Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region

(*This is a revised and preprinted verision, the print version copyrights belongs to AIP Journal of Renewable and Sustainable Energy*)

Hao Chen[1], Yngve Birkelund[1], Stian Normann Anfinsen[2], and Fuqing Yuan[1]

[1] Department of Technology and Safety, [2]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway

Abstract

This paper conducts a systemic comparative study on univariate and multivariate wind power forecasting for five wind farms inside the Arctic area. The development of wind power in the Arctic can help reduce greenhouse gas emissions in this environmentally fragile region. In practice, wind power forecasting is essential to maintain the grid balance and optimize electricity generation. This study firstly applies various learning methods for wind power forecasting. It comprehensively compares the performance of models categorized by whether considering weather factors in the Arctic. Nine different representative types of machine learning algorithms make several univariate time series forecasting, and their performance is evaluated. It is demonstrated that machine learning approaches have an insignificant advantage over the persistence method in the univariate situation. With numerical weather prediction wind data and wind power data as inputs, the multivariate forecasting models are established and made one hour to six hours in advance predictions. The multivariate models, especially with the advanced learning algorithms, show their edge over the univariate model based on the same algorithm. Although weather data are mesoscale, they can contribute to improving the wind power forecasting accuracy. Moreover, these results are generally valid for the five wind farms, proving the models' effectiveness and universality in this regional wind power utilization. Additionally, there is no clear evidence that predictive model performance is related to wind farms' topographic complexity.

Key words: wind energy, machine learning, power forecasting, numerical weather prediction, Arctic

1 INTRODUCTION

To prevent global average temperatures from rising 1.5°C above pre-industrial level, the renewable energy percentage must increase from 20% to 67% of global energy production from 2018 to 2040 [1]. Wind energy is one of the fastest-growing renewable energy sources. It is considered an attractive alternative to conventional electricity sources generated from fossil fuels. Wind power is extensive, and its capacity has surged from 9,936 MW in 1998 to 564,347 MW in 2018, with an annual growth rate of 22.4% in the last 20 years [2]. Along with the electricity grid adds wind power penetrations, the unstable grid factors are also increased, which are undesirable to the power system practical and safe operations. So, it is crucial to use proper methods to understand wind power production and harness proper methods to make forecasts of the electricity generated by the wind parks.

Norway has a cold climate and a 25,148km coastline, both of which are generally characterized by an abundance of wind energy resources, and it is with a complex terrain consisting of mountains, valleys, and fjords, making the wind change dramatically and unpredictably.

Wind power prediction can be divided into ultra-short-term prediction, short-term prediction, medium-term prediction, and long-term prediction [3]. Ultra-short-term forecasts are predictions made from few minutes to 30 minutes in advance; the short-term are forecasts made from 30 minutes to 48 hours ahead, the medium-term refers to predictions made days, weeks, or months earlier, and the long-term is made years in advance.

In wind engineering, hourly wind power forecasting is an essential part of the short-term prediction, whose main applications are maintaining real-time grid operations and keeping operational security in the electricity market [4].

In this study, five wind parks in the Norwegian Arctic regions are taken as the target. Table 1 serves as a summarized comparison in terms of installed capacity, location, and site ruggedness (RIX) [5] of the five sites.

1.1 Related work

In literature, there is much research on wind power forecasting using multiple analysis methods from many perspectives. A preliminary study on wind energy forecasting considered the use of statistical methods. Still, there is a trend of using machine learning algorithms for the forecast. Machine learning is an emerging artificial intelligence approach that attempts to provide learning capabilities for computers or other equipment without clear operations [6]. It aims to develop strategies and algorithms that learn patterns from training data and make predictions. It can be an alternative tool in wind engineering, apart from the statistics and physical methods [7] to forecast wind power with historical wind data. In particular, deep learning, which has emerged in recent years, offers a promise of automating pattern recognition and solving problems such as complex wind power predictions. However, there is a well-known rule called the No Free Lunch (NFL) theorem in the context of supervised machine learning, which states that averaged over all optimization problems, all non-resampling optimization algorithms perform equally well [8]. Due to geographical and engineering reasons, the most suitable machine learning algorithms for wind power prediction for different wind farms vary.

Ref. [9] did a systematic literature review and found that artificial neural networks are used more frequently to predict wind energy and provide better results than other methods, as demonstrated with more than 180 references in the five years. Specifically, Ref. [10] focused on a wind farm in north Iran at 5-min time interval predictions and found that the adaptive neurofuzzy inference system outperforms the other five data mining algorithms: random forests, M5Rules, k-nearest neighbor, support vector machine, and multilayer perceptron. Ref. [11], based on the Portuguese wind power data throughout 2010–2014, also showed that the adaptive neural fuzzy inference system was the best performer. The artificial neural networks and the radial basis function network RBFN-OLS also delivered strong performances. Ref. [12] demonstrated that the proposed hybrid artificial neural network is effective and efficient for wind power forecasting in a Danish dataset. Ref. [13] used an approach combining the infinite feature selection with the recurrent neural networks and proved its edge in a dataset from the National Renewable Energy Laboratory. Ref. [14] investigated five years of wind observation data of Nigde, Turkey, and found that eXtreme gradient boost, support vector regression, and random forests algorithms are powerful in forecasting long-term daily total wind power and the absolute shrinkage selector operator is the worst algorithm due to its linear basis. Ref. [15] mixed basic Multi-Layer Perceptron to complex deep learning neural networks to conduct the power prediction of a wind farm located in the Ecuadorian mountains. The hybrid model is shown to be more advantageous than a single model.

Notably, this journal emphasizes the basis and the state-of-the-art of wind power forecasting. Ref. [16] offered a detailed adaptabilities analysis of the support vector machine, genetic algorithm backpropagation, and radial basis function for wind power forecasting based on three wind farms in China. Ref. [17] noted the essence of deep learning in predictions; sometimes, the forecasts did not need models based on truly deep neural networks, but they offer a sound workflow for correctly developing a proper forecast model.

Meanwhile, much research concerns a particular class of machine learning algorithms, such as kernel methods and neural networks-based methods. There are a few studies that make comparisons between types of algorithms. The reason is applying machine learning always needs tuning the hyperparameters, which makes the comparison rather sophisticated. However, choosing an algorithm less sensitive to parameters or using a suitable method of adjusting parameters, and scale-up and diversification data set can help deal with the problem. Ref. [18] proposed a two-stage wind power forecast method with meteorological factor and fault time and compared the method performance with support vector machines, artificial neural network, generalized regression neural network, and radial basis function and found the edge of the first algorithm. Moreover, there is a lack of complete comparative research on both univariate and multivariate time series forecasting with data science for wind energy prediction in the Arctic region characterized by dense air and excellent wind resources.

1.2 Objective and contributions

The objective and main contributions of this study can be summarized as follows:

1. The paper does a systematic study of the time series forecast for five wind parks generating power with sufficient wind potentials in the Norwegian Arctic region. We mainly focus on investigating the multivariate wind power forecasting models by considering Numerical Weather Prediction (NWP) data.

2. For brief experimental univariate power forecasts. The persistence model and nine machine learning benchmarking algorithms are researched in forecasting models and compared their performance from an algorithm perspective. We find the persistence model performs almost equally to machine learning models in our cases. The result also proves conclusions from Ref.[19]; those classical methods may dominate univariate time series forecasting. However, we find that its performance drops more quickly with the forecast time step rises. Considering the contingency of parameters tuning and computational complexity of the learning algorithm, it is suggested that statistical modeling methods should be primarily considered in the forecasting.

3. The multivariate models with mesoscale NWP wind data, although the data resolution scale is larger than the wind parks area, as inputs can slightly gain prediction accuracy compared with the univariate models with the same algorithm. Moreover, the multivariate models reduce performance slower than the univariate models, which indicates the informative complementary role that the weather data play in the model

4. These five wind farms have different complex terrains and climates. The wind park with complex terrains implies that the NWP wind results are not as accurate as their counterparts of plain landscapes. However, there is no significant evidence that prediction results are related to the ruggedness index of wind parks from our results.

2 DATA DESCRIPTION AND PREPARATION

2.1 Wind power locations

Northern Norway has a complex terrain consisting of fjords, mountains, and valleys that goes from the coast into a moderately high inland along the border to northern Sweden and Finland. The terrain elevation around each wind park is also shown in Fig.1, and their coordinates and heights are listed in Table 1. Nygårdsfjellet wind park is located in a valley, far from the open sea, that reached approximately 450 meters elevation. The mountains south and north of the valley limit the main wind direction to be west-east, and high wind events are expected during the winter season [20]. Havøygavlen, Kjøllefjord, and Fakken wind parks are located close to the open sea and on relative flat hills where large nearby fjords affect both wind direction and speed. Raggovidda wind park is also located near the open sea but on a flat mountain that does not have any vegetation. This location is well known for adequate wind resources and produced power with relatively high capacity factors (the ratio between the real and designed power production) over several years.

2.2 Norwegian Meteorological Institute (MET Norway) numerical weather prediction

The Scandinavian weather institutions use a weather model for weather forecasts named MEPS (Ensemble Prediction System). The weather model makes ensemble forecasts, starting from a composition of several forecasts and quantifying the outcome space of possible weather developments, which depends on the weather itself rather than looking at a single estimate [21].

The NWP model is a complex mathematical model of the atmosphere that divides the earth surface into grids [22]. The spatial resolution of the grid determines how to simulate meteorological processes with different accuracy levels, limiting the quality of forecasts.

A study conducted by MET Norway has demonstrated that the regional NWP models with higher resolution did not result in better wind power forecasts for some Norwegian wind farms [23]. Therefore, in this study, we use the NWP data with 2.5km horizontal resolution, which is regarded as a relatively coarser resolution in wind forecasting.

2.3 Data description and scaling

The hourly power data of five wind farms, measured hourly, used in the research is provided by the Norwegian Water Resources and Energy Directorate (NVE). We choose the wind power data from 0:00 1st January 2017 to 23:00 31st December 2017; the measured data are 8,760 for each wind farm. The total number of wind power data is 43,800. The location, annual mean powers, the standard deviation, and the capacity factor of the five wind farms in 2017 are also shown in Table 1.

Table 1. The location and statistics of power data

Wind Park	Location °N / °E	Height [m]	RIX	Designed power [MW]	Mean power [MW]	Standard deviation [MW]	Capacity factor
Nygårdsfjellet	68.504 / 17.879	410	0-5	32.2	11.132	11.833	34.57%
Fakken	70.098 / 20.081	95	5-10	54.0	15.239	15.858	28.22%
Raggovidda	70.769 / 29.094	440	0-5	45.0	21.782	16.869	48.40%
Kjøllefjord	70.922 / 27.268	280	10-20	39.1	12.349	12.786	31.58%
Havøygavlen	71.012 / 24.589	220	5-10	40.5	10.311	11.037	25.46%

The NWP wind forecast data are extracted from MET Norway operational weather forecast model MEPS and considering that forecasts need two hours to be calculated as usual. The forecasts are all initiated at 00, 06, 12, and 18 UTC.

Wind power generating is mainly affected by wind speed, wind direction, and air density, which is impacted by temperature and air pressure [24]. In this study, we use the variables acquired at time t; such as measured generating wind power, NWP wind speed, NWP wind direction (radian system), NWP surface air pressure, and NWP 2 meters above the ground temperature to predict the wind power generated at $t+n$, where n, ranging from 1 to 6, is the time delay in hours. The NWP wind data are summarized and shown in Table 2. All the items show variables with the *mean value* (*standard deviation*) form, and the relative air pressure means the real local air pressure minus the standard atmospheric pressure (101,325Pa).

Table 2. The statistics of the original NWP data

Wind park	Speed (m/s)	Direction	Temperature	Relative air pressure (Pa)

		(radical system)	(°C)	
Nygardsfjellet	8.096 (5.038)	-0.065 (0.431)	0.045 (7.441)	-5795.564 (1246.119)
Fakken	6.948 (3.885)	0.151 (1.032)	4.193 (5.109)	-1091.373 (1284.892)
Raggovidda	9.49 (5.101)	0.011 (0.855)	-0.91 (6.256)	-5148.793 (1277.989)
Kjøllefjord	7.9 (4.213)	0.15 (0.962)	1.23 (5.763)	-2848.796 (1292.669)
Havoygavlen	8.335 (4.434)	0.136 (0.872)	2.953 (5.33)	-1750.36 (1309.583)

Data scaling is a standard approach to normalize data. An important reason for data scaling is that the algorithm converges faster with feature scaling than without it [25]. And it is convenient to compare the model performance with similar data scales. The wind power data is scaled with min-max normalization between 0.2 and 0.8.

$$x' = a + \frac{(x-min(x))(b-a)}{max(x)-min(x)} \qquad (1)$$

where a and b are the minimum and maximum values of the normalization scale.

2.4 Stationary test

The power data for five wind farms can be treated as five univariate time series sequences. Time series can be divided into stationary and non–stationary data sequences. Whether or not a time series is stationary has long been a question of major interest in the field of time series analysis [26]. Statistical regression processes can analyze stationary time series; meanwhile, the non–stationary time series change their statistical properties with time [27]. So, the forecasting for non–stationary time series is more problematic than for stationary ones. Augmented Dickey-Fuller test (ADF) is a widely used method for testing the null hypothesis that a unit root is present in a time series sample. Its principle is to check whether a unit root is present in a sequence. If no unit root presents, the sequence is stationary; otherwise, it is nonstationary. The ADF test is a standard method to test the stationarity of economic time series [28]The ADF test utilizes the autoregressive process and optimizes its parameters for various lag values. A null hypothesis test can conduct the ADF test application in testing the stationary of a time series sequence.

H_0: the time series is nonstationary, which means it shows a time-dependent structure.

H_a: the time series is stationary.

The ADF test is conducted on the five wind farms power data. The results show that all the null hypotheses are rejected with critical values that are much lower than 5%, demonstrating the five-time series power data are stationary. They also show the power data sequences do not have trends and seasonality, which means there is no need to divide the annual wind power data into monthly or seasonal sequences in forecasts.

3 FORECASTING ALGORITHMS FOR WIND POWER

For each utilized forecasting model, numerous changes are proposed by researchers [29], and it is impossible to conduct all of the existing differences in models. Therefore, our strategy is to consider each benchmarking model for different algorithms for hourly wind power forecasting. The ten prediction models, one baseline statistical model, and nine machine learning models are chosen because they are commonly used models.

1. Persistence Model (PE)
2. Support Vector Regression (SVR)
3. K-Nearest Neighbor regression (KNN)
4. MultiLayer Perceptron (MLP)
5. Radial Basis Functions (RBF)
6. Classification and Regression Trees (CART)
7. Random Forest (RF)
8. Stochastic Gradient Boosting (SGB)
9. SVR optimized with Genetic Algorithm (GA-SVR)
10. Long Short-Term Memory neural networks (LSTM)

Method 1 is the reference in the performance comparison, whereas approaches 2 to 6 are representative and widely used machine learning algorithms, and 7 and 8 are two representative types of ensemble machine learning techniques. There are two main types of ensemble learning methods: bootstrap aggregating (bagging) and boosting [30]. The rest two are emerging trends of predictive algorithms, representing hybrid methods and deep learning, respectively. The following offers a brief description of each algorithm. Due to the page limitation, more detailed descriptions of these algorithms are available in the references of this article.

The PE model takes that the power at time $t + n$ equals t the power at t, n is the next n steps in time series. It assumes that the atmospheric conditions change stationarily.

SVR is a regression model provided by the support vector machine algorithm, which tries to identify the hyperplane that maximizes the margin between two classes and minimizes the total error under tolerance [31]. SVR conducts a penalty with C (complexity penalization term) and achieves the best trade-off between the empirical error and the model complexity. SVR can perform a nonlinear regression because it provides kernel functions (like linear, polynomial, and Gaussian) that map data from the input space to a high dimensional feature space in which regression is conducted. The value of C is taken from a validation test for $C \in \{0.01,0.1,1,10,100\}$. C is found with a value of 1 corresponding to the best performance with the Gaussian kernel function.

KNN regression focuses on feature similarity determined by distance functions, like Euclidean, Manhattan, or Minkowski distance, measurements for data samples [32]. The K parameter, which implies the input consists of the k closest training sample subsets, determines the performance of the algorithm; a large K value can reduce the noise in the regression process, but it also leads to a risk of overfitting. In the study, we conduct a grid search for K from 1 to 10 in experiments to find an appropriate K value.

MLP is a network of simple neurons named perceptrons. The perceptron forms a linear combination based on its input weights and calculates the output through a nonlinear activation function [33]. MLP is a versatile approach for forecasting; it can find nonlinear structures in a problem and model a linear regression process. MLP is a parameterized model. We can manage the MLP complexity by choosing the number of hidden nodes and the type of activation functions. Specifically, the sigmoid function is usually used as the activation function in MLP regression problems. In the study, the topology of the MLP consists of three layers; namely, the number of nodes for the input layer equals input numbers, a hidden layer with ten nodes, and an output layer with one node.

RBF networks are feedforward and similar in structure to the MLP. The radial basis functions are harnessed as their activation functions. Their output is a linear combination of radial basis functions (radially symmetric around a center) applied to values of inputs neurons. The RBF also has a fast and efficient training process of both linear and nonlinear mappings. In the study, the RBF topology is the same as the MLP model.

CART regression is based on a tree-like recursive partition of inputs [34]. The CART is made of internal decision nodes and end leaves. Given a test data set, the terminal leaves are decided by different training sample properties. Besides, a series of tests are created and utilized in the decision nodes, which can define where the inputs should be classified to specific nodes whose splitting will most significantly reduce the mean square error. Moreover, a final decision tree is realized when the mean square error is smaller than a threshold.

Bagging is a unique variant of the model averaging approach to reduce the prediction variances by repeatedly creating subsets of original data to train the machine learning model. RF is an efficient bagging ensemble algorithm and delivers sound capability and low computational cost. RF is based on the establishment of a multitude of sub learners. Each learner is trained by using a bootstrap sample extracted from the whole training set. The forest of learners produces ensemble regression values. The final result is determined, e.g., by averaging over the ensemble [35]. RF has only one difference from the general bagging decision tree: it uses an improved decision tree algorithm, selecting a random subset of features at each sample selection in the training process.

Boosting is an iterative technique that uses the last classification to adjust the weights of nodes constantly in the learning process. SGB is a trendy and widely used boosting learning algorithm. It constructs regressions by sequentially fitting a base learner to current "pseudo"-residuals by least-squares in each iteration. It can improve the accuracy of gradient boosting and training speed by incorporating randomization into the learning procedure [36].

Genetic Algorithm (GA) is one of the well-regarded evolutionary algorithms. It mimics the Darwinian theory of survival of the fittest and arrives at such configuration via cycles consisting of individual population generation, selection, crossover, and mutation phase [37]. During the process, the population of candidates originates from a combination of the offspring and survivors of the previous generation or a randomly generated configuration. The population then faces two selection phases that decide which candidates do not survive into the next generation and then decide which candidates may produce child candidates. This filtering uses a fitness function. GA-SVR is using the GA in optimizing complexity penalization term C of SVR in the training process and has become a so-called hybrid forecasting method.

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture that is mainly used in the field of deep learning [38]. Unlike standard feedforward neural networks, LSTM has feedback connections, and unique LSTM units consist of input, output, and forget gates. It can process not only single data points but also entire sequences of data (such as speeches or videos). It can also be used in time series forecasting with proper pre-treatment of data. We utilize the Vanilla LSTM that is with one hidden layer, including LSTM units and one output to predict.

4 EXPERIMENTAL SETUP

Multi-steps wind predictions are required in short-term wind energy generation. In the study, we make direct forecasting, which only uses actual measured values as model inputs. It builds *n* different prediction models for *n* steps ahead forecasting. The benefit of direct forecasting is that it does not use previous prediction values to forecast the values with higher steps, which means the prediction is not affected by the cumulative error in the forecasting process.

Wind power prediction from t_1 to t_3 for the univariate forecasting with one-dimensional input wind power at time t_0. t_1 to t_6 is for the multivariate forecasting with five-dimensional input (wind power, NWP wind speed, NWP wind direction, NWP temperature, and NWP air pressure at time t_0) because according to Ref. [7], the weather factors are recommended considered after 3 hours. Namely, we conduct two modeling processes for each wind park. The multivariate model is displayed in equation 2.

$$\hat{P}_{t+n} = f_{t+n}(P_t, W_t) + e_n,\ \ n = 1,2,3,4,5,6 \quad (2)$$

where $\hat{P}_{t+n}$ is the *n* steps wind power forecasting, f_{t+n} is the forecasting model, P_t and W_t represent the wind power and NWP weather data at time t, e_n is the model error.

The source data set is divided into 66% for training the model and 34% for testing the models' performance and carrying out comparisons.

4.1 Cross-validation and evaluation Metrics

One of the most critical and popular validation methods for machine learning is *k*-folds cross-validation; it is more suitable for relatively small and limited data set compared with the train and test split validation approach because it can ensure that there is a chance for every sample in the original data to appear in the training as well as the validation process. In this research, we use $k = 10$ in the implementation of the whole training set. According to a study by Kohavi, this is usually a pretty good choice [39].

Two metrics are used in evaluating the performance of the different kinds of algorithms for wind power forecasting. The first metric is the Mean Absolute Error (MAE); the second metric is the Root Mean Square Error (RMSE). The definitions for MAE and RMSE are shown in equations 3 and 4. Both are negatively oriented metrics, meaning the lower scores are related to better performance [40]. In our cases, if an approach has a low MAE but a high RMSE, it generally predicts smoothly and efficiently but has a higher population of large significant forecasting errors that are weighted significantly by RMSE.

$$MAE = \frac{\sum_{i=1}^{n} |prediction_i - observation_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(prediction_i - observation_i)^2}{n}} \tag{4}$$

4.2 Forecasting procedure

For the procedure of univariate hourly wind power forecasting, the wind power data are normalized and tested with ADF stationary tests. Then we use the PE model and nine machine learning algorithms to do three steps forecasts.

Moreover, to further improve wind power forecasting accuracy and make full use of the NWP data, we established a multivariate forecasting model based on SVR, MLP, RF, and LSTM algorithms, shorten as '*NWP* plus the abbreviation', based on the univariate forecasting results and some recommendations for publications [41, 42]. The performance is compared with their counterparts in univariate cases. The procedure of forecasting is illustrated in Fig. 1.

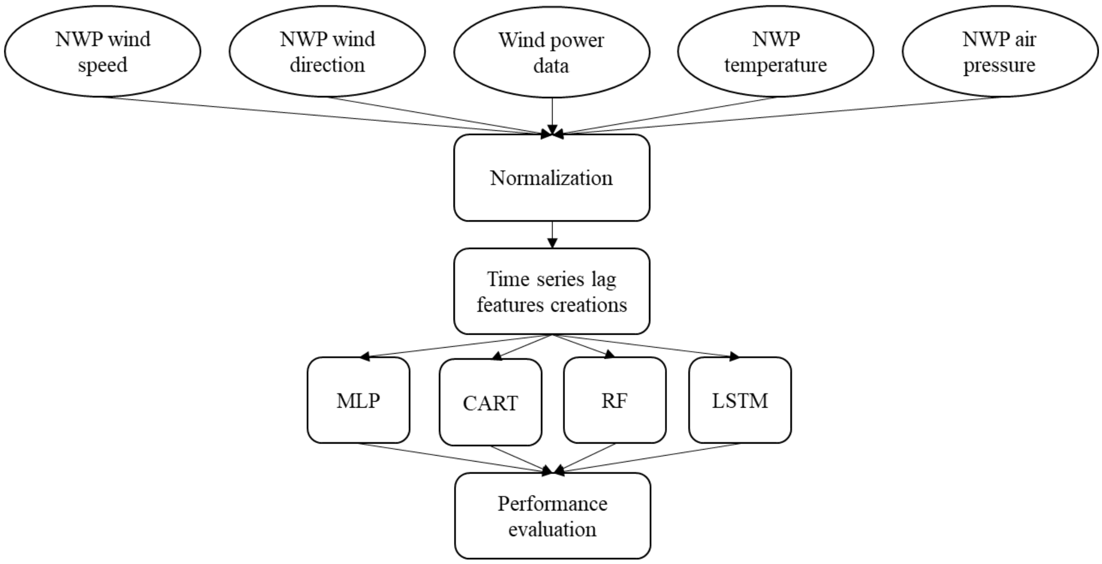

Fig. 1. Procedure for the multivariate hourly wind power forecasting

5 RESULTS

We conduct a univariate wind power forecasting with the aforementioned ten algorithms and adds the NWP wind data as inputs to create new MLP, CART, RF, and LSTM multivariate forecasting models for five wind parks.

For the univariate forecasting, the performance comparison of ten models is briefly conducted. For multivariate forecasting, the four multivariate models' performance is compared with their counterparts in univariate forecasting cases.

5.1 Univariate forecasting

Regarding MAE results, the MAEs of all models increase as the forecasting step adds. The average MAE of three steps predictions is displayed in Fig. 3. The PE models perform similarly with SVR and GA-SVR models in all wind parks. The KNN, MLP, RBF, CART, and LSTM have similar MAEs, which are more unsatisfactory results than PE and SVR and GA-SVR have. The ensemble learning methods show the highest MAEs.

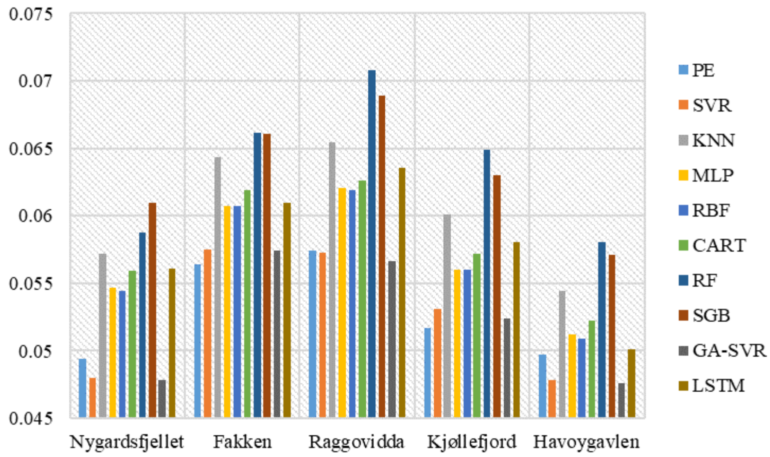

Fig. 2. The average MAE of ten forecasting models for five wind farms

RMSEs of all models have positive correlations with the forecasting step from 1 to 3. The main RMSE of three steps forecasts is shown in Fig. 4. The MLP, RBF, and GA-SVR have the best performance in RMSE, in which MLP has the lowest overall RMSE. The PE, KNN, CART, SGB, and LSTM have similar RMSEs, and nearly all of them have more inferior results than MLP, RBF and GA-SVR do. The RF model has the highest RMSE as it does in MAE analysis.

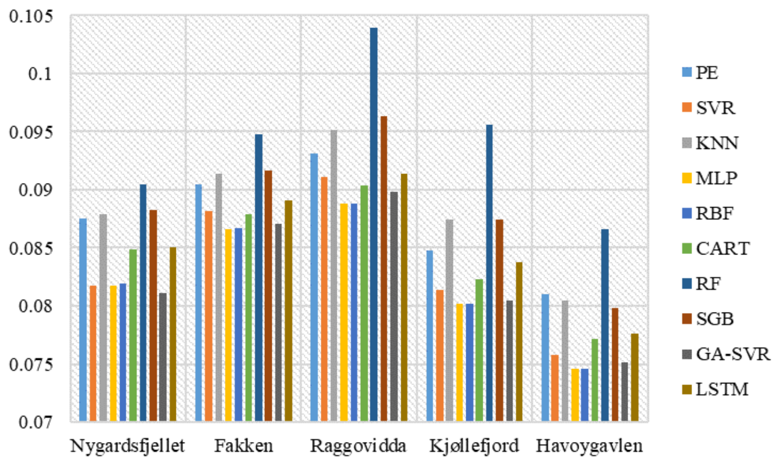

Fig. 3. The average RMSE of ten forecasting models for five wind farms

5.2 Multivariate forecasting

The MAE of the multivariate forecast for the five wind parks is displayed in Fig. 5. We predict one hour to six hours of wind power. As the forecasting step increases, the MAE of all models increases, and the rising speed gets slower. For each forecasting step, the NWP machine learning models have lower MAE than their univariate counterparts, and NWP RF is with the largest improvement. Generally, the edge of multivariate models is incrementally stronger with the raising forecast time. RF and LSTM, which perform unfavorably in univariate forecasts, excel in multivariate predictions. In particular, NWP LSTM dominates all cases, and its domination is reinforced over time. For four of the wind parks, the NWP models perform better in the term of MAE. The exception is Raggovidda wind park, for which the PE model has the lower MAE for all six-time steps than NWP MLP and CART. Meanwhile, the forecasting models produce the lowest MAE for Havoygavlen wind park.

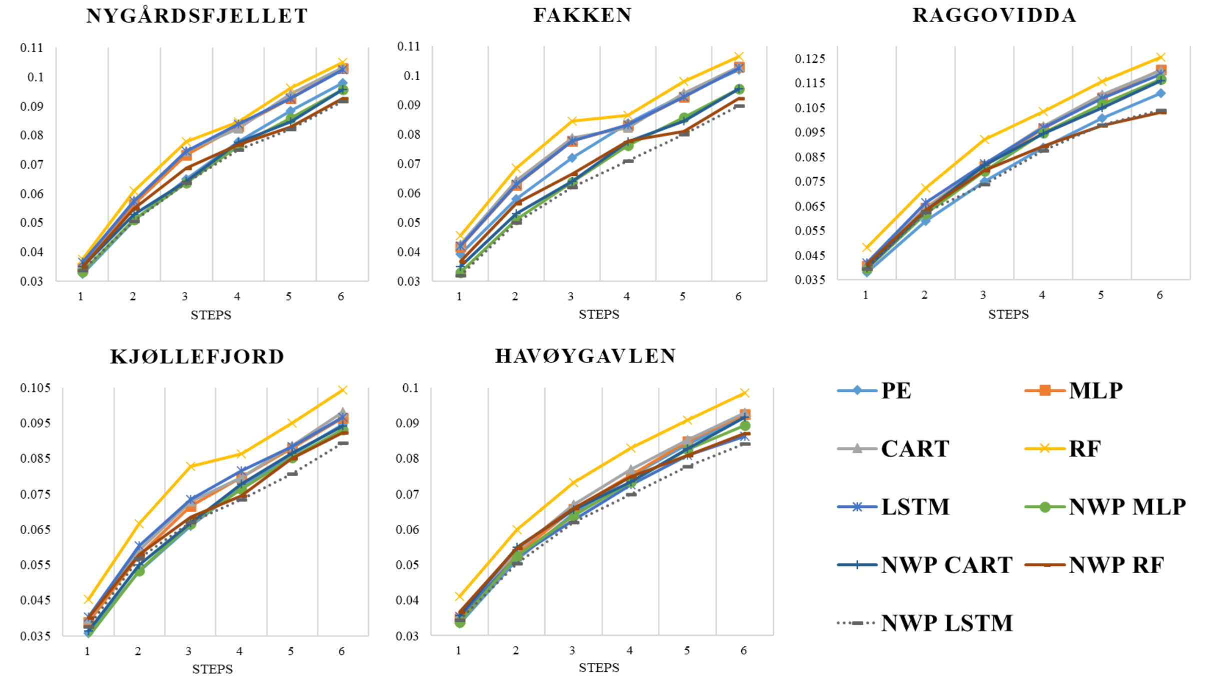

Fig. 4. The MAE comparisons of univariate and multivariate models for five wind farms

Numerical comparisons of average MAE for different models are shown in Table 3. The NWP LSTM model is ranked first for almost all wind parks. Adding the mesoscale NWP wind data can significantly increase forecasting algorithms' performance based on the same algorithm, especially for Fakken wind park, which reduces MAE by 16.14% and 16.54% concerning RF and LSTM, respectively.

Table 3. The comparisons of average MAE for different models

Wind park	PE v.s. NWP LSTM	MLP v.s NWP MLP	CART v.s NWP CART	RF v.s NWP RF	LSTM v.s NWP LSTM	NWP CART v.s NWP LSTM
Nygårdsfjellet	3.78%	8.41%	8.41%	11.02%	11.33%	3.18%
Fakken	14.16%	12.04%	11.94%	16.14%	16.54%	5.99%
Raggovidda	1.53%	2.71%	2.99%	14.89%	9.62%	6.96%

Kjøllefjord	1.81%	5.32%	4.82%	12.98%	8.11%	2.82%
Havøygavlen	5.46%	2.60%	1.92%	10.30%	3.02%	6.32%

The RMSE of the multivariate forecasts displays in Fig. 5. The trends of RMSE are similar to MAE's. Besides, growth rates of RMSE of all machine learning models, especially the NWP MLP, RF, and LSTM models, are lower than the PE model. Concerning each forecasting step, the multivariate models have lower RMSE than corresponding univariate models. For all the five wind parks, the multivariate LSTM performs best in terms of RMSE in nearly all the six predictive steps. Moreover, the Raggovidda wind park still has a higher RMSE compared to other wind parks. Meanwhile, the models still provide the lowest RMSE for Havøygavlen wind park.

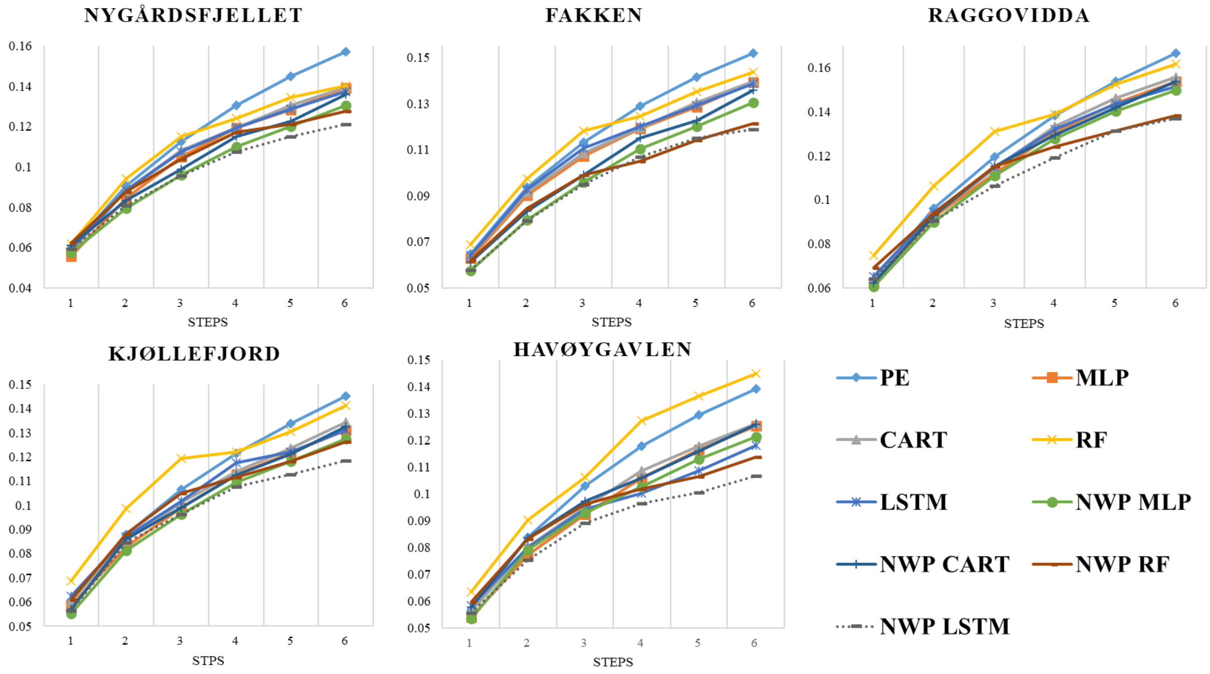

Fig. 5. The RMSE comparisons of univariate and multivariate models for five wind farms

Table 4 shows the comparisons of average RMSE and demonstrates the multivariate LSTM models are the best for all wind parks. It overperforms approximately an average of 14% better RMSE performance than the baseline PE model. The mesoscale NWP wind data provide positive information in the forecast algorithm.

Table 4. The comparisons of average RMSE for different models

Wind park	PE	MLP	CART	RF	LSTM	NWP CART

	v.s. NWP LSTM	v.s NWP MLP	v.s NWP CART	v.s NWP RF	v.s NWP LSTM	v.s NWP LSTM
Nygårdsfjellet	16.69%	6.12%	4.27%	7.44%	9.56%	6.09%
Fakken	17.46%	8.25%	5.61%	14.83%	12.66%	7.21%
Raggovidda	12.14%	2.25%	1.57%	12.20%	7.53%	6.72%
Kjøllefjord	12.03%	2.95%	1.65%	10.33%	7.34%	5.33%
Havøygavlen	16.82%	1.56%	-0.36%	16.09%	6.43%	10.68%

6 CONCLUSION

This paper makes univariate and multivariate short-term wind energy forecasts for five wind parks inside the Norwegian Arctic region. Consequently, the following conclusions can be drawn.

For the univariate time series wind power prediction in these cases, the PE approach and machine learning methods do not have a considerable difference in performance. The SVR and MLP function equally well with the PE model. The machine learning algorithms that perform best in MAE are SVR and GA-SVR, whose average MAE is almost the same (0.18% and 0.10 % lower) as for the PE model. The machine learning algorithm that performs best in RMSE is MLP, whose average RMSE is 5.4% lower than the PE model. SVR, RBF, GA-SVR, and LSTM also have lower RMSE than the PE model has. This generally means the PE model has more large errors in the prediction procedure. Our result also validates the conclusion from research [19] in the wind engineering field. The conclusion is learning algorithms do not deliver on their promise for univariate time series prediction, and the classical statistical methods even perform better. The phenomena may be explained that for univariate series, the complex methods often overlearn the training set and create overfitting in the testing.

For the multivariate wind power forecasting in our cases, the model considers methodological or topographic factors by taking the mesoscale NWP wind data as inputs. Compared to the corresponding algorithm in the univariate case, the multivariate model has a lower MAE and results in a smaller RMSE. When the predictive time increases, the multivariate models are more stable than the PE model, especially in the metric of RMSE. These prove that the multivariate model entirely exceeds the PE model and the univariate model. Furthermore, the sophisticated ensemble and deep learning algorithm demonstrate their superiority in dealing with complex and multivariate pattern recognitions in complicated wind power forecasting problems.

The NWP wind data are generated with mesoscale (2.5km×2.5km), which is larger than the area of our wind parks. However, adding this local weather information can still obviously optimize the performance of forecasting models. The improvements of the penetration of NWP data in wind power prediction can be explained from two aspects: firstly, from the Bayesian theory [43], the introduced NWP wind information can provide a priori probability information to make more precise wind power (corresponding to posterior probability) predictions; secondly, NWP wind data can be regarded as the simulating wind conditions of the whole wind park, which add useful information in the predictive process.

In summary, based on our case studies, it is recommended to use statistical methods for short-term univariate wind power forecasting since learning algorithms involve parameter tuning and larger computational volumes without significantly better performance than statistical ones. It is advisable to include meteorological information, even if the weather data scale is relatively large, into the multivariate predictive models, where the advanced learning algorithms can be really effective with such autoregression combined with meteorological inputs.The replicability of these conclusions is established because the data from the five wind parks are relatively uncorrelated, and the machine learning regressions are built with sufficient considerations for the generalization of the models.

Moreover, we cannot find that each wind park's forecasting results significantly correlate with the site ruggedness. There are two possible reasons for this. First, the decisive, independent variable in the data-driven wind power prediction model is the prior value of power. According to Table 1, the standard deviation of the power time series does not become larger with increasing RIX. Second, because of the relatively large scale of the NWP model and the fact that the concerned wind farms are located near the sea, the effect of complex terrain is mitigated in the NWP grid. Therefore, to further investigate topography influence, the hydrodynamic modeling between topography and wind turbines in wind farms is needed. More interestingly, we find that the performances have some correlations with each wind park's capacity factor; the higher the capacity factor is, the lower the performance. i.e., the most inferior performance (Raggovidda) with the highest capacity factor of 48.40% and the best performance (Havøygavlen) with the lowest capacity factor of 25.46%. This implies that turbulence within the wind farm in complex terrain conditions may be the dominant factor, masking the effects of external large-scale weather.The phenomenon needs further investigations.

ACKNOWLEDGMENTS

The authors would like to thank the Department of Technology and Safety, University of Tromsø-The Arctic University of Norway, for their administration and financial support for this study.

DATA AVAILABILITY

The supporting data for this paper are available on request from the authors.

REFERENCES

1. Tollefson J. IPCC says limiting global warming to 1.5 C will require drastic action. *Nature* 2018; 562: 172-173.
2. Dudley B. BP Statistical Review of World Energy Statistical Review of World. *bp com* 2019.
3. Jung J and Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews* 2014; 31: 762-777.
4. Liu K, Zhang Y and Qin L. A novel combined forecasting model for short-term wind power based on ensemble empirical mode decomposition and optimal virtual prediction. *Journal of Renewable and Sustainable Energy* 2016; 8: 013104.
5. Birkelund Y, Alessandrini S, Byrkjedal Ø, et al. Wind power predictions in complex terrain using analog ensembles. 2018.
6. Samuel AL. Some studies in machine learning using the game of checkers. II—Recent progress. *IBM Journal of research and development* 1967; 11: 601-617.
7. Giebel G, Brownsword R, Kariniotakis G, et al. The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS plus* 2011.

8. Joyce T and Herrmann JM. A review of no free lunch theorems, and their implications for metaheuristic optimisation. *Nature-inspired algorithms and applied optimization*. Springer, 2018, pp.27-51.
9. Maldonado-Correa J, Solano J and Rojas-Moncayo M. Wind power forecasting: A systematic literature review. *Wind Engineering* 2019: 0309524X19891672.
10. Renani ET, Elias MFM and Rahim NA. Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management* 2016; 118: 193-203.
11. Godinho M and Castro R. Comparative performance of AI methods for wind power forecast in Portugal. *Wind Energy* 2020.
12. Wan C, Song Y, Xu Z, et al. Probabilistic wind power forecasting with hybrid artificial neural networks. *Electric Power Components and Systems* 2016; 44: 1656-1668.
13. Shao H, Deng X and Jiang Y. A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network. *Journal of Renewable and Sustainable Energy* 2018; 10: 043303.
14. Demolli H, Dokuz AS, Ecemis A, et al. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management* 2019; 198: 111823.
15. Maldonado-Correa J, Valdiviezo-Condolo M, Viñan-Ludeña MS, et al. Wind power forecasting for the Villonaco wind farm. *Wind Engineering* 2020: 0309524X20968817.
16. Yan J, Gao X, Liu Y, et al. Adaptabilities of three mainstream short-term wind power forecasting methods. *Journal of Renewable and Sustainable Energy* 2015; 7: 053101.
17. Torres J, Aguilar R and Zuñiga-Meneses K. Deep learning to predict the generation of a wind farm. *Journal of Renewable and Sustainable Energy* 2018; 10: 013305.
18. Cao Y, Hu Q, Shi H, et al. Prediction of wind power generation base on neural network in consideration of the fault time. *IEEJ Transactions on Electrical and Electronic Engineering* 2019; 14: 670-679.
19. Makridakis S, Spiliotis E and Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one* 2018; 13: e0194889.
20. Bilal M, Birkelund Y and Homola M. High winds at nygårdsfjell. *Journal of Clean Energy Technologies,()* 2015.
21. Norway M. MetCoOp, https://www.met.no/en/projects/metcoop (accessed 25th, June 2020).
22. Collins SN, James RS, Ray P, et al. Grids in numerical weather and climate models. *Climate change and regional/local responses* 2013; 256.
23. Bremnes JB and Giebel G. Do regional weather models contribute to better wind power forecasts? : The Norwegian Meteorological Institute, 2017.
24. Roy SB. Simulating impacts of wind farms on local hydrometeorology. *Journal of Wind Engineering and Industrial Aerodynamics* 2011; 99: 491-498.
25. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Cornell University Library. 2017.
26. Luo X. Size and power of tests for assessing weak stationarity of time series data: an empirical investigation. *schriftenreihe des studienGAnGs Geodäsie und GeoinformAtiK 2018–* 2018: 187.
27. Ghazali R, Hussain AJ, Nawi NM, et al. Non-stationary and stationary prediction of financial time series using dynamic ridge polynomial neural network. *Neurocomputing* 2009; 72: 2359-2367.
28. Li B, Zhang J, He Y, et al. Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test. *IEEE Access* 2017; 5: 16324-16331.
29. Raschka S and Mirjalili V. *Python machine learning*. Packt Publishing Ltd, 2017.
30. Sagi O and Rokach L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2018; 8: e1249.
31. Smola A. Support vector machines, regularization, optimization, and beyond. MIT press, Cambridge, MA, USA, 2002.
32. Walters-Williams J and Li Y. Comparative study of distance functions for nearest neighbors. *Advanced Techniques in Computing Sciences and Software Engineering*. Springer, 2010, pp.79-84.
33. Panchal G, Ganatra A, Kosta Y, et al. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering* 2011; 3: 332-337.
34. Breiman L, Friedman J, Stone CJ, et al. *Classification and regression trees*. CRC press, 1984.
35. Babar B, Luppino LT, Boström T, et al. Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy* 2020; 198: 81-92.
36. Friedman JH. Stochastic gradient boosting. *Computational statistics & data analysis* 2002; 38: 367-378.
37. Whitley D. A genetic algorithm tutorial. *Statistics and computing* 1994; 4: 65-85.
38. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural computation* 1997; 9: 1735-1780.
39. Kohavi R. Proceedings of the 14th international joint conference on Artificial intelligence. 1995.
40. Chikkerur S, Sundaram V, Reisslein M, et al. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE transactions on broadcasting* 2011; 57: 165-182.42

41. Lahouar A and Slama JBH. Hour-ahead wind power forecast based on random forests. *Renewable energy* 2017; 109: 529-541.
42. Xiaoyun Q, Xiaoning K, Chao Z, et al. Short-term prediction of wind power based on deep long short-term memory. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)* 2016, pp.1148-1152. IEEE.
43. Bernardo JM and Smith AF. *Bayesian theory*. John Wiley & Sons, 2009.