

An Ensemble of Echo State Networks for Predicting the Energy Production of Wind Plants

Sameer Al-Dahidi

Mechanical and Maintenance Engineering Department, School of Applied Technical Sciences, German Jordanian University, Amman 11180, Jordan; sameer.aldahidi@gu.edu.jo

Piero Baraldi

Energy Department, Politecnico di Milano, Milan 20156, Italy. E-mail: piero.baraldi@polimi.it

Eleonora Nigro

Energy Department, Politecnico di Milano, Milan 20156, Italy. E-mail: eleonora.nigro@mail.polimi.it

Enrico Zio

Energy Department, Politecnico di Milano, 20156, Milan, Italy.

MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France.

Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea. E-mail: enrico.zio@polimi.it

Montelatici Lorenzo

Research Development and Innovation, Edison Spa, Foro Buonaparte 31, Milan 20121, Italy. E-mail: montelatici.lorenzo@edison.it

The electricity grid relies on a mixture of conventional (e.g. oil and gas) and renewable (e.g. wind, solar, and geothermal) energy sources. Ensuring the reliability of electric power distribution becomes a fundamental and complex issue due to the stochasticity of the production from renewable sources and the fluctuating behaviour of energy market demand. The necessity to integrate and mitigate these two uncertainty sources requires to tackle the problem of the unit commitment. It is, therefore, fundamental the capability of predicting electrical power output from plants with intermittent energy sources. We propose an approach to predict wind energy production based on an ensemble of Echo State Networks (ESNs) trained with different sets of historical data. A novel Local Fusion (LF) strategy is employed to aggregate the predictions of the individual ESN models. The proposed approach is applied to the prediction of the energy production of a wind plant located in Italy. The obtained results show that the proposed ensemble provides more accurate predictions than a single ESN model and an ensemble approach of literature.

Keywords: Energy prediction, ensemble, echo state networks, local fusion, wind plant.

1. Introduction

The modern electricity grid relies on a mixture of conventional (e.g. oil, gas, chemical, etc.) and Renewable Energy (RE) sources (e.g. wind, solar, geothermal, etc.) (International Energy Agency 2018; Gielen et al. 2019).

The issue of efficiently managing the heterogeneously distributed energy sources to meet the user electricity demand is known as Demand Side Management (DSM) (World Energy Council 2016; D. Li, Chiu, and Sun 2017). The difficulty of this task is due to the stochasticity of the RE sources, which depend on weather conditions (World Energy Council 2016; Viviescas et al. 2019), and the fluctuating behaviour of energy market demand. The need of integrating and mitigating the consequences of these two uncertainty sources requires to inject

flexibility into conventional power plants production and to tackle the problem of unit commitment (Costa et al. 2008).

In this context, the present work is dedicated to the problem of short-term forecasting of energy production from wind plants, using as input data weather forecast. To obtain accurate energy production predictions, the following issues should be taken into account:

- 1) the selection of the weather forecast signals to be used as input of the predictive model;
- 2) the uncertainty on the weather forecast data used;
- 3) the choice of the appropriate modeling approach for the development of the predictive model.

Issue 1) is typically tackled resorting to feature selection techniques, which aim at identifying the best subset of the weather forecast signals for the model prediction. Example of feature selection approaches can be found in (Bolón-Canedo, Sánchez-Marroño, and Alonso-Betanzos 2013). With respect to issue 2), a method based on an ensemble of Artificial Neural Networks (ANNs) has been proposed in (Al-Dahidi et al. 2017) to obtain robust estimation in case of large input uncertainty and to estimate the output uncertainty. In this work, we focus on the development of the predictive approach (issue 3).

Models for the prediction of energy production from wind plants are typically classified as model-based and data-driven (Sideratos and Hatzigiorgiou 2007; Al-Dahidi et al. 2017). The use of model-based approaches has been limited by the difficulty of developing accurate physics-based models of the wind plants. In practice, it is typically necessary to rely on assumptions which in turn lead to a performance reduction (Barbounis and Theocharis 2006; Costa et al. 2008; Al-Dahidi et al. 2017). On the other hand, data-driven approaches, such as ANNs (Ramamamy, Chandel, and Yadav 2015; Al-Dahidi et al. 2017), Support Vector Machines (SVMs) (L. L. Li et al. 2020) and K -nearest neighbors (K -nn) regression (Kramer and Treiber 2015), which rely only on the availability of historical data, have been extensively explored during the last years.

Since energy production from wind plants is characterized by cyclical and periodic behaviors caused by the wind characteristics, in this work we consider the possibility of using a recursive model. In particular, we resort to Echo State Networks (ESNs), which are a type of Recurrent Neural Networks (RNNs).

RNNs have been largely used in the last years in various application areas, such as wind power forecasting (Costa et al. 2008; Soman et al. 2010), estimation of the Remaining Useful Life (RUL) of industrial components (Rigamonti et al. 2017), speech recognition (Graves, Mohamed, and Hinton 2013), generating image descriptions (Karpathy and Fei-Fei 2017), time series anomaly detection (Malhotra et al. 2015), human action recognition (W. Li et al. 2017). ESNs differ from traditional RNNs in the conceptual separation between the reservoir (i.e. a randomly created RNN used as nonlinear temporal expansion function), and a linear recurrence-free readout, which combines the expansion and produces the desired output. Since only the readout needs to be trained, ESN training is less computational intensive than that of traditional RNNs.

In particular, we consider the use of an ensemble of ESNs to increase prediction accuracy

and robustness (Polikar 2006; Khosravi et al. 2011; Al-Dahidi et al. 2017).

The novelty of the present work is two-fold:

- i) the use of an ensemble of ESNs trained using different datasets and initial weights. This allows obtaining diversity among the ensemble base models, which has been shown to greatly increase the overall model accuracy (Rigamonti et al. 2017).
- ii) the aggregation of the outcomes of the ensemble base models by using a properly defined Local Fusion (LF) strategy. The contribution of the base models to the ensemble outcome is proportional to their local performance in the neighborhood of the test pattern under analysis.

The proposed ensemble approach is applied to an industrial case study concerning energy production prediction of a wind power plant. The obtained results are compared to that obtained by using a single ESN and a Global Fusion (GF) aggregation strategy of literature.

The remainder of this work is organized as follows. In Section 2, the problem of predicting wind energy production is introduced. In Section 3, the proposed ensemble method is illustrated. In Section 4, the application of the proposed ensemble to a real case study of a wind plant is presented and the obtained results are compared with those obtained by methods from literature. Finally, some conclusions are drawn in Section 5.

2. Problem Statement

The objective of the present work is the prediction of the energy production of a wind plant, P , with a time horizon Δt , i.e. at the present time t we want to predict the plant production, $P(t + \Delta t)$. The future production of a wind plant is an uncertain quantity which depends on several factors, such as plant configuration, degradation of its components and environmental conditions.

In this work, we focus on the effects of the environmental conditions on wind plant production. We assume to have available a weather forecast model, W , which predicts the environmental conditions, $\vec{E}(t + \Delta t)$, on the basis of input data, $\vec{I}(t)$, collected at the present time t (Figure 1).

Since the model W is an imprecise representation of the reality, its predictions $\vec{E}(t + \Delta t)$ is affected by an error \vec{e} with respect to the true environmental conditions, $\vec{E}(t + \Delta t)$, at time $t + \Delta t$:

$$\vec{E}(t + \Delta t) = \vec{E}(t + \Delta t) + \vec{e} \quad (1)$$

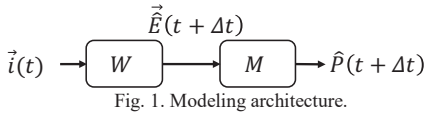


Fig. 1. Modeling architecture.

The objective is, then, to build a model M that receives in input the predicted environmental conditions, $\vec{E}(t + \Delta t)$, provided by model W , and provides in output the predictions of the plant production, $\hat{P}(t + \Delta t)$ (Figure 1 (right)). We assume to have available a dataset containing the historical predictions of the environmental conditions and the corresponding real plant production data:

$$D = \{\vec{E}(t + \Delta t), P(t + \Delta t)\} \quad (2)$$

3. Method

The proposed method for the prediction of the wind power production is based on an ensemble of ESNs. Section 3.1 briefly introduces ESNs, whereas Section 3.2 describes the proposed ensemble model.

3.1 Base prediction models

ESNs have been chosen as base models of the ensemble for the following reasons:

- their ability of modeling the temporal behaviour of dynamic complex systems;
- their ability to preserve information about the input time history. This is obtained by using feedback connections between the neurons of a layer and those of the preceding layers, which give memory to the ESN, i.e. the system output depends on the observed input time history (Herbert Jaeger 2001; Moustapha and Selmic 2007).
- the smaller computational cost of the ESN training with respect to that of traditional RNN. This is due to the conceptual separation between the reservoir (i.e. a randomly created RNN used as nonlinear temporal expansion function), and a linear recurrence-free readout for synthesizing the expansion and producing the desired output (Figure 2). Notice that only the readout needs be trained (Lukoševičius and Jaeger 2009).

The ESN considered in this work has L input units corresponding to L different weather forecasting signals:

$$\vec{E}_t = [E_t^1 \ E_t^2 \ \dots \ E_t^L] \quad (3)$$

The reservoir is characterized by N internal network units whose internal states are represented by the vector $u_t = [u_t^1 \ u_t^2 \ \dots \ u_t^N]$ and

one output unit producing the output signals, $y_t = P_t$.

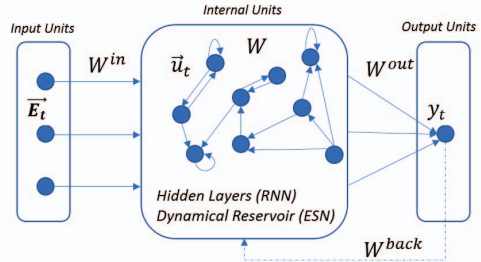


Fig. 2. ESN architecture (adapted from (Peng et al. 2012)).

The activation of internal units u_t at time t is obtained using:

$$\vec{u}_t = f \left(W^{in} \vec{E}_t + W \vec{u}_{t-1} + W^{(back)} y_{t-1} \right) \quad (4)$$

where $f = (f_1, \dots, f_N)$ are the internal units activation functions, which are typically sigmoidal, $W^{in} = (w_{ij}^{in})$ is the $N \times L$ input weights matrix, $W = (w_{ij})$ is the $N \times N$ internal weights matrix, and $W^{back} = (w_{ij}^{back})$ is the vector of the output feedback weights. The input weights, W^{in} , and the output feedback weights, W^{back} , are randomly generated from a uniform distribution. Both W^{in} and W^{back} are properly scaled: the scaling of W^{in} (IS) and shifting of the input (IF) depend on how much nonlinearity is needed. If the inputs are close to 0, the sigmoidal neurons tend to operate with activations close to 0, where they are essentially linear, whereas if the inputs are far from 0, the neurons tend to operate toward saturation, where they exhibit more nonlinearity.

The same idea drives the choice of the output scaling (OS) and shifting (OF), whose values affect the range of the trained W^{out} and might lead to an unstable condition. Finally, the scaling of W^{back} (OFB) is, in practice, limited by a threshold at which the ESN starts to exhibit an unstable behavior, i.e. the output feedback loop starts to amplify the output entering into a diverging generative mode (H. Jaeger 2001). The output provided by the ESN is given by:

$$y_t = f_{out} \left(W^{out} (\vec{E}_t, \vec{u}_t, y_{t-1}) \right) \quad (5)$$

where f_{out} is the output unit activation function, which is typically linear, and $W^{out} = (w_j^{out})$ is the $(L+N+1)$ output weights vector. The ESN training aims at finding optimal values of W^{out} and is performed through a Least Squares linear regression with the objective of minimizing the error between the network output and a target signal on a set of training data. Once the ESN has been trained, it can be used to predict the output,

$y_t = P_t$, applying Eq. (4) and Eq. (5) to the input \vec{E}_t (Rigamonti et al. 2017).

The ESNs are characterized by the echo state property (Herbert Jaeger 2001), which states that the effect of a previous state \vec{u}_t and a previous input \vec{E}_t on a future state \vec{u}_{t+j} should vanish gradually as time passes, and not persist or even get amplified. This property is practically assured if the reservoir weight matrix W is scaled so that its spectral radius (SR) $\rho(W)$ (i.e. the largest absolute eigenvalue of W) satisfies $\rho(W) < 1$.

3.1.1 ESN hyperparameter setting

One major difficulty when developing an ESN model is the setting of the architecture and its hyperparameters, which requires a certain level of expertise (Rigamonti et al. 2017). To overcome this difficulty, we first prioritize the importance of ESN hyperparameters according to (Rigamonti et al. 2017), as reported in Table 1.

Table 1. The two levels of importance of the ESN hyperparameters.

First Order Hyperparameters (<i>FOPs</i>)	Second Order Hyperparameters (<i>SOPs</i>)
Size of the reservoir (S)	Input scaling (IS)
Spectral radius (SR)	Teacher scaling (TS)
Connectivity (C)	

Then, a Hierarchical Grid Search approach (HGS) (Pontes et al. 2016) is used to set the ESN hyperparameters. The objective is the minimization of the Mean Absolute Error (*MAE*) over a test set of data:

$$MAE = \sum_{l=1}^{N_{test}} |\hat{P}_l - P_l| / N_{test} \quad (6)$$

where N_{test} is the number of patterns in the test set, and \hat{P}_l and P_l are the predicted and the real energy production values, respectively.

The HGS procedure is based on the following three steps:

- 1) optimize *FOPs* keeping *SOPs* fixed at reference values found in literature;
- 2) fix *FOPs* at their best values found in 1) and optimize *SOPs*;
- 3) verify if the same *FOPs* are obtained when *SOPs* are fixed at their best values. If this is not verified, repeat the procedure.

3.2 Ensemble of ESNs

Ensembles of models have been used in many different research fields to improve the prediction accuracy of a single model (Polikar 2006). An

ensemble is constituted by multiple regression models (hereafter called base models) whose outcomes are aggregated into a final aggregated prediction. The basic idea is to diversify the base models to leverage their strengths and overcoming their drawbacks (Rigamonti et al. 2017). The aggregation of the outcomes provided by the base models of the ensemble has been shown to provide more accurate predictions when compared to each individual base model (Polikar 2006; Baraldi et al. 2013; Al-Dahidi et al. 2017).

The development of an ensemble requires (Rigamonti et al. 2017; Al-Dahidi et al. 2017):

- i) to generate diverse base models (Section 3.2.1);
- ii) to define a strategy for the aggregation of their outcomes (Section 3.2.2).

3.2.1 Generation of Diverse Models

Diversity among ESN-based regression models can be generated by using models that are (Polikar 2006; Rigamonti et al. 2017; Al-Dahidi et al. 2017):

- i) trained with different training datasets;
- ii) trained using different initial weights;
- iii) fed by different sets of input features

In this work, we adopt strategies i) and ii). Since a training dataset of a recursive model must be formed by temporally consecutive patterns, the traditional methods for the creation of diverse datasets, such as bagging (Breiman 1996; Al-Dahidi et al. 2017), cannot be applied. The novelty introduced in this work consists in using training sets formed by consecutive patterns with different starting points and durations. The starting point is randomly sampled from an uniform distribution between $[0, T_h/2]$, where 0 represents the first available historical data, and T_h the total number of training patterns, whereas the length of the dataset is randomly sampled from an uniform discrete distribution in the range $[T_h/2, T_h]$. The total number of sampled training sets will be indicated by I . Then, A different ESNs are trained using different initial weights for each training set to add further diversity to the ensemble of models. The total number of generated ensemble models is, therefore, $A \times I$.

3.2.2 Aggregation of the outcomes of the individual models

Aggregation strategies are typically classified into Global Fusion (GF) and Local Fusion (LF) methods (Xue, Subbu, and Bonissone 2006; Rigamonti et al. 2017; Al-Dahidi et al. 2017). GF methods, such as median and Globally Weighted Average (GWA), assume that the contribution of a given base model to the ensemble outcome is

always the same, regardless of the input pattern under study. Contrarily, LF methods assume that the contribution of the base models to the ensemble outcome is proportional to their local performance in the neighborhood of the test pattern under analysis. LF methods, which are also called dynamic weighting methods, have shown performance superior to that of global methods in various industrial applications, such as efficiency and load predictions in power plants (Xue, Subbu, and Bonissone 2006; Costa et al. 2008) and RUL predictions of industrial components (Rigamonti et al. 2017). For this reason, a LF method is adopted in this work.

Figure 3 shows the flowchart of the adopted LF method, which is based on the following three steps:

- 1) the test set is divided into periods characterized by similar environmental conditions, e.g. a month;
- 2) the $A \times I$ available models are applied to the first S patterns of the period and the best H models are identified by computing their local accuracy. To this purpose, the MAE metric (Eq. (6)) is used:

$$MAE_{i,a} = \frac{\sum_{k=1}^S |\hat{P}_{i,a}(s) - P(s)|}{S} \quad (7)$$

where $P(s)$ and $\hat{P}_{i,a}(s)$ indicate the true output and the corresponding prediction of the s -th test pattern.

- 3) the aggregated ensemble outcome is the median of the best H models.

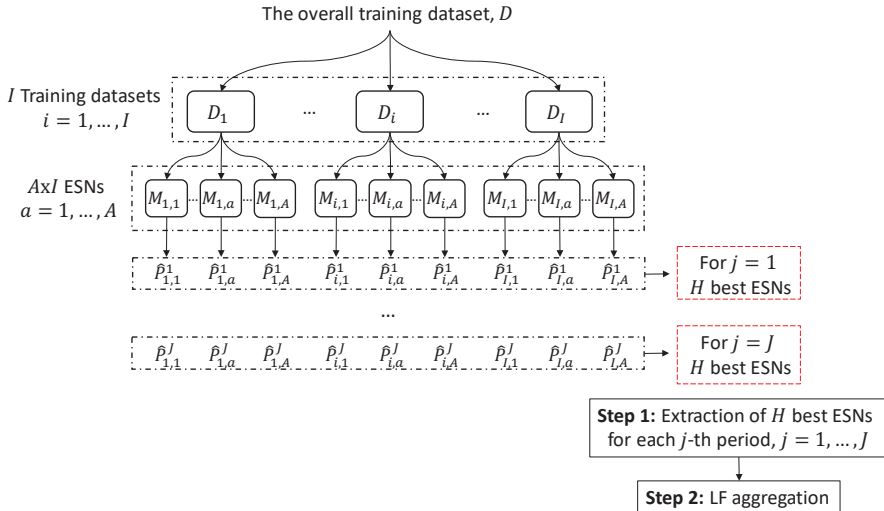


Fig. 3. Flowchart of the LF aggregation method.

4. Application of the proposed ensemble for energy production prediction in wind plants

We consider a wind plant located in Italy for which weather forecast data and corresponding energy production values collected from 2011 to 2013 (Al-Dahidi et al. 2017) are available. The weather forecast signals, \hat{E} , include $F = 19$ features. The forecast horizon is $\Delta t = 24$ hours, i.e. at a given time t_o , the weather forecasts of the following $\Delta t = 24$ hours are available (Al-Dahidi et al. 2017).

Figure 4 shows the one-day ahead wind speed forecast (top) and the associated energy productions (bottom) of two different days in winter (circles) and summer (squares) conditions. It is interesting to notice the large variability of wind speed in different seasons and at different hours of the day (Al-Dahidi et al. 2017).

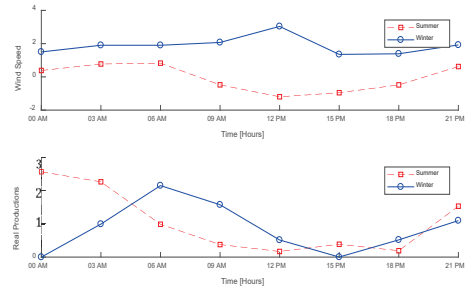


Fig. 4. One day-ahead wind speed forecasts (top) and associated energy productions (bottom).

The 2011-2013 dataset is divided as follows:

- the data collected in the period [2011, 2012] are used to build the base models of the proposed ensemble and setting their hyperparameters;
- the data collected during 2013 are used to evaluate the performance of the proposed ensembles.

Table 2 reports the setting of the ESN hyperparameters obtained by using the hierarchical grid search (Section 3.1.1).

Table 2. Setting of the hyperparameters.

Size	SR	C	IS	TS	MAE
466	0.02	0.13	$10^{-1.44}$	10^{-3}	3.8802

Once the base model architecture is defined, the ensemble model is developed. To this aim, $I = 100$ training datasets are sampled by applying the procedure described in Section 3.2.1 to the 2011-2012 data and used to train a number of $A = 5$ ESN, thus originating $A \times I = 500$ different ESNs, which constitute the base models of the ensemble.

Figure 5 shows the production predictions provided by the 500 base models of the proposed ensemble (thin lines) on 50 test data patterns, together with the true productions (bold line).

Notice that there are time periods where all the base models (thin lines) underestimate or overestimate the real energy production (bold line), and other periods where the base models commit error of different signs.

Among the 500 base models of the ensembles, the best $H = 50$ models are selected on the basis of their performances using the MAE metric. The test set is divided into 12 periods of 1 month and $s = 35$ patterns (corresponding to the first week of weather previsions) are used for the local performance computation.

4.1 Application results

Table 3 reports the average MAE values obtained by a single ESN model and two ensemble of ESN models aggregated using a Global Fusion (GF) and a Local Fusion (LF) strategy. It is worth mentioning that the standard deviations obtained by the GF and LF ensembles when 5000 base models are considered are not reported due to the excessive computational burden.

Table 3. MAE and their standard deviations.

Single ESN	Ensemble of 500 ESNs		Ensemble of 5000 ESNs	
	GF	LF	GF	LF
4.169 ± 0.108	4.039 ± 0.005	3.872 ± 0.005	4.039	3.789

One can easily recognize that the ensemble approaches are superior to the single ESN model. Also, the LF strategy outperforms the GF strategy using both 500 and 5000 base models and the prediction performance increases when more base models are used.

Figure 6 shows the evolution of the MAE as a function of the number H of the best models selected by the LF aggregation strategy.

Notice that the best setting of H is 3 for the ensemble of 500 ESNs, and 7 for the ensemble of 5000 ESNs (Table 4).

Table 4. Best setting of the number of selected ESNs and the correspondent MAE.

	LF 500 ESNs	LF 5000 ESNs
Number of selected ESNs	3	7
MAE	3.793	3.743

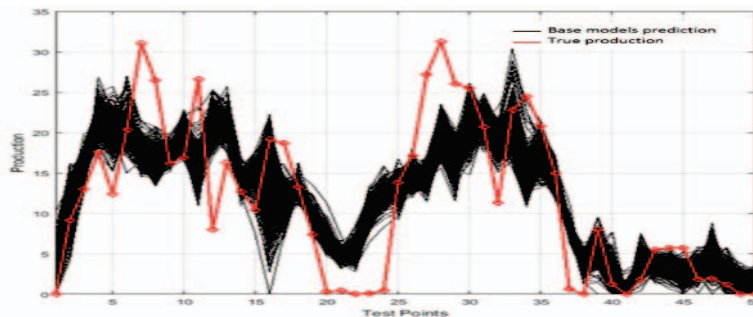


Fig. 5. Diversity among models of the proposed ensemble.

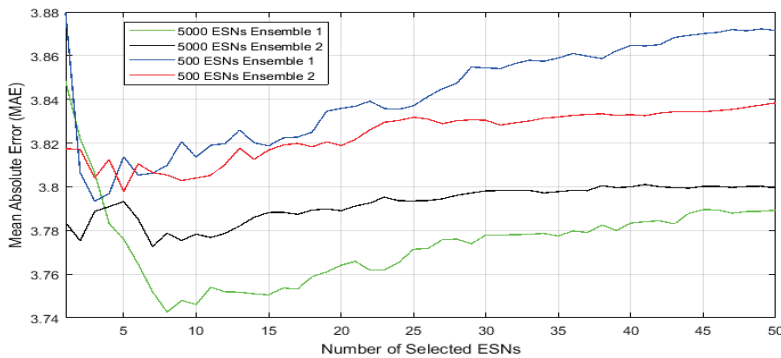


Fig. 6. MAE as a function of the selected number of ESNs in the LF method, for the proposed ensemble (with 500 and 5000 ESNs base models).

5. Conclusions

An ensemble of Echo State Networks (ESNs) has been developed for the prediction of the energy production of a wind plant. The ESN has been chosen as base prediction model due to its intrinsic memory and computational efficiency. A hierarchical grid search approach has been implemented in order to set the optimal parameters architecture of the ESN. The main novelty of the proposed method concerns: 1) the sampling of the datasets used to train the ensemble base models and the proposed local fusion strategy for the aggregation of the outcomes provided by the base models.

The method has been applied to an industrial case concerning the prediction of the energy production of a wind power plant located in Italy. The proposed method has been demonstrated able to provide better prediction performance with respect to an ensemble based on a Global Fusion (GF) aggregation strategy and a single ESN model.

References

Al-Dahidi, S., P. Baraldi, E. Zio, and E. Legnani. 2017. "A Dynamic Weighting Ensemble Approach for Wind Energy Production Prediction." In *2017 2nd International Conference on System Reliability and Safety*, 296–302. Milano.

Baraldi, P., A. Balestrero, M. Compare, L. Benetrix, A. Despujols, and E. Zio. 2013. "A Modeling Framework for Maintenance Optimization of Electrical Components Based on Fuzzy Logic and Effective Age." *Quality and Reliability Engineering International* 29 (3): 385–405.

Barbounis, T. G., and J. B. Theoharis. 2006. "Locally Recurrent Neural Networks for Long-Term Wind Speed and Power Prediction." *Neurocomputing* 69 (4–6): 466–

96.

Bolón-Canedo, V., N. Sánchez-Marroño, and A. Alonso-Betanzos. 2013. "A Review of Feature Selection Methods on Synthetic Data." *Knowledge and Information Systems* 34 (3): 483–519.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24 (421): 123–40.

Costa, A., A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. 2008. "A Review on the Young History of the Wind Power Short-Term Prediction." *Renewable and Sustainable Energy Reviews* 12 (6): 1725–44.

Gielen, D., F. Boshell, D. Saygin, M.D. Bazilian, N. Wagner, and R. Gorini. 2019. "The Role of Renewable Energy in the Global Energy Transformation." *Energy Strategy Reviews* 24: 38–50.

Graves, A., A. Mohamed, and G. Hinton. 2013. "Speech Recognition with Deep Recurrent Neural Networks." In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6645–49.

International Energy Agency. 2018. "2018 World Energy Outlook: Executive Summary." *Internal Energy Agency (IEA)*. Paris. <https://www.iea.org/reports/world-energy-outlook-2018>.

Jaeger, H. 2001. "Short Term Memory in Echo State Networks." Sankt Augustin.

Jaeger, Herbert. 2001. "The Echo State Approach to Analyzing and Training Recurrent Neural Networks." *Technical Report GMD Report 148*.

Karpathy, A., and L. Fei-Fei. 2017. "Deep Visual-Semantic Alignments for Generating Image Descriptions." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (4): 664–76.

Khosravi, A., S. Nahavandi, D. Creighton, and A.F. Atiya. 2011. "A Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances." *IEEE*

- Transactions on Neural Networks* 22 (9): 1341–56.
- Kramer, N. A., and O. Treiber. 2015. “Evolutionary Feature Weighting for Wind Power Prediction with Nearest Neighbor Regression.” In *2015 IEEE Congress on Evolutionary Computation (CEC)*, 332–37. Sendai: IEEE.
- Li, D., W.Y. Chiu, and H. Sun. 2017. “Chapter 7 - Demand Side Management in Microgrid Control Systems.” In , edited by Magdi S B T - Microgrid Mahmoud, 203–30. Butterworth-Heinemann.
- Li, L.L., X. Zhao, M.L. Tseng, and R.R. Tan. 2020. “Short-Term Wind Power Forecasting Based on Support Vector Machine with Improved Dragonfly Algorithm.” *Journal of Cleaner Production* 242: 118447.
- Li, W., L. Wen, M. Chang, S.N. Lim, and S. Lyu. 2017. “Adaptive RNN Tree for Large-Scale Human Action Recognition.” In *2017 IEEE International Conference on Computer Vision (ICCV)*, 1453–61.
- Lukoševičius, M., and H. Jaeger. 2009. “Reservoir Computing Approaches to Recurrent Neural Network Training.” *Computer Science Review* 3 (3): 127–49.
- Malhotra, P., L. Vig, G. Shroff, and P. Agarwal. 2015. “Long Short Term Memory Networks for Anomaly Detection in Time Series.” In *23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings*, 89–94. Bruges (Belgium): i6doc.com publ.
- Moustapha, A.I., and R.R. Selmic. 2007. “Wireless Sensor Network Modeling Using Modified Recurrent Neural Networks: Application to Fault Detection.” In *2007 IEEE International Conference on Networking, Sensing and Control*, 313–18.
- Peng, Y., H. Wang, J. Wang, D. Liu, and X. Peng. 2012. “A Modified Echo State Network Based Remaining Useful Life Estimation Approach.” In *2012 IEEE Conference on Prognostics and Health Management*, 1–7.
- Polikar, R. 2006. “Ensemble Based Systems in Decision Making.” *Circuits and Systems Magazine, IEEE* 6 (3): 21–45.
- Pontes, F.J., G.F. Amorim, P.P. Balestrassi, A.P. Paiva, and J.R. Ferreira. 2016. “Design of Experiments and Focused Grid Search for Neural Network Parameter Optimization.” *Neurocomputing* 186: 22–34.
- Ramasamy, P., S.S. Chandel, and A.K. Yadav. 2015. “Wind Speed Prediction in the Mountainous Region of India Using an Artificial Neural Network Model.” *Renewable Energy* 80 (March 2014): 338–47.
- Rigamonti, M., P. Baraldi, E. Zio, I. Roychoudhury, K. Goebel, and S. Poll. 2017. “Ensemble of Optimized Echo State Networks for Remaining Useful Life Prediction.” *Neurocomputing* 281: 121–38.
- Sideratos, G., and N.D. Hatzigiorgiou. 2007. “An Advanced Statistical Method for Wind Power Forecasting.” *IEEE Transactions on Power Systems* 22 (1): 258–65.
- Soman, S S, H Zareipour, O Malik, and P Mandal. 2010. “A Review of Wind Power and Wind Speed Forecasting Methods with Different Time Horizons.” In *North American Power Symposium 2010*, 1–8. <https://doi.org/10.1109/NAPS.2010.5619586>.
- Viviescas, C., L. Lima, F.A. Diuana, E. Vasquez, C. Ludovique, G.N. Silva, V. Huback, et al. 2019. “Contribution of Variable Renewable Energy to Increase Energy Security in Latin America: Complementarity and Climate Change Impacts on Wind and Solar Resources.” *Renewable and Sustainable Energy Reviews*.
- World Energy Council. 2016. “World Energy Perspectives Variable Renewables Systems: How to Get It Right.” *World Energy Perspective Renewable Intergration*. London.
- Xue, F., R. Subbu, and P. Bonissone. 2006. “Locally Weighted Fusion of Multiple Predictive Models.” In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2137–43.