

Predicting Wind Energy generation with Recurrent Neural Networks^{*}

Jaume Manero¹, Javier Béjar², and Ulises Cortés^{3,4}

¹ Universitat Politècnica de Catalunya - BarcelonaTECH, Barcelona, Spain

`jaume.manero@upc.edu`

² `bejar@cs.upc.edu`

³ `ia@cs.upc.edu`

⁴ Barcelona SuperComputing Center, Barcelona, Spain

Abstract. Decarbonizing the energy supply requires extensive use of renewable generation. Their intermittent nature requires to obtain accurate forecasts of future generation, at short, mid and long term. Wind Energy generation prediction is based on the ability to forecast wind intensity. This problem has been approached using two families of methods one based on weather forecasting input (Numerical Weather Model Prediction) and the other based on past observations (time series forecasting). This work deals with the application of Deep Learning to wind time series. Wind Time series are non-linear and non-stationary, making their forecasting very challenging. Deep neural networks have shown their success recently for problems involving sequences with non-linear behavior. In this work, we perform experiments comparing the capability of different neural network architectures for multi-step forecasting in a 12 hour ahead prediction. For the Time Series input we used the US National Renewable Energy Laboratorys WIND Dataset [3], (the largest available wind and energy dataset with over 120.000 physical wind sites), this dataset is evenly spread across all the North America geography which has allowed us to obtain conclusions on the relationship between physical site complexity and forecast accuracy. In the preliminary results of this work it can be seen a relationship between the error (measured as R^2) and the complexity of the terrain, and a better accuracy score by some Recurrent Neural Network Architectures.

Keywords: Time series · Recurrent Neural Networks · Multi-Step prediction · Seq2Seq · Wind Forecast · NREL Dataset · Wind Energy Prediction.

^{*} This work is partially supported by the Joint Study Agreement no. W156463 under the IBM/BSC Deep Learning Center agreement, by the Spanish Government through Programa Severo Ochoa (SEV-2015-0493), by the Spanish Ministry of Science and Technology through TIN2015-65316-P project, and by the Generalitat de Catalunya (contracts 2014-SGR-1051)

1 Introduction

Wind power generation is already a critical contributor to the electrical supply systems in many countries around the world. We can cite some nations with high wind penetration in their electricity generation mix (as a percentage of total production in 2016) like Denmark (36,8%), Ireland (27%), Portugal (24,7%), Spain (19%) or Germany (16%) [13]. This penetration, already relevant, will see a steep increase in the next few years due to the increased efforts being performed by most countries in order to accelerate the transition to a CO₂ free energy model [12] trying to reduce and stop the negative impact of the greenhouse gases in our atmosphere.

The integration of renewable energy into the grid is complex due to their inherent intermittent nature. The electricity grid has to assure continuous supply using all the different generation technologies available, and for this reason developing forecasting techniques (at demand and production sides of the system) is critical for its stability. Forecasting the energy generation output for the renewable generation assets becomes a core task in the energy management.

Wind energy forecasting has not only value for its contribution to the system stability but also has additional potential for savings in the overall wind-production life-cycle. In this direction, it has been established that a small increase of 10% in the quality of prediction would be able to generate savings of 140 million US\$ in the United States overall system [7].

Many methods have been designed for wind prediction, in two major families: methods based in meteorology input (NWP) and Time Series methods that use only the information from observations in the turbine sensors. This work deals with the second category, the wind time series data. Commercial systems use combinations of both families of methods, but this work will focus only in the learning that can be obtained from Wind Time series past observations. This work deals with the application of Recurrent Neural Networks to the wind time series.

In order to explore these series, large datasets are required. In this work the Wind dataset from the U.S. National Renewable Laboratory (NREL) has been used, A dataset with 120,000 wind points that is the largest available wind dataset in the world.

2 The Nature of Wind Time Series

Wind turbines are provided with hundreds of sensors that generate information in real time, creating a stream of data that is stored and analyzed. This information has many dimensions like energy generated, performance of the internal engines or meteorological data about the environment (wind speed, temperature, pressure, etc.). These data are used to monitor the functioning of the turbine and are a valuable input for prediction. For mid-term prediction the information of the sensors is converted into time series with readings every 5 to 10 minutes.

Typically, a wind time series will be a time-stamped sequence of several measures that can be related to wind. The dimensions are usually; wind power

(MW), wind direction (*degrees*), air pressure (Pa), wind speed (m/s), temperature (C or K), air density (kg/m^3), relative humidity (%). All these observations can be generated at different heights (floor, hub height, half height). As the wind at 100 meters high (hub turbine height) is the one that moves the blades, it is probably the measure with the highest relevance, while wind direction is important to understand how the dominant winds might impact wind patterns and intensity. In [Fig.3] a summary of one-year data from the Sotavento wind park is shown in the wind rose, the dominance of E/NE and W/SW winds is clear on this site.

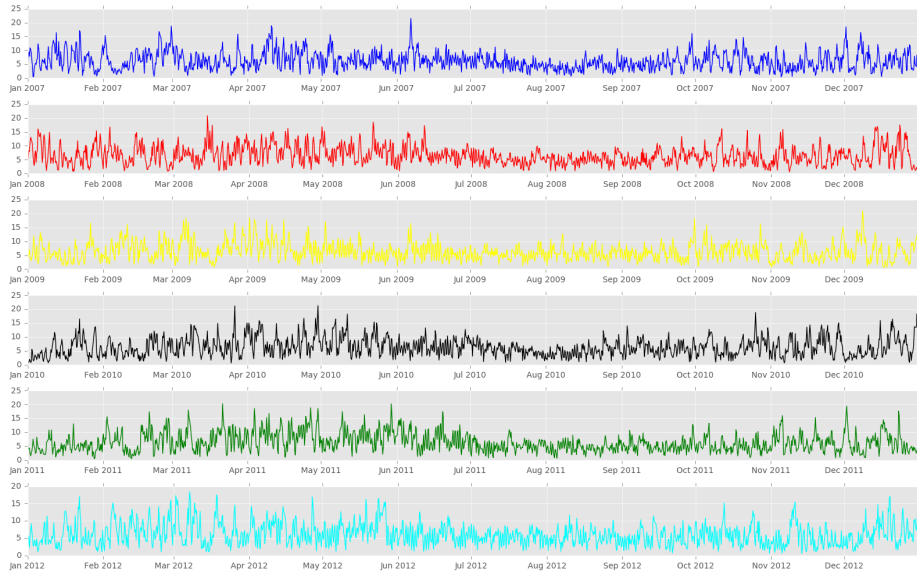


Fig. 1. Wind speed time series in site located in Techado New Mexico USA, vertical axis wind speed in m/s , horizontal axis time. Data from NREL Dataset [3]

Wind is a natural phenomenon that is created by various forces applied to the atmosphere at the same time, namely: the pressure gradient force, the frictional force, the Coriolis force and the gravitational force. For the energy forecast task in wind turbines, only winds close to the surface are studied, and those are impacted by the frictional force, which will depend on the specific orography of the site [6]. It is well-known that wind may vary in two locations not far away. It can be seen in a wind park the different speed of the blades in similar turbines or some turbines idle (no wind) while some others are turning, this shows empirically the wind variation in very closed locations.

But not only orography is relevant for the wind formation. The earth science has already stated that wind is the combination of periodical phenomena like day/night or summer/winter, a result of low/high-pressure variations and all of

them combined with temperature, air density and pressure. The combination of all these factors is of high complexity and the result, over time, is the wind as we know it. For this reason, it is quite usual that in a wind time series all these factors are overlapped (a storm in summer at night from the north), and extracting each factor is of high complexity (if possible at all).

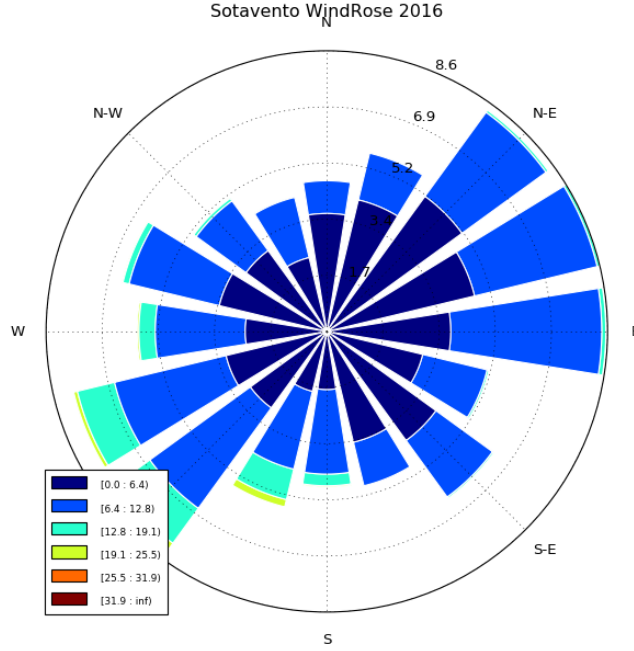


Fig. 2. Wind direction dimension in one year of time series measurements in the Sotavento Park located in Galicia, Spain 2016 [2]

2.1 Some statistical properties of the wind time series

Stationarity in a time series is understood as the property where the statistical characteristics such as mean, variance, autocorrelation, etc. are all constant over time, or repeat over time in some sequences (seasonal, day/night,...).

There are several tests widely used to analyze the stationarity of a time series. The Dick-Füller (ADF) test (and its evolution the augmented ADF) are the most common. The ADF looks for a *unit root* in a time series sample. A *unit root* is a statistical feature that determines randomness in the series. Applying these tests to the dataset we obtain non-stationarity results.

Linearity is another relevant property to be found in the wind time series. Linearity will allow the use of linear forecasting methods and non-linearity needs

of more complex methods (non-linear) have to be used to obtain accurate predictions. The validation of linearity in a time series is not an easy and straightforward task. The surrogate data method, described by Theiler in [11] is a powerful tool to validate linearity. This test applied to wind time series shows that linearity can be found in some wind datasets but not in all of them [4].

If the wind is nonlinear, how can linear models be used for forecasting? The answer lies in the fact that the wind series contains inner structures that might be linear. The best forecasting methods will extract this information (*learn*) the shape of these internal structures to produce more accurate results.

3 The Wind Energy Generation Forecasting Task

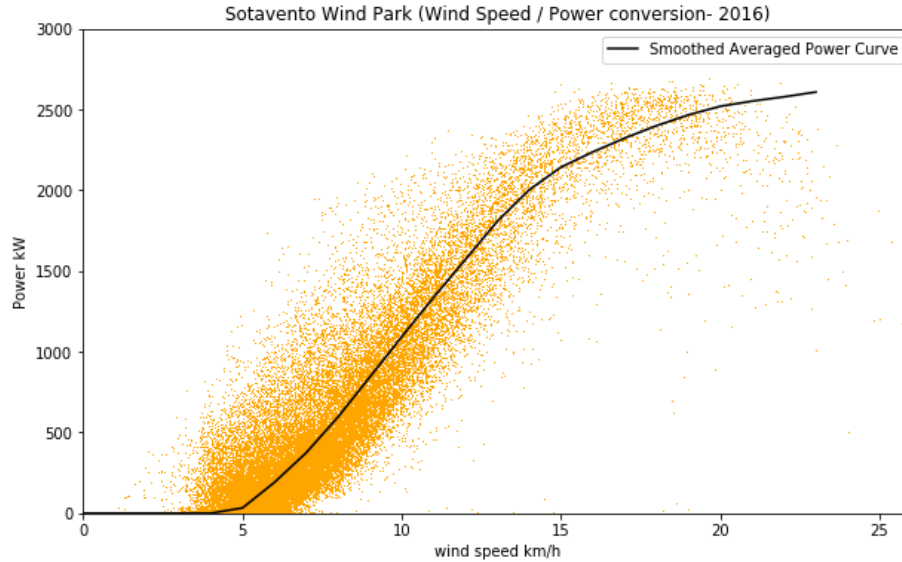


Fig. 3. Energy Generated in the Sotavento wind park (from observation data) [8]

Energy in the turbines is generated from the kinetic energy of wind. The action of wind moves the blades and generates a rotational effect which produces electricity (by the Faraday law). In the field, wind turbines are usually grouped in wind parks that can range from a few turbines up to hundreds of them to leverage areas where the wind is steady and strong over the whole year. The power generated by a wind turbine (see equation [eq:1]) is directly dependent on the swept area of the Blade (A), or and in the Air density (ρ), but mainly on the airspeed (v^3).

$$Energy = \frac{1}{2}\rho A t v^3; \quad (Power = \frac{E}{t}) \quad (1)$$

In figure [Fig:3] a graphical representation of the transformation of wind speed into power can be observed. The red-colored points are tuples of energy generated with wind speed (in 10 min periods). In this graph, the points concentrated along the original theoretical power curve (each turbine has its own). Additionally, it can be seen the existence of two relevant points: (a) the cut-in which defines the speed at which the turbine starts generating energy and, (b) the cut-off which is the threshold where there is no additional power generated (it is quite usual that this point triggers safety mechanisms to avoid the blades to be damaged by strong winds). From all this points, it can be deducted that in order to forecast energy, it is mandatory to predict wind intensity, as it is the critical feature to convert wind to energy.

For the experiments the National Renewable Energy Laboratory (NREL) is used, which is the largest available dataset of wind points [3]. This large dataset offers production and meteorological data (wind speed, wind direction, temperature, humidity and energy) synthesized from Meteorological global models for over 120,000 sites evenly found in the US geography.

4 Methodology and experiments

Two problems have been addressed in the experiments. First the multi-step prediction, and second the methodology to compute the prediction which can be defined as a regression problem.

According to the literature there are several approaches for a multi-step forecast, based on how the future time steps are obtained [10]. The first, and simplest, is the recursive approach, where a data window $[t_{i-k}, t_{i-1}]$ and prediction for time t_i are used for predicting t_{i+1} iteratively. This method is not very good for mid/long term prediction as there is a compounding error effect as the predicted data is reused. Another issue is the lack of the possibility to add exogenous variables in the prediction, which is a handicap for our approach.

The *direct* approach obtains predictions by computing a regression for each of the time points on the future horizon. This has the advantage of not reusing predicted data but is more computationally expensive given that multiple models have to be obtained. Another approach is to use multiple regression or sequence to sequence prediction methods [9]. This means that all the future time steps are obtained at the same time without reusing predictions and with a unique model. Both techniques are used in our experiments.

For solving the regression problem, we have chosen regression support vector machines (SVR) [1] and two neural network methods, multi-layer perceptron (MLP) and recurrent networks (RNN) [5]. The SVR will be used as a baseline and only for direct prediction. MLP and RNN will be used for direct prediction and multiple regression.

A systematic exploration of the parameters for all methods was performed. For SVR, different kernels will be used, namely Radial Basis Function (RBF), linear, quadratic and cubic polynomial with a wide range of values for the C parameter and the bandwidth for the RBF kernel.

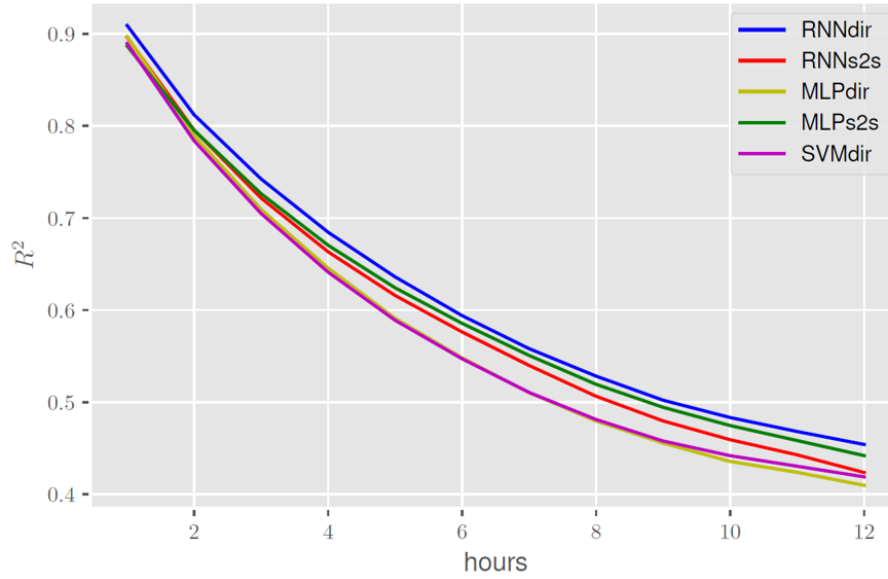


Fig. 4. Mean of R^2 for the 20 best experiments for each architecture, 12 hours horizon.

For the MLP experiments, architectures from one to three layers of different sizes with a linear output were tried, with different activation functions (sigmoid and ReLU) and different values of dropout on each layer. The direct approach had one output neuron and the multiple regression as many outputs as the prediction horizon.

For the RNN architecture a structure from one to three layers of Long Short Term Memory units (LSTM) or Gated Recurrent (GRU) units with different sizes were used, with tanh, sigmoid and ReLU activation functions for the output of the recurrent units and different levels of recurrent dropout. For the direct approach, a MLP with linear output was used.

For the multiple regression, an encoder-decoder architecture was used [9], where the recurrent layers were used as a first stage, performing the encoder task. The state obtained from the encoder was used as input for another recurrent network, acting as a decoder, which generates a sequence with the length of the prediction horizon. Each time step of the decoder had direct access to the state from the encoder additionally to the state from the previous step.

5 Results

5.1 Main results from preliminary experiments

The raw data in the dataset is characterized with five minutes sampling but in order to reduce computational cost and assuming a realistic forecasting scenario

of hourly predictions, the data was reduced to hourly sampling by averaging the measures every hour. The data used for the prediction included the wind speed and direction at 100 meters height plus barometric pressure and air density. The hour and month of the data were added as complementary variables in the time series.

The first four years of the time series was used for training, the fifth year was used as a test set for tuning the model parameters and the sixth year was used as the validation set. The training dataset has a size of 40912 values and the test and validation a size of 10228 values.

In order to analyze the algorithms behavior with different parameters and combinations, different windows lengths were used as input, ranging from three to 36 previous measures. The forecast was the wind speed from one to 12 hours ahead.

To compare the accuracy of the different method results the determination coefficient measure (R^2) was chosen. The data was z-normalized, so the coefficient is equivalent to $1 - MSE$. Figure [Fig.4] shows the averaged R^2 for the best 20 results for each architecture for a specific site. The RNN model with direct prediction is consistently the best model followed by multiple regression with MLP. RNN with seq2seq performs similarly to the multiple regression MLP for the short term, but the mid-term predictions decay faster. The MLP and SVR with direct prediction have very similar results far from the rest.

Performing a two-sided Kolmogorov-Smirnov test for equality of distribution for the R^2 values for each time step of the prediction horizon for the best two architectures (RNN direct and MLP multi-regression) all distributions have less than $1e^{-4}$ as p-value, indicating that their distributions are different. The difference of the means for the different hourly prediction of the 20 best results is in the range 0.007 to 0.022.

The best RNN architectures have two layers of GRU units with ReLU activation functions, drop out of around 0.3 with a window input from 16 to 24 hours. This result shows that the RNN architecture has a superior ability for the prediction task than the simpler MLP or SVR combinations. This shows that the RNN construct have some superior capabilities for this task.

5.2 Relationship with site complexity

An interesting geographical analysis of the errors obtained using different methods shows a measure of the site complexity, related to the errors obtained using different methods.

Representing graphically the error obtained by using persistence and multiple regression (seq2seq) MLP (see Fig.5 and Fig.6) it can be established a relationship between site complexity and prediction error. The most complex sites in the US are located in the western side of the central plains, where high regime winds have high variability and high difficulty for prediction.

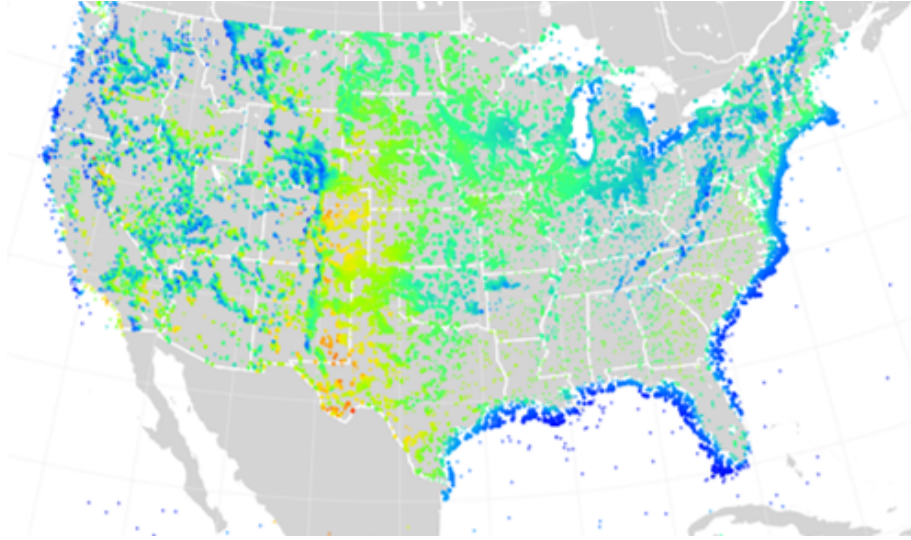


Fig. 5. Error R^2 in persistence over all NREL sites

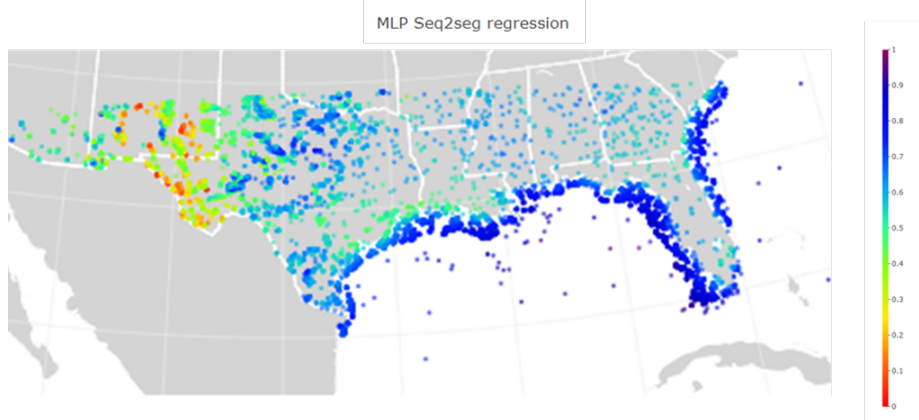


Fig. 6. Error R^2 in multiple regression MLP over NREL sites south Parallel 35°

6 Conclusions and future work

The preliminary results of our experiments show that RNN architectures with direct multi-step prediction can obtain reasonable mid-term predictions of wind speed with consistent accuracy among a limited number of sites. Other approaches show more significant decreases in accuracy the further the horizon of prediction.

The experiments also show that the direct approach for multi-step prediction have better results compared to a multiple regression/sequence to sequence approach.

Other multi-step prediction methods have to be explored, combined with different RNN architectures. For instance, given that it is more important the mid-term prediction, a multiple regression focused only on the more distant future seems more interesting. Also, further experiments using more advanced methods for the RNN sequence to sequence architectures have to be explored, like teacher forcing or attention mechanisms.

Finally as the available data has over 120,000 sites evenly distributed all across the US geography, a set of experiments to identify the best algorithm for each site topology is under way, with the objective to obtain the best algorithm structure that combines the characteristics of the time series with the site geographical characteristics of the site.

References

1. Bishop, C.M.: Pattern Recognition and ML. Springer-Verlag, Berlin (2006)
2. Díaz-Dorado, E., Carrillo, C., Cidras, J., Albo, E.: Estimation of energy losses in a wind park. In: 2007 9th Int. Conf. on Electrical Power Quality and Utilisation. pp. 1–6 (Oct 2007). <https://doi.org/10.1109/EPQU.2007.4424125>
3. Draxl, C., Clifton, A., Hodge, B.M., McCaa, J.: The wind integration national dataset (wind) toolkit. *Applied Energy* **151**, 355–366 (2015)
4. Gan, M., Li, H.X., Chen, C.L.P., Chen, L.: A potential method for determining nonlinearity in wind data. *IEEE Power and Energy Technology Systems Journal* **2**(2), 74–81 (June 2015). <https://doi.org/10.1109/JPETS.2015.2424700>
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
6. Landberg, L.: Meteorology for Wind Energy. John Wiley & Sons (2016)
7. Lew, D., Milligan, M., Jordan, G., Piwko, R.: The value of wind power forecasting (2011), <http://www.nrel.gov/docs/fy11osti/50814.pdf>, nREL
8. López, J., Dorado, A.D., Alvarez, J., Carrillo, C., Cidrás, J.: The sotavento experimental wind park. In: Global Windpower Conference Paris (2002)
9. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. *CoRR* (2014), <http://arxiv.org/abs/1409.3215>
10. Taieb, S.B., Atiya, A.F.: A bias and variance analysis for multistep-ahead time series forecasting. *IEEE trans- on NN and learning systems* **27**(1), 62–76 (2016)
11. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* **58**(1), 77 – 94 (1992). [https://doi.org/10.1016/0167-2789\(92\)90102-S](https://doi.org/10.1016/0167-2789(92)90102-S)
12. UNFCCC. Secretariat: Report of the conference of the parties. Paris agreement 2015 (2015), united Nations Office at Geneva
13. Wind Europe organization: Wind in power 2016 European Statistics (2017), windeurope organization, Belgium