

Forecastability Measures that Describe the Complexity of a Site for Deep Learning Wind Predictions

*Jaume Manero*¹ , *Javier Béjar*^{1,2} 

© The Authors 2021. This paper is published with open access at SuperFri.org

The application of deep learning to wind time series for multi-step prediction obtains good results at short horizons. The accuracy of a wind forecast is highly dependent on the specific structure of wind in the specific location, as many local features influence wind behaviour. The characterization of the complexity of a site for wind prediction is defined as forecastability or predictability and can be obtained from the inner structure of the meteorological time series observations from a site. We analyze the time series structure searching for properties that have a high correlation with the prediction result, properties that can create measures that have the potential to describe the forecastability of a site. The best measures will show a high correlation with the accuracy of the predictions. In this work, we analyze wind time series from 126,692 wind locations in the US, where we apply several deep learning methods first, and then we verify several forecastability descriptors with the accuracy deep learning results. We require High-Performance Computing (HPC) resources for this task as the deep learning algorithms have sensible resource requirements and are applied to a large set of data. The measures defined and explored in this work are based on several techniques that decompose or transform the wind time-series. By combining several of these measures, we can obtain better predictors of the site complexity, which will allow us to evaluate the future error of a prediction on this site. Forecastability measures can contribute to a wind site multi-dimensional description, becoming a valuable tool for wind resource analysts and wind forecasters.

Keywords: wind forecasting, time series, wind time series, deep learning, CNN, convolutional networks, forecastability.

Introduction

Wind-generated electricity is becoming a relevant component of the generation mix. While it keeps growing, some projections of future *CO₂ free* generation mix estimate a high wind generation dependency with percentages close to 40% in most countries [10]. These projections, which are a couple of decades away, are ambitious, but we can find countries which operate electricity systems with over 20% of wind-generated electricity, like Denmark (41%), Ireland (28%), Portugal (24%), Germany (21%) or Spain (19%) [22]. The largest global economies, the United States and China, have an installed capacity of 105 GW and 236 GW, respectively, accounting for over 50% of the total worldwide wind energy generation capacity, which has been 650 GW by the end of 2019.

Renewable energy has an intrinsic property, its intermittency, that challenges the stability of electrical systems. In a stable grid, demand must balance generation, which requires continuous forecasting to orchestrate the multiple energy generation origins to match the predicted demand. To predict wind energy output, we must predict the intensity of wind in the future, being wind a complex weather feature to model. Its formation depends on many local features, requiring meteorological models that work with a very high resolution not widely available today. When evaluating the potential of a wind generation location, we need to assess many dimensions, like the wind strength, the wind seasonality, the site accessibility, or the connectivity to the existing grid.

¹Technical University of Catalonia, Barcelona, Spain

²Barcelona Supercomputing Center, Barcelona, Spain

The basic and possibly, essential feature is the wind intensity as it defines the potential amount of energy that can be generated. However, for the energy generated to become economically viable, it needs to be predictable. Otherwise, its value decreases. In a scenario where the balancing of demand and production is complex, and demand-response markets are appearing, the prediction requirements for a site are critical. There is noticeable economic value from accurately assessing the predictability of a geographical location for wind electricity generation [1].

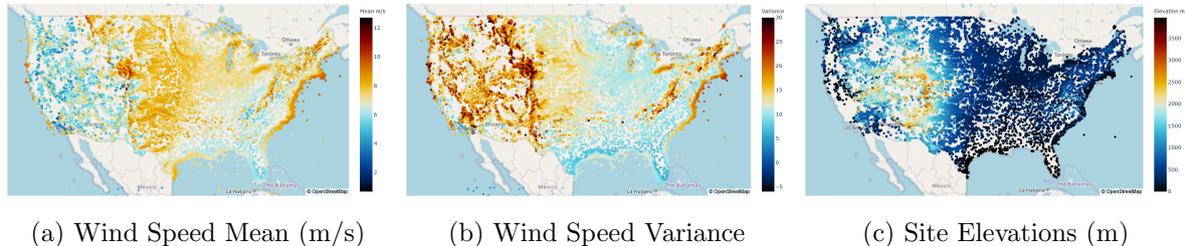


Figure 1. Geographical representation of wind speed mean, wind speed variance and site elevation for the NREL dataset

Previous Works in Forecastability

Wind generators must commit their electricity generation in advance, and for this reason, they predict their park outputs several time-points in the future (like 6, 12, 24 hours). If the real output is not aligned with the prediction, the park owner is economically penalised by the electricity system operator. A wind park easy to forecast will obtain better returns than one that is difficult, and in this sense, we can affirm that good forecastability increases the returns of a wind site.

Forecastability has appeared in the literature on wind prediction literature recently. The definition of this novel term can be found in an article from Rogers [19], where he defines several measures that correlate with the wind prediction. In this article, the measures are defined as Measure-correlate-predict (MCP) algorithms, a definition that evolves in subsequent papers as forecastability. We can find more recent works in this area like:

Girard et al. analyzed in [9] how the forecast error economically impacts wind parks in Denmark, using as source information the penalties incurred by the different wind farm owners provided by the Danish transmission system operator (TSO). In this work, they conclude that, in the assessment phase, forecastability has low value for the producer in an isolated wind park but increases its value if several locations are grouped together or if the wind park is already in production. This work is focused on the producer side and does not analyze the impact that forecastability may have on the overall electricity system. The market imbalance cost reduction impacts the site assessment and the operation phases. This research found that it can also impact maintenance and downtime for off-shore wind farms, as a low prediction quality increases the complexity to find episodes of good weather to access the wind turbines at sea.

Sanz Rodrigo et al. in [11] go more in-depth in applying forecastability properties to the assessment phase of a wind park finding that the capacity factor is significantly more relevant than predictability. However, it has value to determine the stability of the system and thus its resilience. The study uses Denmark and Ireland data but proposes to include sites located in a much broader area creating a virtual clustering. In the example, by adding sites from France and Spain, the prediction error is reduced 10%.

A site can be described by a time series generated from past meteorological measures. As predictability depends on the local wind features and wind can be described by a time series, some inner properties of the time series can describe the forecastability, forming the basis for the approach followed in the next two works.

Feng et al. in [7] analyse the characterisation of the time series structure by applying decomposition, linearity analysis of entropy. This approach is used in a subsequent work [8] where the defined characterisations are applied to wind sites in North America, analysing the relationship of the uncertainty in forecastability to spectral entropy using regression approaches.

Article Structure and Objectives

This article explores and analyses some novel forecastability measures and analyses how their combination improves the characterisation of the prediction error. The prediction is obtained using several deep learning approaches at different time horizons.

The defined forecastability measures are defined from different time series properties, like series decomposition, spectral analysis or aggregation methods. We work with 103 individual measures, and we validate their forecastability by comparing them with deep learning predictions on the US geography. The data used is from the NREL Wind dataset, which has 126,692 wind sites in North America [6]. The amount of data used for predictions and the multiple validations have required the use of High-Performance Computing infrastructure.

This article offers several contributions, as follows:

1. A proposal of forecastability measures that can have a practical use to determine the forecasting complexity of a site.
2. Verify the accuracy of the proposed measures using 126,692 representative sites from all different typologies of wind locations.
3. Obtain combinations of measures with higher correlation and potential regression ability to each time series.

The remainder of the paper is organised as follows. Section 1 describes the National Renewable Laboratory (NREL) dataset. Section 2 reveals the deep learning prediction models applied to the data. Section 3 develops different forecastability measures and their mathematical foundation, and in Section 4 we describe the experimental organisation with the measure exploration approach, regression strategy and individual measure contribution to the predictions. The article closes with some conclusions and with some insights on future work.

1. The Wind Toolkit Dataset from the National Renewable Laboratory (NREL)

The NREL dataset is a wind dataset that contains wind time series, composed of several weather variables and created by the NREL laboratory in the US. Comprises 126,692 sites distributed evenly in the US geography, which are located in a grid of 2x2 km. Each time series is seven years long, from January 1st 2007 to December 31st 2013, sampled every 5 minutes (over 730,000 steps long) and contains the following observations for each site: wind speed, wind direction, humidity, pressure and temperature [6].

The dataset contains a high diversity of wind patterns as they cross many climatic regions, ranging from the seas (Atlantic and Pacific) to the mountain peaks in the Rockies. The highest site is number 125,659, located at the North Star Mountain in Colorado, at latitude:

39.38160216, longitude: -106.101776 , above 3,000 m of altitude. Figure 1 illustrates a geographical representation of the wind-speed mean, wind-speed variance, and site elevation for the sites in the dataset, and we can see how different boundaries are drawn on each map, observing variability of winds or altitudes in the geography.

The NREL is a synthetic dataset, with data obtained from global meteorological models with some post-processing and cross-verified with real observed data. Its value lies in the large number of sites that allow experimentation on a wide diversity of wind time series. It is a large dataset and requires sizeable amounts of computing resources, but it provides the validation of results using a wide diversity of wind locations.

2. Deep Learning Methods Applied to Wind Prediction

Wind prediction can refer to the wind speed or electricity generated forecast (if applied to wind electricity generation turbines). There are many methods and approaches used for wind prediction. We can group them into two major categories, one set of methods based on weather prediction modelling and another set based on time series. For short-term prediction, time series are more accurate, and for longer-term ones, weather modelling is preferred.

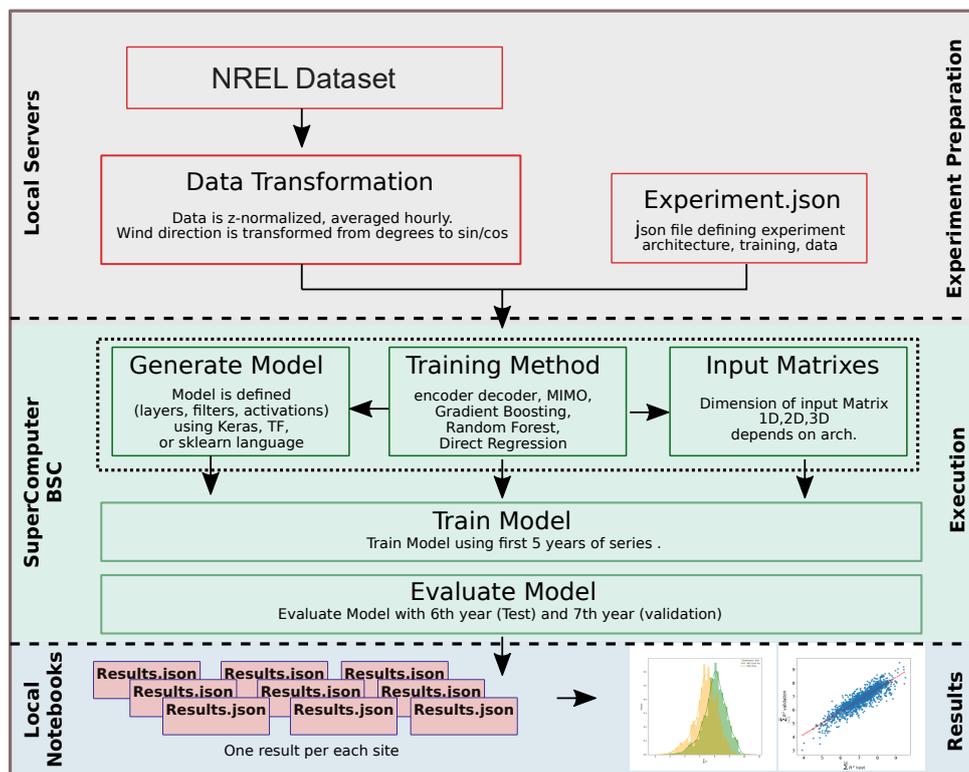


Figure 2. Experimental framework designed to apply several deep learning architectures to the NREL dataset

For this work, we will use deep learning prediction trained on wind time series. As wind is a complex phenomenon, the time series reflect this complexity, showing non-linearity and non-stationarity properties, which imply that the best performing forecasting algorithms must cope with both properties. Deep learning neural networks can model non-linear functions becoming a good candidate for the wind modelling.

There are several published results on the application of neural networks methods to wind forecasting using Multi-Layer Perceptron models (MLP) [12, 16, 20], Recurrent Neural Networks (RNN) [3, 13], and some are using Convolutional Networks (CNN) [23] (for a complete literary review see [15]).

In this work, we use a multi-step prediction approach. Multi-step consists of predicting a set of results in the future in one algorithm execution. The alternative, and more common, is the single-step that predicts a single point in the future. Multi-step prediction is not as widely researched as single-step but has some practical advantages over a single step [2].

We define an experimental methodology and a framework to design and validate the different deep learning architectures on the NREL dataset illustrated in Fig. 2. This methodology is based on three main steps or phases. First we transform the data for better neural network ingestion (z-normalisation, averaged) and we define an input model (based on .json structure) to define the architecture description (layer structure, kernels, regularisation strategies, neurons, training data structure).

Then we develop the implementation of each deep learning architecture using open-source deep learning frameworks (see Section 2.1) and then we train each one with five years of data and validate with two. The evaluation is made using different methodologies like MSE, R^2 or RMSE for each wind site and each predicted step (for example for 12 hours ahead we obtain 12 error measures), for an experiment, we obtain 126,692 values that combined are considered as a probability distribution.

In the last phase we analyse and validate each experiment for a site. For an horizon H is the sum of the individual error values for each prediction step, $\sum_1^H \hat{Y}_i$, and the result of an experiment across all the dataset is the average of the values from all sites (see Tab. 1)³.

Table 1. R^2 (cumulative for all steps) for deep learning models multi-step prediction at different horizons (1 hour, 6 hours and 12 hours). Models abbr. are: CNN: Convolutional Neural Network, CNN-sep: Convolutional separable, MLP: Multilayer perceptron, RNN: Recurrent neural network

Model	Description	1h	6h	12h
MLP	Multi Layer Perceptron	0.82	4.39	7.25
RNN	Recurrent Neural Network	0.81	4.35	7.15
CNN	Convolutional Neural Network	0.81	4.40	7.23
CNN-sep	Convolutional non-separable Network	0.82	4.43	7.32

The deep learning architectures belong to one of the three main categories, MLP, CNN and RNN. The CNN combines convolutional layers with fully connected layers to generate the output sequence. RNN have several typologies, using encoder-decoder constructions, or just combining the RNN with fully connected layers (similar to the CNN).

The results and conclusions for the experimentation are the following (for a detailed description of the experiments see [16]).

³The code and some evaluation notebooks are available at <http://github.com/castorgit/Articles-2020> and http://github.com/castorgit/Wind_code

- **RNN:** Recurrent Neural Networks are the architecture of choice for sequences, however, the best models use small input sequence windows of fixed size, and for this problem the CNN and MLP obtain better results.
- **MLP:** Multi-layer perceptrons obtain good results with architectures 2 or 3 layers depth with 512–1024 neurons each. Deeper architectures do not obtain better results.
- **CNN:** The convolutional 1-dimensional standard operation obtains better performance than the MLP, and the separable convolution shows the best results (see [4]), with filters (3x3, 9x9) and 2 or 3 layers depth.

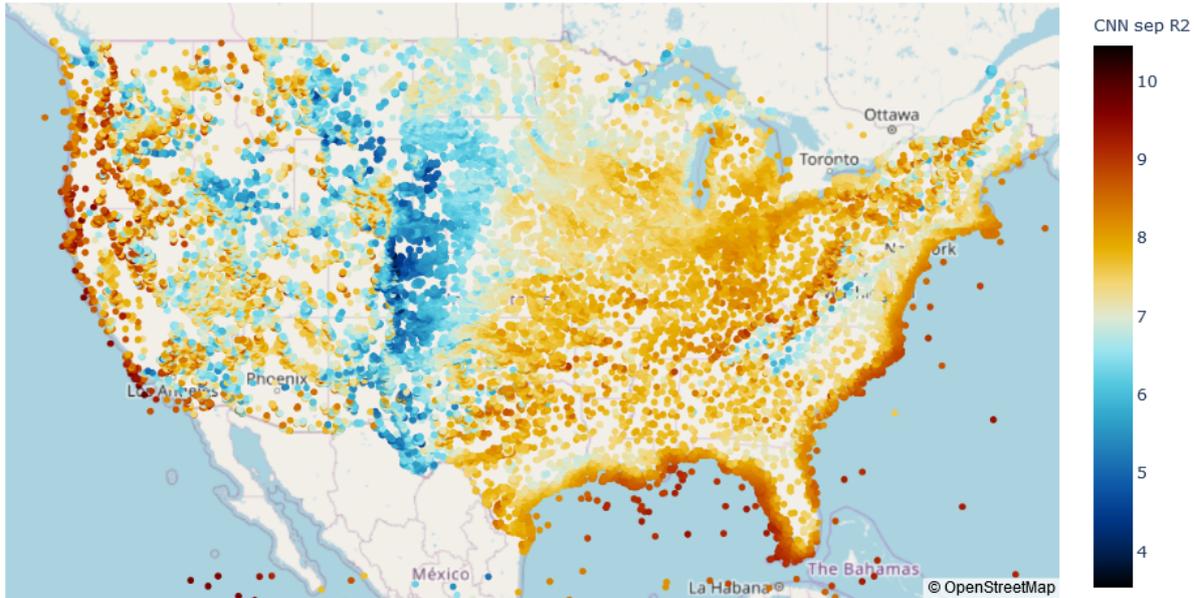


Figure 3. Accuracy of convolutional separable networks applied to the NREL dataset - blue less accurate, dark red more accurate

Table 1 shows the results of 4 main (RNN, MLP, CNN and CNN separable) architectures, that have been validated for prediction on 3 different horizons (1, 6 and 12h ahead).

The results, in this article, are always represented using R^2 , which is the coefficient of determination, which is a good measure to rate the accuracy of a regression, it is a number in the interval $[-\infty..1]$ which is the ratio between the explained variation and the total variation from the regression.

$$R^2 = 1 - \frac{\sum_i^N (y_i - \hat{y})^2}{\sum_i^N (y_i - \bar{y})^2} \quad (1)$$

An important conclusion from the experiments with the DL experiments is the high correlation between the accuracy results from the different architectures. The Pearson correlation between the deep learning methods is > 0.95 , which shows that the site wind pattern structure is more important than the DL specific approach to determine the complexity of a site or, formulated in a different way, that the DL architectures have better results in easy sites and worse results in difficult places. In this research we will try to understand what makes wind-time series complex for prediction.

To improve clarity in the analysis in the next sections we use as representative the results from the best performing architecture, which correspond with the CNN separable architecture that uses a 1-dimensional separable convolution in 2 layers.

2.1. Technical Assessment and HPC Requirements

From the software point of view the deep learning experimental architectures have been developed from scratch using Python 3.6, and using some add-on standard packages, being the most relevant:

- machine learning platform: Tensorflow 1.14.0;
- deep learning library: Keras 2.2.4;
- Scikit-learn machine learning library: 0.21.4;
- statistical support: statsmodels 0.11.

The number of sites (126,692) and the length of the time series (7 years of data) are optimal for an extensive experimentation with deep learning, however, the computer requirements recommended the use of an HPC resource. Thanks to the availability of such infrastructure we performed the experimentation on almost one-hundred alternative DL architectures.

The experiments have been supported by a NVIDIA GPU based computer at the Super Computing Centre [17]. The HPC resource used is the Minotauro cluster which is build with BULL and it has 39 bullx R421-E4 servers, each composed of:

- 2 Intel Xeon E5-2630 v3 (Haswell) 8-core processors (each core at 2.4 GHz, and with 20 MB L3 cache);
- 2 K80 NVIDIA GPU Cards;
- 128 GB of Main memory, distributed in 8 DIMMs of 16 GB – DDR4 @ 2133 MHz - ECC SDRAM;
- 1 PCIe 3.0 x8 8GT/s, Mellanox ConnectX-3FDR 56 Gbit;
- 4 Gigabit Ethernet ports.

The full HPC machine provides a Peak Performance of 250.94 TERAFLUPS distributed as 226.98 TERAFLUPS (K80) + 23.96 TERAFLUPS (Haswell).

3. Basic Forecastability Measures

In this section we describe forecastability measures that come from four major categories, basic statistics Section 3.1.1, time series decomposition Section 3.1.2, aggregation and stability Section 3.1.3 and spectral analysis Section 3.1.4. First we describe the theoretical foundation in Section 3.1 of each one and then we describe the experimentation in Section 3.2.

3.1. Theoretical Foundation of the Basic Forecastability Measures

If we plot the prediction error (measured in R^2) from the convolutional separable architecture across all US sites, we obtain Fig. 3. In this illustration, we can see how the accuracy results clearly define different geographical wind regions. We can observe how the rocky mountains split the map in two, with a turbulent area in the plains and more stationary wind regions by the seas (either in the Pacific or in the Atlantic oceans). With this geographical mapping in mind, we can see how prediction defines groups of sites, pointing to a relationship between the site time series structure and the complexity (or accuracy) of the prediction. This idea guides this work as we try to find the relationships between the internal structural elements in the time series and the prediction error obtained with the DL algorithms.

We perform an exploration of measures in four steps, first we use some basic statistic characterisation measures, like mean or variance, then we verify the use of time series decomposition,

we design some new approaches using aggregation and combination models and we finalise with entropy analysis.

We split this chapter in two sections, the first section contains the theoretical descriptions for each measure, and the second section describes the experimentation using these methods and the results obtained on the NREL dataset.

3.1.1. Basic statistic descriptive indexes

We consider basic descriptive indexes as the ones that have a simple statistical formulation, like mean, median, variance or standard deviation σ . We include elevation in this group, to see how the location altitude influences the prediction result.

In Fig. 1 we can see a geographical plot of Mean, Variance and Elevation of wind speed. In this map we can see how the measures define boundaries between different wind regions, but we need to understand how these measures are related with the prediction accuracy.

3.1.2. Measures based on time series decomposition techniques

If a weather phenomena repeats in cycles, it generates some kind of seasonal pattern in the wind time series, like a day-night flow, or heat-cold changes, or seasonal modifications of wind over the years. A good forecastability predictor will identify these patterns and, if for locations with high seasonality, the predictability measure can be quite effective.

In a time series, seasonality is identified by finding sub-sequences with a large auto-correlation coefficient. To find if a series has seasonal patterns, the best way is to decompose the series in components, using methods that split the series into different components.

Time series decomposition can be additive or multiplicative, depending on the operation (sum or multiplication) used to combine the individual elements of the series as can be seen in (2, 3).

Additive decomposition is the best choice if the time series does not hide a growth or decrease trend over time. When the variation of the pattern is proportional to the mean or level of the series then the multiplicative can be more appropriate, this happens with economic time series which show increases and decreases proportional to time.

Additive decomposition is more aligned with wind time series, as the mean of a wind time series is usually constant over time. For this reason the model applied in this work is the additive approach.

The objective in the decomposition is to split the series W_t in the three components, seasonal (S_t), trend (T_t) and residual (E_t). Depending on the type, the series can be formulated like:

$$W_t = S_t + T_t + E_t \quad \text{additive} \quad (2)$$

$$W_t = S_t \times T_t \times E_t \quad \text{multiplicative} \quad (3)$$

There are several decomposition methods. The classical method is the moving average, based on using the moving average on the series to isolate the trend-cycle, after removing the seasonality in the series (by applying filters or other approaches). In this work we name this method as the ‘classical’ approach W_{tc} , S_{tc} , T_{tc} , E_{tc} .

The STL (Seasonal and Trend decomposition using LOESS) is another widely used method for time series decomposition [5]. The LOESS method (locally estimated scatterplot smoothing)

uses the LOESS smoothing algorithm to extract smooths estimates of the three components, this method requires to specify the seasonal period.

A refinement of the decomposition consists in calculating the strength of the trend f_T or seasonality f_S in the time series, following the framework proposed by Wang et al. in [24].

For data with strong trend, the seasonality adjusted data (deseasonalized data) $W_t - S_t$ has more variation than the residual component, for data without trend the two variances would be the same. In this sense we define the strength of trend as:

$$f_T = 1 - \frac{\text{var}(E_t)}{\text{var}(W_t - S_t)}. \quad (4)$$

Strength of seasonality is defined in the same way, but in this case the variance used is the detrended data, $W_t - T$

$$f_S = 1 - \frac{\text{var}(E_t)}{\text{var}(W_t - T)}. \quad (5)$$

A time series with seasonality strength f_S equal to 0 has no seasonality, and when f_S is close to 1 it is an indicator for strong seasonality.

3.1.3. Measures based on aggregation and stability properties

Other measures that can be defined on sub-series combination features. In this section we describe methods that are based on tiled (non-overlapping) windows.

The first one is stability *Stab*, consisting of the variance of the means, The second one is lumpiness *Lump* or the variance of all the windows variances.

Stab and *Lump* are calculated obtaining sub-series of length *period* from the complete series (usually 12h, 24h, 3m, 6m), where *Lump* is the mean of each sub-series and *Stab* is the standard deviation.

$$Lump_{(P)} = \frac{1}{(N - P)} \sum_{j=1}^{N-P} \left(\frac{1}{P} \sum_{i=j}^{j+P} x_i \right), \quad (6)$$

$$Stab_{(P)} = \frac{1}{(N - P)} \sum_{j=1}^{N-P} \left(\sqrt{\frac{1}{P} \sum_{i=j}^{j+P} (x_j - \bar{x})^2} \right), \quad (7)$$

$$\text{where } N = \frac{\text{Series length}}{\text{Period}} \quad (8)$$

In this experimentation we calculate Lump and Stab for periods of 12, 24 hours, one week, one, three months and six months.

3.1.4. Spectral entropy analysis measures for wind time series

Entropy is a measure of the uncertainty of a random variable. In time series it can be used as a property to quantify the series, and it has been used in wind time series to decompose the signal and improve forecasting [21].

In this work we have used two different entropy calculations approaches, sample entropy and spectral entropy.

Sample entropy (named *SampEnt* in this work for convenience) assesses the complexity of the time series, a large value indicates high complexity while a small one indicates low complexity or more regular series. It was initially created for physiological time series signals [18].

$$H(x, m, r) = -\log \left(\frac{C(m+1, r)}{C(m, r)} \right), \quad (9)$$

where m is the embedding dimension (= order), r is the radius of the neighbourhood (default = $0.2\text{std}(x)$), $C(m+1, r)$ is the number of embedded vectors of length $m+1$ having a Chebyshev distance inferior to r and $C(m, r)$ is the number of embedded vectors of length m having a Chebyshev distance inferior to r .

Spectral entropy (*SpecEnt* in this article) is defined to be the Shannon Entropy of the Power Spectral Density (PSD) of the data:

$$H(x, sf) = - \sum_{f=0}^{f_s/2} PSD(f) \log_2[PSD(f)], \quad (10)$$

where PS is the normalised PSD, and f_s is the sampling frequency.

3.2. Experimentation with the Basic Forecastability Measures

We have performed a set of experiments.

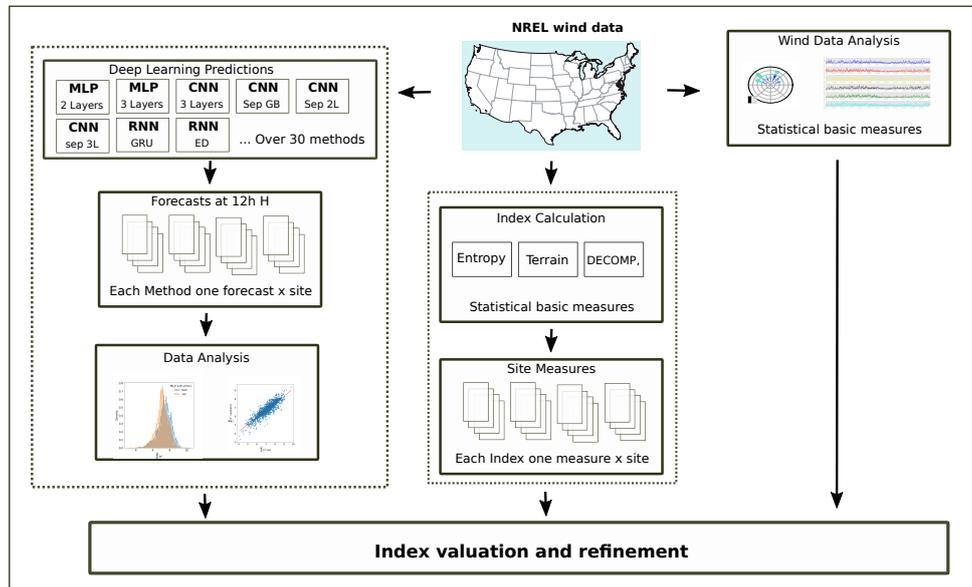


Figure 4. Experimental framework for forecastability measures validation

3.2.1. Correlations with the basic statistical measures

In Tab. 2 we show the correlations between three basic measures (wind-speed mean, wind-speed variance and site elevation) and the DL prediction.

The more significant correlation is with the wind-speed mean, and at lower level with the variance and the site elevation.

Table 2. Correlations between basic statistical measures and a DL prediction with a 12h ahead horizon

Measure	Pearson Correlation with DL prediction
Wind-speed variance	0.3113
Wind-speed mean	-0.5913
Terrain elevation (site altitude)	-0.2759

Representing the relationship between the measures and the predictions we obtain three scatterplots (see Fig. 5) where we can observe visually how the different measures adjust to the prediction.

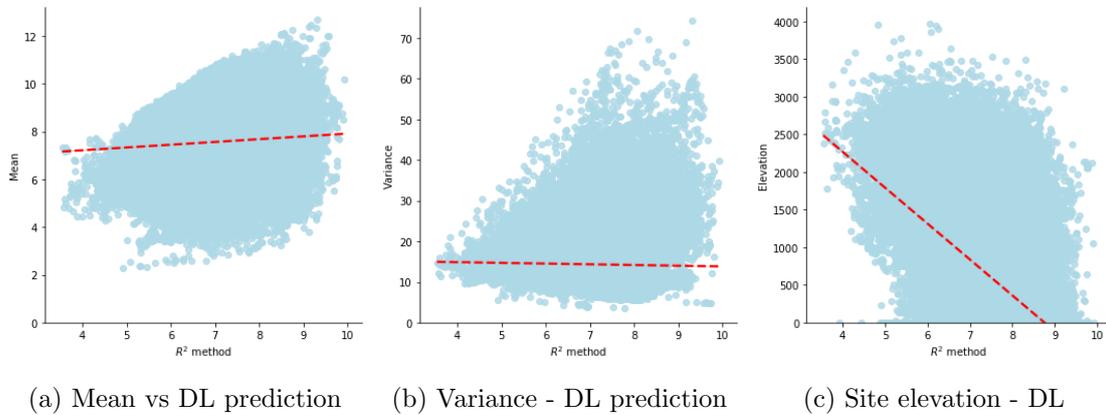


Figure 5. Scatterplots of Wind-speed mean, Wind-speed variance and site elevation with DL (best model convolutional separable network)

3.2.2. Correlation results between decomposition aggregations and spectral measures and DL prediction

We apply the different method approaches to each site, and then we calculate the Pearson correlation between each measure and the deep learning prediction. The results from all the different approaches are presented in Tab. 3.

We use two different decomposition methods, the STL (using LOESS) or the Least Squares. Both methods obtain quite similar results with some better correlations using the STL, because the smoothing method is more sophisticated and fits better the specific characteristics of wind data. We can see the comparison between the two decomposition methods in Fig. 6, where the scatterplots for f_T and f_S are presented side by side. In both cases using the STL method obtains better adjustment to the prediction, with higher correlation. The best correlations are obtained with 12 and 24 hours periods.

It is remarkable how the strength of Trend obtains a correlation of 0.888 which is extremely high, which points very good forecastability properties.

In Fig. 6 we illustrate with a scatterplot the relationship between the reference Deep Learning prediction and the f_S and f_T calculated with the LOESS and Least Squares methods, and in Tab. 3 the values are calculated on the NREL dataset.

Table 3. Pearson Correlations between Time Series decomposition measures (f_t strength trend and f_s strength seasonality), time horizons and DL Prediction across 126,692 sites

	Decomposition					
	STL			Least Squares		
	12h	24h	1m	12h	24h	1m
Strength Trend f_T	0.834	0.888	0.473	0.812	0.841	0.420
Strength Seasonal f_S	-0.199	0.153	-0.042	-0.151	0.110	-0.012
	Aggregation / Combination					
	12h	24h	1w	1m	3m	6m
<i>Lump</i>	0.247	0.282	0.324	0.352	0.346	0.117
<i>Stab</i>	-0.282	-0.058	0.110	0.075	0.038	-0.086
Spectral Analysis						
Spectral Entropy <i>SpecEnt</i>						-0.835
Sample Entropy <i>SampEnt</i>						-0.641

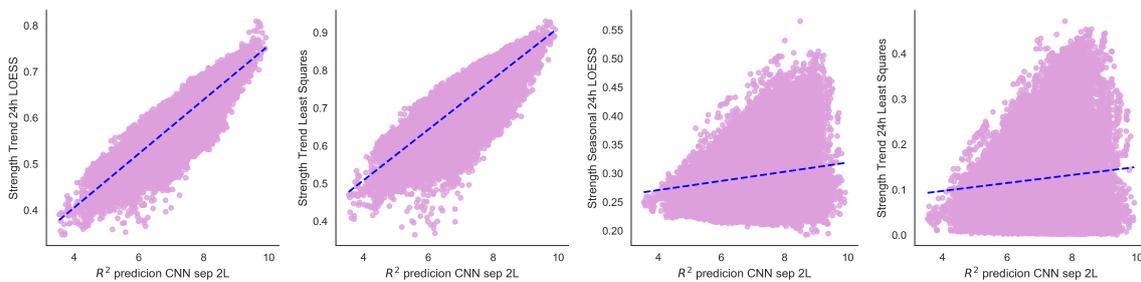
In the scatterplot figures we can observe how the LOESS model generates a cloud of points slightly better adjusted than the Least squares, and how the strength trend S_t obtains a higher correlation using the LOESS method versus the strength seasonality in a 24 hours period using the least squares method (Pearson correlation 0.888 vs 0.841) (see Fig. 7).

When it comes to aggregation measures we find different values depending on the period, being the most effective periods 1 week, 1 month and 3 months.

With Spectral analysis we obtain a high correlation with Spectral Entropy *SpecEnt* where the observed correlation is -0.835 and for *SampEnt* (sample entropy) is -0.641 (see Tab. 3 and Fig. 7).

3.2.3. Discussion on the basic forecastability measures results

We can observe the existence of correlations between basic measures and prediction, the strongest correlations are found in f_T and *SpecEnt*. The correlations are high and close to 0.8. However, we can see that both of them have a set of sites that seem not to be inside the cloud



(a) f_t STL vs DL (b) f_t L.Squares vs DL (c) STL f_s vs DL (d) L.squares f_s vs DL
Figure 6. Scatterplot of decomposition measures for f_T and f_S using two alternative methods, the STL and Least Squares. The X axis shows the DL prediction

points, that can be seen as a tail in the scatterplot figures (see Fig. 6 and Fig. 7). Analysing the outlier points we can see that the points do not belong to a particular area in the map, but are randomly distributed over the map. We propose some further analysis on these points to identify what makes them to have low correlation.

Aggregating a combination of measures obtains results with lower correlations, contributing to the definition of better forecastability measures. The following sections analyse this contribution and how the combination by regression can obtain better predictability.

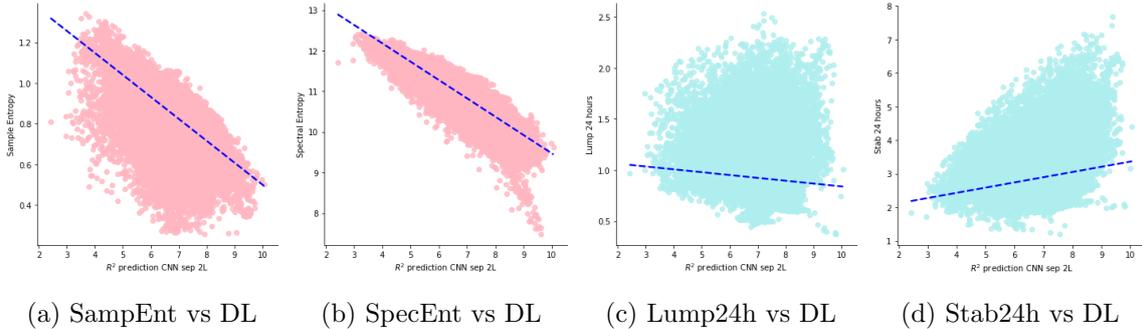


Figure 7. Spectral Analysis measures compared to deep learning prediction (CNN separable 2 layers) with horizon 12 hours ahead

4. Defining Complex Forecastability Measures

In this section we combine elementary measures to improve the forecastability of the results. To identify the best combinations we use the following methodology. In Section 3.2 we analyse the individual correlations, discussed in Section 3.2.3, after this preliminary analysis, in this section, we explore all the possible individual correlations in Section 4.1 and explore the measure importance using Principal Component Analysis (PCA) in Section 4.2. Then we build combinations using regression that obtain the highest predictive value in Section 4.3.

4.1. Analysis of Correlations with All the Individual Measures

In Section 3.2.1 we have analysed the correlation between the wind-speed variable and its transformation with the prediction. Now in the analysis we include all the possible defined measures which are the result of combining the measures *Lump*, *Stab*, f_T , *SpecEnt* with different periods (24h, 12h, 1 month, 3 months, 6 months), for each variable (wind_direction_sin, wind_direction_cos, temperature, pressure, density, wind_speed). The correlation matrix or confusion matrix can be seen in Fig. 8. We visualise high correlation clusters in this figure, The wind speed measures show a cluster, and the rest of measures (wind_direction, pressure, temperature and density in another one. Dark green areas show negative correlation,

Our conclusion, after analysing this figure, is that we can complement the wind_speed measures with other, but due to high correlation between pressure, density, direction or temperature, we do not need to choose the whole set, with a reduced representation we can obtain all the information required for the combined measure exercise.

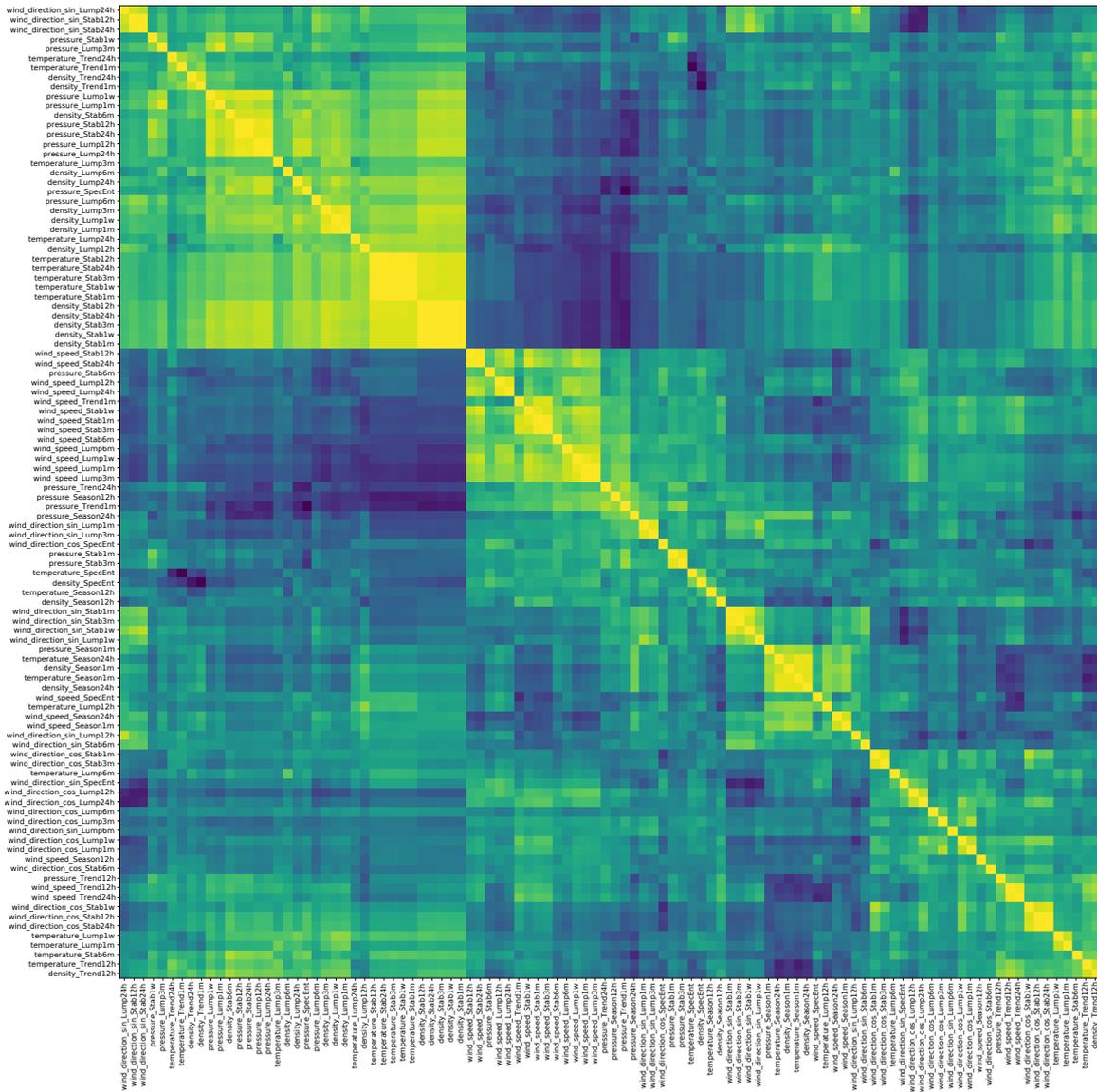


Figure 8. Correlations between measures

4.2. PCA Analysis and Measure Importance

To identify the variance contribution we perform a PCA analysis using the dimensions wind-speed, temperature and density, and using the prediction at 12h ahead horizon for the graphical representation (see Fig. 9). As a further exploration for the measure combination we perform a Principal Component Analysis (PCA), this method allows to reduce the dimensionality of the measures (considered as features). In this exploration we use measures based on wind speed, temperature and wind density. The PCA component decomposition in 3 dimensions reduces the dimensionality of the measures into 3, the PCA components show the underlying structure in the data by finding a set of axis (2-dimensional or 3-dimensional) that when applied to the data we maximize the data variance.

The amount of variance explained by the PCA on the three dimensions is [0.68504753 0.26162715 0.03701629], which amounts for a 94.67% with the first two components and 98.37% with the three components, showing that just with the two dimensions PC1 and PC2 we cover most of the variance.

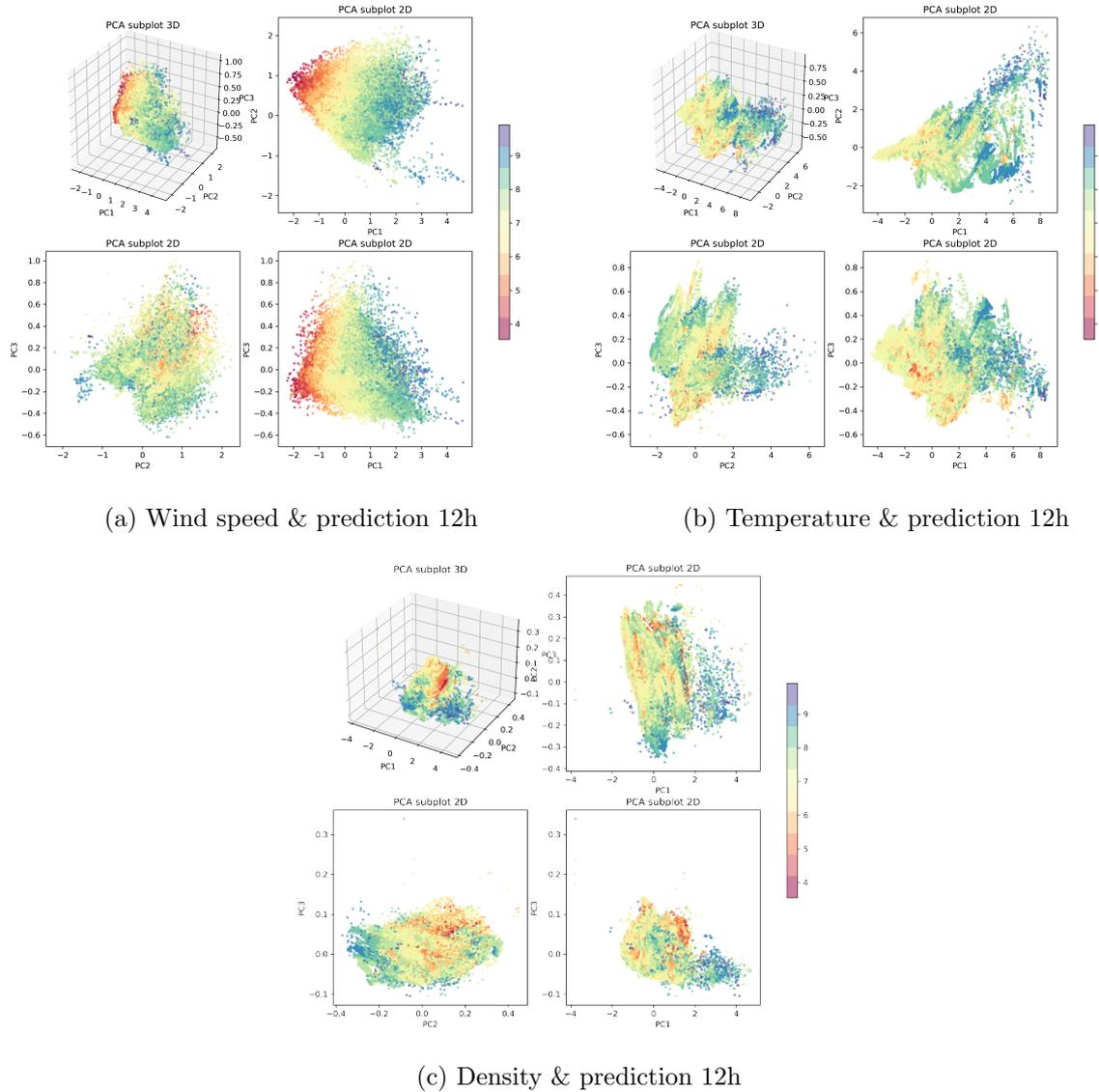


Figure 9. 3D PCA scatterplots using wind_speed, temperature, density, with several measures (*Lump*, *Stab*, f_T , *SpecEnt*) to identify the variance contribution on each dimension

We will use the PC1, and PC2 components in some of our regressions, discarding PC3 as it has low informative value.

4.3. Regression based on Multiple Measures

The next step consists of building a predictive measure, by combining the best features, that will have the highest forecastability capability of the wind prediction result.

This predictor allows to characterise the time series ‘forecastability’, where a high value determines that the series has potential for an accurate prediction using deep learning and a low result points to a time series that will have low accuracy prediction.

We consider the predictor as a function built using individual features or characterisation measures (based on decomposition or spectral analysis). This function can be considered a regression that combines each one of the individual indexes. To develop this function we try several regression models (see Tab. 4) using different techniques. Firstly we try regressions using several measures as regression variables, these models use one or several measures. Secondly we

we introduce some non-linearity by applying a support vector machine approach. We test all the models with the deep learning representative (CNN convolutional with separable convolutions) crating several models, each one with more information and variables.

The first regression model is based on the two measures that show the highest correlation with the prediction ($SapEnt$ and f_T) (see Fig. 10a).

We can consider a multi-variable x_i regression model to be formulated as:

$$y = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} + \varepsilon, \tag{11}$$

where the measures correspond to the x_i variables. To obtain the coefficients β_i , we transform the the objective into an optimisation problem where the goal is to reduce the residual ε term (sum of squares between the target and the regression). The specific optimisation technique defines the nature of the regression method, and it may influence the accuracy of the result. For this work we try two methods, the OLS (Ordinary Least Squares) and the LOESS (Locally Estimated Scatterplot Smoothing). The first optimisation algorithm applied is the OLS (see Fig. 10a). Using this approach and combining the two most relevant measures f_t and $SpecEnt$ we obtain results that are between 0.82 and 0.84 R^2 on the different horizons (see Tab. 4). Using only the two best measures we obtain a prediction result with a correlation over 0.9 and $R2$ of 0.84.

Correlation is not a good measure to validate a regression, and other measures are better tailored for this task. In this case we are using the R^2 or coefficient of determination (described in 1. We keep regression to have a reference with the individual measures, but the quality of the regression must be analysed using the R^2 in this section and in Tab. 4.

An R^2 result of 0.84 is very high, and can be interpreted as a relevant approximation. We illustrate this result in Fig. 10a with a scatterplot where the x axis is the regression result, and the y axis is the original prediction. We observe that the obtained cloud of points is quite adjusted to the optimal line, showing a better fit than the plot of the individual measures (see Fig. 7).

The second model is obtained using an alternative regression approach, the LASSO approach with two variables, in this case the result is very similar, with a correlation and R^2 equal to the least squares. LASSO is widely used and performs the regression by penalizing the sum of the absolute values of the weights β_j , it has the property to reduce the weight of some features, simplifying the model, however in this application as we use only two the results are similar to the OLS. As we can see in Fig. 10a and in Fig. 10b the scatterplots are quite similar.

For the third approach we use the LASSO properties for feature selection, to use all the measures and restrict the less relevant. We include in this regression measures with lower correlation

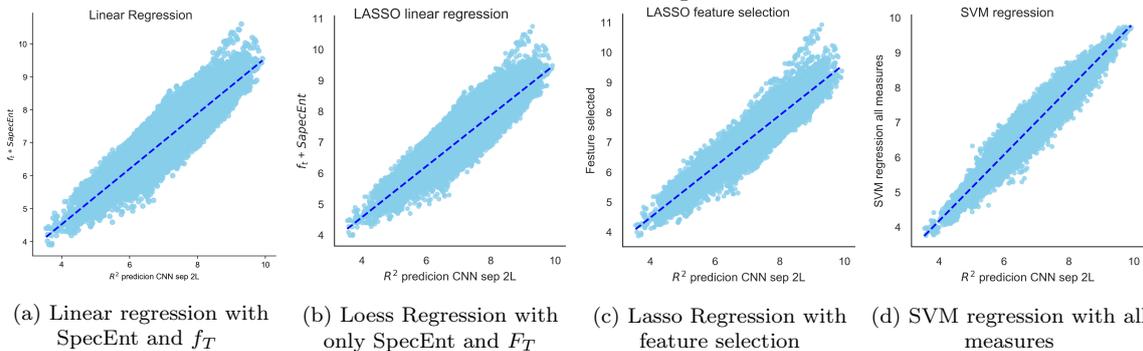


Figure 10. Comparison of regressions using the best individual features

(like *SampEnt*, *Lump* or *Stab*) and the LASSO models the regression with all the features, the final model assigns low weights to the less relevant features, (*Lump1w*, *Trend12h*, *Stab1w*) but extracts information from the rest.

This result obtains a correlation of 0.94 with an R^2 of 0.88, showing results that have very high predictability character. The scatterplot in this case (see Fig. 10c) shows a slightly shranked cloud of points.

As the regression does not improve by linear methods, we use a SVM Support Vector Machine. SVM are well known models proposed by Vapnik able to represent non-linearity by representing the data into a higher dimensional space. To do so we use a kernel function that performs the transformation. The kernel used for this regression is a Gaussian RBF (Radial Basis Function). We have very good results obtaining a coefficient R^2 of 0.950 and a correlation of 0.975 (for 12 hours ahead horizon), being the best regression by far. In Fig. 10d we illustrate the scatterplot of this regression where we can see how the cloud of points adjusts closer compared to the linear regressions.

Table 4. Pearson Correlations and R^2 values for regressions and measures (only wind speed)

Prediction Evaluation	1h		6h		12h	
	corr	R^2	corr	R^2	corr	R^2
Strength Trend f_t	0.92	-	0.90	-	0.83	-
<i>SpecEnt</i> Spectral Entropy	-0.71	-	-0.77	-	-0.83	-
Regression 2 measures	0.90	0.82	0.92	0.85	0.92	0.84
LASSO 2 measures	0.91	0.82	0.92	0.85	0.92	0.84
LASSO all measures	0.96	0.92	0.94	0.89	0.94	0.88
SVM	0.98	0.95	0.98	0.97	0.97	0.95

Conclusions and Future Work

In this article we propose several forecastability measures based on time series properties. We propose these measures as predictors of the accuracy of a method used on this site (as wind time series are originated on a single location). We propose the different measures and we build a regression model combining them that shows an almost perfect fit with the prediction (R^2 of 0.95).

We consider predictions performed using deep learning architectures and the comparisons are made with the best performing approach the convolutional separable models (see [14]). The predictions and measures are performed in the largest wind dataset available, the NREL dataset (126,691 sites), which allows to process a wide set of wind examples, but requiring sizeable resources that have been provided by an HPC infrastructure.

The conclusions from this experimentation are:

- Time series decomposition is a powerful tool to create measures that are correlated with prediction, in this way we find the f_T index based on the series trend that has a very high correlation with the predictions.

- Applying Signal Analysis decomposition we obtain a measure *SpecEnt* that has a very high correlation, and another measure *SampEnt* which has shown some valuable correlation with prediction.
- Two new measures defined using entropy analysis, *Lump* and *Stab*, have some correlation with the predictions.
- The combination of several measures using a linear regression approach, or a non-linear SVM obtains correlations very close to 1 with R^2 results over 0.95.

With these results, we can conclude that there are forecastability measures highly adjusted to deep learning predictions. In this way, we can obtain a prior evaluation of a time series predictability, that allows us to understand the complexity of the prediction for this site and therefore the economic value for energy commitment to the grid. Wind forecasters, wind farm developers and power system operators can benefit from the use of these predictor measures that characterise time series from a specific location. Adding these measures to the tool-set offers a new powerful characterisation tool. We can rate how easy or complex is a site before any prediction is performed.

As future work we propose to evaluate new combinations of individual indexes using linear and non-linear modelling and to apply finance modelling to the indexes to convert forecastability measures into economic values that allow the wind speed industry to take decisions based on site evaluation wind predictability.

Another possibility is to work in the definition of deep learning architectures that adapt themselves depending on the forecastability results for a specific location, in this way the prediction algorithm self-adapts to the data characteristics thus improving the result accuracy.

Acknowledgements

The authors would like to thank the Barcelona Supercomputing Center (BSC) for the usage of their resources and the United States National Renewable Laboratory (NREL) for the use of its Wind Toolkit (wind datasets). We would also like to thank the anonymous reviewers for providing valuable comments that helped to improve the quality of this paper.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Bastian, J., Jinxiang Zhu, Banunarayanan, V., Mukerji, R.: Forecasting energy prices in a competitive market. *IEEE Computer Applications in Power* 12(3), 40–45 (1999), DOI: 10.1109/67.773811
2. Ben Taieb, S., Bontempi, G., Atiya, A.F., Sorjamaa, A.: A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications* 39(8), 7067–7083 (2012), DOI: 10.1016/j.eswa.2012.01.039
3. Cao, Q., Ewing, B.T., Thompson, M.A.: Forecasting wind speed with recurrent neural networks. *European Journal of Operational Research* 221(1), 148–154 (2012),

DOI: 10.1016/j.ejor.2012.02.042

4. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, Honolulu, HI, USA. pp. 1800–1807. IEEE Computer Society, Los Alamitos, CA, USA (2017), DOI: 10.1109/CVPR.2017.195
5. Cleveland, W.S.: Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74(368), 829–836 (1979), DOI: 10.1080/01621459.1979.10481038
6. Draxl, C., Clifton, A., Hodge, B.M., McCaa, J.: The Wind Integration National Dataset (WIND) Toolkit. *Applied Energy* 151, 355–366 (2015), DOI: 10.1016/j.apenergy.2015.03.121
7. Feng, C., Chartan, E.K., Hodge, B.M., Zhang, J.: Characterizing time series data diversity for wind forecasting. In: Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, 5-8 Dec. 2017, Austin, Texas, USA. pp. 113–119. Association for Computing Machinery, New York, NY, USA (2017), DOI: 10.1145/3148055.3148065
8. Feng, C., Sun, M., Cui, M., Chartan, E.K., Hodge, B.M., Zhang, J.: Characterizing forecastability of wind sites in the United States. *Renewable Energy* 133, 1352–1365 (2019), DOI: 10.1016/j.renene.2018.08.085
9. Girard, R., Laquaine, K., Kariniotakis, G.: Assessment of wind power predictability as a decision factor in the investment phase of wind farms. *Applied Energy* 101, 609–617 (2013), DOI: 10.1016/j.apenergy.2012.06.064
10. Jacobson, M.Z., Delucchi, M.A., Bauer, Z.A., et al.: 100% clean and renewable wind, water, and sunlight all-sector energy roadmaps for 139 countries of the world. *Joule* 1(1), 108–121 (2017), DOI: 10.1016/j.joule.2017.07.005
11. Javier, S.R., Frías Paredes, L., Girard, R., et al.: The role of predictability in the investment phase of wind farms. In: *Renewable Energy Forecasting: From Models to Applications*, chap. 14, pp. 341–357. Woodhead Publishing Series in Energy, Elsevier - Woodhead Publishing (2017), DOI: 10.1016/B978-0-08-100504-0.00014-7
12. Li, G., Shi, J.: On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* 87(7), 2313–2320 (2010), DOI: 10.1016/j.apenergy.2009.12.013
13. Liu, Z., Gao, W., Wan, Y.H., Muljadi, E.: Wind power plant prediction by using neural networks. In: *IEEE Energy Conversion Congress and Exposition (ECCE)*, 15-20 Sept. 2012, Raleigh, NC, USA. pp. 3154–3160. IEEE (2012), DOI: 10.1109/ECCE.2012.6342351
14. Manero, J.: Deep learning architectures applied to wind time series multi-step forecasting. Ph.D. thesis, Technical University of Catalonia UPC. Department of Computer Science (2020), <http://hdl.handle.net/2117/328183>
15. Manero, J., Béjar, J., Cortés, U.: Wind energy forecasting with neural networks. a literature review. *Computación y Sistemas* 22, 1085–1098 (2018), DOI: 10.13053/CyS-22-4-3081

16. Manero, J., Béjar, J., Cortés, U.: “Dust in the Wind...”, Deep Learning application to Wind Energy time series forecasting. *Energies* 12(12), 2385 (2019), DOI: 10.3390/en12122385
17. Martorell, J.M.: Barcelona Supercomputing Center: Science accelerator and producer of innovation. *Contributions to Science* 12(1), 5–11 (2016), DOI: 10.2436/20.7010.01.238
18. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 278(6), H2039–H2049 (2000), DOI: 10.1152/ajpheart.2000.278.6.H2039
19. Rogers, A.L., Rogers, J.W., Manwell, J.F.: Comparison of the performance of four measure-correlate-predict algorithms. *Journal of Wind Engineering and Industrial Aerodynamics* 93(3), 243–264 (2005), DOI: 10.1016/j.jweia.2004.12.002
20. Shi, J., Guo, J., Zheng, S.: Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renewable and Sustainable Energy Reviews* 16(5), 3471–3480 (2012), DOI: 10.1016/j.rser.2012.02.044
21. Sun, W., Wang, Y.: Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management* 157, 1–12 (2018), DOI: 10.1016/j.enconman.2017.11.067
22. Walsh, C., Pineda, I.: Wind energy in Europe in 2018 trends and statistics. <http://windeurope.org/about-wind/statistics/european/wind-energy-in-europe-in-2018/>, accessed: 2021-01-09
23. Wang, J., Zong, Y., You, S., Trholt, C.: A review of Danish integrated multi-energy system flexibility options for high wind power penetration. *Clean Energy* 1(1), 23–35 (2017), DOI: 10.1093/ce/zkx002
24. Wang, X., Smith, K., Hyndman, R.: Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery* 13(3), 335–364 (2006), DOI: 10.1007/s10618-005-0039-x