Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Deep Learning is blowing in the wind. Deep models applied to wind prediction at turbine level

To cite this article: Jaume Manero *et al* 2019 *J. Phys.: Conf. Ser.* **1222** 012037

Deep Learning is blowing in the wind. Deep models applied to wind prediction at turbine level

Jaume Manero[1], Javier Béjar[1], Ulises Cortés[1,2]

[1]Knowledge and Machine Learning Group - Computer Science Department.Universitat Politècnica de Catalunya - BarcelonaTECH. Barcelona
[2]Barcelona SuperComputing Center Barcelona

E-mail: jaume.manero@upc.edu, bejar@cs.upc.edu, ia@cs.upc.edu

Abstract. Wind Energy generation depends on the existence of wind, a meteorological phenomena intermittent by nature, with the consequence of generating uncertainty on the availability of wind energy in the future. The grid stability processes require continuous forecasting of wind energy generated. Forecasting wind energy can be performed either by using weather forecast data or by projecting (or regressing) the past time-series data observations into the future. This last method is the statistical or time series approach. Wind Time Series show non-linearity and non-stationarity properties, and these two properties increase the complexity of the forecasting task using statistical methodologies. In this paper we explore the use of deep learning techniques, which can represent non-linearity, to the wind speed prediction using the largest public wind dataset available, the Wind Toolkit from the National Renewable Laboratory of the US. Several deep network architectures like Multi Layer Perceptrons, Convolutional Networks or Recurrent Networks have been tested on the 126,692 wind-sites and with the results obtained valuable comparisons and conclusions have been obtained. The distribution of the wind sites across the North American Geography has allowed to include in the analysis relationships between terrain, wind forecast complexity and deep methods.With the developed testing workbench and with the availability of the Barcelona Supercomputing Center new architectures are being developed. This work concludes with the feasibility of deep learning architectures for the wind and energy forecasting.

1. Introduction

The energy industry is changing its dependence on fossil fuels by investing heavily in the generation using renewable sources. This transformation is a consequence of the Climate Change impact in our economies, and is strongly pushed by international organizations and agreements like the Paris Climate Agreement in 2015 [1].

This transformation has many challenges, being one of the them to cope with the inherent intermittence of the renewable sources which it is related to the inner nature of the natural phenomena that creates them (sun, wind, etc.) as they appear and disappear depending on actual weather, day-night cycles, combination of temperature, clouds and humidity. To cope with this intermittence it is required to develop forecasting strategies that will allow to understand the future patterns of energy generation and to develop mitigating strategies for the different output levels [2]. One specially important forecasting window is 12-24 hours, as it has intrinsically high value (it is used as load confirmation from most network operators), but this forecasting horizon has a high level or uncertainty due to the physical complexity of the wind generation dynamics at local level [3].

Research and implementation of Wind forecasting methods has been extensively developed in the industry, as the requirement has been around for the last couple of decades and massive investments have been done in developing the right tools for the problem [4], however, the use of Machine Learning and Deep Learning techniques is still nascent in this area [5][6][7].

The application of Machine Learning to Time Series forecasting has shown good results in some applications [8], and is currently a topic of research. The representation capabilities of these techniques for non-linear modelling are being used with series that can't be modelled with traditional statistical models [9].

In this work deep learning models are applied to the U.S. National Renewable Laboratory (NREL) Wind Toolkit, the largest publicly available wind dataset. This dataset contains 126,692 wind sites [10] with wind speed, humidity, pressure, wind direction measures sampled every 5 minutes during 7 years (2007-2013). This work has developed experiments with Several deep learning architectures and applied them to all the wind sites, which has allowed to obtain conclusions from a very large resource of data. This wealth of data has allowed to develop geographical interpretations of the results.

This article is organized as follows. First the Time Series Forecasting task is presented as a Regression problem, then the deep learning families of algorithms developed are analyzed, and finally the experimentation is described in some detail. The final conclusions and future work actually in place is described in the last section.

2. Wind Time Series and Forecasting

A wind time series is a set of several observations (wind speed at different heights, wind direction, temperature, pressure, etc.) recorded in a geographical point or site. Usually they are obtained from the extensively sensorized turbines that generate the wind energy. A wind can series is then composed of observations, defined formally as an observation tuple O_i

$$O_i = \langle O_{temperature}, O_{wind-direction}, O_{pressure,\cdots} \rangle \tag{1}$$

There can be many elements in a observation tuple in one time stamp, like wind speed at different heights, humidity, sun radiation and others. Then a wind time series $\mathbf{X}$ can be described as a sequence of n observation tuples.

$$\mathbf{X} = \langle O_1, O_2, \cdots, O_n \rangle \; or \; \langle \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \rangle \tag{2}$$

A forecast $\hat{\mathbf{Y}}$ is a synthetic generated time series obtained by using inference methods in the past observations or by integrating information from other sources (like Weather Meteorological Data for instance). A forecast $\hat{\mathbf{Y}}$ will have two dimensions, an horizon H and a number of predictions. Horizon defines a point in the future H steps ahead. A forecast can have be a single step prediction (only for the horizon) $\hat{\mathbf{Y}} = \hat{\mathbf{y}}_{\mathbf{n+H}}$ or multi-step where a prediction is obtained for each or all the future steps inside the horizon interval $\hat{\mathbf{Y}} = \langle \hat{\mathbf{y}}_{\mathbf{n+1}}, \hat{\mathbf{y}}_{\mathbf{n+2}}, \cdots, \hat{\mathbf{y}}_{\mathbf{n+H}} \rangle$.

The wind forecast problem can be considered as a Regression learning problem. A model $\mathfrak{M}$ is developed that, using the most recent observations $\mathbf{X}$ as an input generates predictions (with an horizon H) as an output $\langle \hat{y}_{n+1}, \hat{y}_{n+2}, \cdots, \hat{y}_{n+H} \rangle$. The objective is to develop a function (or model) $f : \mathbb{R}^n \to \mathbb{R}$ that generates predictions from the existing observed data.

In Deep learning the regression function is obtained by learning the mapping $f : \mathbb{R}^n \to \mathbb{R}$ from examples, the learning is acquired by the model in the training phase of the algorithm. Training a Network consists in presenting a large set of data and the network, through its internal optimization mechanisms, learns the relationships between inputs and outputs. Training in a network has to converge to an optimum value, and this convergence is a key element in the

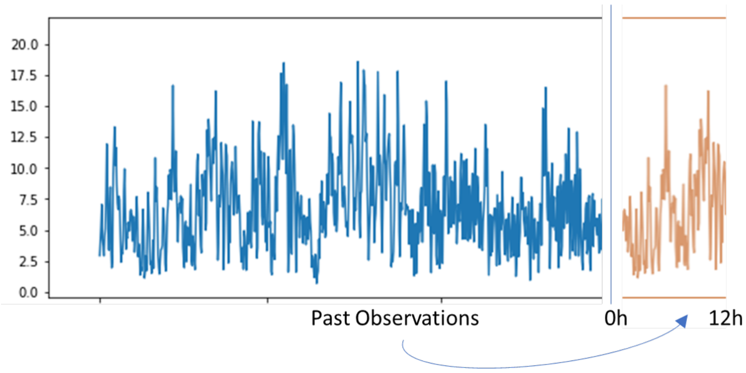

Figure 1. wind time series and prediction

definition and development of network models. There are many optimization algorithms and techniques for Network training [11]. In this work the optimizer used has been Adamax and elements like early-stopping have been included in the architectures [12].

This model is then applied to real data and generates a prediction result that needs to be rated based on the accuracy observed, and in our case we have chosen the R^2 or coefficient of determination which is a very informative measure for linear and non-linear regression problems [13].

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3}$$

For linear models R^2 can be interpreted as the explained variance (of the regressed function) divided by the existing variance. A value of 1 shows a very good fit, but can become negative if the resulting function does not have good results.

The purpose of this work is to predict energy, but the kinetic energy of wind is transformed to Energy through the energy formula $E = \frac{1}{2}\rho A t v^3$, however this transformation depends on many factors, which make the transformation a more complex function like $f : \mathbb{R}^n \to \mathbb{R}$ where the inputs are wind speed, direction, turbine model, technical loses, etc. There is a theoretical Power curve developed by the manufacturer but practically the transformation is not linear and has a level of uncertainty [14] [15]. There are some works that support the view of separating the wind-speed problem from the energy conversion, problems that are of different nature [4]. In this article, we have focused only in the wind speed forecast problem as we have considered it as a more general. To produce the conversion from wind speed into energy the specifications of the sites and turbines need to be used and this conversion requires an approach based on the real power curve function of the site. This specialized approach would use much more local information about the site (geographically, turbine models used, transformation infrastructure, technical loses, etc.). From now on this article when we refer to forecasting we understand forecasting wind speed.

Wind Time series forecasting is a specific breed of general Time series prediction due to the inner nature of the wind time series which are dependent on the physical characteristics of the wind formation, like weather, terrain or even the application of the Coriolis force on air [16], [3]. Linearity is the property of a time series defined by the possibility to adjust the time series by a linear model like: $\mathbf{X}_t = \alpha_1\mathbf{X}_{t-1} + \alpha_2\mathbf{X}_{t-2} \cdots \alpha_p\mathbf{X}_{t-p} + \epsilon_t$ Stationarity is understood as the statistically property defined by observing a constant mean, variance or covariance in a time series.

Wind Time Series present non-linearity and non-stationarity properties this added to the fact that linear models applied to non-linear time series offer poor results, makes mandatory the use of non-linear modelling for wind time series. There are many non-linear models used for wind time series in the literature [17], but there is limited application of Machine Learning or Deep Learning models to the wind forecasting area [5].

Deep Learning and Machine Learning methods are able to represent non-linear relationships and intuitively appear to be a good alternative for Wind Time Series Forecasting, defining the motivation for this work.

3. Forecasting wind with Deep Learning

One practical example of Deep Learning application is the image classification task, where the algorithm is trained with a set of images or examples. In this way the algorithm learns patterns that exists in the images and constructs a model for the classification of new images. With the Deep Learning techniques the models built have shown accuracy levels superior to the human ability classifying images, specifically when the image is clear, [18], [19].

Regression can be considered a variation of classification, in this case the task consists in adjusting a function or a model developed by learning all the examples in a training dataset. The resulting model will apply the learned structures to new data in order to perform a forecast or a regression of the curve (wind speed in this application).

The Deep Learning algorithm will require to be presented with a set of examples Z where each example will be a series of length lag, $Z : (Z_1, Z_2, \cdot, Z_n)$. Each example will be a tuple with the main variable (wind speed) and all the auxiliary variables that the dataset provides us (wind direction, temperature, humidity, density), $Z_i = (Z(\text{wind speed})_i, Z(\text{T})_i, Z(\rho)_i, \cdots,)$. All the algorithms described below use this Examples dataset for its training.

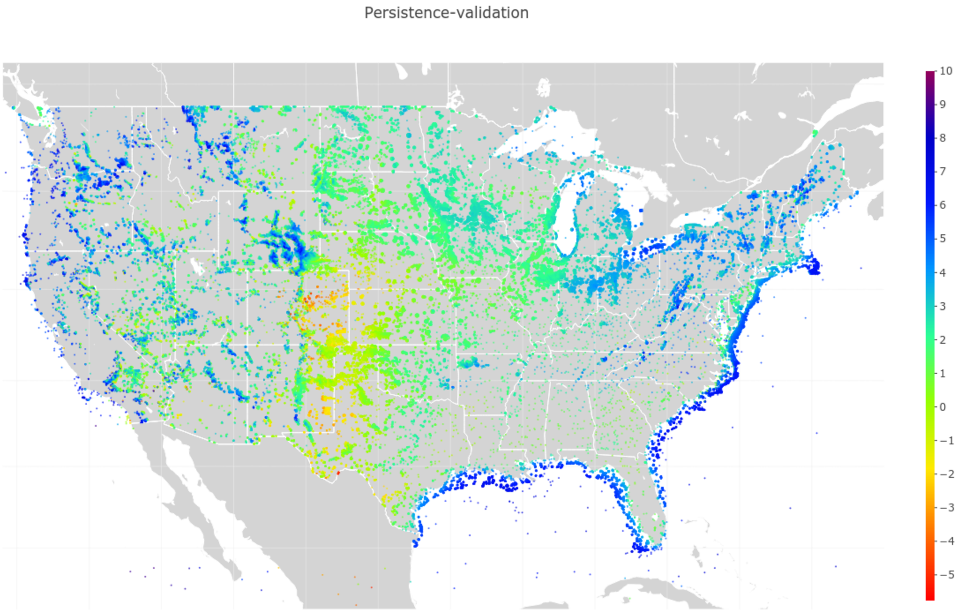

Figure 2. Persistence method applied to all the wind sites. The worse rated sites (yellow-red) mark high wind variability

3.1. Multi Layer Perceptron MLP

Multi Layer Perceptron or Feed Forward Network can be considered as the traditional Neural Network. All the neurons are fully connected to the next layer and the information flows in one direction. The activation function most widely used is the ReLu function [11]. In order to use the MLP architecture two possible approaches can be used. A Direct Regression approach, where the network forecasts a single data point in one execution. Or a sequence to sequence approach where the network forecasts a time series from $n+1$ to k where k is the horizon in one execution. The Direct Regression approach is very expensive computationally wise, as multiplies k times the number of executions and impacts training as well, some preliminary comparisons between direct regression approach and multi-step approach have been performed and the multi-step has shown better accuracy. With this conclusion all the models have been developed with multi-step regression or sequence to sequence. The sequence to sequence approach was defined for Recurrent Neural Networks (RNN) initially and has proven to be very effective for language and voice data [20]. The network is presented with a sequence of values and predicts a sequence of outputs, which is the predicted time series of length 12 (all the work has been done with 12 hours ahead prediction).

The training is performed using backpropagation and standard Gradient Descent with ADAMAX as optimizer [12].

3.2. Convolutional Network CNN

Convolutional Networks are extensively used in the image recognition field. Their main capacity is the application of filters to matrixes of data that can specialize in small patterns in the image.

There is some, but limited, experience on the application of CNN to time Series like [21] or [22], which show applications of convolutional networks to time series with good results.

CNN are feed forward networks in the sense that information flows in only one direction. The structure used in the experiments is a seq2seq CNN where the examples are presented with one dimensional data that is the sequence to be trained on. The CNN generates a sequence as an output which represents the sequence of horizon H to be predicted.

3.3. Recurrent Neural Network RNN

Recurrent Neural Networks are an evolution of feed forward networks, in this case the information is able to flow backward and forward, allowing to process history or long sequences of data. At each time step, RNN will take the input x_i and the output of the previous node h_{i-1}, producing an output h_i.

$$\begin{aligned} h_t &= f(x_i + h_{i-1}) \\ \hat{y}_t &= f(h_t) \end{aligned} \tag{4}$$

RNN can be feed with sequence to sequence like MLP and Convolutional and with encoder-decoder algorithms. An encoder-decoder architecture will take a sequence as an input and a generate a sequence as an output.

There are several variants of RNN networks depending on how the function f behaves, which inputs will process and how to gate them into the function. The two major structures in the neuron are the Long Short Term Memory (LSTM) [20] and the Gated Recurrent Units (GRU) cells [23].

4. Experimental application of Deep Learning to the NREL wind Dataset

The contribution of this work consists in the application of new deep learning architectures on a very large wind dataset. The architecture models described in the previous section have been used in many applications, but there are limited experiences of their use with time-series or with wind data. Regarding data there is not available a study on more than a handful of wind sites using deep learning, for this reason using the NREL database with over 126,692 data sites in North-America [10]

There are some works in the literature using this dataset [24], but, probably due to computing resource availability there are not large exercises available yet. We have used the Super Computing Center in Barcelona [25] as a key computing resource that has made the experiments possible.

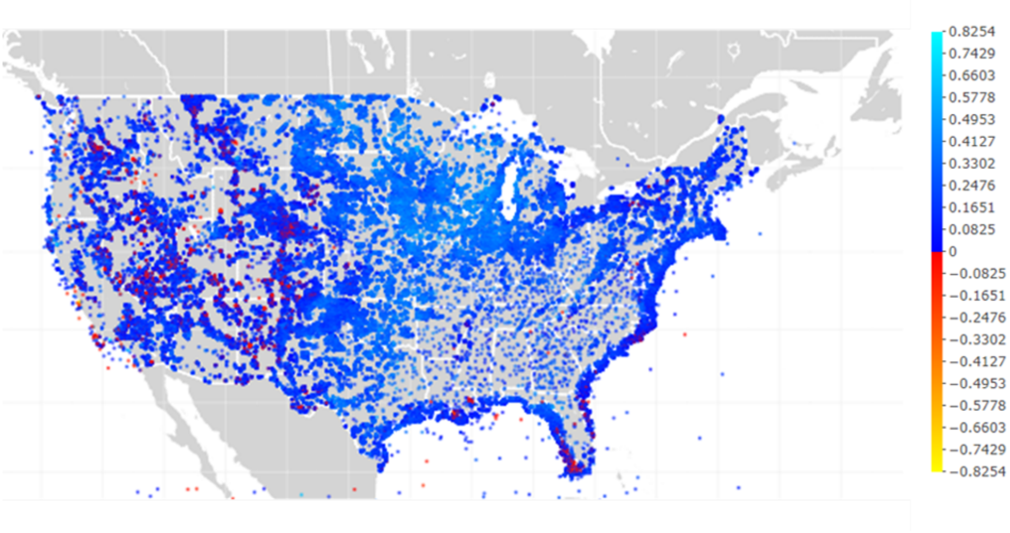

Figure 3. Map of US with the site comparison of MLP vs RNN comparison (blue better MLP, red better RNN)

To perform the experiments we have developed a software framework that allows the application of each architecture to all wind sites under the same conditions in order to obtain comparable homogeneous results.

The algorithms try to learn the wind speed patterns by presentation of different examples, which is accomplished by an sliding window technique.

$$\mathbf{X} = \langle \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n \rangle \tag{5}$$

The Wind Series in (2) is then decomposed in multiple examples of width lag

$$\begin{aligned} \mathbf{X_1} &= \langle \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{lag} \rangle \\ \mathbf{X_2} &= \langle \mathbf{x}_2, \mathbf{x}_2, \cdots, \mathbf{x}_{lag+1} \rangle \\ \mathbf{X_3} &= \langle \mathbf{x}_3, \mathbf{x}_4, \cdots, \mathbf{x}_{lag+2} \rangle \quad \cdots \end{aligned} \tag{6}$$

There are n $\mathbf{X_i}$ possible windows in the series being each one of length lag which are the examples that will be used in the training process by the different architectures. Each model generates a prediction series from n to n+12 steps (hours). The models tested can be seen in table §1.

The data is z-normalized as a pre-process for training so the value obtained for the MSE is also normalized, as a consequence $R^2 = 1 - MSE$, in this sense MSE can be considered an error measure and then R^2 is an accuracy measure, where the closer to 1 the more accurate is the result.

Table 1. Deep Models tested in Experiments

Model	Variant	Description
Multi Layer Perceptron	seq2seq	MLP sequence to sequence
Multi Layer Perceptron	Direct	MLP with Direct Regression
Convolutional Network	seq2seq	CNN sequence to sequence
Recurrent Neural Network	seq2seq	RNN seq2seq
Recurrent Neural Network	encoder decoder	RNN encoder decoder
KNN	KNN neighbors	trained like the Networks

A z-normalization consists on adjusting the scale of the the data obtaining mean of zero and variance (σ^2) of one. This is a usual data transformation process performed if working with neural networks [11] [26].

The accuracy of each method has been calculated using the coefficient of determination $\mathbf{R^2}$. Each one of the models has been applied to the 126,692 wind sites. The experimentation process has followed several steps.

1 Baseline execution

2 MLP Direct Regression vs MLP seq2seq model

3 CNN model

4 RNN models

5 Comparisons and results analysis

The first step has been to perform the baseline models for comparison with the deep learning architectures that will be executed later. The two baseline models chosen have been persistence and k-Nearest Neighbor or k-NN method. Persistence consists in using the x_n value as $\langle \hat{y}_{n+1} = x_n, \hat{y}_{n+2} = x_n, \cdots, \hat{y}_{n+H} = x_n \rangle$ this method obtains reasonable results up to 3 hours and then it deteriorates and in its R^2 becomes highly negative. An interesting finding with this method is that the accuracy of the method marks the wind variability in a site, and the location of these wind variability across the US geography is mostly located in areas with high wind resources and variability like the west part of the central plains (see fig. 2).

The k-NN method is trained with examples 18hours long. The algorithm identifies the nearest neighbors using Euclidean distance. With the 13 nearest neighbors selects the averaged next 12 hours. This algorithm obtains as a sum of the R^2 for the 12 steps ahead a value of 4.9, much higher than persistence which is below 2.5, but lower than all the Deep Learning models which are above 6.

The next step has been to determine which is the best regression strategy, and for this reason an experiment with the MLP has been designed. A model with direct regression (one MLP for each step) and a model for MLP seq2seq (input and output are sequences). Both models are executed in a selected number of sites and compared. The comparison confirms the better accuracy of the MLP seq2seq and the conclusion is to use seq2seq for all the experiments.

The third step has been to test each architecture.Each algorithm has been tested to identify the best parameters for each architecture, and then it has been executed with the same set of parameters with the whole dataset. The data has been divided in three sets, 5 years of data for training (2007-2011), one year for testing (2012) and another year for validation (2013). Each single execution on each site has been stored and used for the analysis.

From the comparison between the different methods we obtain valuable conclusions. The comparisons between the different methods shows a slightly better accuracy of the MLP seq2seq model than both RNN approaches and the CNN, however the four methods obtain very similar ratings. Another conclusion is that there is not something like the best model everywhere. We find sites where the best model is CNN and others where the best is RNN or MLP. In figure 2 we can see the result comparison between MLP and RNN. MLP is better in most sites, but RNN has areas from the Plains to the West Coast where it performs better, there is even a small part in southern Florida where RNN is the best method.

Analyzing the boxplot of each experiment (see 4 we can analyze the performance of the methods comparing the results between validation and test. Two observations can be made, one is the differences between the two sets of data, which probably implies that the model should be retrained in order to cope with the wind regime changes over time, and the second is the difference of accuracy comparison hour over hour. Could it be that there are methods that perform better in shorter time frames than others?

The results from the experiments can be summarized as follows:

(i) Persistence obtains poor results for the horizon of 12 h.(as it was expected).

(ii) Deep learning models are better than statistical baselines (k-NN).

(iii) MLP direct regression has worse results than the seq2seq regression approach.

(iv) MLP with sequence to sequence, as an average, has better accuracy than RNN or CNN models

(v) The best model in a site depends on some characteristics of the site (terrain complexity or wind regimes) showing a dependence between terrain characteristics and the most accurate model (see figure §3).

(vi) There are some differences in forecasting year 6 or year 7, which explains changes in the wind resource year on year. The conclusion is that the models need to be retrained, as the wind is not constant over time.

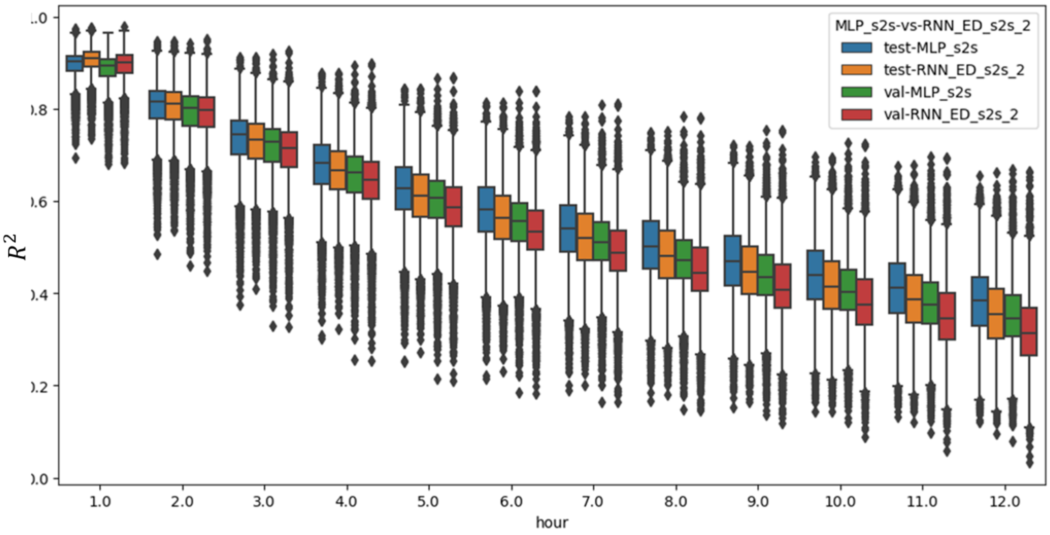

Figure 4. MLP vs RNN boxplot for validation and test datasets

There are narrow differences between the deep learning architectures, and we believe there is room for tuning the deep model structures in order to obtain better results.

5. Conclusions and future Work

The deep learning algorithms tried on the wind toolkit show better accuracy for 12 hour forecast ahead compared to traditional models. The experiments have shown dependency between wind site characteristics and the best performing method. In areas with high and variable winds (Southern plains, Gulf of Mexico and West Coast) the best method is the recurrent neural network, while in other more stable areas the best method is MLP and in some points CNN. This relationship between best method and site characteristics is an objective of future experimentation.

As the best models differ site to site, and there is some model accuracy drifting between the 12 hour steps, the added accuracy of using different models could improve the overall accuracy, with this finding, we can explore the combination of methods in ensembles in order to improve accuracy.

It is common, with the statistical methods, to adjust the forecasting models by heavy manual tuning, as the forecast is usually made at wind park level and not at turbine like in this work. Defining several models and adjusting the best model to each terrain characteristics may generate an improvement in accuracy, and using an ensemble approach will improve the overall results of the architectures tried. Deep Learning ensemble methods applied to the NREL dataset is another area of further work.

The experiments have been made only using past observations. The variables have different levels of information. Wind speed is the most relevant, while temperature or humidity have less information. From another point of view, temperature or humidity are low resolution variables which can be predicted with great accuracy from weather forecasts. We propose a hybrid model where the past observations will be combined with some low resolution variables from weather forecast to improve the overall accuracy.

Finally, even though the 7 years of measures sampled every 5 minutes looks like a lot of data, the experience shows us that Deep Learning exercises require much larger datasets to obtain the best results. It would be desirable to have alrger datasets, and this could be accomplished by using longer time series, but from the experimentation it has been observed that the use longer series presents some challenges, like:

(i) Winds change over time (due to climate change impact) and this fact compromises the the value of longer series of data [27].

(ii) There are limited sources of sites with longer series of data

Due to these challenges it will be interesting to use synthetic data, and one of the areas of future work is how to use Generative Adversarial Networks to develop synthetic data to enrich training and forecast processes like it is already made in other areas and applications[28]. In summary, the authors believe that Deep Learning applied to wind prediction opens a whole new set of tools to improve forecasting for the wind renewable industry.

Acknowledgments

We thank the Barcelona Supercomputing Center for the extensive use of their infrastructure in this project, and the United States National Renewable Energy Laboratory (NREL) for the use of their wind energy datasets.

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)

References

[1] United Nations Framework Convention on Climate Change (UNFCCC) Secretariat 2015 Report of the conference of the parties on its twenty-first session, held in Paris from 30 November to 13 December 2015. addendum. part two: Action taken by the conference of the parties at its twenty-first session united Nations Office at Geneva

[2] Fares R 2015 *Scientific American* – Nature Publishing Group

[3] Landberg L 2015 *Meteorology for Wind Energy* (John Wiley & Sons Ltd) ISBN 9781118913444

[4] Giebel G, Brownsword R, Kariniotakis G, Denhard M and Draxl C 2011 *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd edition* (ANEMOS.plus) project funded by the European Commission under the 6th Framework Program, Priority 6.1: Sustainable Energy Systems

[5] Manero J, Béjar J and Cortés U 2018 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11314 LNCS** 89–98

[6] Hervás-Martínez C, Salcedo-Sanz S, Gutiérrez P A, Ortiz-García E G and Prieto L 2012 *Neural Computing and Applications* **21** 993–1005 ISSN 1433-3058 URL `https://doi.org/10.1007/s00521-011-0582-x`

[7] Torres J M and Aguilar R M 2018 *Complexity* **2018** 10 URL `https://doi.org/10.1155/2018/9327536`

[8] Ahmed N K, Atiya A F, Gayar N E and El-Shishiny H 2010 *Econometric Reviews* **29** 594–621

[9] Ben Taieb S, Sorjamaa A and Bontempi G 2010 *Neurocomput.* **73** 1950–1957 ISSN 0925-2312 URL `http://dx.doi.org/10.1016/j.neucom.2009.11.030`

[10] Draxl C, Clifton A, Hodge B M and McCaa J 2015 *Applied Energy* **151** 355–366

[11] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press) ISBN 978-0262035613

[12] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y and Zheng X 2015 TensorFlow: Large-scale machine learning on heterogeneous systems software available from tensorflow.org URL `http://tensorflow.org/`

[13] Schneider A, Hommel G and Blettner M 2010 *Dtsch Arztebl Int* **107** 776–782 ISSN 1866-0452 21116397[pmid]

[14] Sohoni V, Gupta S C and Nema R K 2016 *Journal of Energy* 18 ISSN 8519785

[15] Shetty R P, Sathyabhama A and Pai P S 2018 *Frontiers in Energy* ISSN 2095-1698 URL `https://doi.org/10.1007/s11708-018-0553-3`

[16] Abkar M and Porté-Agel F 2016 *Phys. Rev. Fluids* **1**(6) 063701 URL `https://link.aps.org/doi/10.1103/PhysRevFluids.1.063701`

[17] Okumus I and Dinler A 2016 *Energy Conversion and Management* **123** 362–371

[18] Geirhos R, Janssen D H J, Schütt H H, Rauber J, Bethge M and Wichmann F A 2017 *CoRR* **abs/1706.06969** URL `http://arxiv.org/abs/1706.06969`

[19] He K, Zhang X, Ren S and Sun J 2015 *CoRR* **abs/1502.01852** (*Preprint* `1502.01852`) URL `http://arxiv.org/abs/1502.01852`

[20] Sutskever I, Vinyals O and Le Q V 2014 *CoRR* **abs/1409.3215** (*Preprint* `1409.3215`) URL `http://arxiv.org/abs/1409.3215`

[21] Qiu X, Zhang L, Ren Y, Suganthan P N and Amaratunga G 2014 Ensemble deep learning for regression and time series forecasting *2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)* pp 1–6

[22] Borovykh A, Bohte S and Oosterlee C W 2017 *arXiv e-prints* arXiv:1703.04691

[23] Chung J, Gulcehre C, Cho K and Bengio Y 2014 *arXiv e-prints* arXiv:1412.3555 (*Preprint* `1412.3555`)

[24] Heinermann J and Kramer O 2016 *Renewable Energy* **89** 671 – 679 ISSN 0960-1481 URL `http://www.sciencedirect.com/science/article/pii/S0960148115304894`

[25] Martorell J M 2016 *Contributions to Science* **12** 5–11 ISSN 0960-1481

[26] Ioffe S and Szegedy C 2015 *CoRR* **abs/1502.03167** (*Preprint* `1502.03167`) URL `http://arxiv.org/abs/1502.03167`

[27] Pryor S and Barthelmie R 2010 *Renewable and Sustainable Energy Reviews* **14** 430 – 437 ISSN 1364-0321

[28] Esteban C, Hyland S L and Rätsch G 2017 *ArXiv e-prints* (*Preprint* `1706.02633`)