CWE-8
June 7-11, 2026
London, ON, Canada

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

Kosuke Yamamoto [a], Hiroshi Hasebe [b]

[a]MS Student, Nihon University, Chiyoda-ku, Tokyo, Japan, csko24016@g.nihon-u.ac.jp
[b]Associate Professor, Nihon University, Chiyoda-ku, Tokyo, Japan 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 the most recent 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 an 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), which possess a recurrent structure, have been widely utilized for time-series forecasting. However, it has been reported that a phenomenon known as the "lagging problem" can occur in short-term predictions (Li et al., 2023). In this situation, the predicted values become almost identical to the immediately preceding input values. Consequently, while the predicted waveform appears strikingly similar to the actual data, it is essentially nothing more than the input sequence shifted by one time step. In contrast, for long-term forecasting, models tend to lose their tracking capability and converge toward the mean value of the training data. It has been reported that such time delays also occur in wind speed forecasting (Xie et al., 2021). Our preliminary study confirmed these behaviors across various models and input datasets. One possible cause is that predictions based solely on recent time series data fail to extract sufficient features to capture future changes, especially for data such as wind speed that fluctuates rapidly over short periods. In this study, we propose a hypothesis regarding the cause of this problem and verify its validity. Furthermore, to address this limitation, we propose a method to improve prediction accuracy by utilizing numerical weather prediction data that incorporates future meteorological information.

2. LONG SHORT-TERM MEMORY

For the forecasting method, we utilize Long Short-Term Memory (LSTM), which is specifically designed for time-series forecasting (Hochreiter and Schmidhuber, 1997). LSTM is an advanced architecture of the Recurrent Neural Network (RNN), developed to overcome the vanishing gradient problem and effectively capture long-term temporal dependencies. As illustrated in Figure 1, LSTM possesses a recurrent structure within its hidden layer, allowing it to process input sequences sequentially and capture temporal dependencies. Furthermore, the hidden layer incorporates three distinct gate mechanisms: the forget gate, the input gate, and the output gate. Through these gates, the model dynamically controls the retention and discarding of information. By learning to selectively process data based on its importance, LSTM achieves both the maintenance of long-term dependencies and an improvement in overall prediction accuracy.

Specifically, the input at time t and the hidden state from the previous time step are combined as shown in Eq. (1).

$$z_t = [h_{t-1}, x_t] \tag{1}$$

where z_t is the combined vector, h_{t-1} is the hidden state from the previous time step, and x_t is the input data at the current time. Next, the combined vector is used to determine the outputs for the forget gate, the input gate, and the output gate. The outputs of the forget gate, the input gate, and the output gate are calculated by Eq. (2), (3), and (4), respectively.

$$f_t = \sigma\left(W_f z_t + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i z_t + b_i) \tag{3}$$

$$o_t = \sigma(W_o z_t + b_o) \tag{4}$$

where f_t, i_t, and o_t are the outputs of the forget gate, the input gate, and the output gate, respectively. W and b are parameters updated during training, with distinct values used for each gate. σ denotes the sigmoid function shown in Eq. (5). This function has a range from 0 to 1 and serves to control the flow of information through each gate.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

where e is Euler's number. Subsequently, the candidate value for updating the cell states from the current input is calculated by Eq. (6).

$$\widetilde{c_t} = tanh(W_c z_t + b_c) \tag{6}$$

where $\widetilde{c_t}$ is the candidate value for updating the cell state. Using the obtained candidate value, the forget gate output, the input gate output, and the previous cell state, the new cell state is calculated by Eq. (7).

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{7}$$

where c_t and c_{t-1} are the new and previous cell state, respectively. $\odot$ denotes element-wise multiplication. The first term governs the information to be retained from the past cell state, while the second term governs the new information to be incorporated from the current input. Finally, the information to be output is controlled by passing the cell state through the tanh function and multiplying it by the output of the output gate, as shown in Eq. (8) to calculate h_t.

$$h_t = o_t \odot tanh(c_t) \tag{8}$$

h_t retains information up to the current time and is passed forward to the calculation of the next time step. Since the h_t output at the final time step can be regarded as encompassing the information from all preceding time points, this mechanism enables forecasting that accounts for temporal dependencies.

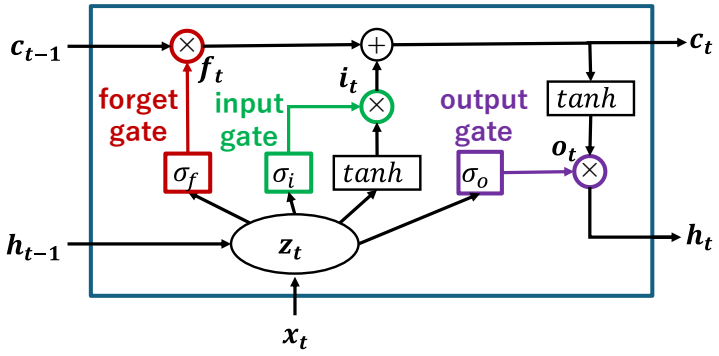


Figure 1: LSTM structure

3. PREDICTION WITH HISTORICAL MEASURED DATA

In this chapter, we conduct forecasting using the most recent wind speed data to verify the occurrence of the lagging problem in short-term predictions and mean convergence in long-term predictions.

3.1. Data

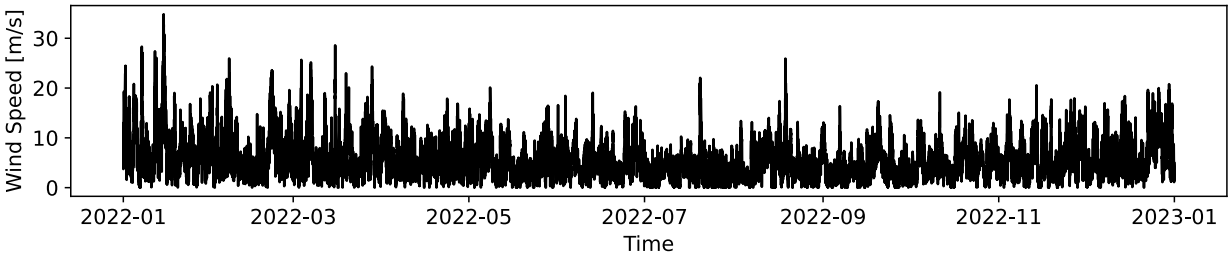
The dataset consists of wind speed observed at a wind turbine in Iitate Village, Fukushima Prefecture, from 2022 to 2024. These data are partitioned into three sets: the 2022 data for training, the 2023 data for validation, and the 2024 data for testing. However, the periods with missing data—from 10:20 on November 6, 2023, to 18:10 on January 9, 2024, and from 11:40 on December 11, 2024, to 16:50 on December 16, 2024—are excluded from the analysis. Figure 2 shows the time-series waveform of the wind speed data.

3.2. Prediction Conditions

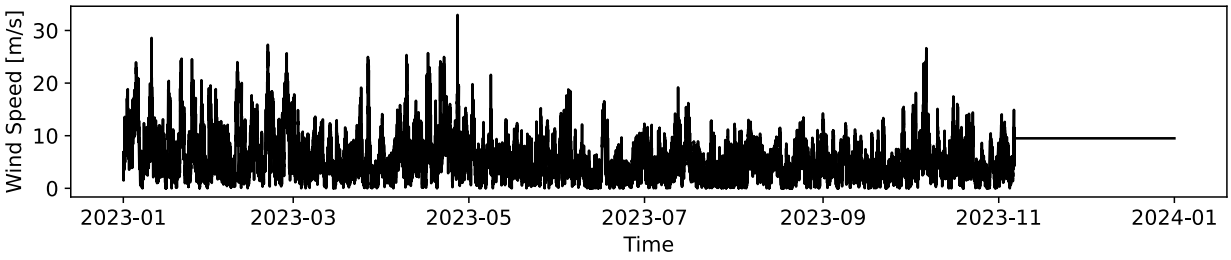
The input data consists of the most recent wind speed time series, with the prediction horizons set to 10 minutes for short-term forecasting and 24 hours for long-term forecasting. The previously described LSTM model is employed for these predictions. The hyperparameters of the LSTM and the input sequence length were optimized using the validation data. The finalized parameter settings are summarized in the Table 1.

Table 1: parameter settings

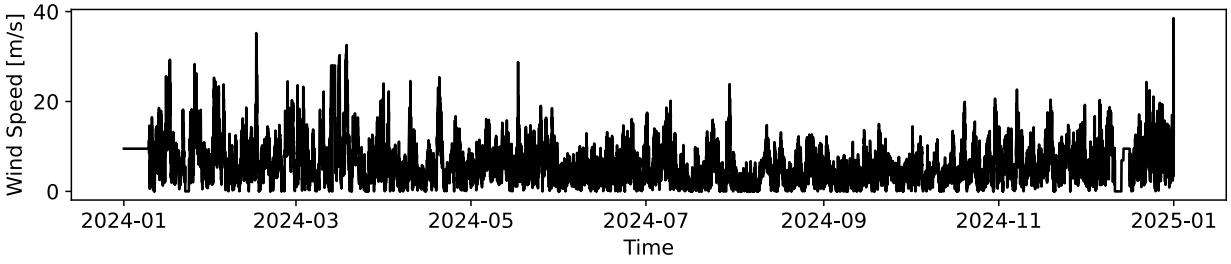
Epochs	20
Batch size	16
Hidden units	16
Number of layers	1
Input sequence length	1 hour



(a) 2022



(b) 2023



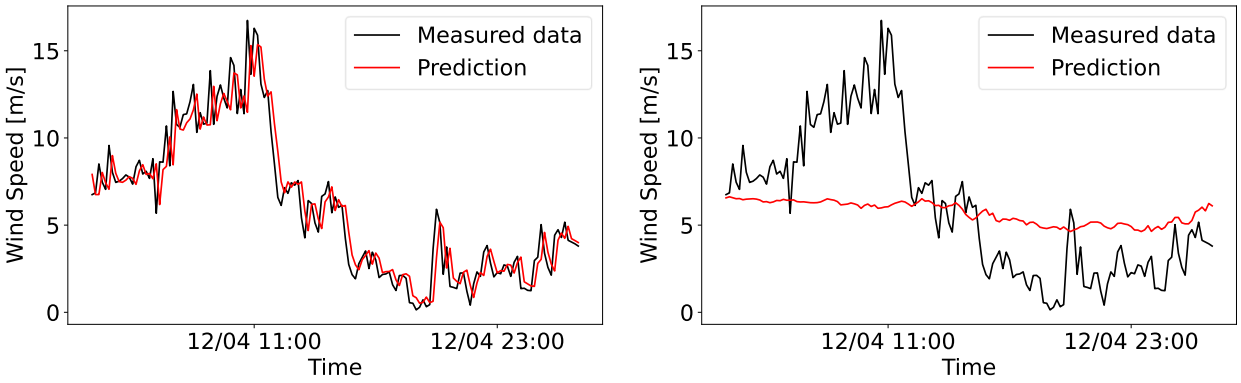
(c) 2024

Figure 2: Wind speed data

3.3. Prediction results

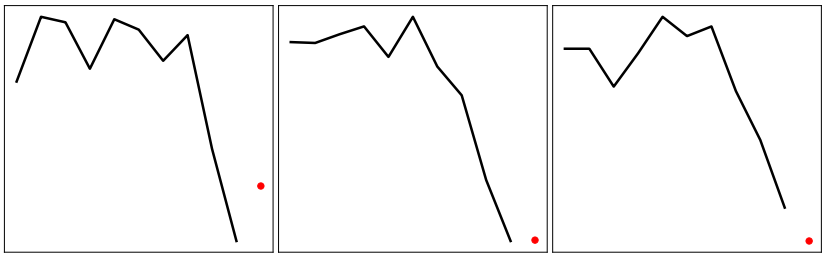
The forecasting results are presented in Figure 3. As shown in the figure, the predicted waveform for short-term forecasting lags behind the actual waveform, indicating that the model primarily tracks the immediate past observations rather than capturing future changes. In other words, a significant lagging problem is occurring. This suggests that the model predicts future values under the assumption that they will not change from the most recent input. To examine this factor in detail, three cases with similar preceding trends were extracted, as shown in Figure 4. In Figure 4 (a), the wind speed increases in the next step; in Figure 4 (b), it remains stagnant; and in Figure 4 (c), it decreases. These cases demonstrate that there is no clear correlation between the most recent trend and the subsequent change. Furthermore, the histogram of wind speed variations in Figure 5 shows that the cases of increase and decrease are almost equally distributed. It is considered that a model trained on such data tends to select an "average" output—effectively predicting no change

(remaining at the same level as the previous time step)—in the process of minimizing the prediction error. On the other hand, in long-term forecasting, the predicted waveform converges to a constant value, which approximates the mean of the training data. Similar to the short-term forecasting issue, this is attributed to the lack of features in the most recent data that can determine changes 24 hours ahead. The reason for the convergence to the mean in long-term forecasting is inferred to be that as the target time becomes more distant from the input, the correlation with the most recent values vanishes and the range of possible values widens. Consequently, the model learns that outputting the mean value is the optimal solution for minimizing the loss function.



(a) 10-minute ahead prediction (b) 24-hour ahead prediction

Figure 3: Prediction using measured data



(a) Increase (b) No change (c) Decrease

Figure 4: Next point behaviour following similar preceding patterns

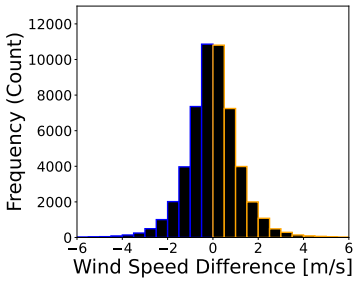


Figure 5: Histogram of wind speed variations

4. NUMERICAL INVESTIGATION INTO CAUSES OF PREDICTION ERRORS

4.1. Sawtooth Wave

To elucidate the mechanisms underlying the "lagging problem in short-term forecasting" and "mean convergence in long-term forecasting" discussed in the previous chapter, we conduct a verification using the sawtooth wave shown in Figure 6. This waveform is designed to have four distinct transition patterns across three consecutive points. In this setup, whether the preceding two points (input sequence) show an increasing or decreasing trend, both "increase" and "decrease" patterns exist for the subsequent point (target). Furthermore, the cases of increase and decrease are distributed almost equally. Additionally, by randomly setting the length of intervals where increases and decreases alternate, values at a point sufficiently distant from a given reference point become random within the range of the sawtooth wave. As described above, this waveform

satisfies the characteristics hypothesized as the causes of the lagging problem and mean convergence: (1) the absence of a causal relationship between preceding and subsequent trends, (2) the equal distribution of increasing and decreasing patterns at the next point, and (3) the property that values at sufficiently distant points can take any arbitrary value within the waveform's range.

4.2. Validation results

Using this waveform, we conducted forecasting with two consecutive points as input. For the verification of short-term forecasting, a one-step-ahead prediction was performed, while a 1,000-step-ahead prediction was conducted for the long-term forecasting. Figure 7 shows the forecasting results. In the short-term forecasting illustrated in Figure 7 (a), the predicted waveform shifted in the positive direction of the time axis relative to the ground-truth waveform, confirming a significant lagging problem. Meanwhile, in the long-term forecasting, as shown in Figure 7 (b), the predicted values converged toward the mean of the sawtooth wave. These results demonstrate the validity of the aforementioned hypothesis regarding the causes of the lagging problem and mean convergence. Next, we conducted predictions under a condition where the data for the actual target time—the future information—was intentionally added to the input data. The purpose of this verification was to confirm whether incorporating future information into the input data would improve forecasting accuracy. Figure 8 shows these results. The accuracy improved in both the short-term forecasting in Figure 8 (a) and the long-term forecasting in Figure 8 (b), confirming that incorporating data containing future information can fundamentally enhance prediction performance. In actual wind speed forecasting, however, the future measured wind speed is unknown and cannot be used as input data. Therefore, this study focuses on the Numerical Weather Prediction (NWP) data. NWP data consist of analytical values of future atmospheric states calculated through simulations based on physical models, thereby providing predictive information on future meteorological dynamics. Furthermore, since these are forecast values distributed in advance, they can be utilized as input data at the time of prediction.

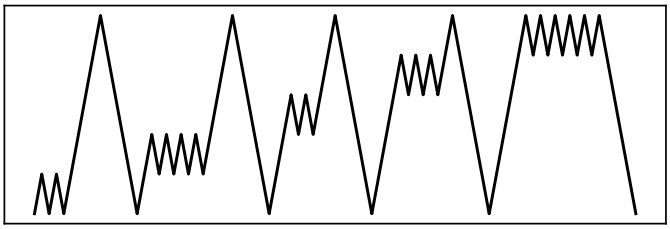
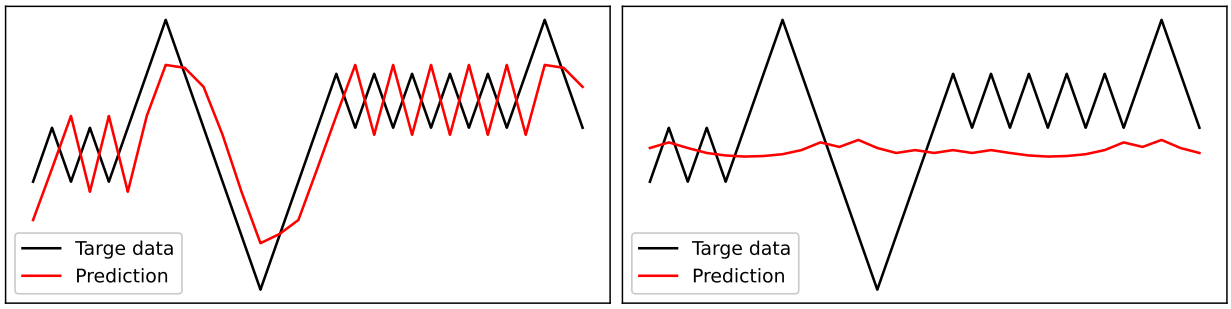
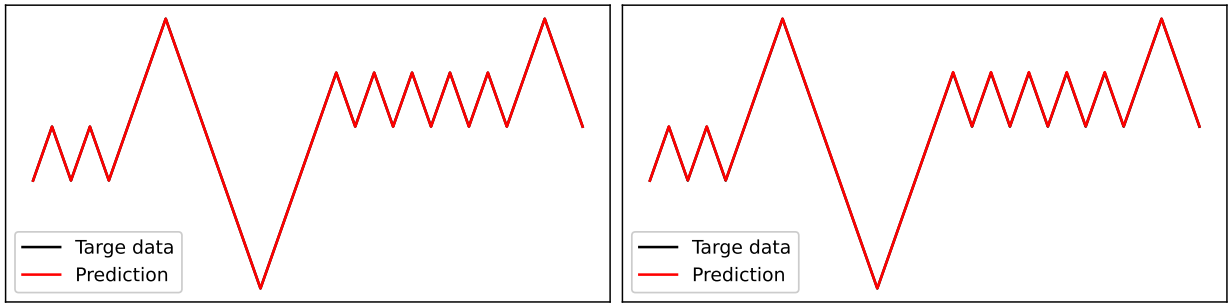


Figure 6: Sawtooth wave



(a) 1-step-ahead prediction (b) 1000-step-ahead prediction

Figure 7: Prediction of sawtooth wave



(a) 1-step-ahead prediction (b) 1000-step-ahead prediction

Figure 8: Prediction of sawtooth wave with target point input

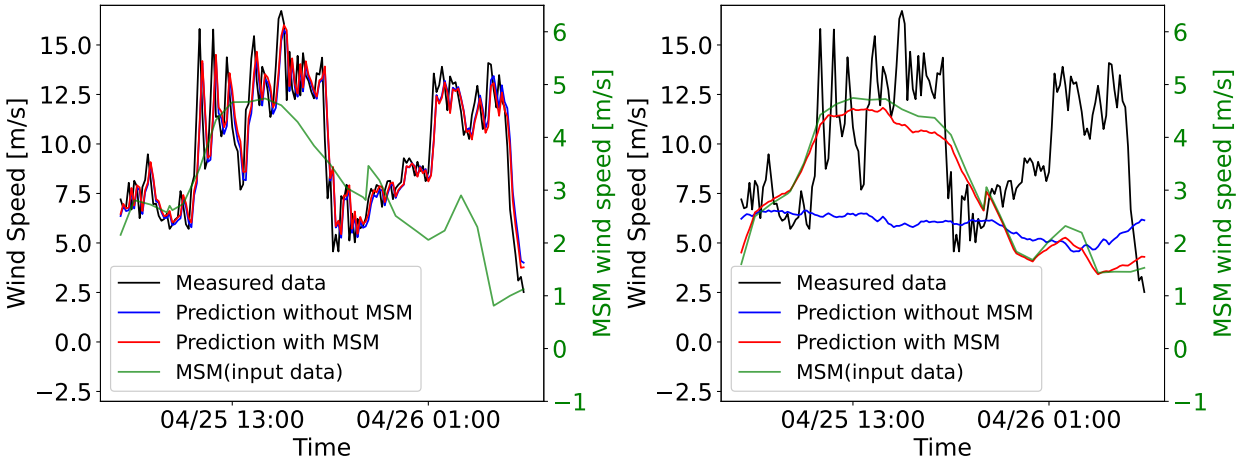
5. PREDICTION WITH NWP DATA

5.1. Prediction using wind speed NWP data

Based on the investigations in the previous chapters, it has been demonstrated that incorporating future information into the model mitigates both the lagging problem and mean convergence. Accordingly, this chapter presents a forecasting approach using numerical weather prediction (NWP) data that contains future meteorological information. For the NWP data, we adopt the MesoScale Model (MSM) provided by the Japan Meteorological Agency (JMA, 2025). The MSM is defined on a grid with a horizontal resolution of approximately 5 km and provides forecast values for up to 51 hours ahead. While various meteorological parameters are available, this study specifically utilizes wind speed data for forecasting. The input data consists of measured wind speed values up to the current time, supplemented by the MSM wind speed forecast corresponding to the target prediction time. For instance, a 10-minute-ahead prediction utilizes the MSM forecast for 10 minutes ahead, while a 24-hour-ahead prediction uses the corresponding 24-hour MSM forecast. Since the MSM data is provided at one-hour intervals, we apply linear interpolation to generate data at 10-minute intervals to match the prediction steps of this study. Furthermore, considering the MSM distribution schedule (9:00 and 21:00 JST daily), we refer to the latest forecast data available at the time of prediction. For example, if the current time is 10:00 and the goal is to predict the wind speed one hour later (11:00), the model utilizes the forecast value for 11:00 from the MSM data most recently distributed at 9:00.

Figure 9 shows a comparison of the forecasting results between the model using only measured data and the model incorporating MSM data. As shown in Figure 9 (a), the lagging problem occurred in the 10-minute-ahead short-term forecasting, even when MSM data was introduced, similar to the case using only measured data. To analyze the cause, we plotted the transition of the MSM wind speed used as input data in the same figure. It was confirmed that the MSM wind speed does not capture fine, short-term fluctuations such as those occurring at 10-minute intervals. The insufficient temporal resolution and the limitations in reproducing local variations in the meteorological forecast data are considered the primary reasons why the short-term forecasting accuracy did not improve. On the other hand, regarding the long-term forecasting results in Figure 9 (b), we focus on two distinct wind speed increase events. For the first event, the introduction of MSM wind speed data successfully suppressed the convergence of predicted values toward the mean value, allowing the model to generally follow the observed trends. This confirmed an improvement in accuracy that was unattainable using measured data alone. However, for the second increase event, the model still failed to capture the rapid fluctuations, and accurate

prediction remained difficult even with the inclusion of MSM data. Upon examining the MSM wind speed (input data) during this specific period, it was found that the MSM data itself did not correctly reflect the actual observed trends. Therefore, the errors in long-term forecasting are inferred to stem from the simulation errors inherent in the MSM forecast values themselves. As described above, the introduction of MSM data partially improved the accuracy of long-term forecasting. Nevertheless, challenges remain, such as resolving the lagging problem in short-term forecasting and further enhancing accuracy in long-term forecasting. Since these issues depend on the accuracy limits of the input MSM wind speed, it is considered necessary to utilize other meteorological parameters.



(a) 10-minute ahead prediction (b) 24-hour ahead prediction

Figure 9: Prediction using MSM wind speed

5.2. Prediction using MSM pressure gradient

To address the degradation in accuracy caused by errors in the MSM wind speed, we introduced new atmospheric pressure data—specifically, the MSM mean sea-level pressure. The rationale for focusing on pressure is that wind is a physical phenomenon driven by pressure gradients, and incorporating these fluctuation characteristics is considered effective for improving predictions. To integrate pressure gradients into the model, we acquired pressure data not only from the target site but also from surrounding locations, utilizing the pressure gradients between these points as input features. Since the spatial relationship between the target site and the surrounding points may influence prediction accuracy, we first investigated this correlation. Specifically, as illustrated in Figure 10, we selected points in eight cardinal and ordinal directions centered on the target site. We then varied the distance from the target site within a range of 1 to 100 grid points, in 1 grid point increments, to compare changes in prediction accuracy. RMSE was employed as the evaluation metric, and the verification was conducted under two conditions: "all samples" and "high-wind-speed samples" (wind speeds of 25 m/s or higher). As shown in Figure 11, the results indicated that the RMSE reached its minimum, yielding the highest prediction accuracy, when using pressure gradients from points located 30 to 80 grid points away from the target site. Based on these findings, we determined the optimal distance for acquiring surrounding pressure data to be within the range of 30 to 80 grid points from the target site.

We conducted wind speed forecasting by adding the pressure gradients between the selected points as new features. The input data consisted of the measured wind speed values up to the current time,

the MSM wind speed forecast for the target time, and the MSM pressure gradients between the target site and the surrounding points at the target time. Figure 12 shows the predicted waveforms, and Figure 13 illustrates the relationship between the prediction horizon and the RMSE. As shown in Figure 12 (a), no significant improvement was observed regarding the lagging problem in short-term forecasting, even with the introduction of pressure gradients. However, in the long-term forecasting shown in Figure 12 (b), the use of pressure gradients information made it possible to predict rapid increases in wind speed that could not be captured by the MSM wind speed data alone. Furthermore, as is evident from the error evaluation results in Figure 13, the method incorporating pressure gradients recorded the lowest RMSE, particularly in the long-term forecasting range, confirming a significant improvement in accuracy. However, as shown in Figure 14, there were still several instances where wind speed fluctuations could not be accurately predicted, even when using pressure gradients. This is likely because the internal calculation errors within the MSM pressure data itself, as well as micro-scale wind speed variations resulting from local topographical effects, cannot be fully predicted using only wide-area pressure gradient information.

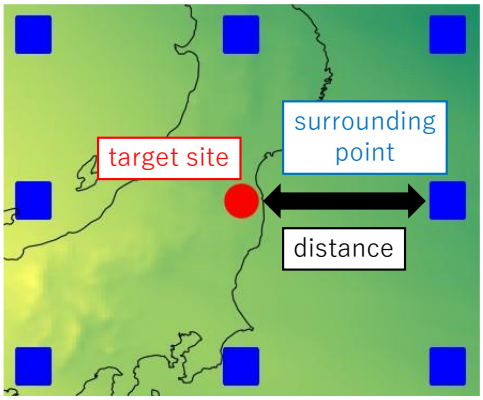


Figure 10: Eight surrounding points

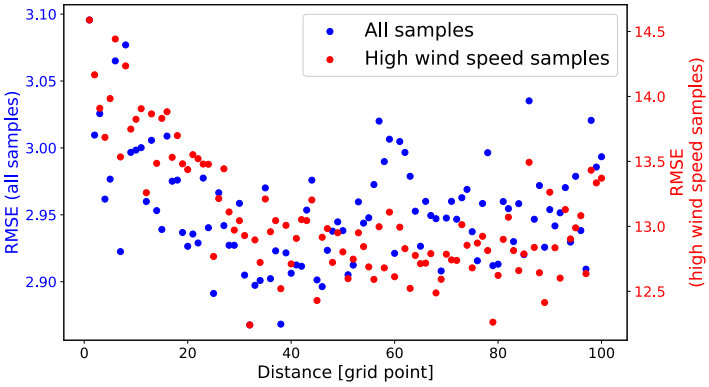
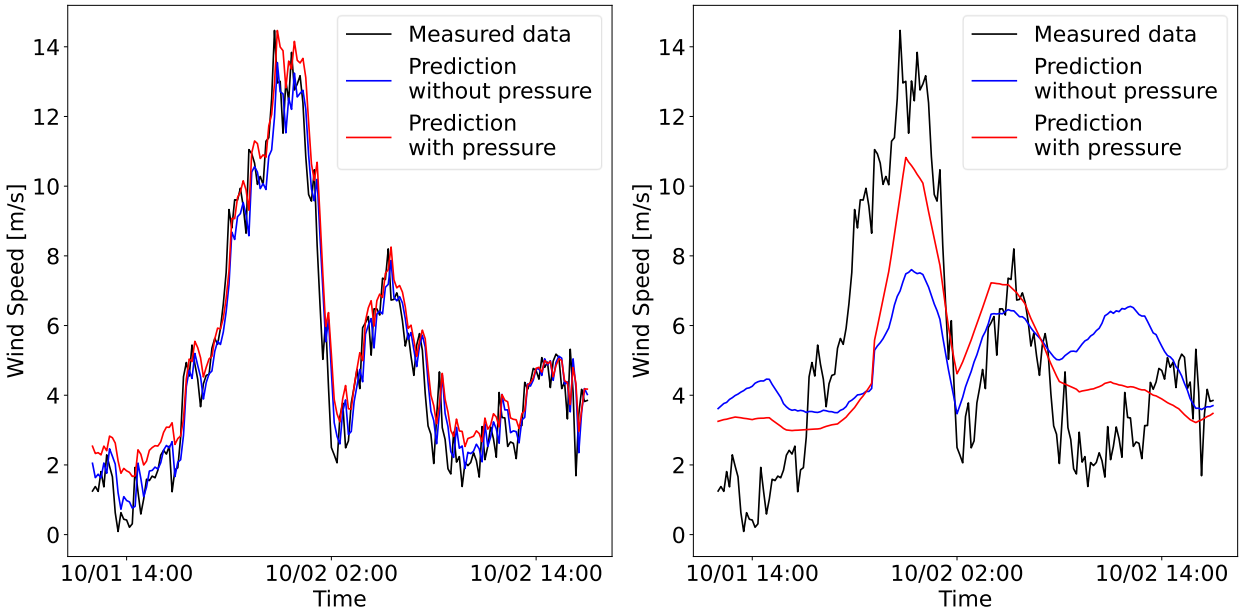


Figure 11: Comparison of RMSE across different distances



(a) 10-minute ahead prediction (b) 12-hour ahead prediction

Figure 12: Prediction using MSM pressure gradient

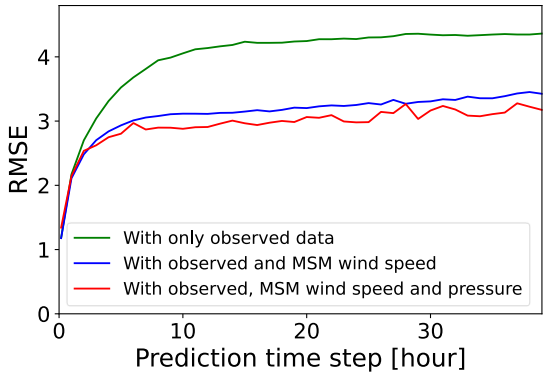


Figure 13: Comparison of RMSE across different prediction horizons

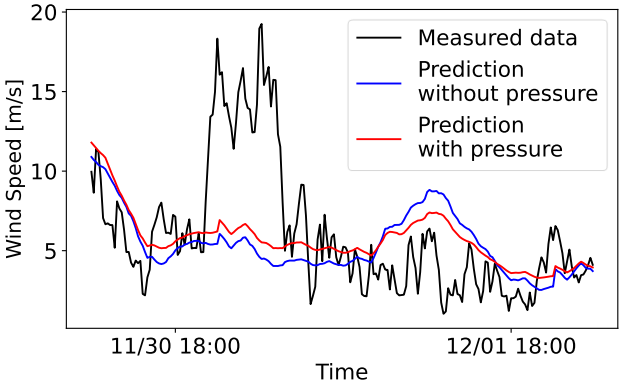


Figure 14: Inaccurate prediction case

6. CONCLUSION

In this study, we conducted wind speed forecasting aimed at maintaining the balance between power supply and demand. First, the mechanisms underlying accuracy degradation in forecasts using only measured data were verified through numerical experiments using a sawtooth waveform. As a result, we identified the causes of the "lagging problem" in short-term forecasting and "mean convergence" in long-term forecasting and derived a countermeasure by incorporating future meteorological information into the input data. Next, we proposed a forecasting method utilizing Numerical Weather Prediction (MSM) data. The results demonstrated that while the lagging problem in short-term forecasting was not fully resolved, the model successfully followed the trends of the actual measured values without converging to the mean in long-term forecasting, thereby improving prediction accuracy. Furthermore, by introducing pressure gradients around the target site as features, we showed that the prediction accuracy for wind speed fluctuations associated with pressure gradients could be further enhanced. However, cases with remaining prediction errors were still observed. Future challenges involve detailed analysis of meteorological factors other than pressure and local topographical effects and integrating them into the model to further improve forecasting performance.

ACKNOWLEDGEMENTS

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

MSM data were collected and distributed by Research Institute for Sustainable Humanosphere, Kyoto University (RISH, 2025)

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.

Xie, A., Yang, H., Chen, J., Sheng, L., & Zhang, Q., 2021. A short-term wind speed forecasting model based on a multi-variable long short-term memory network. Atmosphere, 12(5), 651.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735–1780.

Japan Meteorological Agency, 2025. Mesoscale Model (MSM) dataset. Available at: 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).