

# Overcoming predictive limitations in observation-based wind speed forecasting using NWP data

Kosuke Yamamoto <sup>a</sup>, Hiroshi Hasebe <sup>b</sup>

<sup>a</sup>MS Student, Nihon University, Chiyoda-ku, Tokyo, Japan, [csko24016@g.nihon-u.ac.jp](mailto:csko24016@g.nihon-u.ac.jp)

<sup>b</sup>Associate Professor, Nihon University, Chiyoda-ku, Tokyo, Japan [hasebe.hiroshi@nihon-u.ac.jp](mailto:hasebe.hiroshi@nihon-u.ac.jp)

## SUMMARY

The adoption of wind power has been rapidly increasing in recent years. Maintaining the electricity supply-demand balance for stable power grid operation requires highly accurate wind speed prediction. However, for complex data with weak temporal dependencies, forecasting based on immediate past data suffers from two inherent limitations: short-term lagging and long-term convergence to the mean value. To address this problem, we utilized MSM numerical weather prediction data as input to the LSTM-based model to incorporate future information. This approach successfully improved long-term forecasting accuracy by mitigating the tendency of predictions to revert to the mean value. However, the lagging problem in short-term forecasting remained unresolved. This is because the prediction accuracy relies heavily on the quality of the MSM data itself. Any inaccuracies in the MSM simulations lead to poor wind speed forecasts. Addressing this degradation is identified as a key area for future improvement.

*Keywords:* lagging problem, Wind speed, Machine learning, LSTM, MSM

## 1. INTRODUCTION

The adoption of wind power generation has been rapidly increasing due to its potential as a sustainable energy supply, characterized by low greenhouse gas emissions and renewability. In power grid operations, maintaining the supply-demand balance is critical. Disruptions to this balance can lead to large-scale power outages. To maintain this balance, it is necessary to optimize generation planning, which involves predicting future power output and managing any shortfall or adjusting the generation accordingly. Since wind power output is strongly dependent on wind speed, highly accurate future wind speed prediction is vital.

Recurrent Neural Networks (RNNs) have been widely used for time-series forecasting. However, it has been shown that a phenomenon known as the “lagging problem” or time delay can occur in short-term predictions (Li, 2023). In this situation, predicted values become almost identical to the immediately preceding input values, resulting in a waveform that, while strikingly similar to the actual data, appears to be merely the input sequence shifted by one time step. In contrast, for long-term horizons, predictions often lose their tracking capability and converge toward the mean value of the training data. Our preliminary study observed these behaviors across various models and input data. We identified that the cause lies in the insufficient information contained solely in historical measured data to predict future wind speed changes. To address this limitation, this study proposes a method utilizing numerical weather prediction data which incorporates future meteorological information, to improve prediction accuracy.

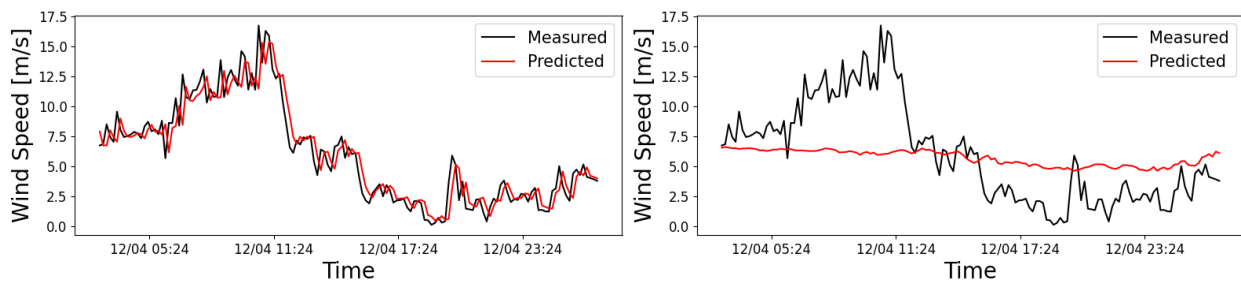
## 2. PREDICTION WITH HISTORICAL MEASURED DATA

Considering the time-series nature of wind speed data, we employ a Long Short-Term Memory (LSTM) network which is a kind of the recurrent neural network architecture optimized for time-

series data. The LSTM has a recurrent structure, enabling the step-by-step processing of sequential inputs to capture temporal dependencies. The dataset used in this study consists of in-situ wind speeds observed at a wind farm located in Iitate, Fukushima Prefecture. The period of the dataset covers three years, from 2022 to 2024. Data from 2022, 2023 and 2024 were allocated for training, validation and testing, respectively. To evaluate the model's performance across different timescales, we conducted short-term (10-minute ahead) and long-term (24-hour ahead) forecasts. In both cases, the input data consisted of a 90-minute historical sequence of measured wind speeds.

Figure 1 shows the concatenated results of sequential predictions for both the 10-minute and 24-hour horizons. For the 10-minute prediction, the figure reveals a notable lagging problem, where the predicted waveform lags behind the actual observed values by approximately 10 minutes. This indicates that the model essentially replicates the most recent data rather than predicting future fluctuations. In contrast, the 24-hour prediction results tend to converge toward the mean value of the training data. These phenomena likely arise because the historical measured data alone do not provide sufficient information to deterministically predict future wind speed changes.

To confirm this hypothesis, we extracted three cases where the input sequence (past 90 minutes) exhibited highly similar patterns, as shown in figure 2. The red dots in each plot indicate the next wind speeds. Despite the similarity in historical trends, the future outcomes vary significantly: the left graph shows an increase, the middle graph shows no change, and the right graph shows a decrease. This demonstrates a lack of a clear dependency between past observations and future changes. For short-term predictions, because wind speed generally does not change drastically over very short intervals, the model learns the feature that 'no significant change occurs'. This manifests as the lagging problem, where the model simply replicates the latest observed value. In contrast, for long-term predictions, even this short-term persistence disappears as the future wind speed can take any value within its historical range. Faced with this high uncertainty, the model fails to capture specific fluctuations and instead learns to output the mean value to minimize overall error, resulting in a loss of predictive variance.



(a) 10-minute ahead prediction (b) 24-hour ahead prediction  
Figure 1: Prediction results using historical measured data

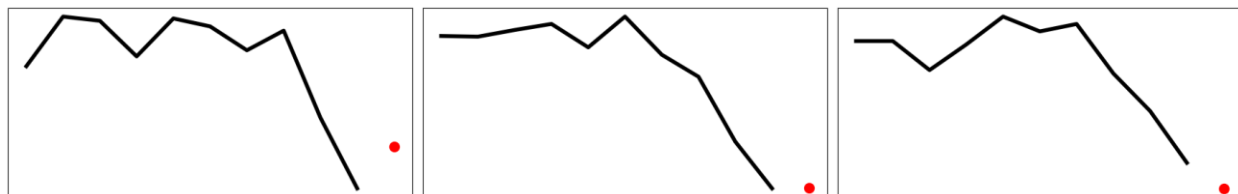


Figure 2: Next point behaviour following similar preceding patterns

To verify the hypothesis that both the lagging problem and the convergence to the mean value occurred from insufficient information regarding future changes, we conducted a validation experiment using a synthetic sawtooth wave illustrated in figure 3. By extracting three consecutive points from this waveform, we distinguished four distinct patterns. As shown in the figure, for any given trend in the first two points (increasing or decreasing), the subsequent third point can essentially move in either direction. This implies that the transition from the input (first and second points) to the output (third point) is not deterministic in the short-term. Furthermore, the sawtooth waveform consists of multiple alternating increasing and decreasing segments, whose lengths are randomly determined. As a result, the value of a point sufficiently distant from two consecutive observations can take an essentially arbitrary value within the waveform’s amplitude range. This indicates that the long-term transition from the input to a distant future point is inherently non-deterministic. Namely, there is no unique relationship between the past trend and the future change.

Using this waveform, we trained the model to perform a one-step-ahead prediction (predicting the third point from the first two) to examine the occurrence of the lagging problem in short-term forecasting. In addition, we conducted a 1000-step-ahead prediction to investigate whether the predictions converge to the mean value in the long-term setting. Additionally, as a control experiment to validate the necessity of future information, we also conducted a prediction where the target point was explicitly included in the input data. The results are shown in figure 4. When using only the first two points as input, a clear lagging problem in short-term predictions and convergence to the mean value in long-term predictions were observed. This empirically demonstrates that these predictive limitations are unavoidable when the input data lack sufficient features to predict future fluctuations. Conversely, when the future point was included in the input data, the predicted waveform coincided perfectly with the actual waveform, effectively eliminating both the lagging problem and the convergence to the mean value. These results lead to the conclusion that incorporating external information regarding future wind behavior is essential to resolve these predictive limitations.

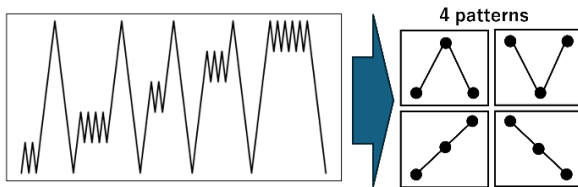


Figure 3: Sawtooth wave

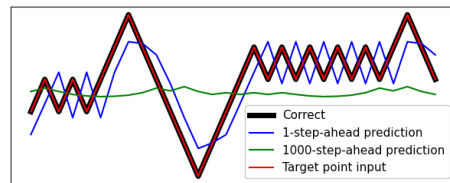


Figure 4: Prediction of the sawtooth wave

### 3. WIND SPEED PREDICTION USING NUMERICAL WEATHER PREDICTION DATA

Building on the finding that incorporating future information eliminates the lagging problem and the convergence to the mean value, we integrated numerical weather prediction data into the model. Specifically, we used forecast data from the Mesoscale Model (MSM) provided by the Japan Meteorological Agency (JMA, 2025) via the Research Institute for Sustainable Humanosphere, Kyoto University (RISH, 2025). The MSM offers meteorological forecasts, including wind speed and temperature with a spatial resolution of approximately 5 km and a forecast horizon of up to 51 hours. In this study, by incorporating the MSM wind speed data into the model, we predicted both short-term (10-minute ahead) and long-term (24-hour ahead) wind speeds. To evaluate the proposed method, we compared an observation-based model using only the preceding 90-minute

wind speeds against an MSM-integrated model that combines those observations with the MSM forecast for the target time.

The prediction results are shown in figure 5. In the observation-based model, short-term predictions tended to replicate the most recent observation data, while long-term predictions converged to a constant value as discussed in Section 2. In contrast, the MSM-integrated model successfully captured future fluctuations in long-term predictions compared to the observation-based model. However, the lagging problem in short-term predictions remained unresolved. This degradation is likely attributed to the accuracy limitations of the MSM data itself. As a grid-based model, the MSM may fail to fully resolve local extremes or rapid intensifications in wind speed.

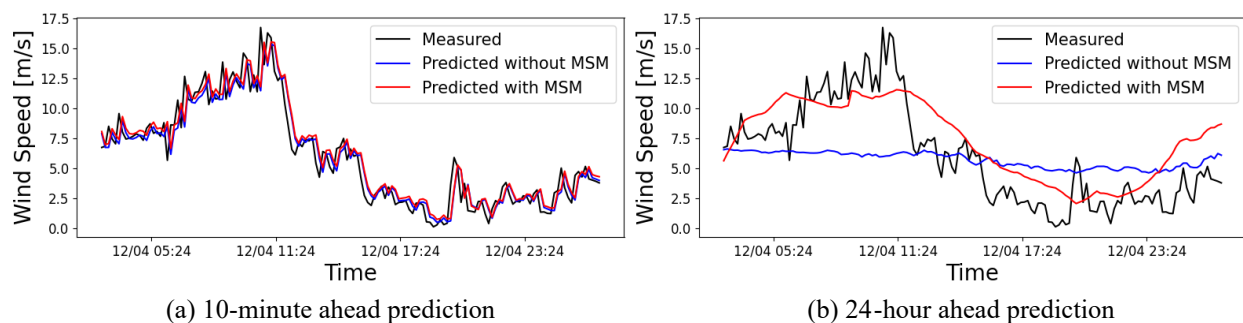


Figure 5: Prediction results using MSM

#### 4. CONCLUSIONS

In this study, we investigated the lagging problem in short-term predictions and the convergence to the mean value in long-term predictions inherent in conventional wind speed prediction models based on only historical observed data. Our investigation revealed that the root cause is the theoretical limitation of historical observed data in capturing future fluctuations due to a lack of information. To overcome these problems, we proposed a prediction model integrating numerical weather prediction data, specifically using the Mesoscale Model (MSM). As a result, the MSM-integrated model successfully mitigated the convergence to the mean value in long-term predictions and significantly improved the prediction accuracy compared to the observation-based model. However, the study also identified a limitation: the lagging problem in short-term predictions remained unresolved, as the model's performance is heavily dependent on the accuracy of the MSM data. Consequently, prediction accuracy tends to degrade when the MSM forecast itself contains uncertainty. Future work will focus on improving robustness against such MSM errors and enhancing prediction accuracy.

#### ACKNOWLEDGEMENTS

The wind speed data were kindly provided by TOKO ELECTRICAL CONSTRUCTION CO., LTD. We hereby express our sincere gratitude for their contribution.

#### REFERENCES

- Li, J., Song, L., Wu, D., Shui, J., & Wang, T., 2023. Lagging problem in financial time series forecasting. *Neural Computing and Applications*, 35, 20819–20839.
- Japan Meteorological Agency, 2025. Mesoscale Model (MSM) dataset. Available at: [https://www.data.jma.go.jp/developer/gpv\\_sample.html](https://www.data.jma.go.jp/developer/gpv_sample.html) (accessed: 2025-12-10).
- Research Institute for Sustainable Humanosphere, Kyoto University, 2025. MSM data download website. Available at: <http://database.rish.kyoto-u.ac.jp/arch/jmadata/data/gpv/original> (accessed: 2025-12-10).