



Multi-Horizon Data-Driven Wind Power Forecast: From Nowcast to 2 Days-Ahead

Vazquez Pombo, Daniel; Göçmen, Tuhfe; Das, Kaushik; Sørensen, Poul Ejnar

Published in:

Proceedings of 2021 International Conference on Smart Energy Systems and Technologies

Link to article, DOI:

[10.1109/SEST50973.2021.9543173](https://doi.org/10.1109/SEST50973.2021.9543173)

Publication date:

2021

Document Version

Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):

Vazquez Pombo, D., Göçmen, T., Das, K., & Sørensen, P. E. (2021). Multi-Horizon Data-Driven Wind Power Forecast: From Nowcast to 2 Days-Ahead. In *Proceedings of 2021 International Conference on Smart Energy Systems and Technologies* IEEE. <https://doi.org/10.1109/SEST50973.2021.9543173>


General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Multi-Horizon Data-Driven Wind Power Forecast: From Nowcast to 2 Days-Ahead

Daniel Vázquez Pombo^{1,3}, Tuhfe Göçmen², Kaushik Das², and Poul Sørensen²

¹ Department of Electrical Engineering – Technical University of Denmark (DTU) - Roskilde, Denmark

² Department of Wind Energy – Technical University of Denmark (DTU) - Roskilde, Denmark

³ Research and Development - Vattenfall AB - Stockholm, Sweden

Email: dvapo@elektro.dtu.dk

Abstract—Generation uncertainty is an obvious challenge posed by renewable energy sources such as wind power with effects spawning from stability threats, to economic losses. Data-driven forecasting methods draw increasing attention due to the amount of data available, flexibility and cost-effectiveness among other factors. However, there are concerns regarding effective feature selection and tuning of these models since common naive approaches focus on Pearson or Shapley. This paper uses the development of an active power forecaster for a wind turbine to conduct a thorough sensitivity analysis addressing how different sampling rates, machine learning (ML) methods, features and hyperparameters influence accuracy. Which is computed with the Root-Mean Squared Error and compared against Persistence. The selected ML-methods are Random Forest and Long-Short Term Memory Artificial Neural Networks. The forecasters are multi-horizon & multi-output model targeting 1 minute, 1 hour, 5 hours and 2 days ahead by using sampling rates of 1 second, 1 minute, 5 minutes and 1 hour respectively. The results show which method is more suitable for which horizon and provides insight into which features reduce RMSE of the best performers, whose average is 10, 13, 17 and 25 % for each horizon respectively. The conclusions of the sensitivity analysis can be applied for regions with highly volatile weather, such as coastal areas.

I. INTRODUCTION

Power forecasting plays a major role in renewable energy production. This is due to the importance of knowing the future power availability of a given plant or system, as this information can be used by different actors. For instance, an utility, or plant owner employs predictions to have an estimation of its future production in order to bid in electricity markets or estimate revenue for the next period (month, year, plant lifetime, etc.) [1]. While, System Operators (SOs) use this information to safely steer the power system operation [2]. In addition, modern hybrid power plants that include energy storage systems require them to effectively account for the available power in order to optimize the charging-discharging periods. In the particular case of Power Systems with very large shares of renewable energy, stability is a major concern. Very-short term forecasts are useful to balance generation and load especially under the assumption of limited reserve availability. On the other hand, wind farm operators rely on two signals to estimate their available power. One is provided by the plant's controller as supplied by the equipment manufacturer, thus a black-box. Which provides an instantaneous value, not a forecast. The second, and only actual forecast, is normally purchased to an external provider, that

is a meteorological station placed nearby. These predictions have variable accuracy, result expensive and have limited availability. [2], [3] The lack of a decent forecast causes direct and indirect economic losses for a plant owner. First, due to unnecessary curtailment if the unit under-bet in the market. Second, due to penalties related to not delivering the promised power or energy. Thus hindering the qualification of the plant as an automatic generation control unit. Furthermore, there is potential to reduce the stress of balancing plants if better forecasting tools are available. Thus, reducing cost of maintenance and extending the equipment life-time. [4]

This research compares two different machine learning methods, Random Forest (RF) and Long-Short Term Memory (LSTM) Neuronal Networks; addressing feature selection, and hyperparameter tuning by using wind power forecast as example. The predictors are built in Python using Open-Source tools and data of a single 11 kW wind turbine. The accuracy is benchmarked against the industry's standard evaluation metric, Persistence. The paper is structured as follows: Section II presents the problem statement and the definition of the different application-dependent prediction horizons. While the origin of the used data is presented in Section III. Then, Section IV presents the role of machine learning of time series forecasting, method selection, and evaluation criteria. Subsequently, Section V covers feature sensitivity and selection. Then, Section VI presents selected results. Lastly, Section VII concludes the paper highlighting its major findings.

II. PROBLEM STATEMENT

Traditional renewable energy forecasting methods use meteorological predictions as input data. Plant owners purchase such predictions and combine them with models of their own generation units in order to estimate their production for a given period. Such methods present varying accuracy as they depend on a number of factors including the distance to the closest meteo-station and the technology used. In addition, they disregard the effects of local orography and phenomena, result expensive in general and are not always available. In any case, their accuracy tends to be quite good, specially for mid-long terms and large size plants as local random noise smooths out. On the other hand, data-driven forecasting methods show potential to become the alternative due to their low cost, flexibility, adaptability and easy deployment. Specially taking

into account that any installation that includes one or more WTs counts with at least one meteo-mast that provides measurements of wind speed, air humidity, temperature, etc. This device is able to report measurements every second or even faster depending on the technology. Particularly, ML ensemble methods which have proven their capacity of self-adapting to generator type, diverse geography, height, etc. Effectively minimising the influence of other factors that weakly influence wind power generation. [1], [3–5]

Different applications require forecasts with diverse horizons and sampling rates. For example, nowcast or immediate horizon (1 to few minutes) and short-term (up to 60 minutes) are needed in control and energy management aspects, while mid-term (few hours) is needed to schedule flexibility and long term (1-2 days) allows to predict economical performance. However, most of the available data-driven forecasting tools focus on larger time horizons. In fact, recent publications pinpoint the need for more research focusing on second and minute scale forecasting for mainly control, grid and economic balancing applications [3]. A summary of relevant prediction horizons with desired resolution along with a brief justification behind its selection is presented in Table I. It should be noticed that the terminology used in this paper does not follow the academic standards, as this usually defines "very-short", "short", "medium" and "long-term" as 8 hours, 48 hours, 7 days and 1 year ahead respectively [6]. Such definitions result relevant when estimating broad economical performance over time, but are unsuitable for the level of optimization required in modern power systems.

TABLE I: Summary of forecast targets.

Term	Horizon	Resolution	Reasons
Immediate	60 sec	1 sec	Nowcast for the immediate controller actions, grid stability
Short	60 min	1 min	Scheduling of flexible loads
Medium	5 hours	5 min	Mid-term planning: load shift, battery management
Long	2 days	1 hour	Long-term planning: Blackout prediction, unit commitment

III. THE DATA IN DATA-DRIVEN

The data employed in this research was recorded in SYS-LAB, an intelligent laboratory for distributed energy resources belonging to the Danish Technical University (DTU) located in Risø, Denmark. Among other equipment, it includes Wind Turbines (WT) and Photovoltaic panels (PV) in the tens of kW range, but also a meteorological mast. The system records 1 Hz values and timestamps them according to epoch time, and stores the information as csv files. Measurements corresponding to the year 2019 in full were selected for this study. The data comes from two different sources, an 11 kW WT and a meteo-mast. The measurement system at the WT reports active power production among other parameters, while the meteo-mast records wind speed, wind direction, temperature, etc. As both devices are approximately 200 meters apart, some spatial correlations might not be captured. The WT is non-dispatchable and has a hub height of 20 meters; due to this, it captures

the extremely noisy wind behaviour that is found close to the ground. Thus, adding further complexity to the forecasting task in very short time horizons. In principle, predicting for larger WTs is easier due to their mechanical inertia and the increasing homogeneity of wind flows with high; which smooths out noise. The data was divided in three sets, training, validation and testing as 70, 20 and 10 %. LSTM uses the first two sets during training, while RF only uses the first. Both their performances are evaluated in the last set.

IV. MACHINE LEARNING & TIME SERIES FORECASTING

The high auto-correlation of meteorological parameters makes necessary to install data-acquisition units on site in order to capture the desired patterns [7]. The development, automation, and deployment of purely ML forecasting tools is not only easier, but also more cost-efficient than their statistical counterparts. These, among other reasons, have motivated extensive work on ML forecasting assessing the suitability of ML and statistic-based methods (e.g. ARIMA) for time series forecasting; generally agreeing on the superiority of ML for short term applications [3], [7]. Particularly [8], a recent review on deep learning applications for renewable energy forecasting highlights the unique potential of certain ML methods to discover inherent nonlinear characteristics and high-level invariant structures in data. Research focused on short-time horizon tends to use sampling rates in the minute range, being 10 or 60 minute averages the typical employed values [9], [10]. On a different angle, many studies claim that the higher the data resolution, the better the performance of the forecasting tool [3], but this is not always true as the measurements might contain noise leading the predictions to under-perform [11] or result too computationally expensive as to allow training. Another limiting factor is the insufficient data employed in the studies or its lack of transparency regarding training/validation distribution, which questions the statistical significance of the results [10]. Lastly, many publications address feature selection in a very shallow manner by using correlation as the only criteria for inclusion [12].

Time series forecasting usually focuses on one-step ahead predictions. However, in this work, the focus is on Multi-Output models. That is, predicting an horizon with a certain sample rate using a single ML model. These are relatively complex models that learn the dependence structure between inputs and outputs but also among outputs. They are relatively slow to train, and risk over-fitting if the available data is not enough, while their execution is extremely fast. Their formal definition can be addressed as equation 1, where P, ML, V, Hor, t, and n stand for Prediction, ML model, Feature Observation, prediction horizon, time, and number of samples respectively.

$$P(t+1), P(t+2), P(t+Hor) = ML \begin{bmatrix} V1(t), V1(t-1), \dots, V1(t-n) \\ V2(t), V2(t-1), \dots, V2(t-n) \\ Vn(t), Vn(t-1), \dots, Vn(t-n) \end{bmatrix} \quad (1)$$

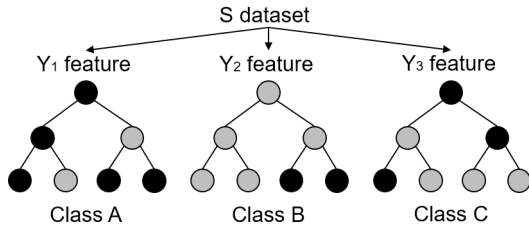


Fig. 1: Simplified Structure of an RF

A. ML algorithm/method/engine selection

Among the available ML algorithms available, RF and LSTM algorithms have gained reputation in modern literature to be suitable for short-term forecasting. Therefore they were selected and compared. The models were developed in Python with the scikit-learn and tensorflow libraries [13], [14].

Random Forest: is an ensemble learning method suitable for regression (although its main application is on classification). The model learns relationships between features and target during training by randomly building "decision-trees". Each tree can be used to make a prediction, but they tend to over-fit the data and are not very good at generalising for new information outside their training set. To overcome this limitation, a number of trees are built forming a decision forest, or RF. Each tree in the forest draws a fixed number of random samples with replacement in order for each tree to use different data of the same size. Thus, increasing diversity leading to a more robust model, as each prediction is made as an average of the predictions of all the individual trees in the forest. Presented in Figure 1, decision trees constitute sets of if-then-else rules aiming to split the data into subsets with homogeneous values. This homogeneity is evaluated in terms of entropy as:

$$E_S = \sum_{i=1}^c -p_i \log_2 p_i \quad \forall i \in [0, 1, 2, \dots, c] \quad (2)$$

Where E_S , p_i , and c stand for entropy of the training data at a given node, probability of class i , and number of classes. Subsequently, we can use Information Gain IG to check whether entropy can be reduced by splitting the data according to different Y subsets of the training dataset S as:

$$IG(Y, S) = E(Y) - E(Y|S) \quad (3)$$

We then choose the subset with largest IG as decision node, creating a new branch. This is an iterative process that continues until the predefined depth has been reached.

Neuronal Networks: Suitable for both classification and regression, they are a group of supervised learning algorithms used to model complex relationships between inputs and outputs or to find patterns in data by weighting inputs according to non-linear approximations. The models are conformed by an input and output layer; with a variable number of hidden layers in between. Particularly, Recurrent Neural Network (RNN) are

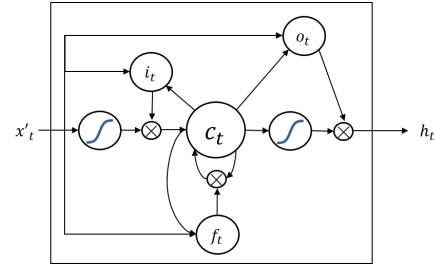


Fig. 2: Neuron Structure of an LSTM, reproduced from [15]

a subset of NN able to establish temporal sequences; making them best suited for time-series forecasting. Lastly, LSTM, are a subset of RNN, whose particularity is to reduce the incidence of the vanishing gradient problem. This causes NN to "get stuck" during training due to very low gradients in the optimisation solution. In LSTM however, each cell vector has the ability to forget part of its previously stored memory adding new information. Figure 2 presents the LSTM neuron structure, where there is no direct connection between input i_t and output o_t . In this way, the information is forced through the memory unit or cell state c_t . This is regulated by a memory preservation coefficient f_t which is defined as in equation 4. Where σ , W_{fi} , and W_{fo} , stand for the sigmoid gate, weights of the forget gate's input and output tensor. While x'_t , h_{t-1} , and b_f represent the current state's input tensor, previous state of the cell's output tensor and the bias vector respectively. The information is then transferred to the input gate i_t and forwarded to the cell state c_t as to save it in the long-term memory as in equation 5.

$$f_t = \sigma(W_{fi} x'_t + W_{fo} h_{t-1} + b_f) \quad (4)$$

$$i_t = \sigma(W_{ii} x'_t + W_{io} h_{t-1} + b_i) \quad (5)$$

where W_{ii} , W_{io} , b_i stand for the weights of the input gate's input and output tensor and the bias vector respectively. Then, the previous cell state c_{t-1} can be updated as equation 6 leading to the neuron's actual output as in equation 7:

$$c_t = \sigma(f_t c_{t-1} + i_t) \quad (6)$$

$$o_t = \sigma(W_{oi} x'_t + W_{oo} h_{t-1} + b_o) \quad (7)$$

where W_{oi} , W_{oo} , b_o stand for the weights of the output gate's input and output tensor and their related bias vector respectively. The definitive output of the cell results:

$$h_t = o_t \tanh(c_t) \quad (8)$$

LSTM algorithms should, in theory, perform better than RF in time series forecasting, since they capture temporal dependencies between observations [16]. A recent review on deep learning applications for renewable energy forecasting concluded that deep recurrent neural networks such as

LSTM outperforms other methods on Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) indices [8]. However, only for large enough datasets, since LSTM must train more parameters. As a rule of thumb, it is assumed that the available samples must be a minimum of 7 times the number of tunable parameters for LSTM to perform as expected. On the other hand, RF can potentially outperform LSTM for smaller datasets or those where random noise is more present than actual tendencies. In the latter case, it might be better to consider each sample as independent since there might not be any relevant time correlation.

B. Monitoring Learning Quality

In the case of RF, R^2 was used as score metric to monitor learning quality. Which is defined as the proportion of the variance in the dependent variable that is predictable from the independent variables [17]. This metric should be equal between training and testing periods as that would imply that the model is able to predict with the same accuracy on both sets. In general, R^2 is slightly better for the training period than for validation. A large difference is a clear sign of over-fitting; since this implies that the model fits very well the training data but fails at making new predictions. On the other hand, LSTM uses an optimization function to improve the training, whose score is evaluated with a certain metric such as MAE or RMSE. This score is monitored during each iteration (epoc) of the optimisation for both training and validation. Then again, these should be compared. The most typical sign of over-fitting is an error increase in the validation dataset with a simultaneous reduction in the training.

C. Performance Evaluation Criteria

In order to evaluate the quality of the predictions, the RMSE is used as defined in equation 9, where n corresponds to the number of samples. This metric allows to understand and evaluate how well the model predicts for a particular period. However, in order to get statistical significance of the results a large number of horizons must be simulated. The calculated values of different models were compared with the performance of the Persistence method; which is the simplest forecasting method, as it consists on selecting a sample at time t_1 and keep it as prediction until t_{1+H} where H is the prediction horizon. Despite its simplicity, this method is actually very difficult to beat for relatively short horizons with high sampling rates. This is due to the fact that introduces the concept of memory in the forecast and, for short horizons, there are usually not extremely large variations in the measured value [18]. Consequently, it became the standard in the industry as any developed ML-based method is tested against it. However, some authors consider that it is so difficult to beat that it might hide potentially useful forecasting methods as it makes the proposed forecasting method compete not only in terms of general accuracy, but also added value [18].

$$RMSE = \sqrt{\frac{\sum_i^n (Observation_i - Prediction_i)^2}{n}} \quad (9)$$

V. FEATURE SELECTION & SENSITIVITY

A number of different features were evaluated in order to select the combination yielding the best possible predictions. Despite most authors using high correlation coefficients as the only selection criteria [12], in this work, we considered others that intuitively have influence even if it is not captured by a correlation coefficient. Then, a number of models were trained covering many of all the combinations. It is complex and time consuming to drag general conclusions about which features provide the best predictions. This is due to the fact that, features' usefulness can't be analysed individually, but as a combined fully trained model. Meaning that the same feature will present different value depending on the group of metrics combined in the model. Briefly, feature X might not hold any importance when combined with Y and Z , but it could be important if combined with W and Z . Implying the need for testing a large number of combinations in order to get a clear picture of the behaviour. On top of that, the methods available to evaluate a feature's importance are different for RF and LSTM; but also the training parameter tuning affects the results along with the number of previous states employed. Most of the available literature recommends using a metric called *feature importance* to evaluate which features add more information to a prediction in RF. However, we found this metric quite misleading, as it's value is related to the internal weighting system of the random forest and does not provide any information regarding the prediction accuracy. Therefore, the quality of predictions was solely evaluated by comparing a given model against persistence in terms of RMSE.

The conclusions of the feature selection sensitivity are drawn based on a thorough iterative process of training and evaluation and post processing for both methods. In general, features identified as important by RF for a particular sampling rate and horizon were still relevant for LSTM. Thus, they can be used as a solid starting point. The complete list of features analysed in this work is presented in Table II along with a summary of the sensitivity analysis' findings. Briefly, despite the large amount of different configurations tested for RF, the feature combination E-F-G was consistently a top performer (although not always the best). In the case of LSTM, that same combination would lead to fairly strong predictors, but not necessarily the best. Nevertheless, this conclusions must be taken cautiously, as the influence of one feature over a model can only be assessed for that a fixed set of features, previous samples, tuning parameters and dataset. Lastly, regarding the number of previous samples to include; RF's best performance was achieved with 3 previous samples for all the horizons, while LSTM uses 30 for all the horizons except for 5 hours in 5 minutes resolution where the best results were achieved using 5 previous samples. A summary of the different tuning configurations employed can be found in Table III. Note that features F and I, are defined as Turbulence Intensity (TI) for either wind speed or active power; which is defined as:

$$TI = \frac{\sigma}{\epsilon} \quad (10)$$

TABLE II: Feature Sensitivity Summary

ID	Definition	Sensitivity
A	Active Power Measurement	As the intrinsic variable, must be included.
B	Mean of the previous samples' active power	It does not add new information, as it is just a recombination of the fed data-set, to the model.
C	Set of artificial weights aiming to increase the importance of recent samples	Only relevant for RF. The idea is to artificially integrate a sense of memory. In practice, it harms accuracy as it alters the way in which the RF algorithm is supposed to weight the variables
D	Time of the day in minutes, that is 0 to 1440	Three different versions were used, number of seconds, minutes and hours, all of them related to one day. That is ranging from 0 to 8640, 1440 and 24. Minutes proved to be the best.
E	Mean of the previous prediction horizon	Improves significantly the prediction as it captures the tendency of the previous horizon
F	Active Power TI as in Equation 10	Improves significantly the prediction as it provides a sense of how noisy was the recent period
G	Wind Speed Measurement	In general, it improves the prediction, but sometimes requires to be together with K, which is easily explained by the distance between the WT and the meteo-mast
H	Mean and Standard Deviation of the previous horizon's wind speed	It has a limited effect and always had to be combined with others such as E, G or K
I	Wind Speed TI as in Equation 10	Only relevant for low sampling rates and in combination with others such as E, G or K
J	Standard Deviation of Previous Horizon	Limited contribution better for low sampling rates than for high ones
K	Wind direction of the previous horizon	Improves the prediction when fed aggregated as 8 different values and combined with others as it helps capture local phenomena such as terrain shadowing and temporal trends

TABLE III: ML Tuning Summary

LSTM	# Layers	1, 2, 3
	# Neurons	25, 50, 75, 100, 125, 150
	Previous Samples	0, 3, 5, 7, 9, 11, 13, 15, 20, 60, 100
	# Epochs	15, 30, 60, 100
	Batch Size	15, 30, 60, 100
R	Loss Function	MAE
	Optimizer	Adam
F	# Trees	50, 75, 100, 125, 150
	Previous Samples	0,1, ..., 10

Where σ and ϵ stand for standard deviation and mean respectively of the previous 10 minutes. TI is, by definition, applied solely to wind speed as in I. However, it also provides valuable information when applied to active power F. The reasoning was not only behind the distance existing between the meteo-mast and the turbine, which affects the correlation between both metrics. But also on the idea that the dynamics of the particular wind turbine (e.g. inertia, vibrations, etc) will be better captured by the active power directly, as it will absorb part of the noise present on the wind speed signal.

VI. SELECTED RESULTS & DISCUSSION

It is infeasible to test every single possible combination of features, number of previous samples, and different tuning parameters for each ML model. Therefore, a subset of the models presenting the best performances, compared with Persistence in terms of RMSE, are presented in Figure 3. Note that, the horizons and sampling rates correspond to those of Table I, letters match features of Table II, Pre indicates the amount of previous samples and the numbers correspond to the neurons per layer in LSTM. Briefly, RF is the best performer for immediate term as it is better able to deal with the extreme noisiness of second sampling. Due to hardware limitations it was unfeasible to train a full RF model with a whole year worth of data in second sampling, so a number of subsets were tested. The RF model presented in Figure 3a corresponds to a training period of just 25 days, against the full year used for LSTM. Not only RF predicts better, but also its training is much faster, as it takes just a few minutes versus the 10-12 days necessary for an LSTM. For the rest of the

forecast targets, LSTM was consistently better than both RF and Persistence. This is explained by its capacity of detecting time-dependent correlations and the noise smothering provided by the different averages. The sensitivity of RF to the inclusion of different number of previous samples and features was quite consistent, while LSTM resulted more difficult to analyse. Briefly, it was necessary to obtain the number of layers, its topology and the features; along with tuning other parameters that control the training. It is clear from Figure 3, that the LSTM complexity is reduced both in number of layers and neurons as less samples are available as it is to be expected. The fact that RF is able to predict better for the highest sampling even despite using less data for the training is quite interesting. This highlights how a forecasting tool should use RF for noisy and highly sampled data, and LSTM once the noise has been reduced by applying averages as to forecast for lower sampling rates. Also, this research also proves that persistence represents a decent back-up forecasting alternative if any problem was to arise with a predictor. Lastly, it must be noted that LSTM should be retrained as more data becomes available; that is to schedule periodic model retraining. Which is easily automated and cheap as requires minimum.

VII. CONCLUSIONS

Given the general trend to transition from dispatchable sources to stochastic, reducing uncertainty is of the utmost importance. Forcing both SO and utilities to use power predictions as to operate properly. Thus, the focus of this paper is to develop a sensitivity analysis of common data-driven wind power forecast for short horizons. After presenting the 1-year long secondly dataset of a single WT and a meteo-mast installed in Risø, Denmark; two ML models suitable for time-series forecasting were presented, RF and LSTM. They were employed in an extensive sensitivity analysis applied to feature selection. Ultimately leading towards the presentation of the best performers obtained for each forecasting target compared in terms of RSME against Persistence. The results clearly show how RF is able to predict better than LSTM for extremely noisy data, where temporal correlations lose importance, even if using 93 % less data. While LSTM consistently outperforms

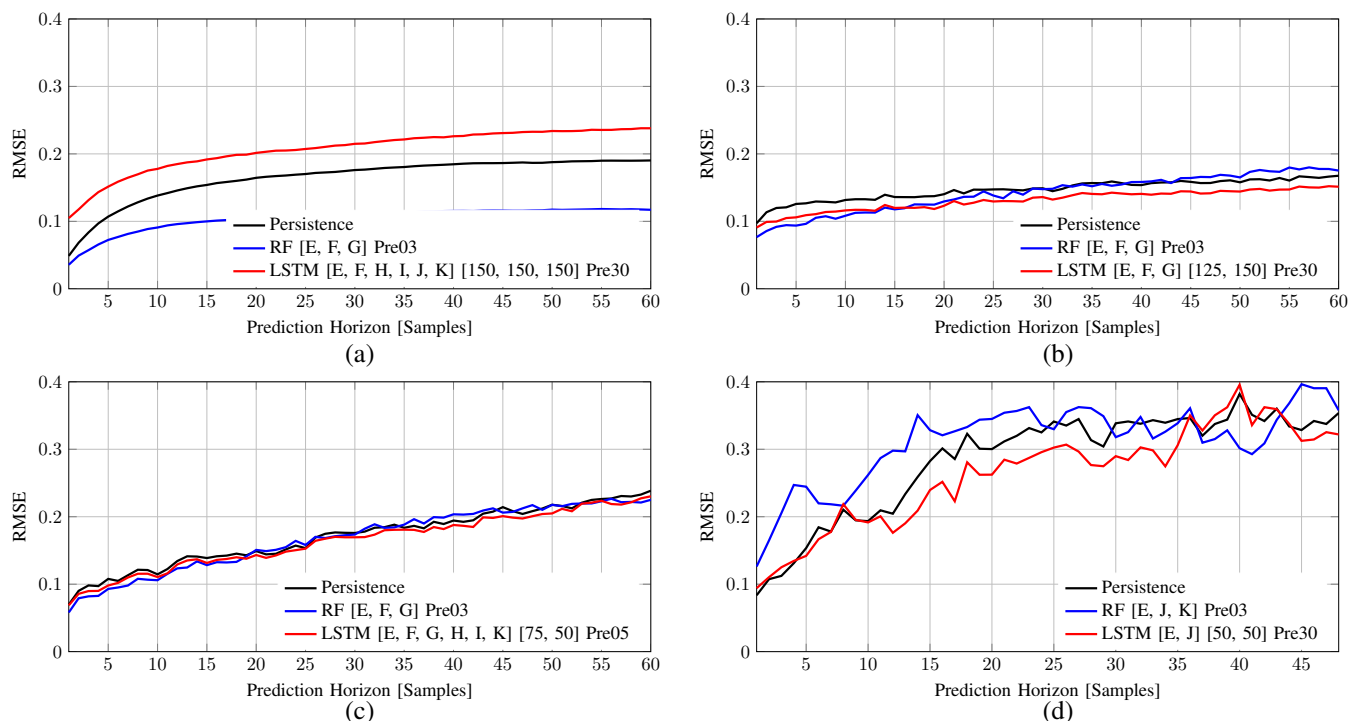


Fig. 3: Results as Sampling - Horizon: a) 1 sec - 60 sec. b) 1 min - 60 min. c) 5 min - 5 hour. d) 1 hour - 2 days.

RF as the noisiness is smoothed out by reducing the sampling. In fact, the best predictions are obtained for the 1 min sampling 60 min horizon, where noise is reduced, the amount of data available is large and the temporal correlations significant enough for LSTM to excel. ML-based forecast represent an inexpensive alternative to purchased weather forecasts or to persistence, as all the employed tools are open-access and easily automated. The provided sensitivity analysis allows to identify useful features for the considered horizons and resolutions in areas with highly volatile weather. Future work, involves a similar analysis for solar forecasting. The obtained results are available to the interested reader at [19].

REFERENCES

- [1] C. Otero-Casal *et al.*, "Development of a high-resolution wind forecast system based on the wrf model and a hybrid kalman-bayesian filter," *Energies*, vol. 12, no. 16, p. 3050, 2019.
- [2] C. Peng *et al.*, "Uncertainty management in power system operation," *Journal of Power and Energy Systems*, vol. 1, no. 1, pp. 28–35, 2015.
- [3] I. Würth, L. Valldecabres, E. Simon, C. Möhrlen, B. Uzunolu, C. Gilbert, G. Giebel, D. Schlipf, and A. Kaifel, "Minute-scale forecasting of wind power—results from the collaborative workshop of iea wind task 32 and 36," *Energies*, vol. 12, no. 4, p. 712, 2019.
- [4] J. Manero *et al.*, "“dust in the wind...”, deep learning application to wind energy time series forecasting," *Energies*, vol. 12, no. 12, p. 2385, 2019.
- [5] Q. Xu *et al.*, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Transactions on sustainable energy*, vol. 6, no. 4, pp. 1283–1291, 2015.
- [6] X. Zhao *et al.*, "Review of evaluation criteria and main methods of wind power forecasting," *Energy Procedia*, vol. 12, pp. 761–769, 2011.
- [7] M. Das and S. K. Ghosh, "Data-driven approaches for meteorological time series prediction: a comparative study of the state-of-the-art computational intelligence techniques," *Pattern Recognition Letters*, vol. 105, pp. 155–164, 2018.
- [8] H. Wang, Z. others Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019.
- [9] M. Khan, T. Liu, and F. Ullah, "A new hybrid approach to forecast wind power for large scale wind turbine data using deep learning with tensorflow framework and principal component analysis," *Energies*, vol. 12, no. 12, p. 2229, 2019.
- [10] Y. Kim and J. Hur, "An ensemble forecasting model of wind power outputs based on improved statistical approaches," *Energies*, vol. 13, no. 5, p. 1071, 2020.
- [11] K. Yeo and I. Melnyk, "Deep learning algorithm for data-driven simulation of noisy dynamical system," *Journal of Computational Physics*, vol. 376, pp. 1212–1231, 2019.
- [12] T. Kaur, S. Kumar, and R. Segal, "Application of artificial neural network for short term wind speed forecasting," in *2016 Biennial international conference on power and energy systems: towards sustainable energy (PESTSE)*. IEEE, 2016, pp. 1–5.
- [13] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [14] M. Abadi, A. Agarwal, P. Barham *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [15] T. Göçmen, A. Meseguer Urbán, J. Liew, and A. W. H. Lio, "Model-free estimation of available power using deep learning," *Wind Energy Science*, vol. 6, no. 1, pp. 111–129, 2021.
- [16] E. Mussumeci and F. C. Coelho, "Large-scale multivariate forecasting models for dengue-1stm versus random forest regression," *Spatial and Spatio-temporal Epidemiology*, vol. 35, p. 100372, 2020.
- [17] D. Zhang, "A coefficient of determination for generalized linear models," *The American Statistician*, vol. 71, no. 4, pp. 310–316, 2017.
- [18] M. P. Mittermaier, "The potential impact of using persistence as a reference forecast on perceived forecast skill," *Weather and forecasting*, vol. 23, no. 5, pp. 1022–1031, 2008.
- [19] D. V. Pombo, "Wind power forecaster: data, model, etc." Dec 2020. [Online]. Available: <https://data.dtu.dk/account/articles/13286336>