LSTM for Grid Power Forecasting in Short-Term from Wave Energy Converters

LSTM for Grid Power Forecasting in Short-Term from Wave Energy Converters

Rafael Natalio Fontana Crespo*, Alessandro Aliberti*, Lorenzo Bottaccioli*,
Enrico Macii*, Giorgio Fighera‡ and Edoardo Patti*
*Politecnico di Torino, Turin, Italy. Emails: name.surname@polito.it
‡ENI, Milan, Italy. Email: name.surname@eni.com

***Abstract*—In recent times, the consistent growth of wave energy makes it one of the most promising forms of renewable energy. Due to the intermittency and non-stationary nature of waves, the grid integration of these renewable energy sources involves a series of complex power conditioning stages to deliver grid electric power that meets the corresponding quality standards. Furthermore, to enable optimal management and operation of a smart grid power system, forecasting the wave power delivered to the grid is essential. In this paper, we present a novel approach based on Long Short-Term Memory Neural Network to forecast the wave power delivered to the grid of a Wave Energy Converter (WEC) - the ISWEC, which is a device able to harvest sea energy by exploiting the inertial effect of a gyroscope - in short-time horizons (e.g. 1min). The data for the analysis was obtained from a simulator that combines a model of the ISWEC device and the power conditioning grid integration for this particular WEC. In addition, to investigate the effectiveness of downsampling, we compared the performance behavior of the raw dataset and downsampled versions of it. The results showed that as the downsampling increases, so does the forecasting accuracy: the forecasting performance of the raw dataset returned the worst results, while the one of the dataset with the biggest downsampling studied returned the best.**

***Index Terms*—Renewable Energy, Wave Energy Converter, Energy Forecast, Neural Networks**

I. Introduction

The global rise in environmental issues and the drawbacks of using fossil fuels as energy sources have led to a significant increase in the utilization of renewable energy sources (RES). In recent times, numerous projects have been developed worldwide to address global warming, including the European Green Deal Project [1], which aims to fully replace all fossil fuels with renewable energy sources by 2050. Consequently, research on renewable energy is a rapidly growing field nowadays. Indeed, solar and wind energy technologies have undergone significant improvements in recent years, making them major contributors to the generation of renewable energy [2], [3]. Within this framework, wave energy has emerged as a serious contender in the renewable energy field [4] due to its high power density, potential [5], and greater predictability compared to solar and wind energy [6]. However, a Wave Energy Converter (WEC) is required to capture the kinetic energy of ocean surface waves and convert it, first, into mechanical energy in the Power Take Off (PTO) and later into electrical energy [5]. WECs are typically classified based on their installation location (i.e. onshore, nearshore, and offshore), size (i.e. point absorber, attenuator, and terminator), and operation principle (i.e. pressure differential, floating structures, overtopping, and impact devices) [7].

The increasing use of RES in recent years has led many countries to develop more advanced power grids, known as Smart Grids. Smart Grids can use network information to improve network stability and management [8]. However, as with other RES such as wind and solar energy, the power generated from wave energy fluctuates due to the intermittent and stochastic nature of waves [9]. This poses a significant challenge for the grid integration of WECs, requiring Smart Grids to be adjusted to overcome these fluctuations. Recent research has suggested that the intermittency problem of RES can be addressed by implementing smart system operating procedures such as state estimation [10] and real-time forecasting [11]. State estimation is the process of evaluating the state of the grid by comparing system measurements and predicted ones at a particular time step k. State estimation allows the detection of possible system failures and measurement errors. To estimate electric power generation - an important grid state variable - it is essential to forecast the power delivered from WECs.

In the wave energy field, physical models have been used to make predictions since the late 1960s. However, to reduce errors, these models have become increasingly complex over time, which has led to a significant increase in their computational cost [12]. One of the fields that have arisen and bloomed to make predictions in the area, with minimal time and cost, is the use of Machine Learning (ML) and Deep Learning (DL) algorithms [13]. These include the use of Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Recurrent Neural Networks (RNN). Recently, the Long Short-Term Memory (LSTM) method, which is derived from RNN, has demonstrated the ability to tackle a variety of sequential learning-related problems, making it one of the most suitable methods for addressing prediction tasks [14]. Several Authors [7], [13], [15]–[17] have implemented LSTM algorithms to forecast the generated power of WECs. In addition, in [13], [15] the original dataset resolution is reduced aggregating, i.e. downsampling [18], the signals in bigger time steps (e.g. 15min). Downsampling is a technique often used to smooth out noise and improve the quality of power signals. Nonetheless, to the best of our knowledge, there is no previous reference in the literature for the forecasting of the power-to-grid delivered of a WEC. The grid integration of a WEC is a

critical step in the wave-to-grid power generation process that present several difficulties (e.g. power conditioning processes) and must be taken into account. Yet, real-world data that accounts for this step is rare or does not exist.

The scientific novelty of our methodology is the use of a model that implements the complete wave-to-grid power generation process, capable of generating a realistic dataset of the power delivered to the grid by the Inertial Sea Wave Energy Converter (ISWEC), an inertial floating, offshore point absorber [19]. This wave-to-grid model is the combination of two previously validated models: the ISWEC model [19]–[22], and the power conditioning stage for this device [23]. What is more, our methodology comprises the application of LSTM algorithms to predict the delivered power to the grid in short-term horizons (1min). We designed and optimized LSTM neural networks by exploiting a dataset consisting of 30 hours of real-world data on the electric power delivered to the grid, collected with our wave-to-grid model. Last but not least, to investigate the effectiveness of the usage of downsampling for wave power forecasting, we compare the prediction performance obtained using raw data and aggregating data in different time steps.

The rest of the paper is organized as follows. Section II reviews the literature solution for Wave Energy Forecasting. Section III presents the proposed methodology for power-to-grid forecasting of the ISWEC device. Section IV discusses the experimental results. Finally, Section V discusses the conclusion remarks.

II. Related Work

For efficient grid integration, wave energy should be estimated and forecasted through proper methods. Consequently, numerous studies were proposed in the literature. Several authors [7], [12], [15], [16], [24]–[26] implement different supervised Machine Learning techniques using input parameters from the WEC. [12], [16], [24] used as input the position (Latitude and Longitude) and the power outputs of an array of WECs of 4 different wave farms to forecast the power output. [12], [24] implemented different configurations of a simple Multi-Layer Perceptron model for predicting the output power of each farm separately. Different accuracies were obtained depending on the nature of each wave regime. In [16], a comparison between different kinds of LSTM models (Vanilla LSTM, Stacked LSTM, Convolutional Neural Network-LSTM, and Bi-Directional LSTM) is performed. The Bi-Directional model showed the best accuracy.

In literature, a common practice is to try to reduce the difficulty of the forecasting problem by studying the correlation between the input data employing Principal Component Analysis (PCA). In [7], a sequence-to-sequence LSTM neural network is implemented, using as input the feature parameters obtained with PCA that represent 90% of the original information. The model is then compared with other traditional ML algorithms such as Support Vector Machine (SVM), Regression Tree (RT), Gaussian Process Regression (GPR) and Ensembled Trees (ET). The results showed that the developed model outperformed the other algorithms. Moreover, [15] used the same inputs, but implemented a sequence-to-one architecture to predict the output power generation of a WEC. Different models were tested (SVM, Neural Networks and LSTM), being the LSTM neural network the most accurate. In [25], the input dataset consists of the accelerations and angular velocities among the 3 axes of the WEC as well as the magnitudes. The data was then downsampled into different window sizes. These features were input to PCA to reduce the dimension of the dataset. Two different approaches were developed for the forecast: one implementing a classifier [27] plus a prediction model, and the other just the prediction model. The prediction models tested were: SVM, Random Forest (RF), and Artificial Neural Networks (ANN). The results showed that the models implementing a classifier combined with the regressor outperformed the others. What is more, the model's accuracy varies according to the window size adopted.

In [26], a Neural Network Autoregressive with Exogenous Input (NNARX) was implemented to predict the wave-induced excitation torque of the ISWEC device, using as input the angular speed of the PTO, the angular speed of the hull and the PTO torque.

Other studies were inspired by the correlation of wave energy power and weather conditions (e.g. wave height and wave period), employing weather-related exogenous inputs for their models. In [28], significant wave height is used as input data and, after a conversion to power using a device power matrix, employed as input to a Non-Linear Autoregressive Recurrent Neural Network. In [13], an LSTM neural network is developed for forecasting the power output of a WEC using as exogenous inputs wave height and wind speed. The results showed a clear relationship between wind speed and the output power of the studied WEC. In [17], the input is composed of wave height, wave direction, wave energy period and wave energy flux. Different ML algorithms were tested, including adaptive neuro-fuzzy inference system (ANFIS), feed-forward neural network (FFNN), polynomial neural networks (PNN), Vanilla LSTM, stacked LSTM, and Bi-LSTM. This last outperformed the other algorithms. In addition, to identify the most dominant wave parameters for predicting the energy flux, different input combinations were tested, giving as best input parameters the wave height and wave direction.

In literature, the proposed models considered the forecasting problem of the instantaneous output power of the WEC PTO, neglecting the grid integration of the device. This last is a critical stage in the wave-to-grid power conversion process and cannot be neglected. Nonetheless, real-world data for this case is rare or does not exist. Our methodology presents a model that implements the whole wave-to-grid power conversion process of the ISWEC device and generates a realistic dataset of the power delivered to the grid. In addition, our methodology comprises an LSTM neural network for forecasting the delivered power to the grid in the short-term (1 min) to meet the requirements of further algorithms needing these results, such as state estimation. Lastly, to study the effectiveness of the usage of downsampling for forecasting the wave power

delivered to the grid, we compare the prediction performance obtained using raw data and aggregating the data in different time steps.

III. Methodology

In this section we present our methodology, which schema is shown in Fig. 1. To the best of our knowledge, real-world data for the case under study is rare or does not exist. Consequently, we simulate the behaviour of the ISWEC and generate a realistic dataset using the *ISWEC Wave-to-Grid Simulator*. This block combines two models: the *ISWEC Wave-to-PTO model* [19]–[22] and the *PTO-to-Grid Electric System model* [23]. The output of this block is the electric power delivered to the grid by this particular WEC. We use this power as input for our forecasting algorithms. Then, the dataset is split into *training* and *test set*. The *training set* is employed to train the *LSTM models*. To study the effectiveness of downsampling, we implement two approaches: using the raw data as input and downsampling the dataset in different time steps before feeding the ML algorithms. During the *test phase*, the *trained LSTM models* are fed with the *test set* accordingly (raw or downsampled data). Lastly, the *trained LSTM models* are evaluated and compared based on their prediction performance on the *test set*.

A. ISWEC Wave-to-grid Simulator

ISWEC is a gyroscope-based floating WEC [19]. Externally, it appears as a monolithic hull constrained to the seabed by a slack (geographic) mooring. The ISWEC's gyroscope mechanism is its main component. A gyroscopic torque is generated around the PTO axis due to the combination of the flywheel's rotational speed, and the wave-induced pitching motion of the floating hull. The PTO can use this torque to generate electrical power [29].

We use the *ISWEC Wave-to-grid Simulator* to simulate the behaviour of the ISWEC and collect the power delivered to the grid. This block consists of two models in literature: the *ISWEC Wave-to-PTO model* [19]–[22] and the *PTO-to-Grid Electric System model* [23].

The *ISWEC Wave-to-PTO model* comprises two main phenomena: the hull hydrodynamics and the mechanics of the gyroscope. The model's input are wave forces, which force the dynamics of the whole system, including the hull. The hull kinematics combined with the gyroscopic effect known as gyroscopic precession generates a torque in the PTO shaft which is converted first into mechanical power and later into electrical power (more details in [19]–[22]).

The *ISWEC Wave-to-PTO model* is then combined with the *PTO-to-Grid Electric System model*. This system can smooth the oscillating power coming from the PTO into a more steady electric power flux to be later delivered to the grid (more details in [23]). The whole system outputs the power delivered to the grid that has been used to generate a realistic dataset (see next Section III-B).

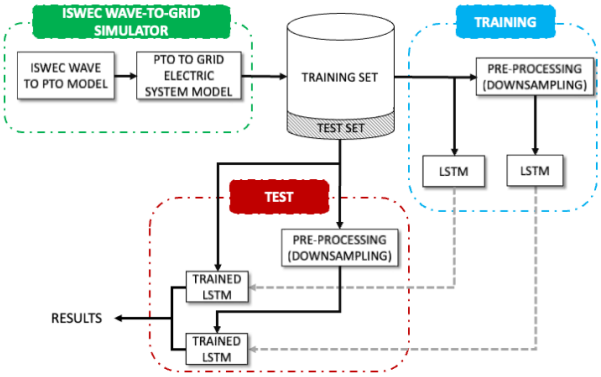


Fig. 1. Schema of the proposed methodology

TABLE I
Training and test set size of the different datasets

	Dataset					
	Raw Data	**D-1s**	**D-2.5s**	**D-5s**	**D-10s**	**D-30s**
Training Set	810000	81000	32400	16200	8100	2700
Test Set	270000	27000	10800	5400	2700	900

B. Dataset and Pre-processing

We exploit a dataset sampled every 0.1s (10 Hz) for 30 hours of delivered power to the grid obtained with the *ISWEC Wave-to-Grid simulator*. The original dataset contains measurements of the device's surge, sway, heave, roll, pitch and yaw position, angular velocities and accelerations - totalling 18 variables - and the delivered power to the grid. For the case under study, we only consider the delivered power to the grid, resulting in 1080000 samples. This work aims to compare the forecasting 1 minute ahead of the delivered power using raw data versus downsampled versions of the same dataset. For downsampling the dataset, we aggregate the data in bigger time steps using the sliding non-overlapping moving average window technique. Regarding the window size - i.e. the time steps in which the data is temporally aggregated - there is not a general agreement on which one to use [30]. We downsampled the data in time steps of 1s, 2.5s, 5s, 10s and 30s. We refer to these datasets as *D-1s*, *D-2.5s*, *D-5s*, *D-10s* and *D-30s*, respectively.

Moreover, to evaluate the prediction performance of the proposed models, we split the dataset of each case into *training* (75% - first 22.5 hours) and *test set* (25% - last 7.5 hours). Furthermore, we use the last 10% of the *training set* for validation during the *training phase*. Table I shows the details of the number of data points for the *train* and *test phases* of the raw and the downsampled datasets. As it can be seen, for increasing time steps the number of data points diminishes, being the *D-30s* dataset the one with least amount of data points both for *training* and *testing*. Its number of data points is 300 times smaller than the ones of the *Raw Dataset*.

Last but not least, feature scaling is required for optimizing and speeding up the training process of ML algorithms [31]. Consequently, we scaled the input dataset (power delivered to the grid) in a range between 0 and 1 using *Min-Max* normalization:

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

TABLE II
SELECTED MODELS

Dataset	Hidden Layers	Batch Size	Time Lag	Neurons Layer 1	Neurons Layer 2	Train Time [s]
Raw Data	2	128	60	64	64	401.45
D-1s	2	64	120	32	32	350.58
D-2.5s	2	16	48	16	16	153.24
D-5s	1	16	48	64	-	126.98
D-10s	2	32	12	64	64	18.57
D-30s	2	8	8	128	64	15.17

where *x* is the vector of values to be scaled, *min(x)* and *max(x)* are the minimum and maximum values of that vector.

C. Long Short-Term Memory Neural Network

The Long Short-Term Memory Neural Network (LSTM) is a RNN developed to solve the "vanishing gradient" problem [32]. Due to its structure, the LSTM presents a significant well performance for time-series forecasting, since it can make connections over 1000 time steps as well as prevent errors from backpropagating between time and layers [33]. For the hidden layer, we decided to use the hyperbolic tangent ($tanh$) activation function because it is a popular choice and provides good results [34]. Instead, for the output layer, we used a linear activation function. The multi-step ahead prediction is implemented with a multi-output approach [35], i.e. a single network with *n* outputs, where *n* is the number of steps ahead to predict. For the different datasets (raw and downsampled ones), the prediction horizon is equal to one minute.

For the general architecture of the network, we considered different configurations employing one or two hidden layers as done in the literature [7], [15], [16]. Regarding the hyperparameters, i.e. batch size, number of regressors (time lag), and number of LSTM units in the hidden layer, there is no established mathematical model to determine the ideal parameters. Therefore, we adopted a trial-and-error approach to find the best hyperparameters. For the batch size, we investigated values ranging from 4 to 128. Furthermore, we studied the number of regressors from 2 to 120. Moreover, for the number of LSTM units, we searched in the range of 2 to 192. For the cases implementing two hidden layers, we arbitrarily decided to consider two different versions for the second hidden layer: one implementing the same number of units as in the first hidden layer, and the other with half the units. The optimization algorithm used for training in all cases is the Adaptive Moment Estimation (Adam Optimizer) [36]. In addition, to prevent overfitting, we employed an early-stopping strategy [37]. Early-stopping is a technique in which, during the training phase, it interrupts the training if there is no improvement in the validation set after a certain number of steps. Apart from preventing overfitting, early-stopping also reduces significantly the training time.

The hyperparameters of the selected model for each dataset are reported in Table II. As presented, all the models employ two hidden layers except for the *D-5s* model. For this reason, this model only implements neurons in the first layer. Except for the *D-10s*, it can be seen a decreasing trend in the batch size for the selected models as the downsampling increases, which can be related to the decreasing size of the datasets.

IV. RESULTS AND DISCUSSION

In this section, we present the results obtained with the proposed methodology. Firstly, we explain the statistical indicators used to analyze and compare the different models. Then we describe the results obtained for the different datasets. To train and validate the models, we run our simulations in a PC equipped with a CPU Intel I3-9100F 4x3.60GHz, GPU Nvidia Quadro P2200 and 32 GB of RAM.

A. Performance Metrics

To evaluate the performances of the different implemented models, we employed three metrics commonly used in literature to measure the similarities between predicted and observed time series [38]: i) *Mean Absolute Difference* (MAD) between predicted and observed values; ii) *Root Mean Square Deviation* (RMSD) measures the standard deviation of the difference between the predicted and the observed values; iii) *Coefficient of determination* (R^2) determines the percentage of the variance in the observed values that is explained by the predicted ones.

The three metrics mathematical expressions are described by the following equations:

$$MAD = \frac{100}{\overline{y}_{test}} \frac{\sum_{i=1}^{n} |y_{pred,i} - y_{test,i}|}{n} \tag{2}$$

$$RMSD = \frac{100}{\overline{y}_{test}} \sqrt{\frac{\sum_{i=1}^{n} (y_{pred,i} - y_{test,i})^2}{n}} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{test,i} - y_{pred,i})^2}{\sum_{i=1}^{n} (y_{test,i} - \overline{y}_{test})^2} \tag{4}$$

where y_{pred} are the predicted values, y_{test} are the observed values, n is the number of predictions, and $\overline{y}$ indicates the mean value. RMSD and MAD are expressed in percentage. A lower value for RMSD and MAD indicates a lower error, and thus, better performance. Instead, R^2 determines the correlation between observed and predicted values, where a value of 1 denotes a complete correlation, whereas smaller values indicates a weaker correlation.

B. Model Evaluation

For the LSTM neural network of each dataset - i.e. raw and downsampled in different time steps - we chose the model showing the best performance. Then, we compared the model of each dataset with the other models according to their prediction performance in the test set.

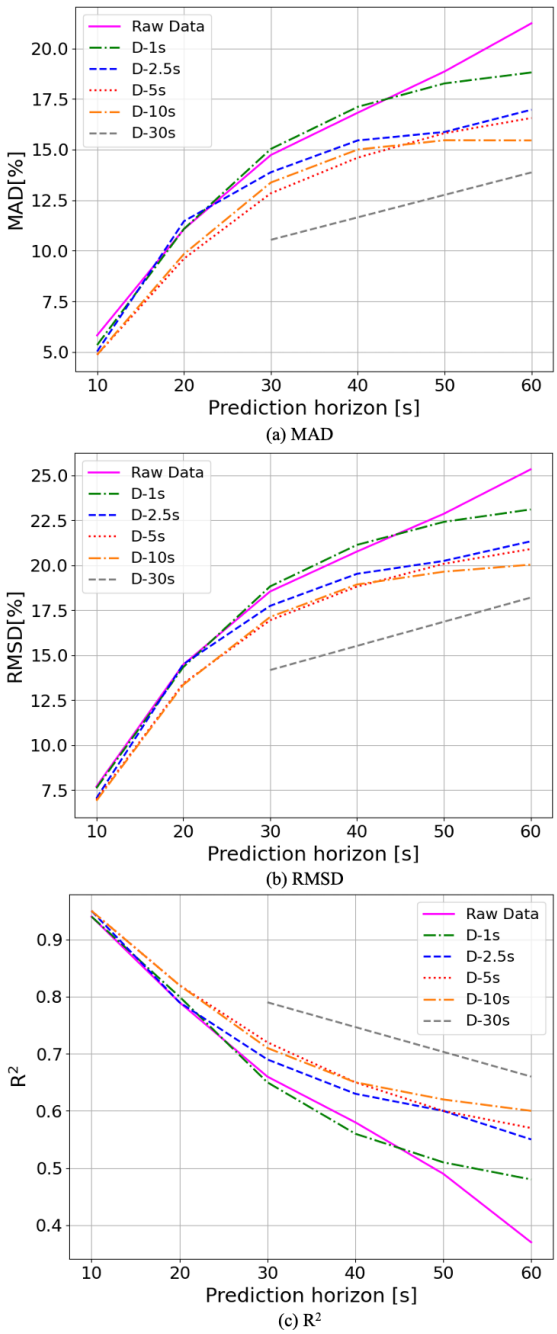
The prediction errors, based on the metrics previously introduced, are visually compared in Fig. 2 and reported in Table III. As expected, the three metrics present a growing trend in the model's forecasting errors, i.e. the prediction errors grow together with the forecasting horizon. The *D-30s* model outperforms the other models for the three metrics. It is important to highlight that, for this model, it was only possible to compare the results for a prediction horizon of 30s and 60s due to the time step size (30s).

TABLE III
MODELS PERFORMANCE BASED ON MAD[%], RMSD[%] AND R^2

Metric	Dataset	Prediction Horizon [s]					
		10	20	30	40	50	60
MAD	Raw Data	5.82	11.07	14.73	16.82	18.86	21.24
	D-1s	5.34	11.08	15.03	17.11	18.27	18.81
	D-2.5s	5.02	11.46	13.88	15.45	15.87	16.97
	D-5s	**4.86**	**9.61**	12.84	**14.60**	15.81	16.56
	D-10s	**4.86**	9.84	13.37	15.00	**15.46**	15.45
	D-30s	-	-	**10.54**	-	-	**13.87**
RMSD	Raw Data	7.69	14.47	18.54	20.76	22.86	25.33
	D-1s	7.60	14.33	18.82	21.13	22.41	23.10
	D-2.5s	7.04	14.49	17.74	19.52	20.24	21.33
	D-5s	6.96	13.42	16.94	**18.81**	20.08	20.90
	D-10s	**6.90**	**13.35**	17.12	18.93	**19.64**	20.03
	D-30s	-	-	**14.17**	-	-	**18.20**
R^2	Raw Data	0.94	0.79	0.66	0.58	0.49	0.37
	D-1s	0.94	0.80	0.65	0.56	0.51	0.48
	D-2.5s	**0.95**	0.79	0.69	0.63	0.60	0.55
	D-5s	**0.95**	**0.82**	0.72	**0.65**	0.60	0.57
	D-10s	**0.95**	**0.82**	0.71	**0.65**	**0.62**	0.60
	D-30s	-	-	**0.79**	-	-	**0.66**

In the very short term (10s), the prediction errors of the models are similar, particularly for the R^2 metric, where the results are around 0.94-0.95. The performance of the *Raw Data* and the *D-1s* models in terms of the three metrics are quite similar for short prediction horizons (10s and 20s). As the prediction horizon increases, the performance of the *Raw Data* and the *D-1s* models decreases considerably. The *Raw Data* model performance degradation above the 50s prediction horizon is notorious in terms of the three metrics, presenting the worst performance for the largest prediction horizon studied (60s): MAD - 21.24%, R^2 - 0.37 and RMSD - 25.33%. The *D-2.5s* model presents the worst performance degradation among all models between the 10s and 20s forecasting horizon, being the worst performing model for a forecasting horizon of 20s according to the three metrics (clearly depicted in the MAD plot in Fig. 2): MAD - 11.46%, R^2 - 0.79 and RMSD - 14.49%. Although its performance decreases for larger prediction horizons, it closes the error gap with the other datasets models. For the 50s prediction horizon, the performance of the *D-2.5s* model in terms of R^2 is the same compared to the *D-5s* with a value of 0.6. Moreover, the *D-5s* model error in terms of RMSD and R^2 is almost equal to the *D-10s* model until a prediction horizon of 40s. Instead, in terms of MAD, the *D-5s* model slightly outperformed the *D-10s* model until the prediction horizon of 40s. For larger prediction horizons (50s and 60s), the performance of the *D-10s* model in terms of the three metrics is better than the *D-5s* model. However, the significantly less training time of the *D-10s* model (18.57s) versus the *D-5s* model (126.98s) makes the first a good alternative.

Last but not least, as the downsampling increases, so the performance of the models does. This difference is more evident for larger prediction horizons (50s - 60s). In addition, the training time also decreases for higher downsampling rates (see Table III): there is a significant improvement for the *D-10s* and *D-30s* models compared to the others.



(a) MAD
(b) RMSD
(c) R^2

Fig. 2. Comparison of models based on MAD, RMSD and R^2

V. CONCLUSIONS

In this work, we presented a novel methodology capable of generating a realistic dataset of the power delivered to the grid of the ISWEC device. In addition, it predicts this quantity in short-time horizons up to 1 minute using an LSTM neural network. Furthermore, we investigated the effectiveness of the downsampling technique for forecasting the power delivered to the grid. To this end, we compared the performance of an LSTM model employing the raw dataset as input, and others using downsampled versions of it. For each model, we search for the best hyperparameters. The results highlighted that the

forecasting performance increases for bigger downsamplings. In addition, the biggest downsampling tested (*D-30s*) significantly outperformed the other models. On the other hand, the *Raw dataset* model showed similar results compared to the other models for very short-term prediction horizons (10s and 20s). However, its performance decreased considerably for bigger horizons, especially above 50s where its performance degradation was notorious.

For wave energy to consolidate as a reliable contender in the RES field, the prediction of the power delivered to the grid is an important aspect. Our work presented the first step in this area. Further works may include the study and comparison of the results obtained with other downsampling techniques (e.g. exponential weighting) as well as the implementation of bigger time steps aggregations for longer forecasting horizons. What is more, we aim to identify other relevant input variables that may enhance the forecast. In addition, once real-world data become available, we plan to replace the simulated data and, by exploiting also transfer learning techniques, make the forecasting. To this end, we plan to use the models pre-trained with the simulated dataset, decreasing the amount of real-world data required.

ACKNOWLEDGMENT

This work was partially funded by ENI S.p.A and by the National Recovery and Resilience Plan (NRRP) from Ministero dell'Università e della Ricerca (MUR - Grant number PE0000021)

REFERENCES

[1] "A european green deal." [Online]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

[2] A. Clement *et al.*, "Wave energy in europe: current status and perspectives," *Renewable and Sustainable Energy Reviews*, vol. 6, no. 5, pp. 405–431, 2002.

[3] I. Lopez *et al.*, "Review of wave energy technologies and the necessary power-equipment," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 413–434, 2013.

[4] P. A. Østergaard *et al.*, "Sustainable development using renewable energy technology," *Renewable Energy*, vol. 146, pp. 2430–2437, 2020.

[5] B. Jiang *et al.*, "Performance analysis and tank test validation of a hybrid ocean wave-current energy converter with a single power takeoff," *Energy Conversion and Management*, vol. 224, p. 113268, 2020.

[6] G. Reikard, B. Robertson, and J. R. Bidlot, "Combining wave energy with wind and solar: Short-term forecasting," *Renewable Energy*, vol. 81, pp. 442–456, 2015.

[7] C. Ni, X. Ma, and J. Wang, "Integrated deep learning model for predicting electrical power generation from wave energy converter," in *Proc. of ICAC 2019*. IEEE, 2019, pp. 1–6.

[8] L. T. Berger and K. Iniewski, *Smart grid applications, communications, and security*. John Wiley & Sons, 2012.

[9] G. Reikard, "Integrating wave energy into the power grid: Simulation and forecasting," *Ocean Engineering*, vol. 73, pp. 168–178, 2013.

[10] A. Fiaz *et al.*, "Distribution system state estimation-a step towards smart grid," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2659–2671, 2018.

[11] H. Sugihara, T. Funaki, and N. Yamaguchi, "Evaluation method for real-time dynamic line ratings based on line current variation model for representing forecast error of intermittent renewable generation," *Energies*, vol. 10, no. 4, 2017.

[12] D. Nalamati, "Forecasting power output of wave farm using machine learning: Multilayer perceptron."

[13] S. M. Mousavi, M. Ghasemi, M. D. Manshadi, and A. Mosavi, "Deep learning for wave energy converter modeling using long short-term memory," *Mathematics*, vol. 9, no. 8, p. 871, 2021.

[14] S. Srivastava and S. Lessmann, "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.

[15] C. Ni, "Data-driven models for short-term ocean wave power forecasting," *IET Renewable Power Generation*, vol. 15, no. 10, pp. 2228–2236, 2021.

[16] D. Nalamati, "Forecasting power output of wave farm using machine learning: Lstm model," 2021.

[17] M. Neshat *et al.*, "Wave power forecasting using an effective decomposition-based convolutional bi-directional model with equilibrium nelder-mead optimiser," *Energy*, vol. 256, p. 124623, 2022.

[18] 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 *2014 IEEE PES General Meeting— Conference & Exposition*. IEEE, 2014, pp. 1–5.

[19] G. Bracco, E. Giorcelli, and G. Mattiazzo, "Iswec: A gyroscopic mechanism for wave power exploitation," *Mechanism and Machine Theory*, vol. 46, no. 10, pp. 1411–1424, 2011.

[20] G. Bracco *et al.*, "Experimental validation of the iswec wave to pto model," *Ocean Engineering*, vol. 120, pp. 40–51, 2016.

[21] A. Battezzato, G. Bracco, E. Giorcelli, and G. Mattiazzo, "Performance assessment of a 2 dof gyroscopic wave energy converter," *Journal of Theoretical and Applied Mechanics*, vol. 53, no. 1, pp. 195–207, 2015.

[22] G. Vissio *et al.*, "Iswec linear quadratic regulator oscillating control," *Renewable Energy*, vol. 103, pp. 372–382, 2017.

[23] G. Vissio, "Iswec toward the sea," Ph.D. dissertation, PhD thesis, Politecnico di Torino, 2017.

[24] B. Burramukku, "Estimator model for prediction of power output of wave farms using machine learning methods," *arXiv preprint arXiv:2011.13130*, 2020.

[25] H. M. Deberneh and I. Kim, "Predicting output power for nearshore wave energy harvesting," *Applied Sciences*, vol. 8, no. 4, p. 566, 2018.

[26] N. Pereira, D. Valério, and P. Beirão, "Control of a wave energy converter using a multi-agent system and machine learning methods," in *Proc. of RENEW 2018, October 8-10, 2018, Lisbon, Portugal*, p. 387.

[27] H. M. Deberneh and I. Kim, "Development of monitoring and classification systems for wave energy," *International Journal of Control and Automation*, vol. 11, pp. 57–66, 01 2018.

[28] 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 *2014 IEEE PES General Meeting— Conference & Exposition*. IEEE, 2014, pp. 1–5.

[29] G. Bracco, E. Giorcelli, G. Mattiazzo, V. Orlando, and M. Raffero, "Hardware-in-the-loop test rig for the iswec wave energy system," *Mechatronics*, vol. 25, pp. 11–17, 2015.

[30] O. Banos, J. M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.

[31] M. Castangia, A. Aliberti, L. Bottaccioli, E. Macii, and E. Patti, "A compound of feature selection techniques to improve solar radiation forecasting," *Expert Systems with Applications*, vol. 178, p. 114979, 2021.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[33] Y. Guo *et al.*, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[34] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized mlp architectures of neural networks," *International Journal of Artificial Intelligence and Expert Systems*, vol. 1, no. 4, pp. 111–122, 2011.

[35] D. M. Kline, "Methods for multi-step time series forecasting neural networks," in *Neural networks in business forecasting*. IGI Global, 2004, pp. 226–250.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[37] P. Lutz, "Early stopping-but when," *Neural Networks: Tricks of the trade*, pp. 55–69, 1998.

[38] C. A. Gueymard, "A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 1024–1034, 2014.