KSME

Springer

Journal of Mechanical Science and Technology 36 (12) 2022

Original Article
DOI 10.1007/s12206-022-1125-3

Keywords:
· Wind energy
· Wind power forecasting
· Numerical weather prediction
· Machine learning

Correspondence to:
Sung Goon Park
psg@seoultech.ac.kr

Citation:
Kim, J., Afzal, A., Kim, H.-G., Dinh, C. T., Park, S. G. (2022). Wind power forecasting based on hourly wind speed data in South Korea using machine learning algorithms. Journal of Mechanical Science and Technology 36 (12) (2022) 6107~6113.
http://doi.org/10.1007/s12206-022-1125-3

Received May 26th, 2022
Revised August 2nd, 2022
Accepted August 11th, 2022

† Recommended by Editor Han Seo Ko

Wind power forecasting based on hourly wind speed data in South Korea using machine learning algorithms

Jeonghyeon Kim[1], Asif Afzal[1,2], Hyun-Goo Kim[3], Cong Truong Dinh[4] and Sung Goon Park[1]

[1]School of Mechanical Engineering, Seoul National University of Science and Technology, Seoul 01811, Korea, [2]University Centre for Research & Development, Department of Computer Science and Engineering, Chandigarh University, Gharuan, Mohali, Punjab, India, [3]Renewable Resource Map Laboratory, Korea Institute of Energy Research, Daejeon, Korea, [4]Department of Aerospace Engineering, Hanoi University of Science and Technology, Hanoi 100000, Vietnam

Abstract Given that wind farms have high initial investment costs and are not easy to move after installation, the amount of energy that can be produced in the desired installation area needs to be predicted as accurately as possible before installation. Four machine learning algorithms are adopted to predict power production based on the daily wind speed average and standard deviation. The actual power output is calculated from the wind data generated by the numerical weather prediction, and its temporal resolution is 1 hour. The R-square (R^2) values of the models range from 0.97 to 0.98 while adopting the average value of daily wind speed as the input data, and it increases by -1 % with the additional input data of the standard deviation of wind speed. The power production is predicted based on the wind data at a relatively lower height of 10 m than the hub height, where the R^2 value ranges from 0.95 to 0.98. The results could provide the possibility of replacing the wind data measurement process at the hub height by that at a relatively lower height, reducing the cost of wind data measurement.

1. Introduction

Fossil fuels are limited as a sustainable energy resource in the future due to problems such as environmental pollution and depletion of energy sources. To reduce dependence on fossil fuels, many countries and researchers are interested in renewable energy with sustainable characteristics. In December 2017, the Korean government announced the Renewable Energy 3020 implementation plan to replace 20 % of domestic electricity generation with renewable energy by 2030 [1]. According to the Korea Energy Agency, the ratio of renewable energy to total power generation in South Korea in 2020 is 6.41 % [2]. Among renewable energy sources, wind power has many advantages, such as low carbon dioxide emissions [3] and little water use for energy production [4] compared with other energy sources. However, because wind farms have high initial investment costs and are not easy to move after installation, accurately predicting the wind power energy source in the wind farm site selection decision-making process is necessary.

A common method for obtaining wind data is direct measurements at weather stations or wind farms. While the direct measurement of the wind speed shows a high accuracy, it has some limitations such as only locally available information or a high measurement cost. To overcome the limitations, many researchers have tried to obtain wind data by adopting numerical weather prediction (NWP) methods. Al-Yahyai et al. [5] reviewed the use of NWP data for wind energy resource assessment. They analyzed some models such as the mesoscale model (MM5) and weather research and forecasting (WRF), showing that NWP can overcome the temporal and spatial limitations of the direct measurement. MM5 is a model to simulate atmospheric conditions with resolutions ranging from 100 km to 1 km [6]. Jimenez et al. [7] compared MM5 and the wind resource assessment program (WAsP) to estimate the wind resource over the German Bight in the North Sea. The MM5 model showed a high prediction accuracy at a low altitude of 30 and 40 m but underestimated the wind speed at a high altitude of 100 m (about a 4 % difference). The WRF is one of the most actively developing models [8-13]. Done et al. [8] investigated the performance of daily convection forecasts from May to July 2003 using the WRF model in central North America. In Korea, WRF-based models have been used for aeronautical and military purposes [9, 10]. Although the NWP method has been widely adopted in many fields, it requires a high computational cost since it considers a complicated mathematical model to solve problems such as fluid movement, thermodynamics, and solar radiation.

Numerous studies have shown that a machine learning technique can be one of the effective ways for predicting productive energy. Wan et al. [14] proposed a probabilistic wind power forecasting based on the extreme learning machine (ELM). ELM is a learning algorithm proposed for learning single-hidden layer feedforward neural networks. The proposed bootstrap-based ELM approach has been tested using the data of a wind farm in Australia. The proposed method could significantly reduce the computation time compared with the existing method. Zameer et al. [15] presented a support vector regression (SVR)-based hybrid model, and it showed better performance and generalization compared with the existing linear regression model. Wang et al. [16] predicted wind power through a genetic algorithm in SVR using seasonal parameters. The proposed method showed higher accuracy than other benchmark models they referenced. Lahouar et al. [17] used a random forest (RF) algorithm for wind power prediction. They found that the RF algorithm is more user-friendly than the artificial neural networks (ANNs) algorithm and showed higher accuracy without going through the process of optimization. Afterward, researchers developed in-house codes rather than general commercial code and compared and analyzed them with existing algorithms [18-21]. Input data optimization was also attempted to improve prediction accuracy [22, 23]. Hail Demolli [24] predicted wind power for four wind farms in Turkey using five machine learning algorithms and showed high prediction accuracy. The used input data is the hourly wind speed data at the hub height of 50 m predicted using the power law. In most studies, machine learning was conducted using wind speed at the height of the wind turbine hub as the input data and showed high accuracy [14-24]. However, measuring the speed at the hub height has some limitations such as a high measurement cost compared with measuring it at a relatively low altitude.

The objective of the present study is to predict wind energy production by adopting machine learning algorithms, where the wind speed data at an altitude of 10 m was used as the input data instead of that at 80 m. To obtain a high prediction accuracy, several machine learning algorithms were adopted, and input data optimization was attempted. Through this study, the validity of using wind data at a low height to predict wind power production is expected to be demonstrated. The remainder of the paper is as follows. The data pre-processing step is introduced in Sec. 2. The four machine learning algorithms are detailed in Sec. 3. The results and discussion are provided in Sec. 4. Finally, the summaries are presented in Sec. 5.

2. Data pre-processing

2.1 Numerical weather prediction

NWP methods use a mathematical model that differentiates the Earth's atmospheric space and consider fluid movement, thermodynamics, and overall solar radiation. It is a technology to predict future weather based on current weather conditions. Most researchers depend on meteorological data generated using NWP methods to predict wind energy production [5-13, 25-29]. The meteorological data used in this study is KIER-WindMap, produced by the Korean Institute of Energy Research (KIER). The wind resource map was produced in 2010 using WRF based on the NWP method. The meteorological data has 1-hour time resolution and 1-km spatial resolution.

2.2 Weather data processing

Fig. 1 shows the position of 14 wind farms in South Korea. The data in the area denoted in blue are used as test data, and the data obtained in the area denoted in red are used for training machine learning models.

The daily wind speed average (WS), daily wind speed standard deviation (STD), and daily maximum and minimum wind speed (MAX/MIN) are derived from the wind speed data at the

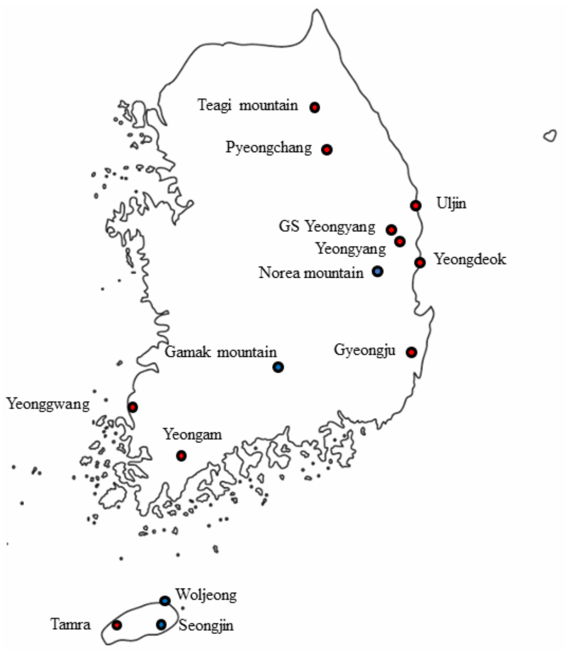

Fig. 1. Position of 14 wind farms in South Korea where data are used for machine learning. The red symbols indicate the training regions and the blue symbols indicate the test regions.

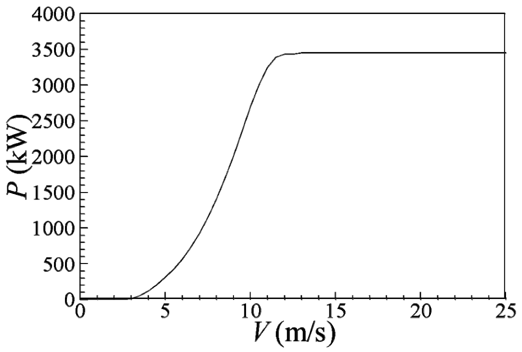

Fig. 2. Power curve as a function of wind speed for 3.45 MW wind turbine.

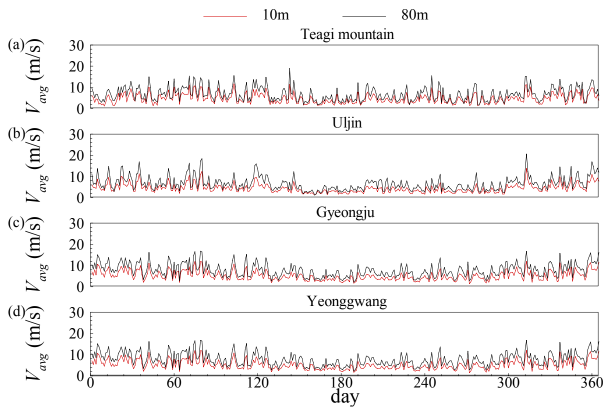

Fig. 3. Daily wind speed average in the areas of (a) Teagi-mountain; (b) Uljin; (c) Gyeonju; (d) Yeonggwang, where the data are used to train the machine learning model. The daily wind speed averages at the heights of 10 and 80 m are denoted by red and black solid lines, respectively.

altitudes of 10 and 80 m of each meteorological data. The derived data are used as input data for machine learning. The total daily power production is calculated by using the hourly wind speed data at the hub height of 80 m obtained from KIER-WindMap and the corresponding power in the 3.45 MW power curve in Fig. 2. Fig. 3 shows the derived WS with a 24-hour time resolution at the two heights of 10 and 80 m, implying that the wind data at the different heights are closely correlated. The daily wind information and the corresponding power in 10 different regions (denoted in red in Fig. 1) are used to train the four machine learning models: artificial neural networks (ANNs), k nearest neighbor (kNN), random forest (RF), and support vector regression (SVR). The daily power production in four regions (denoted in blue in Fig. 1) are predicted by using the daily wind data.

3. Machine learning algorithms

Machine learning is a field of computer science that learns from training data and predicts results [30, 31]. Machine learning is primarily used when the relationship between input and output data is unclear. Many researchers have used machine learning regression algorithms to predict wind resources [14-

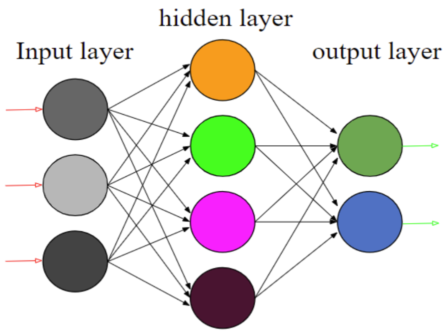

Fig. 4. Schematic diagram of the artificial neurons in each layer.

25]. Machine learning algorithms are grouped into several types, each of which depends on the particular problem. In this study, four machine learning algorithms, namely, ANNs, kNN, RF, and SVR are used. A brief description of each algorithm is given below.

3.1 Artificial neural networks

ANNs are biology-inspired machine learning algorithms. ANNs are designed to simulate how the human brain processes information [32]. ANN detects relationships with patterns in data, collects information, and learns through experience. Many types of neural networks have already been designed and new neural networks are being generated. For all neural networks, the transfer function can be explained by the connection formula with the learning rules. Fig. 4 shows an artificial neuron that receives a signal processing the signal and sending it to the connected neuron. The nodes in each round indicate artificial neurons, and the arrows indicate inputs from one neuron to another. The signal at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. In this study, ReLu is used as the activation function and Adam is used as the optimizer.

3.2 k Nearest neighbor

kNN is an instance-based lazy learning classification algorithm [33]. Machine learning is driven by closeness based on distance measurement. In the classification stage, k is a user-defined constant, and vectors without a classification name are most often assigned a classification name relative to k. Fig. 5 shows an example of kNN classification. Test vectors in the form of circles should be classified as squares or triangles. The test vector for $k = 2$ is classified as a square. However, if $k = 5$, the test vector is classified as a triangle. Choosing the optimal value for k depends on the nature of the data. In general, the larger the value of k, the less the classification noise, but the more unclear the boundaries between the items. In this study, the value of k is 4.

3.3 Random forest

RF is a decision tree algorithm that constructs multiple deci-

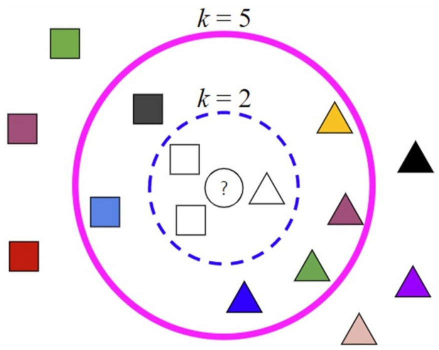

Fig. 5. Example of k-NN classification. The test sample should be classified as either squares or triangles.

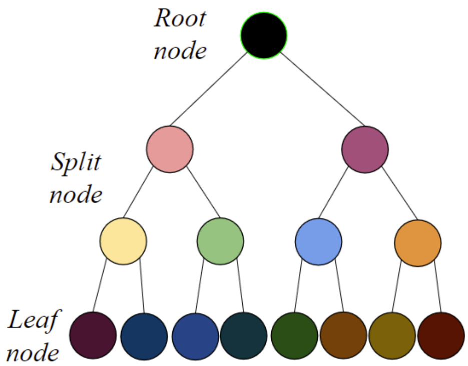

Fig. 6. Diagram of a tree in RF.

sion trees from an input data set. RF divides the input parameters of the dataset into parts to form a decision tree for each part of the function and uses the results of each decision tree to make the final decision [34]. RF consists of nodes and edges in the tree hierarchy. A node has one receiving end and generally has two edges. Decision trees divide the decision process into a simple problems hierarchy. The most influential parameters of RF are the number of trees and the maximum depth. As the number of trees increases, the results are relatively continuous and the generalization ability is excellent, but the execution time increases.

The maximum depth determines the maximum number of nodes to traverse from the root node to the end node in a tree. Setting an appropriate value is important because underfitting occurs if the maximum depth is small and overfitting occurs if the maximum depth is too large. In this study, the number of trees and the maximum depth are determined to be 10 and 8, respectively.

3.4 Support vector regression

SVR is a regression version of the support vector machines (SVM) algorithm [35, 36]. SVM is a map learning model for pattern recognition and data analysis. Given a particular dataset, the SVM algorithm generates a classification model that determines the category to which the new data belongs, based on the given dataset. The generated classification model is

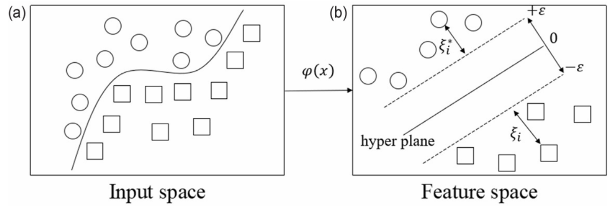

Fig. 7. Transformation process illustration of an SVR model.

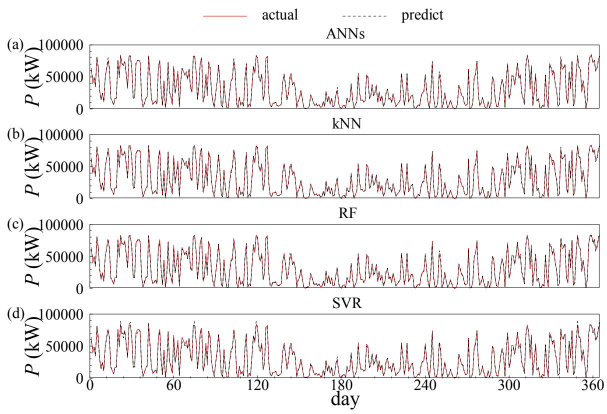

Fig. 8. Total daily power production in Norea-mountain. The red line is the actual value and the dotted line is the predicted value for (a) ANNs; (b) kNN; (c) RF; (d) SVR when the input data are the daily average wind speed and the daily wind speed standard deviation at the hub height of 80 m.

represented by the boundaries of the space to which the data is mapped, and the SVM algorithm is applied to find the boundaries of the maximum width. Non-linear SVR attempts to find a regression function in the input hyperplane. The parameters that affect the result significantly are the type of function, the gamma, and the value of C. The type of function should be carefully selected according to the characteristics of the input data. Gamma and C values affect model overfitting and underfitting. In this study, a radial basis function is used as the function type. The constant values are C = 3000 and gamma = 0.1.

4. Prediction of power production

4.1 Result analysis: input data at the hub height of 80 m

Using the training data of 10 wind farms, the daily power production of four wind farms including Norea-mountain is predicted using four machine learning methods.

The power curve as a non-linear function of wind speed in Fig. 2 implies that the total daily power production is closely related to the input data sets of WS and STD. Fig. 8 shows the total daily power production (red solid lines) in the Norea-mountain region and the predicted values by using the four machine learning algorithms (black dotted). Here, WS and STD at the hub height of 80 m are used as the input data. Fig. 9 represents the actual and predicted power production on a daily basis (totaling 365 data points), implying that the daily

Table 1. R-square (R^2) value of for each machine learning algorithms. The input wind data are obtained at the hub height of 80 m.

		Norea mt.	Gamak mt.	Waljung	Sungjin
80 m WS	ANNs	0.979769	0.979853	0.974831	0.972872
	kNN	0.975430	0.979182	0.970582	0.973228
	RF	0.978898	0.980480	0.974252	0.973211
	SVR	0.978396	0.979332	0.974807	0.971446
80 m WS+ STD	ANNs	0.996541	0.997346	0.996107	0.995687
	kNN	0.996973	0.996513	0.994508	0.994472
	RF	0.994449	0.994607	0.992689	0.992492
	SVR	0.994411	0.996167	0.992812	0.991569

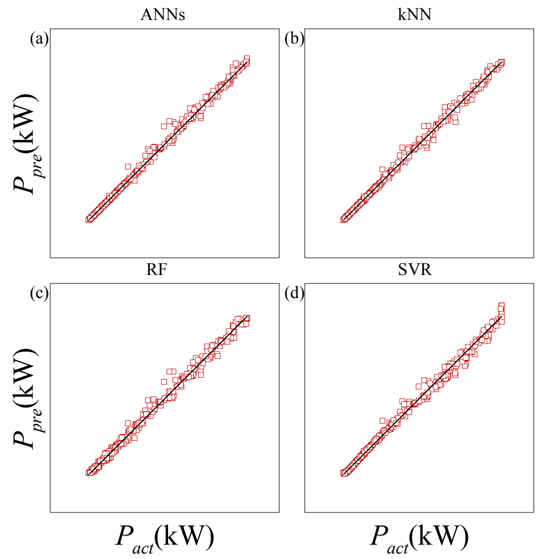

Fig. 9. Comparisons of actual and predicted power production for (a) ANNs; (b) kNN; (c) RF; (d) SVR. The input data are the daily average wind speed and the daily wind speed standard deviation at the hub height of 80 m.

power production can be predicted with high accuracy based on the daily wind information such as WS and STD.

Table 1 shows the R-square (R^2) values for the four machine learning algorithms by using the wind data at the hub height of 80 m. The R^2 values of the machine learning models established by only using the input data of WS observed to be 0.97-0.98, and the best performance appears at Gamak-mountain when adopting RF. The accuracy of the daily power prediction is improved by 1 %-2 % with the input data of WS together with STD data as shown in Table 1, where the R^2 values are observed to be above 0.99 for all the cases. This implies that the wind speed average and standard deviation values at the hub height on a daily basis provide excellent prediction for the daily wind power prediction. The data of WS and STD could be regarded as a useful indicator to estimate the wind power production in the decision-making process of the wind farm site selection.

4.2 Result analysis: input data at 10 m height

To accurately predict daily power production, measuring and

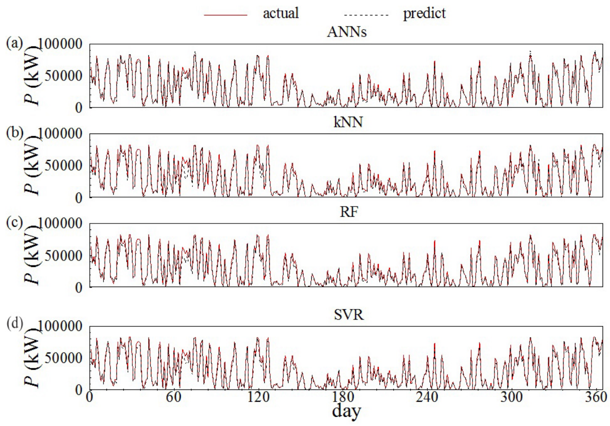

Fig. 10. Total daily power production in Norea-mountain region. The red line is the actual value and the dotted line is the predicted value by using: (a) ANNs; (b) kNN; (c) RF; (d) SVR. The input data are the daily average wind speed, the daily wind speed standard deviation and the daily maximum and minimum wind speed at the height of 10 m.

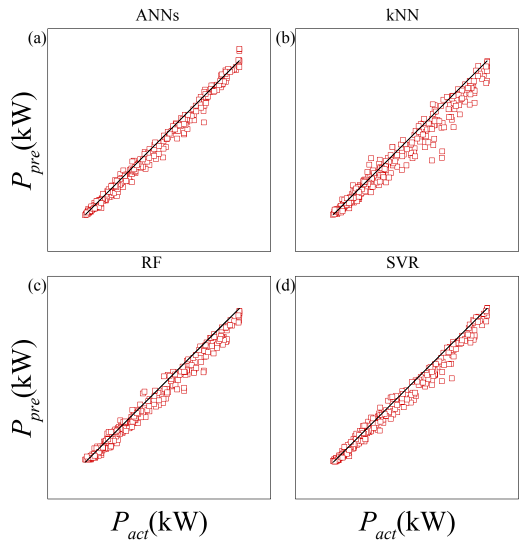

Fig. 11. Comparisons of actual and predicted power production by (a) ANNs; (b) kNN; (c) RF; (d) SVR. The input data are the daily average wind speed, the daily wind speed standard deviation and the daily maximum and minimum wind speed at the height of 10 m.

predicting wind speed data at the hub height of a wind turbine is recommended. However, the speed measurement at a high altitude requires the installation of a tall meteorological tower. The prediction of the wind speed at the hub height by the wind data at a relatively lower height such as about 10 m could help reduce the cost spent in the wind farm economic evaluation process. Fig. 3 implies that the daily wind speed averages at the different heights of 80 and 10 m are closely correlated providing the possibility of wind power prediction at a specific site with the data at the height of 10 m.

The four machine learning algorithms are adopted to predict the daily power production based on the daily WS, STD, MAX/MIN at the height of 10 m. Fig. 10 shows the time history of daily power production in the Norea-mountain region, where

Table 2. R-square (R^2) value of test data for the four different machine learning algorithms. The data at the height of 10 m are used as input data.

	Norea mt.	Gamak mt.	Waljung	Sungjin
10 m WS				
ANNs	0.973067	0.953389	0.931220	0.950386
kNN	0.946871	0.927262	0.921251	0.927302
RF	0.961752	0.943323	0.932798	0.936326
SVR	0.961689	0.945665	0.937055	0.934966
10 m WS+STD				
ANNs	0.987427	0.963027	0.942307	0.973736
kNN	0.967744	0.947589	0.937344	0.949083
RF	0.974506	0.954589	0.945940	0.957167
SVR	0.973973	0.956774	0.950852	0.953709
10 m WS+STD +MAX/MIN				
ANNs	0.987761	0.963469	0.944127	0.975588
kNN	0.957305	0.944367	0.945292	0.952516
RF	0.974790	0.953149	0.950136	0.958807
SVR	0.978299	0.958671	0.951773	0.963051

the red and black dotted lines indicate the time histories of power production based on the wind data at the heights of 10 m, respectively. Fig. 11 shows the comparison of the actual/pre- dicted power production, and the four algorithms show good prediction performance. Table 2 shows the R^2 value for the four algorithms at the four different regions. When adopting only WS as input data, the R^2 value ranges from 0.93 to 0.97. The ANNs algorithm shows the best performance among the four algorithms, except for the region of Waljung, where the SVR algorithm shows the best prediction performance. Waljung is located near the sea area and has a low surface elevation compared with other regions. The R^2 value of Waljung is lower than that of other regions seemingly because the present machine learning models do not capture geographical features. The prediction accuracy is observed to be improved (-1 %) by adopting the data of STD together with WS as input data, as shown in Table 2. The daily maximum and minimum values of wind data are additionally considered the input data as well as WS and STD, where the accuracy is not significantly improved compared with the WS+STD cases in Table 2. Fig. 2 shows the power curve of a 3.45 MW wind turbine, where the cut-in and cut-out speed are set to be 3 and 12 m/s, respectively. The prediction of the power production corresponding to the wind data below the cut-in speed or above the cut-out speed could be adjusted by additionally considering the daily maximum and minimum wind speed, improving the accuracy of the prediction (although its magnitude is not significant).

5. Conclusion

Four machine learning algorithms, namely, ANN, kNN, RF, and SVR, were used to predict the daily wind power production based on the daily average (WS) and standard deviation (STD) values of wind speed. The meteorological data used in the present study were obtained from the numerical weather prediction method, and their time resolution was 1 hour. The actual daily power productions were obtained from the 1-hour wind speed data at the height of 80 m and predicted by using the four machine learning algorithms. The R^2 values ranged from 0.97 and 0.98 when adopting WS at the height of 80 m as input data. By additionally considering STD together with WS as input data, the accuracy of the prediction was improved by -1 %, and the ANNs algorithm showed the best performance, where the R^2 values reached above 0.99. This implies that the WS and STD are good indicators to predict the daily wind power production at a specific location. The prediction of the power production based on the wind data at a relatively lower height than the hub height can reduce the cost spent in the evaluation process of wind farm site selection. The power production was predicted based on the wind data at the height of 10 m, where the R^2 value ranged from 0.95 and 0.98 by adopting WS and STD as input data. The results could provide the possibility of replacing the wind data measurement process at the hub height by that at a relatively lower height, reducing the cost of wind data measurement.

Acknowledgments

This work was supported by the Research Program funded by the SeoulTech (Seoul National University of Science and Technology).

References

[1] Ministry of Trace, Industry and Energy, *Renewable Energy 3020 Implentation Plan*, http://www.motie.go. kr/motiee/.

[2] C. S. Kim, *New and Renewable Energy Statistics 2020*, 2021 Edition, KOREA Energy Agency (2021).

[3] M. Inman, The true cost of fossil fuels, *Scientific American*, 308 (4) (2013) 58-61.

[4] D. Weisser, A guide to life-cycle greenhouse gas (GHG) emissions from electric supply technologies, *Energy*, 32 (9) (2007) 1543-1559.

[5] S. Al-Yahyai, Y. Charabi and A. Gastli, Review of the use of numerical weather prediction (NWP) models for wind energy assessment, *Renewable and Sustainable Energy Reviews,* 14.9 (2010) 3192-3198.

[6] G. A. Grell, J. Dudhia and D. Stauffer, *A Description of the Fifth-generation Penn State/NCAR Mesoscale Model (MM5),* No. NCAR/TN-398+STR, University Corporation for Atmospheric Research (1994) doi:10.5065/D60Z716B.

[7] B. Jimenez, F. Durante, B. Lange, T. Kreutzer and J. Tambke, Offshore wind resource assessment with WAsP and MM5: Comparative study for the German Bight, *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 10 (2) (2007) 121-134.

[8] J. Done, C. A. Davis and M. Weisman, The next generation of NWP: explicit forecasts of convection using the weather re-

search and forecasting (WRF) model, *Atmospheric Science Letters*, 5 (6) (2004) 110-117.

[9] C.-H. Bang, J.-W. Lee and S.-Y. Hong, Predictability experiments of fog and visibility in local airports over Korea using the WRF model, *Journal of Korean Society for Atmospheric Environment*, 24 (E2) (2008) 92-101.

[10] U. Y. Byun, S. Y. Hong, H. Shin, J. W. Lee, J. I. Song, S. J. Hahm, J. K. Kim, H. W. Kim and J. S. Kim, WRF-based short-range forecast system of the Korea Air Force: Verification of prediction skill in 2009 summer, *Atmosphere*, 21 (2) (2011) 197-208.

[11] J. Brioude, D. Arnold, A. Stohl, M. Cassiani, D. Morton, P. Seibert, W. Angevine, S. Evan, A. Dingwel, J. D. Fast, R. C. Easter, I. Pisso, J. Burkhart and G. Wotawa, The lagrangian particle dispersion model FLEXPART-WRF version 3.1, *Geoscientific Model Development*, 6 (6) (2013) 1889-1904.

[12] A. J. Litta, S. M. Ididcula, U. C. Mohanty and S. Kiran Prasad, Comparison of thunderstorm simulations from WRF-NMM and WRF-ARW models over east indian region, *The Scientific World Journal* (2012).

[13] D. C. Wong, J. Pleim, R. Mathur, F. Binkowski, T. Otte, R. Gilliam, G. Pouliot, A. Xiu, J. O. Young and D. Kang, WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results, *Geoscientific Model Development*, 5 (2) (2012) 299-312.

[14] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong and K. P. Wong, Probabilistic forecasting of wind power generation using extreme learning machine, *IEEE Transactions on Power Systems*, 29 (3) (2013) 1033-1044.

[15] A. Zameer, A. Khan and S. G. Javed, Machine Learning based short term wind power prediction using a hybrid learning model, *Computers and Electrical Engineering*, 45 (2015) 122-133.

[16] J. Wang and Y. Li, Short-term wind speed prediction using signal preprocessing technique and evolutionary support vector regression, *Neural Processing Letters*, 48 (2) (2018) 1043-1061.

[17] A. Lahouar and J. Ben Hadj Slama, Hour-ahead wind power forecast based on random forests, *Renewable Energy*, 109 (2017) 529-541.

[18] H. Liu, C. Chen, X. Lv, X. Wu and M. Liu, Deterministic wind energy forecasting, a review of intelligent predictors and auxiliary methods, *Energy Conversion and Management*, 195 (2019) 328-345.

[19] Jørgensen, K. Lau and H. R. Shaker, Wind power forecasting using machine learning: state of the art, trends and challenges, *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*, IEEE (2020) 44-50.

[20] M. Neshat, M. M. Nezhad, E. Abbasnejad, S. Mirjalili, L. B. Tjernberg, D. A. Garcia, B. Alexander and M. Wagner, A deep learning-based evolutionary model for short-term wind speed forecasting: a case study of the Lillgrund offshore wind farm, *Energy Conversion and Management*, 236 (2021) 114002.

[21] J. Heinermann and O. Kramer, Machine learning ensembles for wind power prediction, *Renewable Energy*, 89 (2016) 671-679.

[22] A. U. Haque, P. Mandal, J. Meng and M. Negnevitsky, Wind speed forecast model for wind farm based on a hybrid machine learning algorithm, *International Journal of Sustainable Energy*, 34 (1) (2015) 38-51.

[23] M. Optis and J. Perr-Sauer, The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production, *Renewable and Sustainable Energy Reviews*, 112 (2019) 27-41.

[24] H. Demolli, A. S. Dokuz, A. Ecemis and M. Gokcek, Wind power forecasting based on daily wind speed data using machine learning algorithms, *Energy Conversion and Management*, 198 (2019) 111823.

[25] P. Piotrowski, D. Baczyński, M. Kopyt, K. Szafranek, P. Helt and T. Gulczyński, Analysis of forecasted meteorological data (NWP) for efficient spatial forecasting of wind power generation, *Electric Power Systems Research*, 175 (2019) 105891.

[26] D. Kim and J. Hur, Short-term probabilistic forecasting of wind energy resources using the enhanced ensemble method, *Energy*, 157 (2018) 211-226.

[27] W. Dong, Q. Yang and X. Fang, Multi-step ahead wind power generation prediction based on hybrid machine learning techniques, *Energies*, 11 (8) (2018) 1975.

[28] P. Piotrowski, D. Baczyński, M. Kopyt, K. Szafranek, P. Helt and T. Gulczyński, Analysis of forecasted meteorological data (NWP) for efficient spatial forecasting of wind power generation, *Electric Power Systems Research*, 175 (2019) 105891.

[29] G. Alkhayat and R. Mehmood, A review and taxonomy of wind and solar energy forecasting methods based on deep learning, *Energy and AI* (2021) 100060.

[30] A. L. Samuel, Some studies in machine learning using the game of checkers, II—Recent progress, *IBM Journal of Research and Development*, 11 (6) (1967) 601-617.

[31] Alpaydin, Ethem, *Introduction to Machine Learning*, MIT Press (2020).

[32] S.-C. Wang, Artificial neural network, *Interdisciplinary Computing in Java Programming*, Springer, Boston, MA (2003) 81-100.

[33] Z. Yao and W. L. Ruzzo, A regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data, *BMC Bioinformatics*, 7 (1) (2006).

[34] L. Breiman, Random forests, *Machine Learning*, 45 (1) (2001) 5-32.

[35] H. Drucker, C. J. Burges, L. Kaufman, A. Smola and V. Vapnik, Support vector regression machines, *Advances in Neural Information Processing Systems*, 9 (1997) 155-161.

[36] W. C. Hong, Y. Dong, L. Y. Chen and S. Y. Wei, SVR with hybrid chaotic genetic algorithms for tourism demand forecasting, *Applied Soft Computing*, 11 (2) (2011) 1881-1890.

Sung Goon Park received his Ph.D. in Mechanical Engineering from the Korea Advanced Institute of Science and Technology (KAIST). He is currently an Assistant Professor at the Seoul National University of Science and Technology, Korea. His research interests include computational simulations of fluid-structure interactions and energy systems.