

Downscaling and improving the wind forecasts from NWP for wind energy applications using support vector regression

Ghali Yakoub, Sathyajith Mathew and Joao Leal

University of Agder, Jon Lilletunsvei 9, 4879 Grimstad, Norway

ghali.yakoub@uia.no, sathyajith.mathew@uia.no, joao.leal@uia.no

Abstract. The stochastic nature of wind poses challenges in the large scale integration of wind energy with the grid. Wind characteristics at a site may significantly vary with time, which will be reflected on the wind power production. Understanding and managing such variations could be challenging for wind farm owners, energy traders and grid operators. In this work, we propose models based on support vector regression (SVR) to downscale the speed and direction of wind at a specific site using both historical observed measurements and numerical weather predictions (NWP). Several meteorological variables, considered to have potential influence on the wind, were used in the feature selection for the models. The models are then optimally developed and used to predict the wind speed and direction at the site considered. In view of the two of Nord pool's energy markets namely the intraday and day ahead markets, approaches for short-term forecasts ($t + 1$ hours) and medium-term recursive forecasts ($t + 36$ hours) were developed. The proposed SVR models are found to be accurate and efficient in correcting the NWP information and predicting the wind speed and direction for the short-term forecasts. For medium-term forecasts, the developed models could outperform the NWP, especially for the wind speed predictions.

1. Introduction

Wind is considered as one of the important renewable energy resources, which is inexhaustible, clean and economically competitive. As a result, wind energy has gained significant focus around the world today. For example, with the addition of 51.3 GW in 2018, the total installed wind capacity has reached 591 GW. Global Wind Energy Council's (GWEC) forecasts there would be new installations around 55 GW, each year until 2023. Therefore, wind is one of the fastest growing energy sources for electricity generation, which will play a significant role in the future of clean energy [1]. This development introduces a major challenge in large scale integration of wind generated electricity into the grids due to the stochastic nature of wind. Speed and direction of wind at a given site may significantly vary even within short intervals of time, these fluctuations in the velocity will be reflected in the wind power production in a magnified scale. Understanding and managing such a fluctuating resource is a real challenge for wind farm owners, energy traders and grid operators.

In many countries, wind power producers are required by law to participate in electricity markets in the same way as conventional power generators. A good example is the Nordic power market (Nord Pool), which includes the day-ahead market 'Elsport' and the continuous intraday market 'Elbas'. In 'Elsport', the producers and consumers submit bids and offers covering every hour of the following day, which is cleared before 12 noon. Whereas in 'Elbas', the trade is settled individually between two parties which closes one hour prior to the delivery hour. The balancing market regulates production or

consumption up or down depending to keep the instantaneous balance of the grid [2]. In these markets, the power producers are accountable for the cost of real-time balancing of their deviations from their forward-contracted volumes [3]. Hence, to avoid the financial losses, it is essential to estimate the day ahead and hour ahead productions from the wind farms with an acceptable level of accuracy. Hour ahead power predictions are also important for formulating the dispatch schedules for the grid operation. Hence, reliable short to medium term wind power forecasting has become increasingly important in the efficient management of wind energy farms.

Wind forecasting approaches can be broadly based on physical, statistical and/or artificial intelligence (AI) approaches. Physical methods such as NWP and mesoscale models use physical considerations like the terrain obstacles, air pressure, humidity and temperature to forecast the weather. For instance, NWP models solve numerically the Navier–Stokes equations and mass continuity equation, coupled with the first law of thermodynamics and the ideal gas law, representing the full set of prognostic equations upon which the change in space and time of wind velocity, pressure, density and temperature are described in the atmosphere [4]. Statistical methods, such as autoregressive integrated moving average models (ARIMA), make forecasts by establishing the relationship of the observed wind speed time series, in other words forecasting the wind speed based on historical observed data [5]. Methods based on AI, like artificial neural network (ANN) and support vector regression (SVR), have recently attracted wider attention among the forecasting community [6-8]. Hybrid methods implementing two or more from the previous methods together are also getting popular due to lower errors expected [9].

Physical methods usually provide the satisfactory forecast precision by combining multiple physical considerations and have advantages in long-term predictions. While the statistical methods perform well in short-term prediction, even though this classification is not absolute. Both physical and statistical models are mostly used together, where NWP results are usually regarded as input variables, together with historical data, to train the system on the local conditions according to statistical theories. Thus, the physical method serves as an initial analysis, and its results make the statistical method more efficient [10]. In addition, NWP models have a fundamental advantage over many other scientific methods in which its performance is objectively evaluated daily and globally, so that success and failure of forecasts is accurately known and pathways to improve predictive skill can be effectively tested.

NWP models are operational at many national meteorological agencies such as in Europe (ECMWF), America, and Japan (JMA). These models have been often upgraded due to increasing computer power, as a result, most Global NWP models nowadays have spatial resolution in the range of 10-30 km, while regional NWP models are ranging from 1-5 km. The models might be different in terms of physical parameterization, forecast range, forecast issue routine, and spatiotemporal resolution, as detailed in [11]. In best scenarios, the spatial resolution of NWP models is between 1 and 5 km, which is not accurate enough in case of wind power forecasting considering the complex terrains and challenging site environments. Therefore, an efficient approach to downscale the NWP's parameters to a specific location is highly required.

The use of AI and machine learning (ML) techniques on wind energy, in particular on wind forecasting, are becoming more popular due to its effectiveness, high performance and computational speed. *Shirkhani et al.* [12] developed one simple linear regression and two non-linear regression models (polynomial regression) to link low resolution NWP mean hourly wind velocity to station level wind velocity. The authors stated that linear model outperformed the non-linear ones for that specific station. Similarly, *Watters and Leahy* [13] compared three different statistical downscaling methods particularly, linear, Kalman filter and neural network methods to downscale mean hourly wind speeds and wind directions obtained from Global Environmental Multiscale model. The results, in general, showed that the three methods showed similar performance. Nevertheless, Kalman filter performed better on average and required less amount of data. In the same perspective, *Dupre et al.* [14] developed two linear regression models to forecast the wind speed and wind energy at a sub-hourly time scale, where one model uses only variables from the NWP model and the other uses both local wind speed measurements and variables from NWP. The authors compared the results of the two models with the

persistence method and two benchmarks methods namely, ANN and ARMA. The results from all methods were then applied to a wind farm to predict the power output, showing that the linear regression models outperformed the persistence method and the two benchmark methods. Similarly, *Sfetsos* [15] compared the performance of traditional statistical models, autoregressive moving average (ARMA) and neural networks models using mean hourly wind speed data and found that the later outperformed the other linear and nonlinear statistical models.

Okumus and Dinler [16] extensively reviewed the recent advances in statistical wind forecasting, highlighting that one of the present-day challenges is the drop in the accuracy in wind prediction for forecast horizons above 6 hours. They concluded that this challenge can be overcome by using a combination of several prediction methods. Pre and post processing of data for the specific target site is also important in improving the prediction accuracy.

In this paper, we propose models based on Support Vector Regression (SVR) to downscale and improve the site-specific wind forecasts (both wind speed and direction) from a regional NWP model. By downscaling, the general systematic errors in the NWP forecasts for the site are reduced and the accuracy of the wind forecasts are further enhanced by using previous observations. The goal is set to tackle the day ahead market ($t + 36$ hours) and intraday market ($t + 1$ hours).

2. Study case

To demonstrate the proposed SVR based downscaling method, the geospatial location of a fully instrumented 5 kW experimental wind turbine is considered in this study. The wind turbine is located at Smøla island, on the west coast of Norway (see figure 1), within a wind farm composed of 67 turbines with installed capacity of 148.4 MW. The uniqueness of its location introduces an interesting challenge for the proposed method.

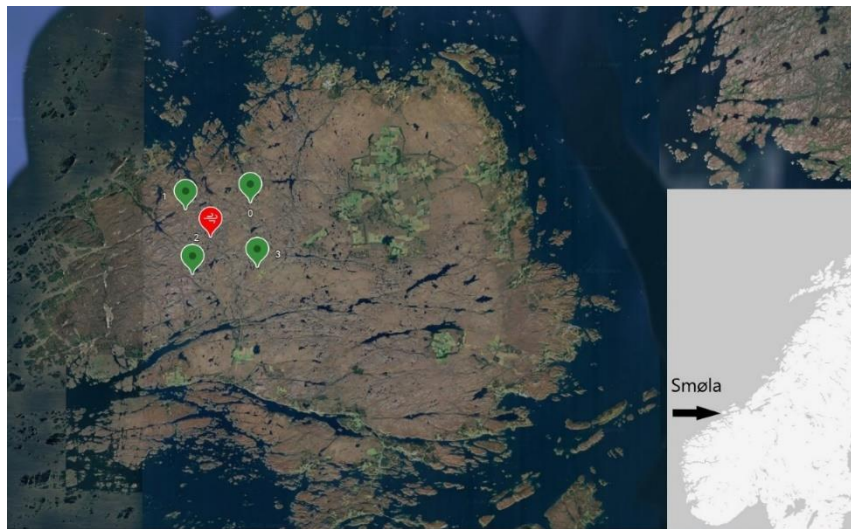


Figure 1. Wind turbine location and the four nearest NWP grid points *source Google earth.*

One-year of data ‘2019’, from two sources, were retrieved for this study. The first set of data were from a met mast installed next the experimental wind turbine at 20 m above ground level (AGL). These observed data were recorded at an interval of 5 min and consist of the wind speed and direction.

The second data set were extracted from the archived historical hourly raw weather forecast at a height of 10 m AGL from the regional NWP model METCoOp Ensemble Prediction System (MEPS). This model is a convection-permitting atmosphere ensemble model covering Scandinavia and the Nordic sea with a horizontal resolution of 2.5 km, 65 vertical levels and 10 members [17]. Several weather parameters like wind speed, wind direction, air temperature, relative humidity and air pressure were extracted from that model and considered in this work. For better understanding of the weather

conditions at the point of interest, weather forecast data were extracted from the four grid cells nearest to the wind mast location as shown in figure 1. The green icons show the locations of the centers of the four NWP grid cells and the red icon shows the location of the target wind turbine at Smøla island.

3. Methodology

3.1. Downscaling models

The development of a robust and efficient machine learning model must go through several steps. A general overview of model development process used in this study is shown in figure 2. The process has mainly three phases: pre-processing, model building and post-processing (see dashed boxes in figure 2).

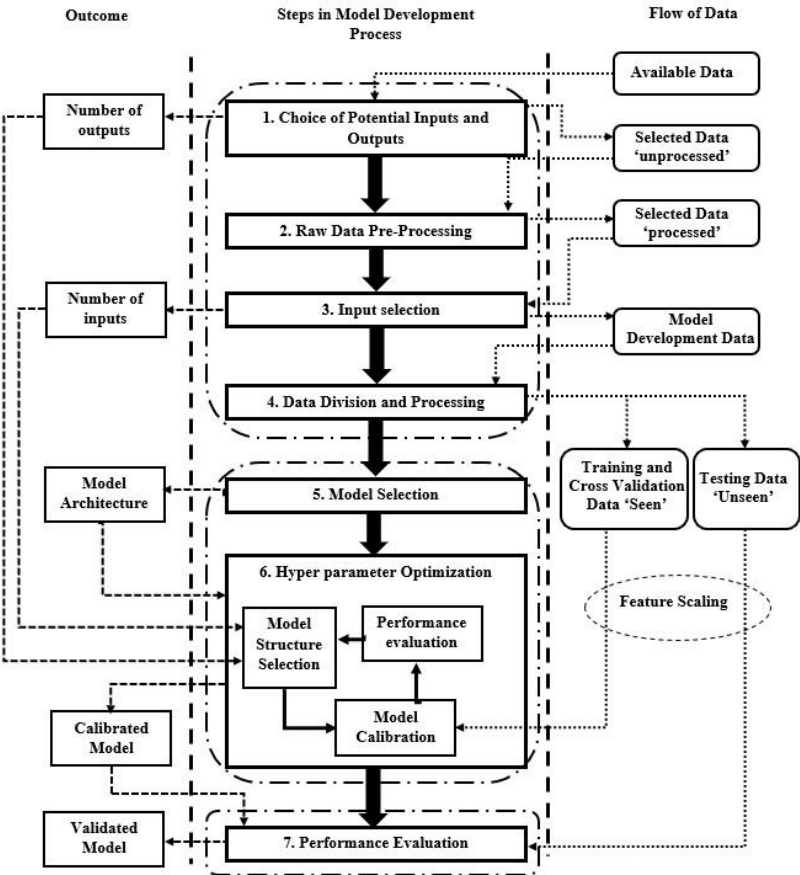


Figure 2. Model development processes flowchart (adapted from Meier et al. [18]).

The pre-processing phase involves utilizing several approaches to identify and prepare the essential data for building the model based on the formulated objectives. In addition to data cleaning techniques, it consists of the following steps: i) input selection and raw data pre-processing, and ii) data division and processing. As input selection is a key step in developing a reliable model, several techniques were used to identify the most relevant model inputs. Starting from the Ad-Hoc (Available data) approach, potential variables for raw data pre-processing procedures (data cleaning) are initially summarized. Then, the dependency between the variables were systematically quantified using Pearson correlation, Kendall and Spearman correlation and Mutual Information Regression (MI) [19, 20, 21]. Afterwards, the selected datasets were divided into predefined percentages of calibration (training) and evaluation (testing) subsets. The data division was obtained by random resampling using supervised trial and error method with manual adjustments to get a satisfactory level of agreement between the statistical

properties of the subsets. Later, feature scaling method is used to avoid the domination of features in magnitude.

The model building phase involves model selection, development of its architecture and optimization of the hyper parameters to achieve high accuracy and generalization ability. In this study, SVR is selected as the supervised learning algorithm to build up the predictive model [22]. SVR has been proven to be an effective tool in real value function estimation. SVR models are trained using a symmetrical loss function, which equally penalizes high and low misestimates. Using Vapnik's ϵ -insensitive approach, a flexible tube of minimal radius is formed symmetrically around the estimated function, such that the absolute values of errors less than a certain threshold ϵ are ignored both above and below the estimate. In this manner, points outside the tube are penalized, but those within the tube, either above or below the function, receive no penalty. One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. Additionally, it has excellent generalization capability, with high prediction accuracy [23]. A stepwise constructive approach, using the Grid Search algorithm, combined with the cross-validation method are used to optimize (tune) different hyperparameters of the SVR architecture.

The post-processing phase involves mainly the evaluation of the model's performance and prediction accuracy. In addition to that, this phase also concerns about utilization ability based on different proposed forecasting schemes. In this study, the evaluation of the models is carried out by using several error metrics and statistical methods such as Root Mean Squared Error (*RMSE*), Normalized Root Mean Squared Error (*NRMSE*), Mean Absolute Relative Error (*MARE*), Pearson correlation (*r*) and Overfitting indicator.

3.2. Forecasting strategies

Two forecasting schemes using the downscaling models are proposed for the intraday ($t + 1$) and the day ahead ($t + 36$) energy markets.

The intraday ($t + 1$) forecast is obtained by using directly the downscaling models developed before setting the time at ($t + 1$), since at the beginning of each forecast we have available all the observed values at time (t) and the latest update NWP forecasts at time ($t + 1$).

For the day ahead ($t + 36$) forecast, to be performed before noon (12:00) each day, it is proposed to use a multi-step recursive forecasting strategy. The strategy consists in running consecutively the downscaling models, starting by forecasting ($t + 1$) exactly as the intraday and then progressing to higher times. Note that for higher times the observed values in the previous hour are not known and therefore are replaced by forecasted ones. It also be noted that for each forecast the NWP values are maintain equal to the ones available at the beginning of the forecast period.

The coefficient of determination (R^2) is used to evaluate the performance of forecasts and compare them with NWP forecasts.

4. Results and discussion

4.1. SVR downscaling models

After pre-processing phase (see subsection 3.1), periods with three or more hours with consecutive missing values due to system failures and maintenance were eliminated and the data set ended up with 8700 points. Consecutive missing values in shorter periods were replaced by averaging the data over the previous instances. As for weather forecast data, there are several archives based on several NWP members (ensemble techniques) and therefore the missing values were replaced with the data from another member. It should be noted that average hourly was used for both weather forecast and observed data as those datasets have different time resolutions.

For input selection, the results using linear and nonlinear methods mentioned in the methodology are presented in figures 3 and 4 for wind speed and direction, respectively. In those figures, WS, WD and WG, stand for wind speed, direction and gust, respectively, *obs* represents observed values, P_i (with

$i = 0, 1, 2$ and 3) represents NWP values at the four nearest grid points, and $(t - 1)$ stands for values at the previous hour.

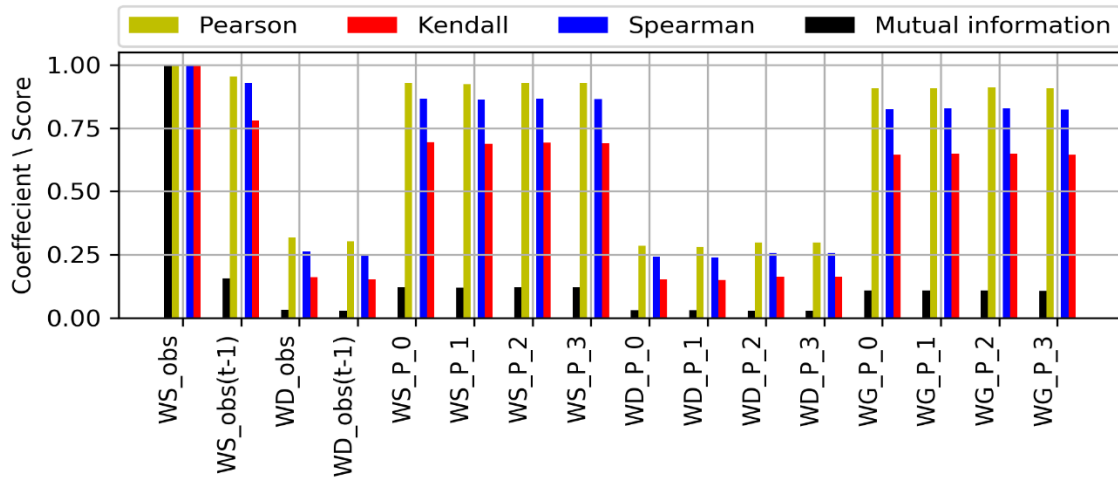


Figure 3. Correlation coefficients and mutual information score between observed wind speed and other variables.

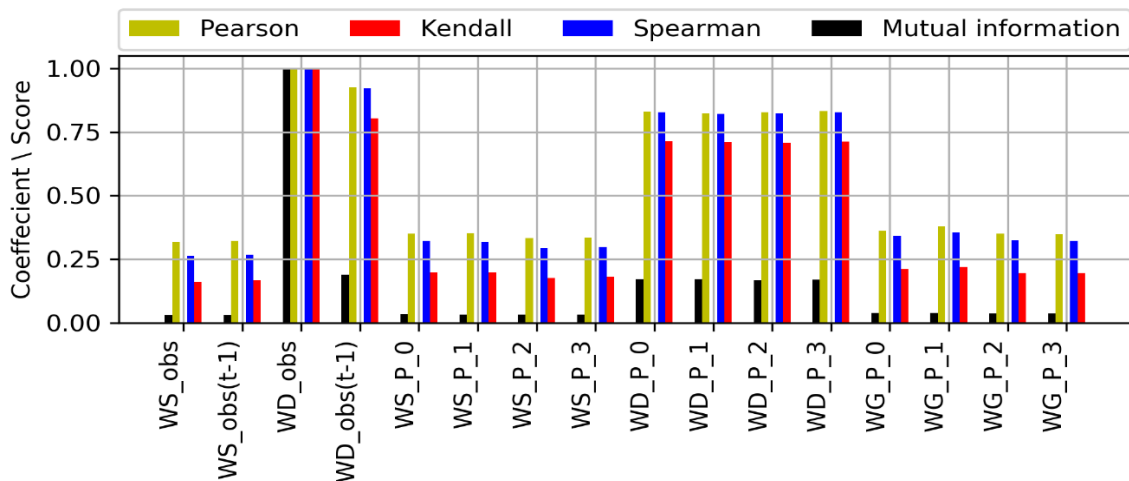


Figure 4. Correlation coefficients and mutual information score between observed wind direction and other variables.

Wind gusts (WG_P0 to WG_P3) and wind speeds (WS_P0 to WS_P3) from the NWP model are highly correlated with observed wind speed at the met mast (see figure 3). Similarly, the wind directions (WD_P0 to WD_P3) from NWP model are highly correlated with observed wind direction (see figure 4). In addition, wind speed and wind direction at a given instant t are highly correlated with the respective values corresponding to the previous hour ($WS_{obs}(t - 1)$ and $WD_{obs}(t - 1)$ in figures 3 and 4, respectively). Results for temperature, relative humidity and air pressure are not presented as they showed very low correlation with the output (WS_{obs} and WD_{obs}). Based on these results, the inputs for the models were selected as in table 1. Note that *avg* represents the average value of the four nearest grid points. This was used instead of the information on each point to reduce the redundancy in the model, leading to better generalization.

The data were further divided into two main subsets for training (80%) and testing (20%). These subsets were resampled randomly, keeping average, standard deviation and coefficient of variation similar between subsets (maximum 1% difference). This ensures that similar trends, like seasonal and

diurnal variations, are present in all subsets. The subsets were further scaled to be used for the SVR algorithm using standard scaler, which transforms the data sets to a zero mean and unitary standard deviation.

Table 1. Models inputs and outputs.

Model	Inputs	output
Wind Speed	$WS_{obs(t-1)}, avg_WS_P_{i(t)},$ $avg_WG_P_{i(t)}$	$WS_{(t)}_{downscaled}$
Wind Direction	$WD_{obs(t-1)}, avg_WD_P_{i(t)}$	$WD_{(t)}_{downscaled}$

The optimal SVR configurations, for both wind speed and wind direction models, have the nonlinear kernel RBF (radial basis function) and a penalty parameter of 10. The gamma parameter is 0.06 and 0.01, respectively. The epsilon parameter is 0.3 and 0.2 respectively.

The SVR models, as discussed above, were developed and tested. The results of the performance evaluation are presented in table 2. All performance statistics show that both models improve the NWP forecasts for time (t) at the proposed site. The models have an RMSE of around 1 ms^{-1} and 33° for wind speed and direction, respectively. Corresponding NRMSE, normalized over the range, were around 4% and 10 %. Similarly, MARE, which reflects the effect of the magnitude of the error relatively to the individual observation, and other error indicators, follow the same pattern as shown in table 2. These, together with the high values of the overfitting indicator, demonstrate the robustness of the developed SVR models.

Table 2. Performance statistics of both wind speed and wind direction models.

Performance metrics	Wind Speed			Wind Direction		
	SVR Train-set	SVR Test-set	NWP ^a	SVR Train-set	SVR Test-set	NWP ^a
RMSE	1.019	1.023	3.065	30.594	32.982	51.421
NRMSE	0.037	0.039	0.117	0.090	0.103	0.160
MAE	0.745	0.747	2.740	17.031	17.805	26.640
MARE	0.128	0.123	0.410	0.154	0.164	0.232
r	0.963	0.963	0.930	0.938	0.930	0.848
R ²	0.930	0.928	0.356	0.880	0.864	0.669
Overfitting Indicator	.	0.996	-	.	0.927	.

^a Full record of NWP based on the most recent forecast, i.e. considering the 4 updates per day.

4.2. Forecasting strategies

The utilization ability of the models for wind forecasting were tested with an independent dataset from January 2020. Results for forecasted wind speed and direction are shown in figures 5 and 6, respectively. In those figures, the pink line refers to a new cycle of day-ahead ($t + 36$) predictions starting at noon (12:00) each day, where the last hour observed value is assumed to be known and the NWP forecast values are updated. The dashed pink lines define the day ahead region of interest (ROI). Table 3 summarizes the performance of the models following the two forecasting strategies.

The intraday ($t + 1$) forecasts show high performance ($R^2 = 0.936$ and 0.943 for the wind speed and direction, respectively). It is clear from both figures 5 and 6, that the hour ahead forecast is the one that follows best the actual measured values.

As expected, the multi-step recursive forecast for 36 hours ahead is relatively less accurate ($R^2 = 0.823$ and 0.803 for the wind speed and direction, respectively). From figure 5, both underestimation and overestimation periods can be seen within the ROI. It should be highlighted that, except for the initial step ($t + 1$), the recursive forecasting strategy progressively uses the previous hour's predictions as an input. Hence, the relatively higher error for this forecasting strategy is due to the

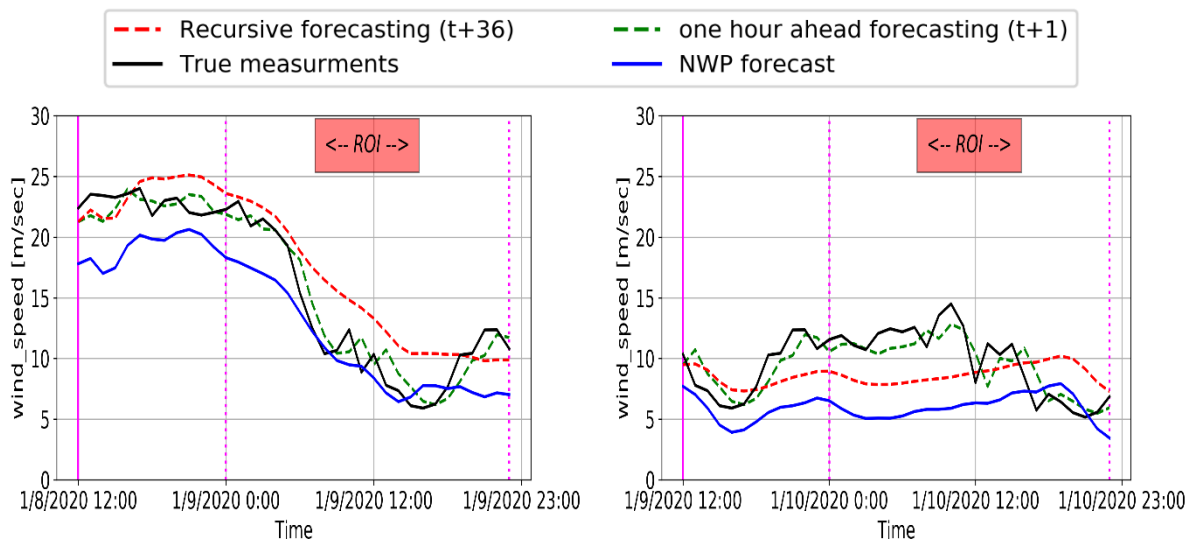
accumulation of errors from the successive hour ahead predictions, which were used as additional inputs. Additionally, those consecutive hour-ahead predictions (multi-steps) use NWP values that are not updated throughout the forecast horizon. Therefore, they are, in principle, worse than the intraday ($t + 1$) forecast, where NWP values are updated each six hours. Nevertheless, it should be noted that the recursive forecasting strategy gives much better results compared to the NWP ($R^2 = 0.578$ for wind speed while NWP outperformed the recursive forecasting having $R^2 = 0.88$ for direction). This is confirmed by looking at the wind speed results within the ROI (figure 5), where the recursive forecast shows to be effective in avoiding the general underestimation of the NWP. For the wind direction (figure 6), that effectiveness is not so evident, which indicates that for medium time horizons the wind direction has a lower degree of temporal dependence than the wind speed.

Table 3. Evaluation of the forecasting strategies using independent datasets.

Propertie s	Wind Speed			Wind Direction		
	Intraday ($t + 1$)	Day ahead ^b ($t + 36$)	NWP ^b	Intraday ($t + 1$)	Day ahead ^b ($t + 36$)	NWP ^b
Inputs	$WS_{obs(t)}$, $avg_WS_P_{i(t+1)}$ $avg_WG_P_{i(t+1)}$	$WS_{obs(t)}^a$, $avg_WS_P_{i(t+1)}$ $avg_WG_P_{i(t+1)}$		$WD_{obs(t)}$, $avg_WD_P_{i(t+1)}$)	$WD_{obs(t)}^a$, $avg_WD_P_{i(t+1)}$)	
RMSE	1.332	2.220	3.424	13.124	24.862	19.160
R^2	0.936	0.823	0.578	0.943	0.80	0.88

^a Only using observed recent value at the first step i.e. ($t + 1$), for next steps using predicted downscaled ones.

^b The day ahead and the NWP R^2 values were calculated based on the region of interest (ROI).



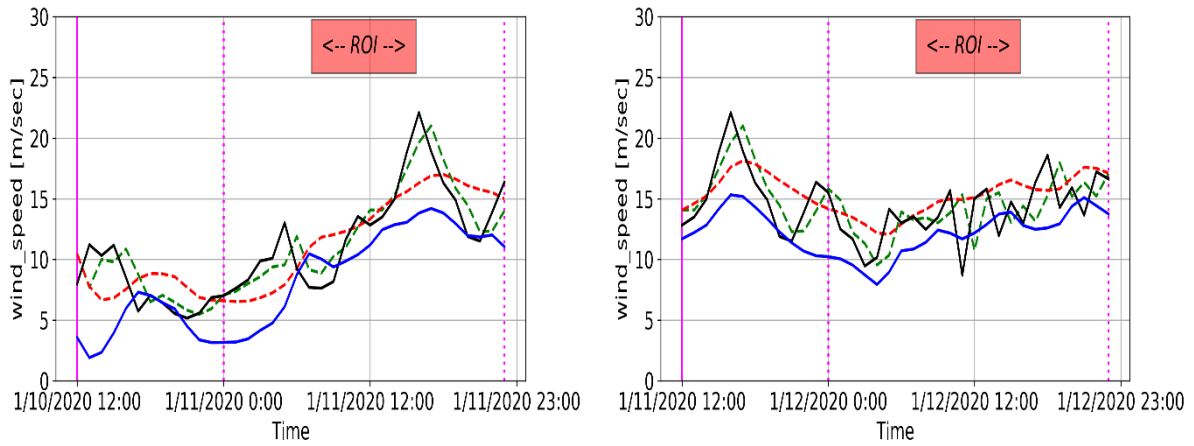


Figure 5. Wind speed forecasts for four days (8th to 12th September 2019).

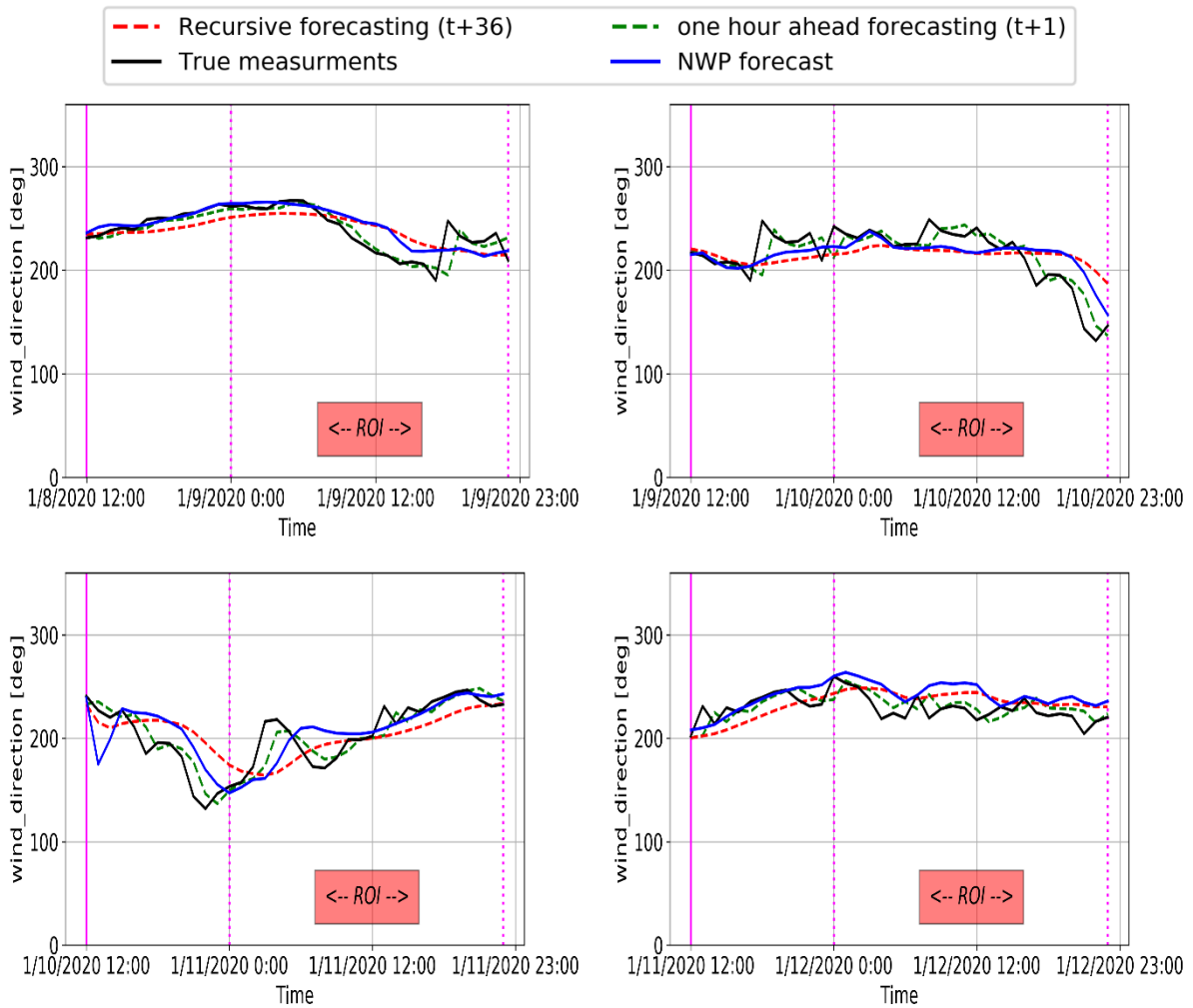


Figure 6. Wind direction forecasts for four days (8th to 12th September 2019).

5. Conclusion

In this paper, we demonstrate that machine learning models based on SVR can be efficiently used for downscaling NWP results to a specific site. The inclusion of observed values in the previous hour allows

accurate correction of wind speeds and wind directions. The downscaling models can be used for short-term (intraday) forecasting of wind available at a specific location of interest, rendering accurate predictions. The downscaling models can be further used for medium-term forecasting (day ahead). Although the prediction accuracy decreases, the recursive forecasting strategy still clearly overperforms the NWP forecasts. The proposed wind forecasting strategies can be effectively used for applications like short- and medium-term wind power forecasting.

Acknowledgements

The Authors would like to thank the National Wind Energy Center in Norway (NVES) for their cooperation in providing the observed data needed for this study. This work was partly funded by the Research Council of Norway (RFF Midt-Norge) through the project number 299452. The first author also acknowledges the Ministry of Education in Norway for the PhD fellowship at the University of Agder.

References

- [1] Global Wind Energy Council 2019, *Global Wind Report 2018*, ed K Ohlenforst et.al, chapter Market Status 2018 pp 22–30
- [2] Mazzi N and Pinson P 2017 *Renewable Energy Forecasting from Models to Applications*, ed Kariniotakis G, chapter 10 pp 259–78
- [3] Skajaa A, Edlund K and Morales J 2015 Intraday trading of wind energy *IEEE Transactions on power systems*, 30 pp 3181–89
- [4] Bauer P, Thorpe A and Brunet G 2015 The quiet revolution of numerical weather prediction *Nature* 525 pp 47–55
- [5] Torres J, Garcia A De Blas M and De Francisco A 2005 Forecast of hourly average wind speed with ARMA models in Navarre (Spain) *Solar Energy* 79 pp 65–77
- [6] Marguan A, Marquez F, Perez J and Ruiz H 2018 A survey of artificial neural network in wind energy systems *J. Applied Energy* 228 pp 1822–36
- [7] Li G and Shi J 2010 On comparing three artificial neural networks for wind speed forecasting *J. Applied Energy* 87 pp 2313–20
- [8] Salcedo-Sanz S, Ortíz-García E, Pérez-Bellido A, Portilla-Figueras J and Prieto L 2011 Short term wind speed prediction based on evolutionary support vector regression algorithms *Expert Syst. Appl.* 38 pp 4052–57
- [9] Mana M, Burlando M and Meissner C 2017 Evaluation of two ANN approaches for the wind power forecast in mountainous site *International Journal of Renewable Energy Research* 7 pp 1629–38
- [10] Lei M, Shi-yan L, Chuan-wen J, Hong-ling L and Zhang Y 2009 A review on the forecasting of wind speed and generated power. *Conf. Proc. Renewable and Sustainable Energy Reviews*, 13 pp 915–20
- [11] Collins S, James R, Ray P, Chen K, Lassman A and Brownlee J 2013 Grids in numerical weather and climate models, ed Zhang Y and Ray P, chapter 4 pp 111–28
- [12] Shirkhani H, Seidow O and Mohammadian A 2013 Downscaling method for wind data: case study of Agadez in Niger *3rd Int. Conf. on Disaster Prevention and Mitigation Canadian Society for Civil Engineering* (Montreal Quebec) pp 1–10
- [13] Watters C and Leahy P 2011 Comparison of linear, Kalman filter and neural network downscaling of wind speeds from numerical weather prediction *10th Int. Conf. on Environment and Electrical Engineering (IEEE)* (Rome) pp 1–4
- [14] Dupre A, Drobinski P, Alonzo B, Badosa J, Briard C and Plougonven R 2020 Sub-hourly forecasting of wind speed and wind energy *Conf. Proc. Renewable Energy* 145 pp 2373–79
- [15] Sfetsos A 2000 A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Conf. Proc. Renewable Energy* 21 pp 23–35

- [16] Okumus I and Dinler A 2016 Current status of wind energy forecasting and a hybrid method for hourly prediction. *Conf. Proc. Energy Conversion and Management* **123** pp 362–71.
- [17] The MetCoOp Team 2017 The MetCoOp ensemble *MEPS HIRLAM Newsletter* **no 8** pp 98–103
- [18] Maier H, Jain A, Dandy G and Sudheer K 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software* **25** pp 891–909
- [19] Guyon I and Elisseeff A 2003 An Introduction to variable and feature selection *J. Mach. Learn. Res.* **3** pp 1157–82
- [20] Ross B 2014 Mutual Information between Discrete and Continuous Data Sets *J. Plos One* **9** pp 1–5
- [21] Kraskov A, Stogbauer H and Grassberger P 2004 Estimating mutual information. *Phys. Rev. E, Statistical, nonlinear, and soft matter physics* **69** pp 1–16
- [22] Drucker H, Burges C, Kaufman L, Smola A, and Vapnik V 1996 Support vector regression machines. *NIPS* **9** pp 155–61
- [23] Awad M and Khanna R 2015 *Efficient Learning Machines Theories, Concepts and Applications for Engineers and System Designers* chapter 4 pp 67–74