Multi-step Forecasting of Wave Power Using a Nonlinear Recurrent Neural Network

Kostas Hatalis, Parth Pradhan, Shalinee Kishore, Rick S. Blum
Electrical and Computer Engineering Department
Lehigh University, Bethlehem, PA

Alberto J. Lamadrid
Economics Department
Lehigh University, Bethlehem, PA

Abstract—Short term forecasting is a vital interest to future implementations of a smart grid, particularly in the reliable integration of renewable energy resources. In this study we focus on multi-step prediction of high resolution wave power. Significant wave height data was first obtained from Belmullet Berth, Ireland and underwent several data preprocessing steps. These include a linear interpolation to fill irregular or missing data points, conversion to power using an interpolated power matrix of a Pelamis Device energy converter, and then exponential smoothing is applied. We utilized a nonlinear autoregressive recurrent neural network for 3, 6, 12 and 24 hour prediction. Our method showed highly accurate results when data has been smoothed, versus raw data, and when compared to previous studies.

Index Terms—Short Term Forecasting, Ocean Wave Power, Artificial Neural Networks, Renewable Energy Integration.

I. INTRODUCTION

A major issue in integrating electricity from renewable sources into power grids is short term forecasting. Due to the volatility in the production of power from renewables such as wind or wave energy, gaps are left in the supply which must be filled by dispatchable resources. Particularly in competitive electricity markets, accurate predictions are of great importance for a variety of reasons. Short-term forecasts are then needed to identify reserves and balancing requirements in deregulated markets. A correct estimate can help to develop well-functioning multi-settlement markets for pricing and selling of electricity.

In this study we focus on wave power. While generally more predictable than other renewable energy sources (RES, e.g. wind), the intermittency of wave energy is a big challenge to implementing this resource as a reliable autonomous source of electric power. Large-scale wave-penetration requires answers to a variety of problems involving real-time grid operations, standards of interconnection, quality of power, capacity of the transmission system and stability and reliability of the power system. But a more pressing issue is the prediction of future levels of wave power for operational scheduling. With the deployment and utilization of more cognizant networks and smart grids, the two way communication necessary to integrate and respond to disturbances from unforeseen events and changes in RES availability can be enhanced. This integration into the existing and new cyber infrastructure can then benefit from machine learning algorithms that provide actionable information in this data-rich environment. Other goals of this vision include the creation of intelligent electricity networks that use energy resources efficiently, reduction in carbon emissions, and increasing robustness of the overall electricity delivery system to prevent failures.

This paper presents a technique to conduct short term multi-step forecasts of wave power directly from observed ocean wave heights. This is achieved through the use of an artificial neural network to recognize hidden patterns in past data and accordingly make a prediction of future values. The rest of this paper is organized as follows. Section II provides a brief overview of different model-based and time series-based prediction techniques applied to renewable prediction. In section III we describe our neural network model while in section IV we outline the experimental setup and data sources. Section V discusses our results and finally conclusions are drawn and future work is proposed in Section VI.

II. BACKGROUND

A number of different forecasting methods have been applied to predict different aspects important to the smart grid, especially as it relates to the goal of reliable RES integration. Diverse domains where short term forecasting is applied include ocean wave characteristics, wind speed and power, electric load, and solar irradiance. Regarding ocean waves specifically, conventionally physical models have been utilized to predict wave heights [1] but most recently a move to using past time series data to predict future heights has seen an increase in popularity. Various methods have been tested from regression to neural networks models. Much of the motivation for our research focuses on the analysis of the characteristics of the power output generated by wave converters. Previously, prediction of wave elevation levels was done by [2] for real time control of wave energy converters. The authors tested several prediction methods
including cyclical models, extended Kalman filters, and autoregressive models, [3] looked at time series and physics-based forecasting models for analyzing wave energy fluxes and power output series of different converters.

Point forecast time-series-based models like the auto-regressive (AR), auto-regressive moving average (ARMA), and auto-regressive integrated moving average (ARIMA) have also been found suitable for prediction in a number of different areas. AR and ARMA models specifically are appropriate for stationary time series, while ARIMA and other models such as Kalman filters are aimed at non-stationary data and long-term series. These models therefore have also found widespread use in the simulation and estimation of the wave characteristics based on historical data [4]. Alternatively, artificial neural networks (ANN) have been found equally useful and even posed to outperform ARMA models [5]. In an ANN, forecasting is done by feeding a sequence of previous observations to the model as input so that it can recognize hidden patterns in the series and consequently estimate future values. Neural networks have been applied to study aspects of waves such as predicting the height of a wave at the time of its breaking and the depth of water in which it breaks [6].

Of related interest, significant wave height (SWH) is defined as the average height of the highest one-third waves over the observation interval. Using an ANNs non-linear technique of modeling, real-time forecasting of SWH was studied over several short lead times using relatively new soft computing tools such as genetic programming [7]. Our paper deals with the forecasting of exponentially smoothed wave power directly converted from measured SWH. Power is forecasted using multilayered recurrent neural networks. Specifically, a nonlinear autoregressive model is used with recurrent algorithms in a backpropagation neural network to predict wave power for lead times of 3, 6, 12 and 24 hours.

III. NEURAL NETWORK MODEL

Varieties of different ANNs have been successfully applied for time series prediction and have been found particularly useful for cases in which the time series exhibits chaotic behavior [8]. A popular ANN model for time series prediction is the Nonlinear Auto Regressive model with eXogenous inputs (NARX) neural network [9]. This is a recurrent dynamic network with feedback connections enclosing the layers along with embedded memory (tapped time delay). It is a combination of a multilayered perceptron, a simple recurrent network, and a feedforward backpropagation network. The output of the NARX network is fed back into the input of the model and can be seen as an estimate of an output of a nonlinear dynamic system. The NARX network has been applied before for electrical load forecasting [10] and even for prediction of SWH involving 3h and 6h lead periods [11].

For our prediction we differentiate ourselves by first utilizing a variation of the NARX net known as a Nonlinear Autoregressive (NAR) neural network. Second, we predict high resolution (ten minute intervals) exponentially smoothed values of wave power instead of SWH, and third we predict values at multiple time steps ahead into the future instead of one time step.

Below we first describe the structure of the NARX model:

\[ y(n) = f(x(n-1), x(n-2), \ldots, x(n-m_x), \ldots, \hat{x}(n-1), \hat{x}(n-2), \ldots, \hat{x}(n-m_{\hat{x}})) \]

where \( x(n) \) and \( \hat{x}(n) \) symbolize the input and output functions of the neural network at time \( n \), while \( m_x \) and \( m_{\hat{x}} \) represent the input and output memory order, and \( f \) is a nonlinear function. The function \( f \) is approximated by a multilayered perceptron. The output at time-step \( n \) is dependent on its past \( m_{\hat{x}} \) values and the past \( m_x \) input values.

The network uses a time-delayed (TDL) architecture with a feedback connection from the output layer to the input layer of the network.

Training for a NARX network can be done in one of two modes. The first one is a parallel mode in which the output is fed back to the input of the neural network with a tapped delay line. The second one is called series-parallel mode in which the actual output is used versus feeding back the estimated output. States in the NARX network are updated as

\[
\begin{align*}
x_i(n+1) = & \begin{cases} 
u(n) & i = m_x \\ y(n) & i = m_x + m_y \\ x_{i+1}(n) & 1 \leq i < m_x \text{ and } m_x < i < m_x + m_y \end{cases} \\
\end{align*}
\]

so that at time \( n \) the tap delays correspond to the values

\[ x(n) = \{ u(n-m_x), \ldots, u(n-1), y(n-m_y), \ldots, y(n-1) \} \]
The MLP approximation of the function $f$ consists of a set of nodes organized into two layers. There are $H$ nodes in the first layer and they performs the function

$$z_i(n) = \sigma \left( \sum_{j=1}^{N} a_{ij} x_j(n) + b_{ij} u(n) + c_i \right) \quad i = 1, \ldots, H$$

where $\sigma$ is the nonlinear transfer function, $a_{ij}$, $b_{ij}$, and $c_i$ are fixed real valued weights. The output layer consists of a single linear node

$$y(n) = \sum_{j=1}^{H} w_{ij} z_i(n) + \theta_i$$

where $\theta_i$ and $w_{ij}$ fixed real valued weights. An example diagram of a NARX network with $m_y = m_z = 2$ and $H = 6$ is shown in Fig 1.

An advantage of the NARX network is that it can converge faster and generalize better than other networks [8]. Our NAR network has an input layer of neurons, an output layer, and one hidden layer. The Lavenberg-Marquardt training algorithm for the NAR model is used with standard backpropagation. Unlike the NARX network described, we do not have exogenous inputs. So during testing, predicted values at each time step are the only inputs fed to the network.

IV. EXPERIMENTAL SETUP

The experimental setup involved a number of data preprocessing steps. After receiving data from the Irish Marine Institute, we applied a linear interpolation to fill in missing data, split the data to ten minute intervals, converted the SWH data to power (kW) using an interpolated power matrix, and then applied exponential smoothing. Results were also taken without smoothing for comparison. The following section describes in detail our data source and these preprocessing steps. We also outline the error measurements we used for analyzing results.

A. Data Source

This study focuses on quantifying the short-term variability of output wave power, which requires wave data at high resolution. The wave data is from Nov. 2011 to Dec 2011, from Belmullet Berth B (Lat 54.23, Long -10.14), a high wave potential region off the north-west coast of Ireland. This location has been identified to be suitable for the deployment of full scale wave energy devices. Fig 2 shows the location on an Ireland map. We extracted SWH and wave period from the sampled raw data providing three values every ten minutes on average.

We applied linear interpolation on the given datasets to create a new dataset with even intervals every minute to correct missing and irregular data points. From this set we then extracted ten minute data points.

B. Power Matrix

The evaluation of wave power uses the power matrix of a wave energy converter for a given SWH and wave period. This approach thus ignores the phase difference between the instantaneous power and the two input parameters. This phase difference occurs due to complex interaction between different hydrodynamic forces and the reaction force of the wave energy device power take-off.

We use the power matrix of a Pelamis P2 Device, an attenuator type oscillating body, with a rated capacity of 750kW [12]. The power matrix is originally designed to estimate the change in power production corresponding to changes larger than 0.5 m and 0.5 s in SWH and wave period respectively. As the study requires capturing smaller variability of the wave power, a bi-linear interpolation method is employed to extend the procedure to smaller variations. Fig 3 shows the interpolated power matrix.

C. Exponential Smoothing

Exponential smoothing is a widely used moving average technique for smoothing out discrete time data series. Exponential smoothing is often used as forecasting method and is fairly popular due to its simplicity and low computation requirements [13]. In this study, we utilize it
instead as a form of data preprocessing to smooth out noise from our data and provide a better picture of power signals. In addition, in exponential smoothing we allow the more recent values of the series to have greater influence on the forecast of future values than the more distant observations. The higher the weight given to the present observation, the lower a weight is given to the immediately preceding observations. We believe this gives the neural network model a bias towards recent data point patterns during validation, while also taking into account patterns far into the past. For instance, we believe it can better identify peaks and troughs in the power series by giving more emphasis on recent data changes leading to a peak. The simple exponential smoothing formula is

\[ y_{i+1} = \alpha y_i + (1 - \alpha) y_{i-1} \]

where \( y_{i+1} \) is the smoothed weighted average from the previous observation \( y_i \), and previous smoothed value \( y_{i-1} \), and where \( \alpha \) is the smoothing factor

\[ \alpha = \frac{2}{k+1}, \quad 0 < \alpha < 1 \]

where \( k \) specifies window size, a constant exponential factor. In our case, the window size is set to 12 h. This basic form of exponential smoothing is also known as an exponential moving average.

D. Error Measurements

Error measurement statistics play a critical role in analyzing forecast accuracy, observing exceptions, and benchmarking methods. In our results we use three error statistics and provide confidence intervals for predicted values. We first use the mean absolute percentage error (MAPE) which measures the size of the error in percentage terms and is a common estimate of error in forecasting problems. MAPE is defined as

\[ MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{x_i - y_i}{x_i} \right| \]

where \( N \) is the total number of predicted values, \( x_i \) is the actual observed value, and \( y_i \) is the forecasted value by the neural network. We also use the mean absolute deviation (MAD) which measures the size of the error in units. MAD takes the absolute value of forecast errors and averages them over the entirety of the forecast time periods. It is calculated as the mean of the unsigned errors

\[ MAD = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i| \]

Lastly we use the Pearson's correlation coefficient (CC) which is a measure of the linear dependency between two variables, in our case the observed and forecasted values, and is calculated as

\[ CC = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \]

A confidence bound for a model or variable is an interval of values within which we expect the true value of the population parameter to be contained. For calculating bounds for our model we use the bootstrapping method [14] which resamples residuals for estimating the variance of forecasted values at each time step. The upper and lower endpoints are specified as

\[ \bar{y} \pm z \frac{\sigma}{\sqrt{n}} \]

with \( \bar{y} \) representing the sample mean (center of the confidence interval), \( \sigma/\sqrt{n} \) being the standard deviation of the sampling distribution, and \( z \) is a constant multiplier set at 1.96 for a 95% confidence interval.

V. RESULTS

The NAR neural network is used to predict results of up to 3, 6, 12, and 24 h forecasts. Each forecast is a multistep prediction, meaning that for 24 hour forecasts, for instance, we are predicting 144 values at once. Fig 4 showcases an example month (November 2011) which we used for experimentation.
This paper structures."

Results show that prediction occurs better when smoothed data is used to predict the next 24 h of wave power shown in Fig 6. Table I shows results averaged from multiple runs for 3 to 24 h forecasts using smoothed data, while Table II shows the same forecasts using unsmoothed data. We see that in Tables I and II that errors increase slightly as the forecast window increases. The MAPE, MAD, and correlation coefficients of predicting smoothed data show superior results versus inputting data into the network which has not been smoothed. Comparing correlation coefficients with [11] who obtained results of 0.95 for 3h, 0.90 for 6h, 0.87 for 6h, and 0.73 for 24h, shows our forecasts in Table I exhibit much higher accuracies despite predicting outward at multiple time steps (note [11] predicted SWH and their data source was different). In Table II we obtained poor to mixed results with predicting non smoothed data. Fig 7 shows an example of a 3 h prediction.

Table I: Table of error statistics of forecasting smoothed

<table>
<thead>
<tr>
<th></th>
<th>3h</th>
<th>6h</th>
<th>12h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.006</td>
<td>0.009</td>
<td>0.027</td>
<td>0.043</td>
</tr>
<tr>
<td>MAD</td>
<td>0.011</td>
<td>0.019</td>
<td>0.062</td>
<td>0.157</td>
</tr>
<tr>
<td>CC</td>
<td>0.999</td>
<td>0.993</td>
<td>0.991</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table II: Table of error statistics of forecasting non-smoothed data

<table>
<thead>
<tr>
<th></th>
<th>3h</th>
<th>6h</th>
<th>12h</th>
<th>24h</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>0.05</td>
<td>0.1</td>
<td>0.12</td>
<td>0.24</td>
</tr>
<tr>
<td>MAD</td>
<td>1.02</td>
<td>1.16</td>
<td>0.75</td>
<td>1.6</td>
</tr>
<tr>
<td>CC</td>
<td>0.08</td>
<td>0.09</td>
<td>0.32</td>
<td>0.5</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

Our method of a NAR neural network with exponentially smoothed data is successful in predicting future ocean wave power levels. Results show that prediction occurs better when data has been smoothed. This paper demonstrates that smoothed data is able to iron out noise, represent accurate power levels, and provide accurate short term multistep predictions yielding high correlation coefficients of up to 0.99.

Figure 6. 24 h Prediction From Training Data in Fig 5.

Figure 7. 3 h prediction using unsmoothed data.

A. Future Work

Our long term objective is to integrate these advanced forecasting methods in the analysis as part of the inputs of a network model, analyzing the technical and economic effects of wind and wave integration. Further work on forecasting is planned to expand the prediction capabilities of the NAR model and provide a deeper computational and error analysis.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Irish Marine Institute for providing high resolution ocean data.

REFERENCES