

# A novel ensemble learning approach to improve wind prediction

Alexander S. Choi

*Bergen County Academies, Hackensack, NJ 07601*

KEYWORDS. Wind forecast, numerical models, ensemble algorithms

BRIEF. Various ensemble algorithms improve wind speed forecasts over existing numerical models.

**ABSTRACT.** Wind farms are one of the most promising methods of renewable energy production. However, the intermittent nature of wind poses a significant challenge to the viability of wide-scale wind farming, underscoring the crucial role of wind forecasting for wind farm operators. Currently, the National Oceanic and Atmospheric Administration (NOAA) uses numerical weather models such as the High-Resolution Rapid Refresh Model (HRRR) and the Rapid Refresh Model (RAP) to generate wind forecasts based on physical equations, but these models often exhibit significant systematic bias. This study aims to reduce bias by combining the HRRR and the RAP into a single forecast model using a simple stacking method based on linear regression and comprehensive bagging and boosting algorithms based on machine learning models. Our findings indicate that these ensemble models, especially bagging and boosting, predicted wind speed more accurately than the existing RAP and HRRR models, as measured by the root mean squared errors (RMSE). We demonstrate that this novel approach of combining existing numerical forecast models with statistical methods generated more accurate wind forecasting over traditional numerical weather models. The findings of this study provide significant benefits to wind farms by enhancing the efficiency and reliability of wind power generation.

## INTRODUCTION.

Wind farms are one of the most promising methods of renewable energy production. However, the intermittent nature of wind poses a significant challenge to the viability of wide-scale wind farming [1]. Consequently, accurate wind forecasting is crucial for a multitude of reasons, including the need for wind farm operators to adjust turbine settings for maximum power output [1], and for power grid dispatchers to rely on power generation forecasts (which primarily rely on wind forecasts) to integrate wind power into different power grids [2].

Multiple models are currently being used to predict wind, differing in the methods they use to generate forecasts [3]. A key category of weather prediction models is the numerical weather model, which solves physics equations that dictate the weather with real-world observation data [4]. Another prominent type of weather prediction model is the statistical model, which uses statistical and machine learning methods to observe patterns in previous wind data and then generate forecasts [5]. However, each type of method has its unique downsides and limitations in wind forecast generation, from the heavy computational resources of numerical weather models to more significant errors in statistical models [3].

Ensemble algorithms, a widely used method in machine learning, combines different models to create a more accurate single model [6]. Ensemble algorithms involve fitting a line of best fit onto a data set using a decision tree and then improving on these original decision trees in one way or another. In a nutshell, a decision tree splits the original set of independent data into smaller subsets (the branches of the tree) based on minimizing the variance of these smaller subsets - in essence, classifying the data into smaller groups with similar properties; the decision tree then repeats this branching until a stopping condition (e.g., depth of the tree, number of branches, etc.) is reached. After these smallest sets, called leaf nodes, are reached, corresponding

dependent variables and independent variables are associated - thus associating different groups of independent variables with a correlated output value (i.e., “training” the models) [6].

The ensemble algorithms we employ in this study are the “stacking,” “bagging,” and “boosting” algorithms. The stacking algorithm initially trains multiple models side by side on randomized training data; afterward, the algorithm combines the predictions of the different models using another machine learning algorithm to produce a final prediction [7]. Similarly to the stacking algorithm, the bagging algorithm involves taking random samples of the training data and training multiple random tree machine learning models with each unique sample set; however, instead of creating a new machine learning model for the purpose of aggregating the multiple models created earlier, it simply outputs the average of the different learning models [7]. The boosting algorithm, in contrast, trains a single model on an original training dataset multiple times, which it then tweaks to focus on errors from previous iterations, ultimately halting when a certain amount of predetermined model iterations is reached [7].

This study introduces a novel wind prediction model that combines two numerical models, the High-Resolution Rapid Refresh Model (HRRR) and the Rapid Refresh Model (RAP) for wind prediction using various ensemble algorithms. This new model has the potential to significantly improve wind prediction accuracy and reduce errors, thereby enhancing the efficiency and reliability of wind power generation in the renewable energy sector.

## MATERIALS AND METHODS.

We first obtain wind observation data as well as weather forecast data from the Global Systems Laboratory Model Analysis Tools Suite (MATS) website (<https://gsl.noaa.gov/mats/>), which compiles and displays data collected by the National Oceanic and Atmospheric Administration (NOAA).

From this website, we select data points at 12 AM each day from January 1, 2024, to July 1, 2024. We further specify data points by using data from the entire domain of HRRR grid points instead of a single station to get complete datasets. This setting averages all wind speed observation data from all observation stations in the HRRR prediction domain. Subsequently, we subject the data to a three-hour average around 12 AM to reduce data variability and took model forecasts that were produced 6 hours in advance.

After a data cleaning process, we are left with three datasets of 168 data observations each - one dataset containing wind observations from multiple weather stations and two datasets containing the RAP and HRRR forecasts from the same stations on the same days.

We then split the dataset into training and testing subsets, using 134 data points (80%) for training and the remaining 34 (20%) for testing. The choice of an 80/20 split is common practice in machine learning literature and is intended to balance model training with the ability to test the model’s generalization capabilities on unseen data. It is crucial to note that while the models contain observations from 134 days as training data, these individual models produced from this training require all available data before the output data to produce an output on that date due to machine learning model constraints. For example, though the boosting algorithm trains its base model with only the ob-

servation data from January 2, 2024, to May 22, 2024 (first 134 training data points), it would need all the wind forecast data from January 2, 2024, to June 16, 2024, to output a wind speed forecast for June 16, 2024. After splitting the datasets, each training dataset is used as input for each ensemble model to produce new forecasts for the corresponding date in the testing dataset.

Finally, after each ensemble model produced forecasts, the accuracy of the ensemble models and the original numerical models (HRRR and RAP) is evaluated by calculating the root mean squared error (RMSE) of each testing dataset relative to the observation dataset.

This process is repeated for datasets forecasting 12 and 24 hours in advance. Each dataset is slightly different in size, with the 6, 12, and 24-hour ahead forecast datasets being 168, 167, and 163 data points, respectively. Still, we take the first 80% of the data points as the training dataset and the remaining 20% as the test dataset.

## RESULTS.

Three ensemble algorithms — stacking, bagging, boosting — are adopted to the wind dataset after splitting the dataset into two subsets: the training set, used to train the models, and the testing set, used to evaluate the models' performance. In a time series plot of observed wind speed vs. forecasted wind speeds, the HRRR, RAP, and machine learning models demonstrate that all ensemble models outperform the HRRR model, and the bagging and boosting models also outperform the RAP model in all forecast horizons. This finding is most evident in the results for the 24-hour forecast horizon even on visual inspection (Figure 1), but also holds true in the results for the 12-hour forecast horizon (Figure 2) and the 6-hour forecast horizon (Figure 3).

Our evaluation metric to quantify the accuracy of the models is the root mean squared errors (RMSE) of the derived forecast relative to the actual observation values. As shown in Table 1, the bagging and boosting models consistently outperform the HRRR and RAP models by the RMSE metric across all forecast horizons, whereas the stacking model has no advantage over the RAP model, although it is still better than the HRRR model. Specifically, for the 24-hour forecast horizon, the bagging model has a RMSE of 0.2502, a 63% improvement over the RAP model (RMSE 0.6926) and a 85% improvement over the HRRR model (RMSE 1.6612). The boosting model has a RMSE of 0.2763, a 60% improvement over the RAP model and a 83% improvement over the HRRR model. The results are similarly favorable for the 6-hour and the 12-hour forecast horizon.

## DISCUSSION.

This study introduces a novel approach that combines predictions from two numerical wind forecast models (HRRR and RAP) using multiple ensemble algorithms (stacking, bagging, and boosting) to produce a single wind forecast. This innovative process significantly enhances current wind forecast models by merging numerical and statistical methods.

The ensemble algorithms employed in this study effectively harness the strengths of various numerical models to reduce both bias and variance in forecasting. By combining predictions from the HRRR and the RAP, model averaging addresses uncertainties in model selection and mitigates the risk of overfitting. This process, which involves aggregating predictions from different models, results in more accurate and robust forecasts.

In addition to the improvement in accuracy brought about by combining distinct base models, these ensemble algorithms also improve on previous models due to the combination of numerical and statistical methods it uses in prediction. By incorporating both methods through ensemble algorithms, the final models benefit from both methods while reducing the problems caused by each base method. Not only

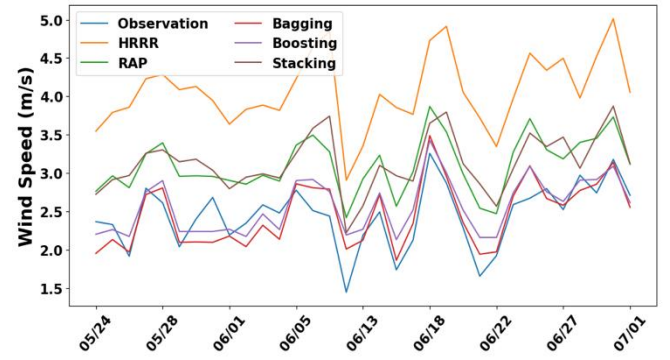


Figure 1. Time series prediction of models in the 24-hour forecast horizon.

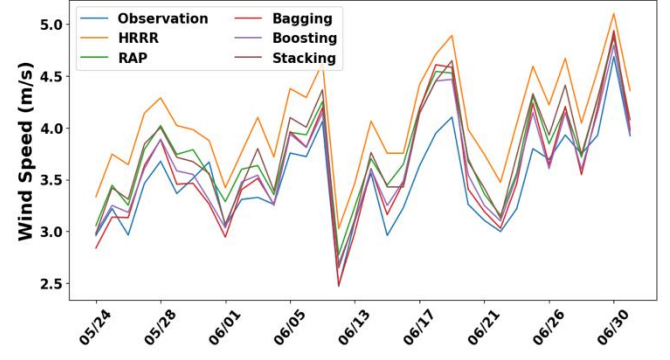


Figure 2. Time series prediction of models in the 12-hour forecast horizon.

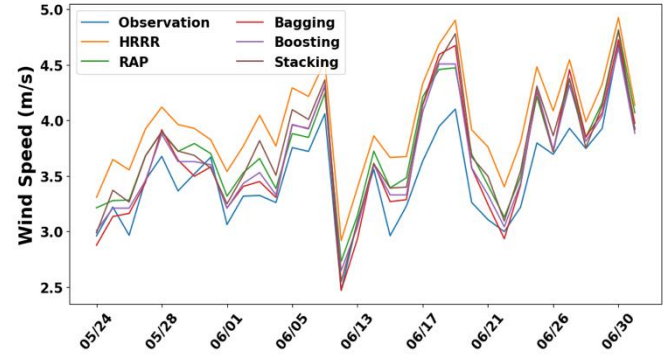


Figure 3. Time series prediction of models in the 6-hour forecast horizon.

Table 1. RMSE of different models for different forecast horizons.

Models	Forecast Horizon		
	24 hours	12 hours	6 hours
HRRR	1.6612	0.5965	0.5150
RAP	0.6926	0.3106	0.2816
Stacking	0.7499	0.3232	0.3125
Bagging	0.2502	0.2444	0.2537
Boosting	0.2763	0.2431	0.2687

does the model ultimately benefit from statistical methods as previously described, but it also begins with a base level of accuracy brought on by the numerical base models, thus allowing it to benefit from numerical methods. These ensemble models do not require additional computational resources compared to the original numerical

models, and the numerical base models do not lose any predictive accuracy due to the addition of statistical methods.

Importantly, this improvement in wind prediction model accuracy is significant for wind power prediction accuracy. With more accurate wind power predictions, less oversight is required to balance wind reserves and supply, thus reducing the cost of integrating wind into preexisting power grids [8].

While ensemble models, especially bagging and boosting, improve significantly on current models, ample room still exists for additional improvement. First, the selection of base models that are combined in the ensemble algorithms can be improved. While the RAP and the HRRR are distinct, they ultimately have very similar physics engines. Consequently, the ability of each ensemble algorithm to take advantage of the differences in the base models is limited. Employing more distinct wind forecast base models or even distinct model types may significantly increase the time adaptability and accuracy of each ensemble model. In certain settings, model types such as persistence models, based on the idea that wind speed will tend to fluctuate around a central average, are more accurate in shorter forecast periods than the numerical models utilized in these studies [9].

Second, it is important to note that all the ensemble models are deterministic, meaning each ensemble model produces a single prediction value for wind speed, unlike many current models that are probabilistic. Probabilistic models provide a range of possible wind outcomes with corresponding probabilities of each prediction occurring. This limitation means that the model offers little value in identifying the overall spread of wind speeds, which is also crucial in wind prediction. While ensemble models excel in producing a single, accurate prediction, it falls short in providing a comprehensive view of the potential range of wind speeds.

## CONCLUSION.

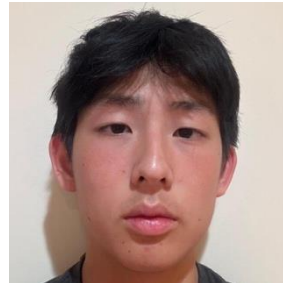
This study proposes a novel method to improve forecasting of wind speed based on the ensemble merging of existing wind forecast models. Among different ensemble algorithms, bagging and boosting performed better than stacking. The potential for further improvement and application of this method is vast, including predictions for other meteorological variables such as temperature or humidity. For future investigation, applying similar ensemble methods with more distinct base models could further enhance wind forecast accuracy.

## ACKNOWLEDGMENTS.

I would like to thank my mentor, Dr. Minwoo Song at New York University, for guiding me through this research.

## REFERENCES

1. E. J. Novaes Menezes, A. M. Araújo, N. S. Bouchonneau da Silva, A review on wind turbine control and its associated methods. *J. Clean. Prod.* **174**, 945-953 (2018).
2. H. Chen, H. Zhang, New energy generation forecasting and dispatching method based on big data. *Energy Rep.* **7**, 1280-1288 (2021).
3. S. M. Lawan, W. A. W. Z. Abidin, W. Y. Chai, A. Baharun, T. Masri, Different Models of Wind Speed Prediction; A Comprehensive Review. *Int. J. Sci. Eng. Res.* **5**, 1760-1769 (2014).
4. P. Bauer, A. Thorpe, G. Brunet, The quiet revolution of numerical weather prediction. *Nature* **525**, 47-55 (2015).
5. E. A. Tuncar, S. Sağlam, B. Oral, A review of short-term wind power generation forecasting methods in recent technological trends. *Energy Rep.* **12**, 197-209 (2024).
6. I. D. Mienye, Y. Sun, A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, **10**, 99129-99149 (2022).
7. B. Efron, T. Hastie, *Computer age statistical inference, student edition: algorithms, evidence, and data science*. (Cambridge University Press, Cambridge, UK, 2021).
8. B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, K. Rohrig, Predicting the Wind. *IEEE Power Energy Mag.* **5**, 78-89 (2007).
9. Y. K. Wu, J. S. Hong "A literature review of wind forecasting technology in the world" in *Proceedings of the IEEE Conference on Power Tech* (IEEE, 2007), pp. 504-509.



Alexander S. Choi is a student at Bergen County Academies in Hackensack, NJ.