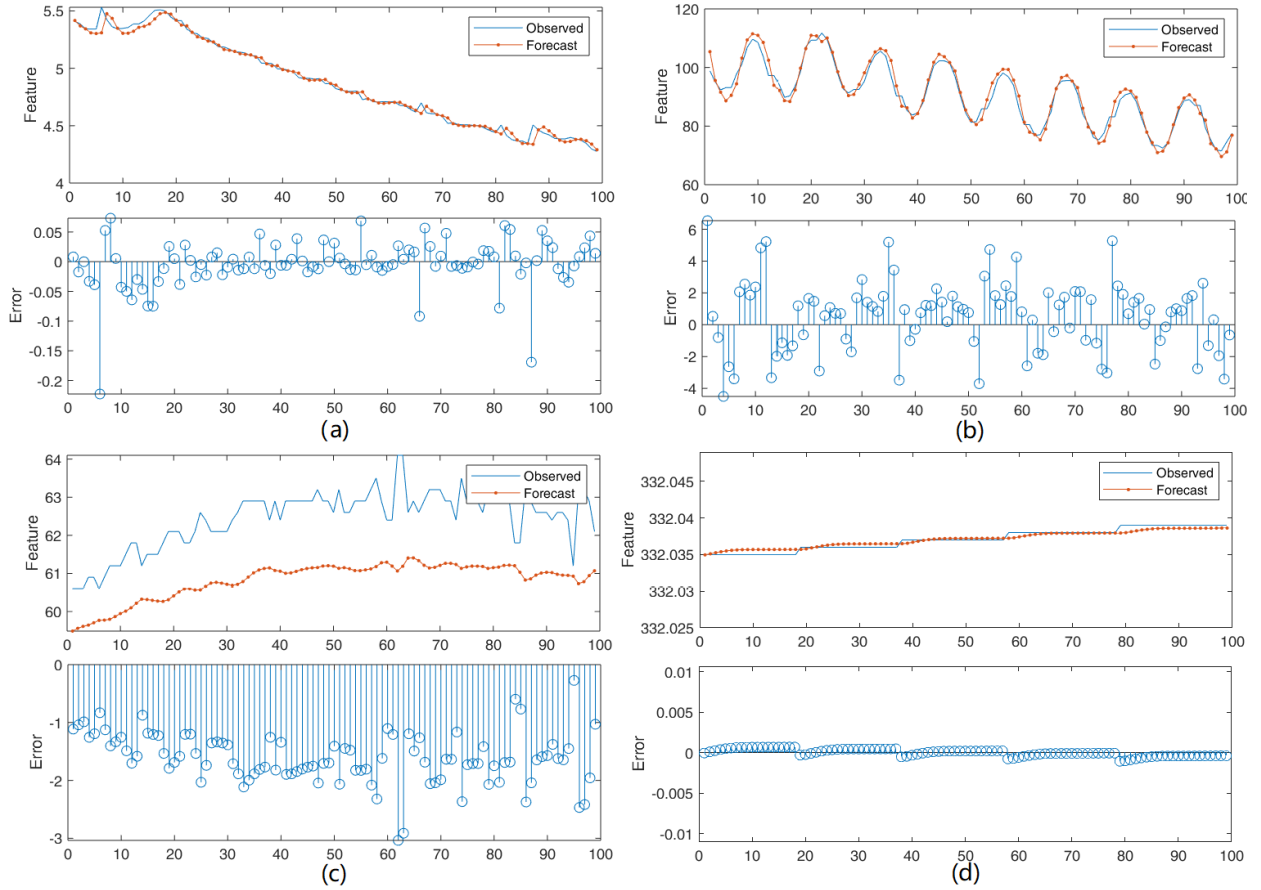


Figure 6. The actual values and predicted values produced by the LSTM model, (a) oil pressure, (b) rotor torque, (c) generator voltage and (d) accumulated power

separately. Figure 10 shows the prediction results of the individual variables. The blue dotted line in each subplot represents the test data while the red line denotes the predicted values. Evidently, the predicted values well match the measurement data, demonstrating the PCA captures the appropriate features from each signal and the LSTM performs a good performance in time-series power prediction. In particular, the oil pressure, rotor torque and accumulated power output have shown a better forecasting result compared with the generator voltage in term of both tendency and magnitude. The results show that the LSTM produces quite different predictions if the individual signal sequence is imported into the model possibly because the observation data of the voltage exhibit more complex features and variations during the whole timespan of observations.

Table III gives the scores of metrics for the model evaluation. The R-squared value for oil pressure, rotor torque and generator power reaches above 0.9, indicating a very good forecasting accuracy. The voltage performs worst among them, which needs further investigations.

D. Method comparison

For comparison, other AI methods are also performed using the same collected signals in order to test and prove the performance of the data-driven model proposed in this paper. The support vector machine (SVM), regression tree (RT), Gaussian process regression (GPR) and ensembles of trees (ET) are applied and the average metrics of performance are given in Table IV.

E. Discussions

Theoretically, the LSTM is able to remember the long and short states within the time series data and performs the prediction more accurately. Generally speaking, the LSTM models can keep track of arbitrary long-term dependencies in the input sequences and solve the vanishing gradient problem [17]. In some cases, if the sequential data have long-term relationships, the LSTM may not able to keep the long-term dependencies sufficiently to perform the prediction. Take the WEC into account, the wave resources, which are finally converted into electrical power, can be affected by a number of factors such as these operation parameters described above, among of which could have the long-term connection even for days and weeks when the WEC is in operation .

In view of the results, the prediction results are affected not only by data quantity but also by data quality. The LSTM seems to adapt perfectly to the treatment of the gentle and regular data than those tremble data, as can be seen from Figure 6. The actual values of the voltage are poorly matched with the predicted ones probably because the actual voltages oscillate irregularly and it is hard to follow the inner connections between former and later time-series data. Furthermore, the voltage signals from the WEC always behave irregularly and have not built close connections between neighboring observations in time. By contrast, the oil pressure and rotor torque signals change gradually due to the nature of its fluid and mechanical properties.

TABLE IV. THE PERFORMANCE METRICS COMPARISON WITH DIFFERENT METHODS

Methods	RMSE	MAE	R ²	Time(s)
SVM	17.679	12.784	0.86	426.9
RT	20.681	15.460	0.81	6.9
RPG	16.366	11.974	0.88	653.4
ET	17.444	13.417	0.87	11.9

There are two modes of data input to apply LSTM, namely, sequence-to-one and sequence-to-sequence respectively. The sequence-to-sequence mode deals with multiple time series inputs to predict multiple time series problems while the sequence-to-one mode is used to predict one time-series of observations.

In this case, the sequence-to-sequence mode gives acceptable results and therefore appears to have advantages in predicting parameters for later classification. The sequence-to-one mode is of cardinal significance in accurate regression and prediction of future time step of a sequence if a suitable input is given. Besides, it is appropriate for predicting the value of the next time step. If the actual values of time step between predictions become available, the network state can be undated with the observed values instead of the predicted values, thus eventually improving the accuracy.

V. CONCLUSIONS

This study proposes a deep learning network to deal with wave power forecasting by training and testing real signals collected from an open sea testing WEC. The LSTM is used to avoid long term independences during the forecasting while the PCA is used to extract principal components from a set of input signal sequences. A case study of forecasting wave power is presented based on the proposed PCA-LSTM model.

The results show that the proposed network is able to perform accurately prediction of the wave power parameters. The R-squared value reaches highest at 0.988 for oil pressure, demonstrating that the predication results are very well matched with the actual values. For comparison, traditional machine learning algorithms such as SVM, RT, RPG and ET are also applied and the results reveal the proposed network outperforms these machine learning algorithms and also consumes less computation time. The PCA-LSTM method can determine which features to remember and which needed to discard and hence performs forecasting properly both in short term and long term.

The work also raises some issues to be addressed during the future work. Due to the stochastic and intermittent nature of the wave energy, the power parameters seems to behave irregularly, such as high-frequency oscillations and long term dependency as shown in the voltage signals. Hence, future work is required to develop an updated and integrated LSTM model to tackle these problems, thus improving the robustness of the model.

ACKNOWLEDGMENT

This research is supported from SFMRE (Special Funds for Marine Renewable Energy) from MNR (Grant number

GHME2019ZC01) and S&T innovation project at NOTC (91700414)). The authors thank those colleagues for their valuable comments and suggestions that have helped improve the manuscript.

REFERENCES

- [1] Y. Saadat, N. Fernandez, A. Samimi, M.R. Alam, M. Shakeri, R. Ghorbani, "Investigating of helmholtz wave energy converter", *Renew. Energy*, Vol. 87, pp. 67–76, 2016.
- [2] International Energy Agency (IEA), "Ocean-potential", Archived from the original on 2015-05-22. Retrieved 2016-08-08.
- [3] C. Ni, X. Ma, Y. Bai, "Convolutional Neural Network based power generation prediction of wave energy converter", *the 24th International Conference on Automation & Computing*, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2018.
- [4] M. Browne, B. Castelle, D. Strauss, R. Tomlinson, M. Blumenstein, "Near-shore swell estimation from a global wind-wave model: spectral process, linear, and artificial neural network models", *Coast. Eng.*, Vol. 54, No. 5, pp. 445–460, 2007.
- [5] J.M. Torres, R.M. Aguilar, K.V.Z. Niga-Meneses, "Deep learning to predict the generation of a wind farm", *J. Renew. Sustain. Energy*, Vol. 10, 013305, 2018.
- [6] S. Shikhar, L. Stefan, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data", *Solar Energy*, Vol. 162, pp. 232-247, 2018.
- [7] Y. Huang, S. Liu, L. Yang, "Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM", *Sustainability*, Vol. 10, 3693, 2018
- [8] A. S. Qureshi, A. Khan, A. Zameer, A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning", *Applied Soft Computing*, Vol. 58, pp. 742–755, 2017.
- [9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M. Lew, "Deep learning for visual understanding: A review", *Neuro computing*, Vol. 187, pp. 27–48, 2016.
- [10] A. Felix, S. Jurgen, C. Fred, "Learning to forget: Continual Prediction with LSTM", *Neural Computation*, Vol. 12, No. 10, pp. 2451–2471, 2000.
- [11] Y. Wang, D. Xie, X. Wang, Y. Zhang, "Prediction of wind turbine-grid interaction based on a principal component analysis-long short term memory model", *Energies*, Vol. 11, 3221, 2018.
- [12] G. S. Afan, H. Yaya, A. Edi, S. Wayan, "Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting", *Procedia Computer Science*, Vol. 135, pp. 89–98, 2018.
- [13] J. Liu, T. Zhang, et al, "TD-LSTM: temporal dependence-based LSTM networks for marine temperature prediction", *Sensors*, Vol. 18, 3797, 2018.
- [14] J. H. Md, J. M. Kim, "Bearing fault diagnosis under variable rotational speeds using Stockwell transform-based vibration imaging and transfer learning", *Appl. Sci.*, Vol. 8, 2357, 2018.
- [15] K. Pearson, "On lines and planes of closest fit to systems of points in space", *Philosophical Magazine*, Vol. 2, No. 11, pp. 559–572, 2010.
- [16] W. Yang, R. Court, J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis", *Renewable Energy*, Vol. 53, pp. 365-376, 2013.
- [17] M. Buccino, D. Vicinanza, D. Salerno, D. Banfi, M. Calabrese, "Nature and magnitude of wave loadings at seawave slot-cone generators", *Ocean Eng.*, Vol. 95, pp. 34–58, 2015.