Cross-location wind speed forecasting for wind energy applications using machine learning based models

Valsaraj Perumpalot\textsuperscript{a,b}, G. V. Drisya\textsuperscript{a}, K. Satheesh Kumar\textsuperscript{a,*}

\textsuperscript{a}Department of Futures Studies, University of Kerala, Kariavattom, Kerala, India - 695 581
\textsuperscript{b}Agency for Non-conventional Energy and Rural Technology (ANERT), Thiruvananthapuram, Kerala, India - 695 033

Abstract

The widespread utilisation of grid-integrated wind electricity necessitates accurate and reliable wind speed forecasting to ensure stable grid and quality power. Machine learning algorithm based wind speed forecasting models are getting increased attention in the literature owing to its superior ability to learn by effectively capturing the changing patterns from the data. Most of the reported wind forecasting models built on machine learning algorithms are location specific and tested against data adjacent to the training data. In this work, we develop the machine learning based wind speed forecasting models and analyse their performance when applied to data from different cross-locations up to a year ahead. Two distinct machine learning models based on Support Vector Machine (SVM) and Random Forest (RF) algorithms have been developed and tested separately for a relatively large geographical area. The results of analysis of 1-hour forecasts obtained at various cross-locations and points of time up to a year ahead show 80\% of predictions within a Root Mean Square Error (RMSE) of 1.5 m/s, 95\% within 2.5 m/s and 98\% within an RMSE of 3.5 m/s. The 75\% of 2-hour predictions are within RMSE of 1.5 m/s, 16-hour predictions within RMSE of 2.5 m/s and 48-hour predictions within RMSE of 3.5 m/s. When applied to the same location of training data, the models generate reliable forecasts for periods up to 22 hours, with the added advantage that the models perform consistently throughout the year ahead horizon, independent of the lead time from the training data. The output of the analysis is highly promising to the wind energy industry in wind forecasting for locations where historical wind speed data are not available for model building and training.

Keywords: wind energy, wind speed forecasting, cross-location forecasting, machine learning, support vector machine, random forest, mutual information, RMSE.

1. INTRODUCTION

As the energy demand multiplies the world over and the conventional energy resources deplete alarmingly, wind energy utilisation has attained greater importance from the perspective of sustainable development, as it is renew-
Cross-location wind speed forecasting

able, clean and comparatively cost effective. As a result, wind energy has already emerged as an essential constituent in the global energy mix and is all set to grow to its maximum potential across countries and regions in the years to come. As wind power is directly proportional to the cube of wind speed, even slightest variations in wind speed will significantly affect the power output from wind energy generators. Since wind is a fluctuating resource in respect of availability and speed, precise wind forecasting becomes essential, especially when wind power penetration grows, for the effective management of electricity grid to ensure quality power supply. Accurate wind forecasts on different lead time scales help wind farms in real-time grid operations, economic load dispatch planning, reserve requirement decisions, market trading, maintenance planning and the like.

Wind forecasting continues to be an area of high research interest owing to its practical relevance in the ever-expanding wind energy industry. Costa et al.¹ have reviewed the research in short-term wind prediction over 30 years, giving attention to forecasting methods, mathematical, statistical and physical models, as well as meteorology. Foley et al.² and Okumus et al.³ have conducted an extensive review of the current methods and improvements in the field of wind power forecasting. Based on the methodology adopted, wind forecasting models are grouped mainly into physical, statistical, data learning and hybrid models. The physical models, which utilize different atmospheric parameters are useful for identifying recurring patterns and making long-term predictions. Statistical models assume that the wind speed fluctuations are stochastic. However, it has recently been demonstrated that the underlying dynamics of apparent random-like fluctuations of wind speed measurements is deterministic, low-dimensional and chaotic.⁴,⁵,⁶ Hybrid models have been developed recently by combining different methods such as physical, statistical and machine learning methods to enhance prediction accuracy.⁷,⁸ Models based on artificial neural network and other data learning techniques have also been gaining increased attention in literature in the recent past. Mohandes et al.⁹ have compared a support vector regression (SVR) approach for wind speed prediction favourably against a multi-layer perceptron (MLP) for systems with orders 1 to 11. Liu et al.¹⁰ attempted short-term wind speed forecasting using wavelet transform and Support Vector Machines, applying a genetic algorithm for parameter optimization. Fugon et al.¹¹ applied linear and non-linear data mining algorithms for the short-term wind power forecasting at three distinct wind farm locations in France. Lahouar et al.¹² tried out Random Forest model for an hour ahead wind power prediction and tested the model with measured wind data and showed good improvement of forecast accuracy compared to classical neural network prediction. Mohandes et al.¹³ conducted a study within Saudi Arabia to estimate the mean monthly wind speed at certain locations using the historic mean monthly wind speed data from a number of other locations and reported good agreement between estimated and measured monthly mean wind speed values. Browell et al.¹⁴ attempted very short-term wind forecasting by incorporating large-scale meteorological information into a vector autoregressive model and showed improved accuracies in forecasting in different case studies conducted in the United Kingdom.

Various machine learning algorithms have been tested successfully to predict wind speed variations. However, almost all reported studies are location specific as training and testing data are sourced from the same location. Apart from that, application of such models
have not been tested for time independence as
the test data considered are adjacent in time
to the training data. In this work, we investi-
gate the possibility of developing time and lo-
cation independent models based on machine
learning algorithms for wind speed forecast-
ing for the wind energy industry. Such mod-
els hold practical value for wind energy indus-
try for locations where sufficient past data are
not available for model training. The first ob-
jective of this study is to investigate the accu-
racy of models applied to data moving away
from the training data set. The second objec-
tive is to analyse the accuracy of cross-location
predictions in which models trained using data
from one location are tested against data from
another location. Proper training of the fore-
cast model with the available data is a cru-
cial factor for its performance. In the case
of wind speed time series, it raises the ques-
tion of deciding the optimum size of past data
required to train the model for producing ac-
curate time ahead predictions, as the duration
of wind speed time series is significant since
the fluctuation characteristics vary remark-
ably over different seasons. The theory of mutual
information has been applied to estimate the
optimum size of preceding data set to be ap-
plied in the training of models and testing of
forecasts. The study reported here investigates
the efficacy of two uniquely trained and tested
machine learning models based on (i) Support
Vector Machine (SVM) and (ii) Random For-
est (RF) algorithms for wind speed forecasting
in the same-location as well as cross-location
scenarios, wherein promising results are ob-
tained.

2. MACHINE LEARNING MODELS

Machine learning is a method of improving
the performance of a computational software
programme by itself with experience. In ma-
chine learning, past experience is fed to the
machine as input, and it gives the output as
a typical model that is capable of solving fu-
ture problems of the same nature. Here, past
experience is collected for the purpose of im-
parting training. An abstract target function
is determined that well describes the relation-
ship between existing input and desired output.
Subsequently, a machine learning model is se-
lected to approximate the target function. In
the end, a suitable algorithm is used to build
the model from the training examples. In this
paper, two different machine learning models,
namely, Support Vector Machine and Random
Forest models have been used for the inves-
tigations. The e1071 package in R is used in
this research to develop, train and test the mod-
els.15

2.1. Support Vector Machine model

Support Vector Machine (SVM) is a su-
pervised learning method derived from Vap-
nik’s work on statistical learning theory which
was initially used for classification problems
and later generalized for regression.16,17 It
is an optimization technique of finding a sur-
face which maximizes the margin between two
classes based on two main ideas namely, the
maximization of distance between the classi-
yfying surface and the nearest elements called
support vector and the transformation of the
input dimension into higher dimensional space

| Location | Latitude    | Longitude    |
|----------|-------------|--------------|
| L1       | 09°45′30.2″  | 77°10′41.3″  |
| L2       | 09°59′09.5″  | 77°11′50.0″  |
| L3       | 12°01′32.7″  | 75°20′32.4″  |
| L4       | 09°39′11.5″  | 76°53′3.0″   |
| L5       | 10°48′57.7″  | 76°40′10.3″  |

Table 1: Geographic coordinates of the wind masts used for wind data collection.
Cross-location wind speed forecasting

Figure 1: Geographical locations of the wind measuring masts indicated on 3D map (Courtesy: Google Maps).

using a kernel function. SVM method performs classification tasks by constructing hyperplanes in a multidimensional space that separate cases of different class labels. To construct an optimal hyperplane, SVM employs an iterative training algorithm which is used to minimize an error function. In the case of a linear classification problem, the hyperplane can be expressed as

\[ y_i(w \times x_i + b) \geq 1 - \psi_i \]

where \( x_i, y_i \in \mathbb{R}^n \) are the training data pairs and \( w \) the coefficient vector of classification hyperplane, \( b \) the offset of the hyperplane from the origin and \( \psi_i \) are the positive slack variables. The optimum hyperplane is obtained by solving the minimization problem

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i=1}^{n} \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} n \alpha_i \alpha_j y_i y_j (x_i, x_j) \\
\text{subject to} & \quad \sum_{i=1}^{n} n \alpha_i y_j = 0 \text{ and } 0 \leq \alpha_i \leq C
\end{align*}
\]

where \( \alpha_i \) are Lagrange multipliers and \( C \) the penalty. RBF kernel has been used in our present analysis. The SVM can also be used for regression without sacrificing its main features and it is resistant to overfitting.

2.2. Random Forest model

The Random Forest (RF) algorithm is a non-parametric ensemble based learning technique used for both classification and regression. The decision tree algorithm works on a set of rules and the possible outcomes to form a tree-like structure. However, such a system is prone to error propagation contributed by an incorrect rule which adds the impurity to the subsequent nodes. Random Forest algorithm eliminates error diffusion process inherent in decision trees by constructing multiple trees. Random samples of given data set are generated and fed to several tree-based learners to form a random forest. Splitting condition for each node in a tree is based only on the randomly selected predictor attributes which lower the error rate by avoiding the correlation among the trees. The successful application of random forest regression algorithm has already been reported in many fields like chemoinformatics, speech recognition, bioinformatics, classification and prediction in ecology. The random forest regression, which is a non-parametric, captures the functional relationship between dependent and independent variables from the features of the data. From a given data set, algorithm generates a forest of \( n \) trees as \( \{T_1(X), T_2(X), \ldots, T_n(X)\} \), using an \( m \) dimensional vector input \( X = (x_1, x_2, \ldots, x_m) \).
Cross-location wind speed forecasting

Figure 2: The mutual information of hourly wind speed time series as function of delay.

Every tree $T_i(X)$ generates an outcome $W_i = T_i(X)$. The average of all individual outputs is considered as the response of random forest. The bagging process of selection with replacement is done both on the samples and the attribute. Normally two third of the data will be the size of bootstrap samples, and the rest is known as out-of-bag samples. The combined effect of bootstrap and attribute bagging helps the algorithm to reduce misclassification error.

3. RESULTS AND DISCUSSIONS

The hourly wind speeds at the height of 80 m above the ground level at five different windy locations situated in the Indian state of Kerala, represented by L1 to L5 as shown in 3D map given in Fig. 1 and summarised in Table 1 measured using wind masts over a continuous period of two years 2012 and 2013 have been utilised for the analysis in this work. These locations are geographically distributed in such a way that the locations L2, L3, L4 and L5 are at radial distances 25 km, 321 km, 34 km and 130 km respectively from the location L1 and the triangular area formed by the locations L1, L2 and L4 comes to 385 square kilometres. Two different machine learning forecast models built on SVM and RF algorithms have been employed for the analysis. The wind speed data for the first year (2012) have been used for training and validating the models, in order to ensure effective learning of the dynamics of wind flow fluctuations over a complete seasonal cycle. The wind speed data for the second year (2013) have been used for the testing of forecast results.

The extent of dependence of a value in a time series on the previous values can be estimated by the calculation of mutual information between delayed time series. Relying on this, the optimum length of dependency of wind speed values on the past data has been determined by computing values of mutual information and the same knowledge made use of in the training of models and testing of forecasts. The mutual information in the present analysis appears to become negligible by 72 data points, after which the plot starts to level-off, as observed in Fig. 2. It is therefore taken that every wind speed data point is a function of its previous 72 data points. Hence, the training of SVM model using the past data has been done for all the wind speed values in the
Cross-location wind speed forecasting

Figure 4: Comparison between the actual and forecast wind speeds with repeated 5 hour ahead predictions for the same-location forecasting for location L1, using SVM model.

In the first stage, same-location predictions (where data from the same location is used in parts for training and testing) have been examined for all the five locations, for all hours ahead predictions from 1 hour up to 48 hours ahead using SVM model. Each of the one step ahead predictions is generated by inputting the immediately preceding 72 data points into the trained model. In 2 and higher hour ahead predictions, one step ahead predictions are repeated, each time by using the lastly predicted value as the last of the preceding 72 input points. The predicted time series segments have been compared with actual values along the one year test period, and the deviations analysed using the statistical measure of Root Mean Square Error (RMSE). In the next step, all possible combinations of cross-location predictions have been experimented, wherein, SVM model trained with data from one location has been employed to generate wind speed forecasts at the other four locations by inputting test data from those four locations respectively. In the subsequent stage, all the above investigations have been repeated by developing and using the machine learning model of RF and the results obtained with both the models have been analysed and compared.

In the typical prediction scenario we input past 72 wind speed data points and obtain one step ahead prediction. Further values are predicted recursively using one step ahead prediction. Fig. 3 shows a typical 3-hour ahead prediction up to 1000 hours ahead of the training data set at location L1 using SVM model. Similar plot of 5-hour ahead prediction is given in Fig. 4. In both these plots, the predictions are seen remarkably close to the original measured data, except for some over predictions at peaks. With the SVM model trained using the data in 2012, we obtained predictions up to 48 hours ahead corresponding to each data point in 2013 by inputting 72 previous values. These predictions were compared with actual data to find the RMSE. Fig. 5 shows the mean RMSE versus prediction time averaged over predictions with respect to each data point in

Figure 5: Mean of RMSEs of predictions versus prediction time, for predictions at the locations L1 to L5 using SVM model trained with data from the respective locations.
Cross-location wind speed forecasting

Figure 6: Comparison between the actual and forecast wind speeds with 3-hour ahead predictions for the cross-location forecasting for location L1, using SVM model trained with data from location L5.

2013 for all the five locations. It may be noted that the RMSE is less than 2 m/s up to 48-hour ahead prediction for locations L1 and L3 whereas the same is almost less than 3 m/s for other locations. However, up to 22-hour ahead, predictions show RMSEs less than 2.5 m/s in all the locations.

As a next step, we trained the model with data of one location of the year 2012 and applied to predict, corresponding to each data point in 2013, up to 48 hours ahead for each other location. Two sample predictions, 3-hour and 5-hour ahead, up to 1000 hours ahead of the time of training data are given in Figs. 6 and 7. When the above investigation is carried out under the cross-location scenario for the five locations, a total of 5 prediction scenarios, each with predictions for four locations using the SVM model trained with the data from the other location is available for analyses. Fig. 8 depicts one such scenario, wherein predictions are generated for the locations L2 to L5 using the SVM model trained with the historic data from the location L1. In this case, it can be seen that the mean RMSE is less than 2.5 m/s for up to 48-hour predictions even in the worst case of location L4. In a similar fashion, we repeated the cross-prediction sequence by selecting one location for modelling and other locations for prediction. In order to have a better understanding of the efficacy of the cross-location predictions, a percentage-wise representation of all predictions throughout the year 2013 for all possible cross-location prediction situations, that show individual RMSE values less than 1.5 m/s, 2.5 m/s and 3.5 m/s are plotted against the prediction time in Fig. 9. This plot is useful in two different ways as it helps to determine (i) the length of time ahead predictions that is achievable for a given accuracy of prediction and (ii) the accuracy levels for a given length of time ahead prediction. For prediction accuracy with RMSE less than 1.5 m/s 75 % of predictions goes up to 2-hour ahead in time. When the level of prediction accuracy with RMSE is less than 2.5 m/s, up to 16-hour ahead predictions constitute up to 75% of the prediction samples. More than 85% of the 48-hour ahead cross-location predictions are with RMSE less than 3.5 m/s. In another way of interpreting the above plot is to say that, for example, if 10 hour ahead prediction is consid-
Cross-location wind speed forecasting

Figure 8: Mean of RMSEs of predictions versus prediction time, for cross-location predictions at the locations L2, L3, L4 and L5 using SVM model trained with data from the location L1.

Figure 9: Percentage of cross-location predictions with RMSEs below 1.5 m/s, 2.5 m/s and 3.5 m/s versus prediction time, when using SVM model.

...ed, around 94% of prediction samples show individual RSME values below 3.5 m/s, 80% of samples show the same value below 2.5 m/s and 45% of samples show the same below 1.5 m/s.

In the final phase of the research, the above investigations have been repeated by employing RF model in place of SVM model. The results show predicted values to match closely with the actual wind time series dynamics almost as seen as in the case of SVM model, both for same-location as well as cross-location forecasting. The typical cross-location predictions presented in Fig. 6 and Fig. 7 are reproduced in Fig. 10 and Fig. 11 respectively with RF model in place of SVM model for the forecasting, which again show comparable results in terms of wind flow dynamics and forecast accuracy.

In Fig. 12 the scenario in Fig. 9 for SVM model is reproduced with the corresponding results generated by the RF model overlaid. As can be seen from Fig. 12 both models show similar behaviour, especially for lower time ahead predictions. If more stringent levels of forecast accuracy is desired, the SVM model shows marginal supremacy over the RF model in the shorter time ahead predictions and vice versa in the longer time ahead predictions.

4. CONCLUSION

In this work, we investigated the prospect of employing machine learning based predictive models for the cross-location prediction of wind speed variations. We analysed wind speed data for a period of two years from 5 locations separated by a minimum of 25 km and a maximum of 321 km by using both Support Vector Machine (SVM) and Random Forest (RF) models. The dependency of wind speed on past data has been assessed using the theory of mutual information and the same estimate has been made use of in the intelligent training of models with proper time delay embedding matrix and also in inputting proper length of past data in the testing of forecasts. The results indicate that the models developed and trained here can be effectively used in wind speed forecasting on the same time series at far away points in relation to the training data with respect to time. This time-independent characteristic is helpful in avoiding the need in
the usual methods for training of models using the immediate past data every time predictions are attempted. The research further proves that both these models, together with the methods of training and testing followed here, can generate reliable and quality cross-location short-term wind speed forecasts to a duration of 16 to 17 hours from a given point in time across a geographical area as wide as in the present case, with not less than 75% of such time ahead prediction samples from all along the testing period showing individual RMSE values below 2.5 m/s. In the case of one hour ahead predictions along the one year test period of all the 20 cross-location prediction scenarios from the 5 locations, almost 95% of such predictions show RMSE values below the same value. The results obtained further show that in cross-location forecasting, the RF model slightly outperforms the SVM model in the longer time ahead predictions when higher levels of accuracies are desired. The promising results obtained in the cross-location forecasting of wind speed point to the possible existence of certain collective characteristics hidden within the surface wind flow dynamics, which deserve to be studied further. From the practical perspective, the research outcome is very promising for supporting the growing wind energy industry, by being able to help develop hardware instruments embedded with trained models for time as well as location independent wind speed forecasting tasks. The cross-location forecast capability makes it possible to predict wind speeds at newly identified locations where sufficient past data are not available for model-building by employing a model trained with historical wind speed data from a different location within that geographical area.

**ACKNOWLEDGEMENT**

The authors acknowledge with gratitude the use of wind data, measured by NIWE (National Institute of Wind Energy under the Ministry of New and Renewable Energy, Government of India) for and on behalf of ANERT (Agency for Non-conventional Energy and Rural Technology under the state government of Kerala in India), in this research work. The authors are also grateful to the campus
computing facility of University of Kerala set up under DST-PURDE programme for providing computational facilities. The authors also state that they have no conflicts of interest to declare.

REFERENCES

References

1 Alexandre Costa, Antonio Crespo, Jorge Navarro, Gil Lizzano, Henrik Madsen, and Everaldo Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744, 2008.

2 Aoife M Foley, Paul G Leahy, Antonino Marvuglia, and Eamon J McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8, 2012.

3 Inci Okumus and Ali Dinler. Current status of wind energy forecasting and a hybrid method for hourly predictions. *Energy Conversion and Management*, 123:362–371, 2016.

4 RC Sreelekshmi, K Asokan, and K Satheesh Kumar. Deterministic nature of the underlying dynamics of surface wind fluctuations. *Annales Geophysicae-Atmospheres Hydrospheres and Space Sciences*, 30(10):1503, 2012.

5 GV Drisya, DC Kiplangat, K Asokan, and K Satheesh Kumar. Deterministic prediction of surface wind speed variations. *Ann. Geophys.*, 32:1415–1425, 2014.

6 GV Drisya, K Asokan, and K Satheesh Kumar. Diverse dynamical characteristics across the frequency spectrum of wind speed fluctuations. *Renewable Energy*, 119:540–550, 2018.

7 Anbo Meng, Jiafei Ge, Hao Yin, and Sizehe Chen. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Conversion and Management*, 114:75–88, 2016.

8 Qinkai Han, Fanman Meng, Tao Hu, and Fulei Chu. Non-parametric hybrid models for wind speed forecasting. *Energy Conversion and Management*, 148:554–568, 2017.

9 Mohammad A Mohandes, Talal O Halawani, Shafiqur Rehman, and Ahmed A Hussain. Support vector machines for wind speed prediction. *Renewable Energy*, 29(6):939–947, 2004.

10 Da Liu, Dongxiao Niu, Hui Wang, and Leilei Fan. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, 62:592–597, 2014.

11 Lionel Fugon, Jérémie Juban, and Georges Kariniotakis. Data mining for wind power forecasting. In *European Wind Energy Conference & Exhibition EUWE 2008*, pages 6–pages. EWEC, 2008.

12 A Lahouar and J Ben Hadj Slama. Hour-ahead wind power forecast based on random forests. *Renewable energy*, 109:529–541, 2017.

13 Mohamed A Mohandes, Shafiqur Rehman, and Syed Masuir Rahman. Spatial estimation of wind speed. *International Journal of Energy Research*, 36(4):545–552, 2012.

14 Browell J., Drew D. R., and Philippopoulos K. Improved very short term spatio temporal wind forecasting using atmospheric regimes. *Wind Energy*, 0(0):1–12, 2017.

15 David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien, 2017. R package version 1.6-8.

16 Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

17 Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

18 Pijush Samui. Slope stability analysis: a support
Cross-location wind speed forecasting

vector machine approach. *Environmental Geology*, 56(2):255, 2008.

19 Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

20 Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of chemical information and computer sciences*, 43(6):1947–1958, 2003.

21 Peng Xu and Frederick Jelinek. Random forests in language modeling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004.

22 Hongying Jiang, Youping Deng, Huann-Sheng Chen, Lin Tao, Quying Sha, Jun Chen, Chung-Jui Tsai, and Shuanglin Zhang. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC bioinformatics*, 5(1):81, 2004.

23 AM Fraser and HL Swinney. Using mutual information to find independent coordinates for strange attractors. *Phys. Rev. A*, 33:1134–1140, 1986.