Forecasting of Day-Ahead Wind Speed/electric Power by Using a Hybrid Machine Learning Algorithm

Atilla Altıntaş[1(✉)], Lars Davidson[1], and Ola Carlson[2]

[1] Division of Fluid Dynamics, Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 41296 Gothenburg, Sweden
altintas@chalmers.se

[2] Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

Abstract. The amount of energy that has to be delivered for the following day is currently predicted by power system operators using day-ahead load forecasts. With the use of this forecast, generation resources can be committed a day in advance, some of them may require several hours' notice to be ready to produce power the following day. In order to determine how much wind power will be available for each hour of the following day, power systems with large penetrations of wind generation rely on day-ahead predictions. The main objective of this study is to improve the day-ahead forecasting of wind power by improving the forecasting method using machine learning. A hybrid approach, which combines a mode decomposition method, Empirical Mode Decomposition (EMD), with Support Vector Regression (SVR), is used. The results suggest that using Support Vector Regression together with the hybrid method, which includes the Empirical Mode Decomposition to predictions can improve the accuracy of predictions. Higher accuracy forecasting of wind power is expected to improve the planning of dispatchable energy generation and pricing for the day-ahead power market.

Keywords: Wind energy · wind turbine · Empirical Mode Decomposition (EMD) · forecasting · machine learning · renewable energy · grid integration · energy market

1 Introduction

Among available renewable energy sources, wind energy has the largest potential [5,17]. Wind turbines are connected to the medium voltage distribution grid or the regional transmission grid. In order to transport the energy with low losses, the electric grid is usually designed as a high voltage transmission grid that is

This work is supported by the Chalmers University, Energy Area of Advance.

J. L. Afonso et al. (Eds.): SESC 2022, LNICST 502, pp. 3–11, 2023.
https://doi.org/10.1007/978-3-031-33979-0_1

connected to centralized production units. Since high voltages are impractical and dangerous to use, voltage is transformed to lower levels and distributed to the end users via a distribution grid. The main challenge when operating the power system is to keep the system in balance, i.e. to keep the balance in supply and demand. Integration of the wind power into the electric grid is problematic since the varying nature of the wind speed gives fluctuations in the produced wind power [17].

Sweden has the clear goal of 100% renewable electricity production by 2040. The Swedish Energy Agency's assessment shows that wind turbines will produce 60–90 TWh. Many new wind turbines need to be built to achieve that goal and the total supply needs for electric power production is estimated to be about 180 TWh. Although, recent studies show a much larger need of electric power in the future. Currently, 61% of Swedish electricity generation comes from hydro- and wind power. The Swedish Energy Agency estimates that in order to achieve this goal, the nation will need to install an extra 2.5 to 6 TWh of renewable energy capacity per year between 2030 and 2040 [7,10]. The growth of wind power that is connected to the electric grid requires wind farms appear more like conventional power plants and hence it will be necessary to forecast the produced power.

Recently, efforts to improve forecasting methodologies have also included the use of EMD in many areas from wind energy to financial time series ([3,6,9,11,13,15]). EMD is a method which decomposes a complex time series of data into its frequency components i.e so called intrinsinc mode functions (IMFs) ([3,8,9]). EMD divides data into its IMFs, which represent a number of high to low frequency components. The high frequency component corresponds to short-term changes, and low frequency component corresponds to long-term changes. By using different combination of the frequency components of the data we can predict both short and long-term predictions much more accurately compared to using the entire data set. The general idea behind the use of EMD for forecasting purposes is to separate the data into its components which reduces the complexity, separates the trends of different time scales. In this way the accuracy of the forecasting is improved.

In order to determine how much wind power will be available for each hour of the following day, power systems with large penetrations of wind generation rely on day-ahead wind predictions (Piwko and Jordan [12] and Rintamäki et al. [16]). In this study we improve the accuracy of the day-ahead wind speed forecasting by using a hybrid EMD-SVR method. The measured power is used as input data and selected IMFs of the power are used as influence parameters for the SVR regression model.

2 Theory and Method

2.1 Scale Decomposition by Empirical Mode Decomposition

EMD is predicated on the idea that any data signal is made up of a number of basic intrinsic oscillations, with the raw signal being a superposition of these oscillations (refer to Ref. [3] for further analysis). Each mode is referred to as an IMF that satisfies the two conditions; the local extrema and zero-crossing numbers must be equal or differ by one at the most and the mean of the curve that is constructed by connecting the maxima and minima should be zero [8].

2.2 SVR Method

The Support Vector Regression (SVR) is an algorithm for machine learning, which is a variant of Support Vector Machine (SVM) (Qiu et al. [14], Altıntaş et al. [4]). Consider a time-series data,

$$D = (X_i, y_i), 1 \leq i \leq N,$$

where X_i represents the ith element and y_i corresponds the target output data. The SVR function, f, is a linear function that relates the input and output data as $f(X_i) = \omega^T \phi(X_i) + b$, where ω, b, and $\phi(X_i)$ are the weight vector, bias, and the function that maps the input vector X into a higher dimensional feature space, respectively.

Python programming language and scikit-learn 1.1.2 package has been used for SVR. The radial basis function (RBF) is chosen as the kernel function, then Kernel function written as:

$$K(X_i, X_j) = exp(-\gamma X_i - {X_j}^2), \tag{1}$$

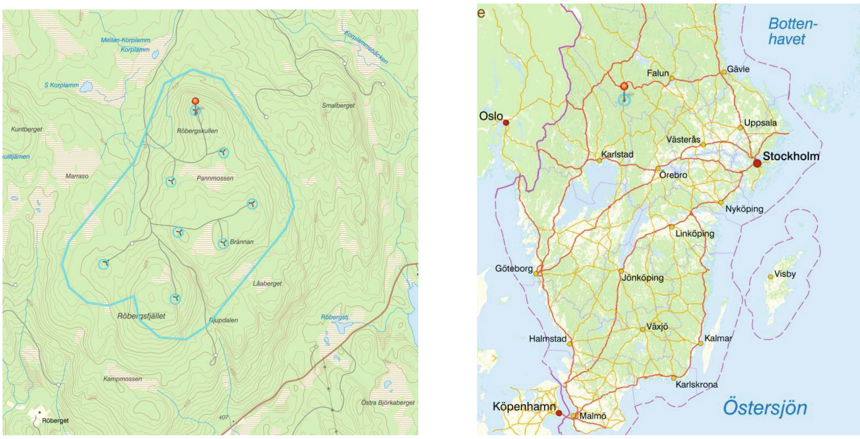
where the parameter γ, defines the degree to which the effect of a single example of training reaches. In this study parameters are set to, $\gamma = 0.96$, $C = 1.0$, which balances the trade-off between the complexity of the model and its generalization ability, and the maximum error, ϵ, is set to 0.03, and are used for all the predictions.

2.3 Wind Power Data

The data are from the Röbergsfjället wind farm which is situated at Röbergskullen in the southernmost section of the Swedish municipality of Vansbro (60160 49.8"N, 14120 59.6"E) (see Fig. 1). The wind farm was constructed in 2007, with its highest point being at 543 m above sea level. There are 284 m between the wind farm's highest and lowest elevations. It consists of eight Vestas V90-2MW horizontal axis wind turbines [2]. The wind turbine that has been used for this study is highlighted with the red pin in Fig. 1(a) and also the area is highlighted in a larger map in Fig. 1(b).

The data consist of a list of records including power, hub direction, pitch angle, rotor RPM, temperature, wind direction and wind speed for the period of 21 June 2017 to 3 February 2019. The data are recorded every second.

Wind turbines measure the wind speed with an anemometer which is installed at a specific location on the nacelle. This anemometer is installed behind the blades thus exposed to created turbulence by the rotor blades. Therefore we



(a) Röbergsfjället wind farm. The data is from the wind turbine pointed with the red pin.

(b) Location of the windfarm Röbergsfjället.

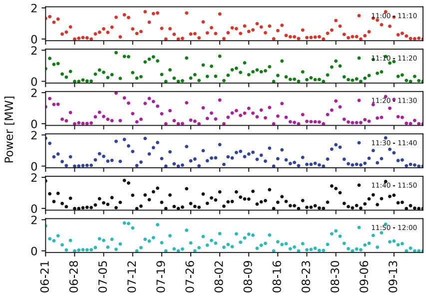
Fig. 1. Wind farm and turbine location [1].

can not trust the wind speed measured in the downstream wake area, and wind direction is also not trusted for the same reason. Moreover, it is a pointwise measurement, however, the wind speed field that creates power is the rotor plane area which is far from homogeneous. For these reasons we can not use wind speed from the anemometer. In this study, output power history data has been used.

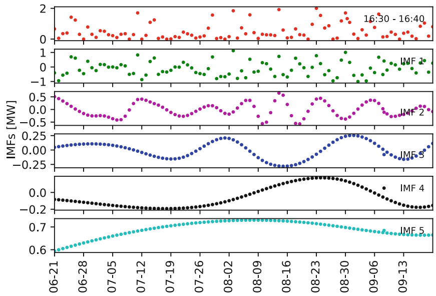
The three months of data has been used from 21 June 2017 to 20 August 2017. Thus a seasonal wind behaviour has been tried to be captured. For the same reason, the data between 11:00 to 17:00 has been used. The power data of the turbine are averaged over 10 min of time windows. For instance, the window 11:00–11:10 represents the data that has been averaged over 10 min in the given interval. There are missing records, meaning that in the given second the turbine has generated no power, which is excluded.

The predictions are made for every 10 min averaged time-window between 11:00–17:00, therefore thirty-five time-windows are used for forecasting. The data have been split into a training and a test part. The last day of the data which is 20 August 2017 is the test part, i.e the part to predict using the training data. The previous days' 10 min time-window has been used as the feature to forecast the next day's 10 min time-window, i.e., time-window 11:00–11:10 for the training days enters the process to forecast the test day's time-window, 11:10–11:20. The part of the data (for 11:00 to 12:00 for 92 days) which is split into ten minutes time windows are given in Fig. 2(a)).

In the EMD-SVR hybrid method, EMD used as a preprocessor to SVR. EMD splits data into IMFs and each IMF is a feature (input) for SVR. IMFs are frequency modes that are obtained by applying EMD to the original data



(a) A part (11:00 to 12:00) of the data.



(b) The upper signal is the raw data, and the subsequent five signals are the IMFs obtained by applying EMD to the raw data.

Fig. 2. Raw data and IMFs.

(raw data). The sum of all IMFs is equal to the original data. In Fig. 2(b), the original data of average power 11:00–11:10 and its IMFs' obtained by EMD are given. The data set has been split into its IMFs by limiting the number of IMFs to five, the fifth IMF is including the residual. Each five IMF have been an input for SVR. The data are scaled by Min-Max scaling method to an interval of $[0, 1]$ before the SVR process. The combinations of the outputs are the predictions. That process is repeated for all thirty-five time-window predictions in the EMD-SVR hybrid method. A process of EMD combined with SVR is given in Fig. 3.

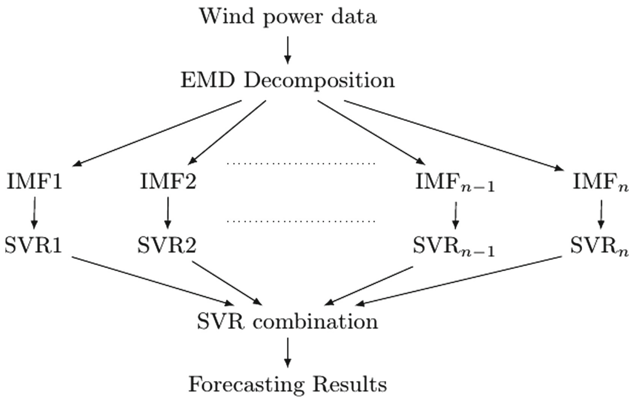


Fig. 3. Process schema of EMD-SVR method.

3 Results

The predictions for the test data which is the last day of 92 days of data are obtained for both the SVR and the EMD-SVR method. We would like to clarify that all the parameters in both SVR and EMD are kept the same for all predictions. SVR is performed by using the original data of the measured power as the feature. In EMD-SVR, EMD is used as a preprocessor to SVR that splits original data into its IMFs. In EMD-SVR, each IMF is a feature for SVR instead of the original data.

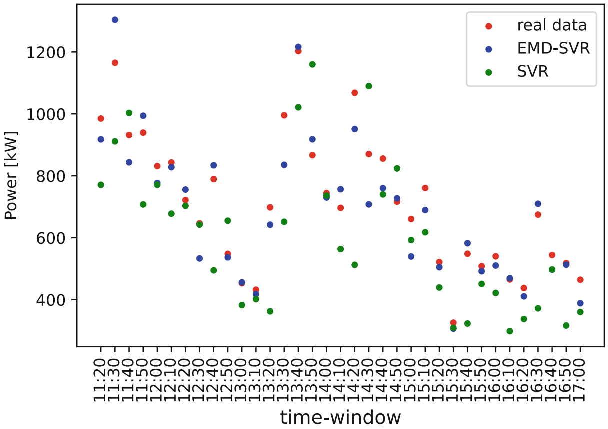


Fig. 4. Power predictions for the day 20 August 2022.

A total of thirty-five 10-min averaged time-window predictions for the hours, 11:00–17:00, for the day 20 August 2017 are given in Fig. 4. Normalized root mean square error (nRMSE) are given in Table 1, where the best approximation is given in a separate column and also highlighted in red.

Only for the time windows, 11:30–11:40, 12:10–12:20, 12:20–12:30 and 13:50–14:00, does raw data predict better than the IMF or IMF combinations. That means that approximately 90% of the cases IMF and combinations provide better prediction than using the raw data as feature (see Fig. 4 and Table 1). The combination of IMF 1 + IMF 2 agrees better with real data than all the other IMF combinations and original data for the total of nine cases, that is approximately 25% of the total cases. The lowest error obtained in the predictions is 0.4% and the maximum error is 18%.

Table 1. The power prediction errors for the time between 11:00–17:00, for the day 20 August 2017, for 10 min averaged time-windows (see Fig. 4). nRMSE = normalized root mean square error.

	nRMSE												
	Original data	**IMF 1**	**IMF 2**	**IMF 3**	**IMF 4**	**Residual**	**Total IMFs**	**IMFs 1+2**	**IMFs 2+3**	**IMFs 3+4**	**IMFs 1+2+3**	**IMFs 1+2+3+4**	**IMFs 4+Residual**	**Best Approximation**
11:10-11:20	0.214	0.372	0.692	0.682	0.734	0.712	0.797	0.066	0.376	0.418	0.246	0.512	0.449	IMFs 1+2
11:20-11:30	0.215	0.561	0.670	0.615	1.02	0.634	0.480	0.234	0.289	0.644	0.145	0.118	0.663	IMFs 1+2+3+4
11:30-11:40	0.074	0.093	0.616	0.604	0.675	0.749	1.250	0.286	0.222	0.280	0.681	1.0	0.423	Original data
11:40-11:50	0.242	0.439	0.724	0.776	0.798	0.516	0.738	0.163	0.502	0.574	0.056	0.257	0.317	IMFs 1+2+3
11:50-12:00	0.070	0.367	0.693	0.643	0.914	0.601	0.773	0.063	0.337	0.557	0.290	0.374	0.515	IMFs 1+2
12:00-12:10	0.193	0.224	0.790	0.582	0.674	0.702	1.01	0.017	0.375	0.259	0.397	0.722	0.380	IMFs 1+2
12:10-12:20	0.023	0.150	0.668	0.504	0.445	0.359	1.86	0.176	0.174	0.044	0.668	1.22	0.192	Original data
12:20-12:30	0.005	0.385	0.782	0.433	0.630	0.570	1.18	0.170	0.218	0.062	0.391	0.761	0.203	Original data
12:30-12:40	0.369	0.843	0.888	0.512	0.697	0.562	0.489	0.731	0.401	0.209	0.243	0.045	0.262	IMFs 1+2+3+4
12:40-12:50	0.165	0.345	0.447	0.494	1.18	0.019	1.50	0.202	0.053	0.676	0.706	0.524	0.202	Residual
12:50-13:00	0.152	0.372	0.309	0.004	0.855	0.372	2.09	0.313	0.694	0.148	1.31	1.46	0.224	IMF 3
13:00-13:10	0.067	0.523	0.505	0.397	0.672	0.361	1.53	0.031	0.090	0.072	0.569	0.894	0.032	IMFs 1+2
13:10-13:20	0.475	0.401	0.734	0.587	0.491	0.300	1.48	0.136	0.322	0.079	0.273	0.780	0.207	IMFs 3+4
13:20-13:30	0.343	0.973	0.719	0.751	0.709	0.547	0.291	0.694	0.473	0.464	0.448	0.158	0.259	IMFs 1+2+3+4
13:30-13:40	0.149	0.639	0.566	0.779	0.803	0.612	0.592	0.207	0.347	0.583	0.009	0.206	0.417	IMFs 1+2+3
13:40-13:50	0.334	0.293	0.913	0.731	0.671	0.931	0.453	0.207	0.644	0.403	0.056	0.384	0.603	IMFs 1+2+3
13:50-14:00	0.008	0.018	0.879	0.542	0.497	0.622	1.43	0.098	0.422	0.040	0.556	1.05	0.120	Original data
14:00-14:10	0.168	0.084	0.602	0.933	0.815	0.605	1.12	0.479	0.538	0.751	0.546	0.728	0.423	IMF 1
14:10-14:20	0.515	0.630	0.925	0.899	0.873	0.774	0.108	0.557	0.826	0.774	0.459	0.334	0.647	Total IMFs
14:20-14:30	0.249	0.182	0.624	0.576	0.839	0.585	1.18	0.187	0.203	0.418	0.610	0.768	0.427	IMF 1
14:30-14:40	0.132	0.556	0.553	0.602	0.574	0.751	0.956	0.109	0.158	0.179	0.283	0.709	0.325	IMFs 1+2
14:40-14:50	0.147	0.477	0.504	0.629	0.480	0.557	1.34	0.013	0.135	0.110	0.382	0.901	0.036	IMFs 1+2
14:50-15:00	0.098	0.308	0.377	0.915	0.595	0.580	1.21	0.311	0.293	0.511	0.395	0.796	0.179	IMFs 4+residual
15:00-15:10	0.172	0.436	0.653	0.624	0.863	0.594	0.818	0.091	0.278	0.488	0.281	0.416	0.458	IMFs 1+2
15:10-15:20	0.151	0.236	0.695	0.451	0.573	0.277	2.22	0.540	0.151	0.029	1.083	1.509	0.140	IMFs 3+4
15:20-15:30	0.046	0.167	0.167	0.052	0.133	0.069	4.14	0.994	0.763	1.07	1.93	3.07	1.20	IMF 3
15:30-15:40	0.404	0.566	0.368	0.072	0.577	0.108	2.52	0.061	0.552	0.342	0.985	1.40	0.537	IMFs 1+2
15:40-15:50	0.109	0.195	0.452	0.577	0.265	0.323	2.17	0.347	0.031	0.152	0.768	1.50	0.405	IMFs 2+3
15:50-16:00	0.215	0.365	0.548	0.511	0.537	0.230	1.79	0.080	0.062	0.051	0.566	1.02	0.226	IMFs 3+4
16:00-16:10	0.349	0.400	0.332	0.273	0.711	0.244	2.51	0.261	0.383	0.008	0.981	1.26	0.530	IMFs 3+4
16:10-16:20	0.223	0.273	0.152	0.340	0.577	0.479	2.16	0.568	0.501	0.080	1.22	1.64	0.058	IMFs 4+residual
16:20-16:30	0.443	0.540	0.781	0.417	0.528	0.640	1.08	0.323	0.199	0.049	0.258	0.725	0.170	IMFs 3+4
16:30-16:40	0.083	0.181	0.359	0.254	0.199	0.083	3.27	0.820	0.377	0.537	1.56	2.36	0.711	Residual
16:40-16:50	0.380	0.327	0.426	0.007	0.533	0.152	2.85	0.239	0.559	0.453	1.23	1.69	0.616	IMF 3
16:50-17:00	0.220	0.714	0.444	0.336	0.539	0.602	3.23	0.157	0.888	0.793	1.17	1.63	1.05	IMFs 1+2

4 Conclusion

In this study, an EMD-based decoupling procedure is applied as a preprocessor to SVR to improve the day-ahead wind power forecasting. First, IMFs are obtained by applying EMD to the original data, each IMF is used as a feature for SVR instead of the original data. The prediction results are compared for combinations of the IMFs and the original data. The data set has been split into 10 min time windows and a previous days' 10 min averaged time-windows has been used as a feature in the forecasting. All SVR parameters are kept the same for all predictions. As a result, for thirty-one out of thirty-five 10 min time-windows, IMF or IMF combinations approximate the real data better than using raw data in the prediction process. With the method we applied we approximate the next day's 10 min averaged power production with a maximum of 18% of error. Twenty-nine out of thirty-five time windows have been predicted with an error of less than 10%, and six of those are predicted with an error of less than 1%. With the results obtained in this study, we suggest that the EMD-based signal decomposition could be beneficial in wind power/speed forecasting by increasing accuracy.

References

1. Vindbrukskollen. https://vbk.lansstyrelsen.se/en. Accessed 19 July 2022
2. Abedi, H., Sarkar, S., Johansson, H.: Numerical modelling of neutral atmospheric boundary layer flow through heterogeneous forest canopies in complex terrain (a case study of a Swedish wind farm). Renewable Energy **180**, 806–828 (2021)
3. Altıntaş, A., Davidson, L., Peng, S.: A new approximation to modulation-effect analysis based on empirical mode decomposition. Phys. Fluids **31**(2), 025117 (2019)
4. Altıntaş, A., Davidson, L., Kostaras, G., Isaac, M.: The day-ahead forecasting of the passenger occupancy in public transportation by using machine learning. In: Martins, A.L., Ferreira, J.C., Kocian, A. (eds.) Intelligent Transport Systems. INTSYS 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 426, pp. 3–12. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-97603-3_1
5. Archer, C.L., Jacobson, M.Z.: Evaluation of global wind power. J. Geophys. Res.: Atmospheres **110**(D12), D12110 (2005)
6. Hong, Y.Y., Yu, T.H., Liu, C.Y.: Hour-ahead wind speed and power forecasting using empirical mode decomposition. Energies **6**(12), 6137–6152 (2013)
7. Hu, X., Jaraitė, J., Kažukauskas, A.: The effects of wind power on electricity markets: A case study of the Swedish intraday market. Energy Econ. **96**, 105159 (2021)
8. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454**(1971), pp. 903–995 (1998)
9. Liu, K., Zhang, Y., Qin, L.: Erratum:"A novel combined forecasting model for short-term wind power prediction based on ensemble empirical mode decomposition and optimal virtual prediction". J. Renew. Sustain. Energy **8**, 013104 (2016)
10. Mauritzen, J.: Now or later? trading wind power closer to real time and how poorly designed subsidies lead to higher balancing costs. Energy J. **36**(4), 149–164 (2015)

11. Nava, N., Di Matteo, T., Aste, T.: Financial time series forecasting using empirical mode decomposition and support vector regression. Risks **6**(1), 7 (2018)
12. Piwko, R., Jordan, G.: Impacts of improved day-ahead wind forecasts on power grid operations: September 2011. Tech. rep., National Renewable Energy Lab. (NREL), Golden, CO (United States) (2011)
13. Premanode, B., Vongprasert, J., Toumazou, C.: Noise reduction for nonlinear non-stationary time series data using averaging intrinsic mode function. Algorithms **6**(3), 407–429 (2013)
14. Qiu, X., Suganthan, P.N., Amaratunga, G.A.: Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines. Procedia Comput. Sci. **108**, 1308–1317 (2017)
15. Ren, Y., Suganthan, P., Srikanth, N.: Ensemble methods for wind and solar power forecasting-a state-of-the-art review. Renew. Sustain. Energy Rev. **50**, 82–91 (2015)
16. Rintamäki, T., Siddiqui, A.S., Salo, A.: Strategic offering of a flexible producer in day-ahead and intraday power markets. Eur. J. Oper. Res. **284**(3), 1136–1153 (2020)
17. Steen, D., et al.: Challenges of integrating solar and wind into the electricity grid. In: Systems Perspectives on Renewable Power, pp. 94–107 (2014)